



**HAL**  
open science

# Segmentation des tumeurs du foie sur des images CT

Daniel Pescia

► **To cite this version:**

Daniel Pescia. Segmentation des tumeurs du foie sur des images CT. Vision par ordinateur et reconnaissance de formes [cs.CV]. Ecole Centrale Paris, 2011. Français. NNT : 2011ECAP0002 . tel-00649030v1

**HAL Id: tel-00649030**

**<https://theses.hal.science/tel-00649030v1>**

Submitted on 6 Dec 2011 (v1), last revised 8 Dec 2011 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



ECOLE CENTRALE DE PARIS

# PHD THESIS

to obtain the title of

**PhD of Science**

of Ecole Centrale de Paris

**Specialty : APPLIED MATHEMATICS**

Defended by

Daniel PESCIA

## Segmentation of liver tumors on CT images

Thesis Advisor: Nikos PARAGIOS, Stéphane CHEMOUNY

prepared at Ecole Centrale de Paris, MAS laboratory

defended on January the 7, 2011

### **Jury :**

<i>President :</i>	Alain LUCIANI	-	INSERM - H. Mondor
<i>Reviewers :</i>	David SARRUT	-	CNRS - Creatis
	William PUECH	-	CNRS - LIRMM
<i>Advisors :</i>	Nikos PARAGIOS	-	INRIA - ECP
	Stéphane CHEMOUNY	-	Intrasense
<i>Examinators :</i>	Eric PICHON	-	GE Healthcare
	Gérard SUBSOL	-	CNRS - LIRMM

2011ECAP0002



# Acknowledgments

First, I would like to thank my two PhD advisers. I am grateful to Professor Nikos Paragios for accepting me in his team. His expert guidance, his patience and support truly contributed to this work. I am also grateful to Stéphane Chemouny for offering me to continue as a PhD student at the end of a traineeship several years ago. Seeing the company grow during my PhD while keeping the good atmosphere of the beginning was a pleasure; this success is particularly due to his personality.

The financial support of this thesis was granted by the ANRT (Association Nationale de la Recherche et de la Technologie), which is an organization aiming at reinforcing public private partnerships towards the innovation. I divided my time between the MAS (Mathématiques Appliquées aux Systèmes) laboratory in the ECP (Ecole Centrale Paris) and the Intrasure company in Montpellier. Despite some unavoidable logistic difficulties I have always been welcome in both places and provided with the material required for my work; for this I am very grateful.

I am grateful to David Sarrut and William Puech for accepting to review this document. I appreciated their valuable comments and constructive feedback about my work. I would also like to thank Alain Luciani who replaced at short notice the president of my jury, as well as Eric Pichon and Gérard Subsol for accepting to be member of the thesis committee; their presence was an honor for me. A particular thank to Gérard Subsol for putting me in contact with Nikos Paragios for this PhD.

Warm thanks to the members of the vision group in MAS Laboratory, their support and the delightful atmosphere made my days more enjoyable. I would like to specially mention Régis Behmo and Panagiotis Koutsourakis for their valuable help with C++ coding and Martin de La Gorce for his help with Matlab. Special thanks to my successive office-mates: Ahmed, Fabrice, Aristeidis, Chaohui, Salma, Loic, the two Oliviers and Pascale with whom I had a nice time and enjoyed many valuable discussions.

I would also like to thank all members of Intrasure for their support during my thesis and for the great atmosphere in the company. In particular, I would like to thank Charles Caderas de Kerleau for introducing me to the available tools in the company and making them evolve to suit some of my requests. I would also like to thank Frédéric Banégas for following my work and the rewarding discussions about planning and priority choices. I am also grateful to Didier Lovera for this help with the graphic elements of this document and Eric Durieux, Olivier Pathy and Christoph Spuhler for their help with English.

I am indebted to all people who have manually segmented or validated liver and lesions. Without their complex and time consuming work, none of this work would have been possible. S. Hemm-Ode, A. René, Dr. L. Pupulim, Pr. B. Gallix, Dr. C. Bonnel, Dr. M. Zappa your help was greatly appreciated.

Many friends have helped with their support, proof reading of this manuscript... I greatly value their help.



# Abstract

This thesis is dedicated to 3D segmentation of liver tumors in CT images. This is a task of great clinical interest since it allows physicians benefiting from reproducible and reliable methods for segmenting such lesions. Accurate segmentation would indeed help them during the evaluation of the lesions, the choice of treatment and treatment planning. Such a complex segmentation task should cope with three main scientific challenges: (i) the highly variable shape of the structures being sought, (ii) their similarity of appearance compared with their surrounding medium and finally (iii) the low signal to noise ratio being observed in these images.

This problem is addressed in a clinical context through a two step approach, consisting of the segmentation of the entire liver envelope, before segmenting the tumors which are present within the envelope.

We begin by proposing an atlas-based approach for computing pathological liver envelopes. Initially images are pre-processed to compute the envelopes that wrap around binary masks in an attempt to obtain liver envelopes from estimated segmentations of healthy liver parenchyma. A new statistical atlas is then introduced and used to segmentation through its diffeomorphic registration to the new image. This segmentation is achieved through the combination of image matching costs as well as spatial and appearance priors using a multiscale approach with MRF.

The second step of our approach is dedicated to lesions segmentation contained within the envelopes using a combination of machine learning techniques and graph-based methods. First, an appropriate feature space is considered that involves texture descriptors being determined through filtering using various scales and orientations. Then, state of the art machine learning techniques are used to determine the most relevant features, as well as the hyperplane that separates the feature space of tumoral voxels to the ones corresponding to healthy tissues. Segmentation is then achieved by minimizing an MRF energy that combines class probabilities and neighbor constraints.

Promising results demonstrate the potentials of our method.



# Résumé

Cette thèse porte sur la segmentation des tumeurs du foie sur des images tomodesitométriques. Ce sujet présente un intérêt certain pour le domaine médical puisque les médecins pourraient ainsi bénéficier d'une méthode reproductible et fiable pour segmenter de telles lésions. Une segmentation précise des tumeurs du foie permettrait en effet d'aider les médecins lors de l'évaluation des lésions (détection, localisation, quantification), du choix d'un traitement, et de sa planification. Les méthodes développées dans ce cadre doivent faire face à trois principales difficultés scientifiques: (i) la grande variabilité de l'apparence et de la forme des structures recherchées, (ii) leur ressemblance avec les régions environnantes et finalement (iii) la faiblesse du rapport signal sur bruit observé dans les images dans lesquelles on travaille.

Ce problème est abordé dans une optique d'application clinique et est résolu en suivant une approche en deux temps commençant par le calcul d'une enveloppe du foie, avant de segmenter les tumeurs présentes à l'intérieur de cette enveloppe.

Nous commençons par proposer une approche basée sur des atlas pour le calcul d'une enveloppe des foies pathologiques. Tout d'abord, un outil de traitement d'image a été développé pour calculer une enveloppe autour d'un masque binaire, afin d'essayer d'obtenir une enveloppe du foie à partir d'une estimation du parenchyme sain. Un nouvel atlas statistique a ensuite été introduit, puis utilisé pour la segmentation à travers son recalage diffeomorphique avec une image. La segmentation est finalement réalisée en combinant les coûts d'appariement des images avec des a priori spatiaux et d'apparence, le tout en suivant une approche multi échelle basée sur des MRFs.

La deuxième étape de notre approche porte sur la segmentation des lésions contenues dans ces enveloppes en combinant des techniques d'apprentissage par ordinateur avec de méthodes basées sur des graphes. Un espace d'attributs approprié est tout d'abord défini en considérant des descripteurs de textures déterminés à travers des filtres de diverses tailles et orientations. Des méthodes avancées d'apprentissage automatique sont ensuite utilisées pour déterminer les attributs pertinents, ainsi que l'hyperplan qui sépare les voxels tumoraux des voxels correspondant à des tissus sains dans cet espace d'attributs. Pour finir, la segmentation est réalisée en minimisant une énergie sous forme de MRF, laquelle combine les probabilités d'appartenance de chaque voxel à une classe, avec celles de ses voisins.

Des résultats prometteurs montrent les potentiels de notre méthode.





# Foreword

Segmentation has well known but different meanings in Computer Vision and for liver anatomy. In computer Vision, segmentation refers to the partitioning of one image in multiple regions that correspond to different anatomical structures in medical imaging. For physicians, the segmentation of the liver refers to the subdivision of the liver in 8 or 9 segments that are defined using the vessels inside the liver. In order to prevent confusion between these two definitions, segmentation will refer to the meaning in Computer Vision, while anatomical segmentation will refer to the physician meaning.



# Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	Vision and segmentation . . . . .	5
1.2	Medical imaging . . . . .	6
1.3	Motivation . . . . .	9
1.4	Challenges . . . . .	10
1.5	Overview . . . . .	11
<b>2</b>	<b>Liver Tumor segmentation</b>	<b>13</b>
2.1	Intro . . . . .	15
2.1.1	Intro . . . . .	15
2.1.2	Overview . . . . .	15
2.2	Medical background . . . . .	16
2.2.1	The liver . . . . .	16
2.2.2	Injection protocols . . . . .	21
2.2.3	Type of lesions . . . . .	23
2.3	Liver Tumor Segmentation, framework and purpose . . . . .	34
2.3.1	Setting the problem . . . . .	34
2.3.2	Medical value . . . . .	38
2.4	State of the art . . . . .	40
2.4.1	Review . . . . .	40
2.4.2	Extracting some knowledge . . . . .	42
2.5	Overview of the proposed method . . . . .	44
2.6	Pre-processing . . . . .	45
2.6.1	Normalization, working on a common basis . . . . .	46
2.6.2	Retained method . . . . .	48
<b>3</b>	<b>Segmenting the liver</b>	<b>51</b>
3.1	Intro . . . . .	54
3.1.1	Liver problematic . . . . .	54
3.1.2	Context . . . . .	55
3.1.3	Worth . . . . .	58
3.1.4	Overview . . . . .	58
3.2	Previous work . . . . .	59
3.2.1	Region growing strategies . . . . .	60
3.2.2	Probabilistic Atlases . . . . .	60
3.2.3	Statistical Shape Model . . . . .	61
3.3	A simple approach: improving existing segmentation . . . . .	62
3.3.1	Intro . . . . .	62
3.3.2	Transformation into a correct envelope . . . . .	64
3.3.3	Filling surface cavities . . . . .	68

3.3.4	Test protocol . . . . .	73
3.3.5	Discussion, results . . . . .	74
3.4	Statistical atlas to represent image variability . . . . .	74
3.4.1	Motivation . . . . .	74
3.4.2	Theoretical background . . . . .	76
3.4.3	Creation of a statistical atlas . . . . .	87
3.5	Segmentation, atlas based . . . . .	95
3.5.1	Intro . . . . .	95
3.5.2	MRF for image segmentation . . . . .	96
3.5.3	Atlas Registration . . . . .	101
3.5.4	Segmentation by registering a statistical atlas . . . . .	107
3.5.5	Test protocol . . . . .	109
3.5.6	Results, discussion . . . . .	111
3.6	Combined segmentation . . . . .	113
3.6.1	Intro . . . . .	113
3.6.2	Introducing prior models . . . . .	114
3.6.3	Segmentation using prior models . . . . .	117
3.6.4	Protocol, method . . . . .	121
3.6.5	Discussion, Results . . . . .	121
3.7	Conclusion, summary . . . . .	125
<b>4</b>	<b>Tumor segmentation inside a liver envelope</b>	<b>127</b>
4.1	Intro . . . . .	130
4.1.1	Context . . . . .	130
4.1.2	Challenge . . . . .	132
4.1.3	Role in the global process . . . . .	133
4.1.4	Overview . . . . .	134
4.2	Machine learning . . . . .	135
4.2.1	Solving complex problems . . . . .	135
4.2.2	State of the Art . . . . .	136
4.2.3	Importance of validation . . . . .	138
4.2.4	AdaBoost, a relevant method for our problem . . . . .	139
4.3	Selection of texture features . . . . .	146
4.3.1	Texture metrics, quantifying local appearance . . . . .	147
4.3.2	Filtering, preparing images before treatment . . . . .	152
4.3.3	Defining texture features . . . . .	162
4.3.4	Feature selection . . . . .	166
4.4	Segmentation of tumoral tissues . . . . .	171
4.4.1	Creation of a classification function . . . . .	171
4.4.2	Segmentation based on pixel classification . . . . .	174
4.5	Protocol . . . . .	177
4.5.1	Defining the classification function . . . . .	177
4.5.2	Implementation issues . . . . .	180
4.5.3	Test . . . . .	186

---

4.6	Discussion, Results . . . . .	188
4.6.1	Straight segmentation of tumors . . . . .	188
4.6.2	Segmentation of tumors using MRFs . . . . .	193
4.7	Conclusion, Summary . . . . .	196
<b>5</b>	<b>Conclusion, Future Work</b>	<b>199</b>
5.1	Contributions . . . . .	199
5.2	Future work . . . . .	200
	<b>Bibliography</b>	<b>203</b>



# List of Figures

1.1	Picture of a scanner. There are two components, a table where the patient lies down and a hollow tube within which the CT images are acquired. . . . .	7
1.2	Principle of the scanner. The absorption of X-rays is measured for a same slice of the body but for diverse directions (a,b). Then, the mean absorption inside a quantization of the scanned object (c) is computed by solving an inverse mathematical problem. The original image of the abdomen comes from the Visual Human Project <sup>®</sup> [NML 1989]. . . . .	8
1.3	Hounsfield scale with some corresponding tissues. Values for air and water are the only stable values as they are used to set the Hounsfield scale. Other correspondences are given as rough estimates because they may vary between the exams. . . . .	9
2.1	Human circulatory system with focus on the liver. The liver receives blood from two networks; the portal vein brings nutrients from the intestine and the hepatic artery supplies the liver with oxygen. Then, the hepatic veins drain de-oxygenated blood from the liver. Relative dimensions of the organs are not kept. . . . .	17
2.2	Anterior view of the blood networks inside the liver. The three blood networks are shown inside the liver. The portal veins are shown in purple (a,b), the hepatic artery in red (a) and the hepatic vein is displayed in light blue (b). Images are provided by CHU St Eloi. Segmentation and rendering were achieved with Myrian <sup>®</sup> . . . . .	19
2.3	Contribution of enhancement for a HCC. A same axial slice is shown before the injection of a contrast product (a) and then for diverse enhancements, arterial phase (b) and portal phase (c). Images are courtesy from Beaujon's hospital. . . . .	22
2.4	Enhancement of some anatomical structures, functions of the time length after injection. The enhancement is given in Hounsfield Unit (HU) for the aorta, the portal vein and the liver. The graph is taken from Blum's book [Blum 2002]. . . . .	24
2.5	Worldwide incidence of liver cancers for males in 2002. Data was provided by CANCERmondial (IARC-WHO) [GLOBOCAN 2002]. . . . .	25
2.6	Appearance of HCC on CT images for arterial and portal/delayed phases. Data is taken from Honda's study [Honda 1993]. . . . .	27
2.7	Appearance of several metastases in arterial phase on CT images, functions of the primary site. The possible appearances of four lesions are given along with the repartition of these appearances as percentages. Data was extracted from a paper by Nino-Murcia <i>et al.</i> [Nino-Murcia 2000]. . . . .	30



2.8	Clinical use of Myrian <sup>®</sup> for lungs and liver. A virtual hepatectomy is first shown, i.e. the planning of the surgical resection of a part of the liver (a). Then, the evaluation of emphysema inside lungs is presented (b). . . . .	37
2.9	Segmentation methods for the liver tumors, ordered by increasing overlap error. The overlap errors are given for a same set of images that was used for MICCAI segmentation challenge [Xiang 2008]. Some information was not available in the papers and will be shown by a dash. . . . .	43
2.10	Detection and characterization methods for the liver tumors. Some information was not available in the papers and will be shown by a dash. . . . .	44
2.11	Histogram matching for two Gaussian distributions. Histograms (a) and cumulative histograms (b) are shown for two Gaussian distributions. A target is shown as a solid blue line, and the source is displayed as a red dashed one. The matching process for a single gray value $I_{src}$ is then shown as the research of a correspondence between the two cumulative distributions (b). . . . .	47
2.12	Histogram equalization for the peppers image. The histogram (c) of the initial image (a) is equalized (d), which gives an image with more contrast (b). Histograms are shown as blue bar graphs and cumulative histograms as red solid lines. . . . .	49
3.1	Variability of shape and volume for several livers. 6 reference segmentations for tumoral liver are shown using Myrian <sup>®</sup> with surface rendering in frontal view. The volume of the hepatic envelope is given for each image. The envelope is displayed on each image in light pink and lesions in purple. Images are courtesy from Beaujon's hospital. . . . .	56
3.2	Variability of appearance for three liver examples with colorectal metastases. Images are courtesy from the CRLC Val d'Aurelle. . . . .	57
3.3	Limits of the computation of a smooth envelope from a liver ROI. Reference ROI is shown in dark gray, missing parts in light gray and the correction with the envelope is displayed as a black line. . . . .	64
3.4	Segmentation of the hepatic envelope from the healthy parenchyma. Ground truth liver is shown in green, the healthy hepatic parenchyma given as input is blue, and the result of the hole filling process is displayed as purple. The hole filling process should fill interior holes (c.1) and cavities on surface (c.2). . . . .	65
3.5	Filling holes on meshes as proposed by Zhao [Zhao 2007]. A flowchart of the proposed method is given (a), along with the corresponding states for a skull (b). The figure is extracted from Zhao's paper. . . . .	67

---

3.6	Principle of the hole filling process for an artificial example. The process flowchart (a) for the hole filling process is presented and illustrated for an artificial example created by the exclusion of an ellipsis and a circle (b). . . . .	68
3.7	Computation of a contour from a ROI, example for a connex component of the liver. An initial segmentation is shown in light blue, along with the reference for the liver in orange (a). A missing part of the liver on the right is due to a tumor lesion. The contour for the ROI is first computed using Freeman coding chain (b). Then, non contributive nodes are removed from this contour (c). Finally, this contour is pruned in order to simplify its representation (d). . . . .	70
3.8	Matching candidate boundaries. The matching process is illustrated for a liver outline shown in medium gray, where candidate boundaries are shown as black x. The direct view constraint is presented on a first example (a), with solid blue line for correct pairs and dashed red ones for the incorrect matches. The contribution of the heuristic value is shown in a second example that describes the heuristic value of the match with colored lines (b). . . . .	72
3.9	Examples of results for the hole filling process. The initial binary ROI and the filled parts are shown as masks on CT images. The former is shown in light pink and the latter in red. . . . .	74
3.10	Nonrigid registration of two slices of the abdomen. The source image (a) is registered on the target image (b). This registration is done through the computation of a deformation field (c). This deformation field is then applied to the source image in order to define a deformed image (d) that better matches with the target image. For this example, registration was achieved with the drop2D software using SAD as similarity measure [Glocker 2009, Komodakis 2009b]. . . . .	77
3.11	Intensity distribution of one GMM along with its three components. The global distribution (solid purple) is defined as a weighted sum of three normal distributions with diverse parameters. Each basis distribution is related to a different type of object (or tissue). . . . .	81
3.12	GMM for clustering. An artificial sample distribution contains spatial locations of two distinct modes shown as blue crosses and red squares (a). By fitting a GMM with 2 components on this data, the two modes may be retrieved (b). . . . .	82
3.13	One iteration for the k-means algorithm. Observations are shown as crosses. Clusters are shown as circles, squares and triangles. The initial observations are shown in (a). First an initial pick of centroids is defined by taking some observations as cluster centroids (b). Then, each observation is assigned to the cluster with the closest centroid (c). Finally, new centroids are computed in (d), which may be different from the existing observations. . . . .	83

3.14	Atlases as mean images created on raw and registered image sets. Elliptic phantoms are considered to create an atlas as a representative image. A mean image computed on the raw samples is first shown (a). Then a mean image is computed on a registered set (b). The registered set was created using the first sample as target image during the registration using drop2D with NMI or SAD as similarity measure [Glocker 2009]. . . . .	88
3.15	Comparison of two definitions for statistical atlases. Two atlases are computed for the previous set of registered ellipses (fig. 3.14). First a statistical atlas defined with one Gaussian pdf per pixel is shown (a). Then an atlas defined with GMMs is displayed (b). For both atlases the intensity distribution is shown for each part of the image. . . . .	90
3.16	Creation of registered volumes for use during atlas construction. Several volumes are registered on a same target. All these volumes are first clipped around the liver region in order to simplify and speed-up the process. Then, they are subsampled by a factor 2 for performance reasons (both because of reduced memory consumption and additional speed). Finally, the subsampled volumes are registered on a same target and in a soft manner. . . . .	91
3.17	Creation of an atlas from a set of registered volumes. Given a set of n volumes registered on a same target, an intensity sample is first extracted for each voxel of the space. Then, each sample is modeled by GMM with exactly 1, 2, 3, 4 or 5 components. The optimal model at each spatial location is finally decided by the MDL measure. . . . .	93
3.18	Segmentation through registration of a representative image. A representative image (a) is registered onto a new image (c) through a transformation defined by a deformation field (b). The segmentation of the new image (e) is then obtained by applying this deformation field on a reference segmentation for the representative image (d). . . . .	95
3.19	Graph cut segmentation of a small image. A graph is defined (b), where each node is one point of the source image (a) and where two terminal nodes are added (S for object and T for background). The edges of the graph are defined either by the connexity inside the image (black) or by additional edges between each node of the image and both terminal nodes (light and medium gray). To each edge corresponds a cost shown by the width of the edge, either as a continuity condition (black edges) or as a similarity measure for a class. Finally a cut is done while minimizing the cost of the cut edges (c). This cut defines two distinct graphs that define the segmentation results. The example is taken from [Boykov 2001b] . . . . .	100

3.20	Registration process for the fusion of two liver slices. The deformed source and target are shown along with related deformation fields for an example previously introduced (fig. 3.10). The deformed source is shown in red and the target image in blue on blended images. Intermediate deformations and blended images are shown for diverse scales. The incremental construction of the deformation is also shown for the coarser scale. . . . .	104
3.21	.Segmentation through the registration of an atlas. An atlas (a) is registered onto a new image (c) through a transformation defined by a deformation field (b). The segmentation of the new image (e) is then obtained by applying this deformation field on a reference segmentation for the atlas (d). . . . .	108
3.22	Segmentation through atlas registration, quantitative evaluation. . .	111
3.23	Segmentation through atlas registration, examples for two livers. Several results of segmentation are given for one metastatic liver an one with HCC. The reference segmentation is shown in blue and the automatic segmentation in red. . . . .	112
3.24	Combined segmentation, solving process. . . . .	120
3.25	Combined segmentation, evaluation on several images. First, the results with atlas only are reminded. Then, the results of combined segmentation are given for two appearance priors, either an intensity distribution, or a classification function based on texture features. . .	122
3.26	Combined segmentation, examples for a metastatic liver. Several results of segmentation are compared for segmentation through atlas registration, and combined segmentation with either a pdf or a texture prior. The reference segmentation is shown in blue and the automatic segmentation in red. The examples are given for a metastatic liver on a CT image with a size of voxel of $1.6 \times 1.6 \times 10\text{mm}$ . . . . .	123
3.27	Combined segmentation, examples for a liver with HCC. Several examples of segmentation are compared for segmentation through atlas registration, and combined segmentation with either a pdf or a texture prior. The reference segmentation is shown in blue and the automatic segmentation in red. The examples are given for a liver with HCC on a CT image with a size of voxel of $1.5 \times 1.5 \times 2.5\text{mm}$ . . . . .	124
4.1	Separation of two noisy classes in the space, the problem of overfitting. An artificial sample is considered, with two classes shown as red squares and blue circles that are linearly separated. However, some noise is added to the sample. A separation is then researched to distinguish these two classes. First, a linear separation is computed (a), and then a separation based on B-splines is researched (b). . . . .	139
4.2	AdaBoost, algorithm of the learning process. . . . .	142

4.3	Contribution of validation for two classification functions obtained while selecting features for the detection of liver colorectal metastases. Sensitivity and specificity are displayed functions of the length of the classifier. . . . .	144
4.4	Comparison of diverse machine learning techniques for the classification of healthy vs. tumoral tissues. Training was done with a set containing 1014 samples described by 144 features. The learned classifier was then evaluated on a test set of 75987 samples. . . . .	146
4.5	Two voxels with the same intensity, but with very different surrounding textures. The first texture (a) seems random, while the second one (b) contains linear patterns. . . . .	147
4.6	Creation of the co-occurrence matrix $P(1, 0)$ . The occurrences of each pair of pixels are counted (c) for a texture patch (a). The count is done iteratively by considering each pair of pixels at distance 1 and for angle 0. A pair of pixels (shown in light red) is considered and used to update the co-occurrence matrix, while ignoring the order of the intensities (b). . . . .	150
4.7	Convolution of an image, practical use. Given an image (a) convolution of the entire image is done pixel by pixel. Treatment of each pixel is done by taking a patch around this pixel first, whose size is that of the kernel, and then by computing the new intensity of the pixel with a weighted sum between patch intensities and weights inside the kernel (b, c). . . . .	153
4.8	Influence of voxel sizes for filtering, sagittal view. The volume used by taking three slices is shown between two vertical lines. The anatomical structures contained inside the 3 slices are very different between images with a slice thickness of 1mm (a) and 5mm (b). . . . .	154
4.9	Impact of mean and Gaussian filters on a noisy image. The pepper image (a) was modified by adding a Gaussian noise (0 mean and deviation 25) (b). The noisy image is then smoothed for diverse sizes of kernel using a mean filter (c,d,e) and a Gaussian filter with diverse standard deviations (f,g,h). . . . .	155
4.10	Impact of the parameters on a Gabor's kernel. The impact of the parameters on a Gabor's kernel is shown by modifying one parameter at a time from an initial set of parameters. For this initial setting, the orientation is $\theta = 0$ , the wavelength $\lambda = 10$ , the spatial aspect ratio $\gamma = 0.5$ , the bandwidth $b = 1$ , and the phase offset $\varphi = 0$ . The intensities displayed in the images were modified to improve the visualization. . . . .	158
4.11	Comparison of median and mean filters on a 1D signal. A random noise is added to a Heaviside step function (a, b). Noise reduction is then compared for mean (c) and median (d) filters using a neighborhood of 3 points. . . . .	159

4.12	Influence of median filtering on a binary image ( $400 \times 400$ ) for diverse sizes of patch. . . . .	160
4.13	Contribution of Nagao's filter on a simple example. A pixel (black square), located on the edge between two regions is filtered using a local neighborhood (black square contours). Different masks that may be considered for Nagao's are shown in medium gray (b,c,d). . .	161
4.14	The 3 kinds of Nagao's masks retained in our study, each type with 3 possible locations on z. Centered masks (a,b,c) are boxes including the central voxel of the patch. Corner masks (d,e,f) and masks for linear boundaries (g,h,i) are respectively designed for regions in contact at the corners and on sides of patches. . . . .	163
4.15	Impact of mean, median and Nagao's filters on a noisy image. The pepper image (a) was modified by adding salt and pepper noise (b). The noisy image is smoothed using a mean filter (c,d,e), a median filter (f,g,h) and Nagao's filter with Nagao's initial masks(i,j,k), all with diverse sizes of kernel [Nagao 1979]. . . . .	164
4.16	Quality of segmentation functions of the size of texture. First, the global quality is given for diverse sizes of texture (a). Then, the underlying metrics, namely sensitivity and specificity are given (b). .	167
4.17	Computation of the parameters for a weak learners. The values of the features are shown on a horizontal axis as bars, whose size shows the weight associated to each sample. The candidate thresholds are displayed as green strokes. . . . .	173
4.18	Creation process of a correct classification function. . . . .	178
4.19	Definition and use of an integral image. The integral image at point (x,y) is defined as the sum of pixel intensities on the upper left part of this point, and shown as a gray region in image (a). Then, sums on squares may be computed using an integral image by using only the values of the integral image on the corner of this square (b). . . .	183
4.20	Straight segmentation of tumors for a texture size of $13 \times 13$ without removing the vessels. Ground truth is outlined in white and the result of the segmentation in black. CT images are courtesy from V. Vilgrain's Department (Beaujon, Paris). . . . .	189
4.21	Segmentation results obtained with the same classifier for diverse tumors. The segmentation of the vessel networks as a first step does not modify the sensitivity, whereas specificity improves. . . . .	192
4.22	Percentage in volume of tumors accurately segmented, functions of the volume of the lesions. The volumes of the texture patches span between $0.61$ and $3.05 \text{ cm}^3$ , with an average volume of $1.54 \text{ cm}^3$ . . .	192

---

4.23	Segmentation of colorectal metastases using MRFs. Results of segmentation and references are given for three volumes. For the first volume, all ROIs obtained by our approach are displayed (a), while only 5 are shown as reference (b). On the second volume is only shown a single non confluent lesion (c,d). For the last volume, the 5 lesions chosen in the medical protocol are shown (f), along with the same lesions for the automatic approach (e). The ROIs of the lesions are displayed with surface rendering using Myrian <sup>®</sup> . . . . .	194
4.24	Segmentation of colorectal metastases using MRFs, results. The quality of segmentation is evaluated using measures from Computer Vision with the exception of specificity that is non significant due to the relative volumes of lesions and images. These measures are given for all lesions, or for all lesions $\geq 3\text{cm}^3$ . . . . .	196
4.25	Segmentation of colorectal metastases using MRFs, results. The quality of segmentation is evaluated using several medical metrics for the evaluation of solid tumors. These measures are given for all lesions, or for all lesions $\geq 3\text{cm}^3$ . . . . .	196

# Notations

Symbol	Meaning
$p(A   B)$	conditional probability of A given B
$\mathbb{N}$	the set of natural integers
$\mathbb{R}$	the set of real numbers
$\bar{\delta}_{x,y}$	inverse Kronecker's delta
$\ \cdot\ $	any distance function
$\ \cdot\ _e$	euclidean distance function
$ \cdot $	absolute value
$ \cdot _c$	cardinality of a set
$\Theta$	a set of parameters
$\mathcal{N}(\mu, \sigma)$	the normal distribution with mean $\mu$ and standard deviation $\sigma$
$\pi_i$	the weight of the $i^{th}$ component of a GMM
$\mathcal{V}$	an image volume
$\mathcal{V}^{(t)}$	a subsampled image volume
$\Omega$	image domain
$\Xi$	the space of GMMs
$\mathbf{x}, \mathbf{y}$	voxels
$\mathcal{A}$	an atlas
$\mathcal{I}$	a set of images
$\mathcal{W}$	a set of registered images
$\mathcal{P}_n$	a neighborhood patch
$\mathcal{B}_{Nagao}$	a set of Nagao's mask
$c_i$	a class of tissues
$x_i$	an observation
$(x_i, y_i)$	a sample and its expected class
$\chi$	a training set
$h_t$	a weak learner
$\alpha_t$	an AdaBoost weight
$D_t$	an AdaBoost distribution
$\gamma$	threshold for the weak learner definition
$\mathcal{G}$	a graph
$\mathcal{N}$	a set of edges for a graph
$V$	a set of nodes for a graph
$u_p$	the label at node p
$V_p(c)$	the singular potential at node p for the label c
$V_p^{(t)}(u_p)$	the singular potential at node p for the label $u_p$ and for the $t^{th}$ step of an iterative process
$V_{p,q}(u_p, u_q)$	the pairwise potential between nodes p and q for the labels $u_p$ in p, and $u_q$ in q



---

Symbol	Meaning
$\mathcal{C}$	a labeling, i.e. a list of labels for each node of a graph
$\mathcal{C}^*$	an optimal labeling
$\mathcal{L}$	the set of possible labels
$\mathcal{T}$	a transformation
$\mathcal{T}^*$	the optimal transformation
$\mathcal{T}^{(t)}$	transformation in an iterative scheme
$\mathcal{D}$	displacement field in (sect. 3)
$\mathcal{M}_c$	a spatial prior
$p_{app}(\cdot   i)$	the probability of belonging to a class $c_i$ for an appearance prior
$\mathcal{H}_{c_i}$	a histogram for the class $c_i$
$\mathcal{CH}_i$	a cumulative histogram for the class $c_i$
$\mathcal{V}_{\mathcal{M}}$	an optimal representative volume
$\sigma_{\mathcal{M}}$	a variance map
$\mathcal{F}$	a bank of filters
$\mathcal{D}$	a bank of texture descriptors in (sect. 4)
$\phi$	a feature
$\mathcal{S}_{bin}$	a binary ROI as mask
$\mathcal{Q}$	a mesh

# Acronyms

<b>Acronym</b>	<b>Meaning</b>
2D	2-dimensional
3D	3-dimensional
AFM	Advancing Front Mesh
AI	Artificial Intelligence
AH	Adenomatous Hyperplasia
CCC	Cholangiocarcinoma
CPU	Central Processing Unit
CT	Computed Tomography
CV	Computer Vision
DSC	Dice Similarity Coefficient
EM	Expectation Maximization
FFD	Free Form Deformation
FN	False Negative
FNH	Focal Nodular Hyperplasia
FP	False Positive
GMM	Gaussian Mixture Model
HAP	Hepatic Arterial Phase
HBV	Hepatitis B Virus
HCC	Hepatocellular Carcinoma
HCV	Hepatitis C Virus
HU	Hounsfield Unit
ICC	Intrahepatic Cholangiocarcinoma
KPCA	Kernel Principal Component Analysis
LBP	Loopy Belief Propagation
LiTT	Laser-induced Interstitial Thermotherapy
LUT	Look Up Table
M.D.	Medicinæ Doctor (Doctor of Medicine)
MDL	Minimum Description Length
MI	Mutual Information
MRF	Markov Random Field
MRI	Magnetic Resonance Imaging
MWA	Microwave Ablation
NURBS	Non-Uniform Rational Base Spline
PA	Probabilistic Atlas
PCA	Principal Components Analysis
pdf	Probability Density Function
PEI	Percutaneous Ethanol Injection

---

<b>Acronym</b>	<b>Meaning</b>
PVP	Portal Venous Phase
ROI	Region of Interest
RFTA	Radiofrequency Thermal Ablation
SAD	Sum of Absolute Differences
SSM	Statistical Shape Model
SVM	Support Vector Machine
TACE	Transartérial Chemoembolization
TAE	Transcatheter Arterial Embolization
TN	True Negative
TP	True Positive

# Introduction

---

## 1.1 Vision and segmentation

The ability to see and to understand what is seen has been the subject of many debates and many theories since the ancient Greeks. Sight is indeed the primary sense that humans use to understand and interact with the environment. Thus, understanding the biological phenomena behind the vision is particularly interesting. Vision is often divided in two processes, a sensing step and an interpretation step, respectively achieved by the eyes and the brain. While extremely complex, sensing with the eyes is simple compared to the interpretation of what is seen. The eyes can indeed be described as organs that transform light into an electric signal. On the opposite, the interpretation by the brain cannot be easily described as a single process. Visual perception indeed involves a large percentage of the human brain (50%) and is still not well understood. Moreover, vision is not something that is taught, but is learned during childhood. Consequently, explaining what one sees and how one can understand what he is looking at are two difficult tasks. For example, explaining how to distinguish a cat from a dog will not be easy, and neither will be the description of an elephant to a blind man.

In Computer Vision, one precisely tries to teach how to see and interpret what is seen, i.e. to extract information from an image. Besides, the explanations are not meant for a human, who could take advantage of his experience or his other senses, but are meant for a computer that does not have any intelligence. In the computer world, like for human, the visual perception remains divided in a sensing and an interpretation step. However, both steps are radically different in computer vision compared to the human vision. The sensing is indeed carried out by a machine such as a video camera or a CT-scanner that outputs an image as a 2D or 3D function. This image is defined by a value, or intensity, for each position inside a discrete quantization of the space. Then, the interpretation step is the problem that is handled in Computer Vision. However, the extraction of information from images has still to improve, as computers are still unable to analyze complex scenes or images. Due to the complexity of the problem, and the way the human brain interprets the image signals, Computer Vision has relations with many other fields such as mathematics, statistics, machine learning, AI, signal processing, and image processing. . .

The extraction of information from an image may take many forms, such as identification of particular features, recognition of objects, or segmentation. . . This latter may be seen as the research of a partition of an image in a number of regions with

homogeneous features. Such regions are named Regions of Interest (ROIs). These ROIs correspond to structures that are imposed by the choice of a homogeneity condition suited to a specific problem. Two main approaches may be followed during segmentation, either a boundary or a region based approach. The first approach defines the problem as the research of boundaries between different structures. And the second approach restates the problem as the research of voxels spatially and visually close inside an image. This latter approach may also be seen as a labeling problem, where a label is assigned to each voxel of an image depending on the structure it is the more similar to. In particular, region based approaches fit better for the segmentation of liver tumors, as there are no marked differences for many boundaries of the tumors.

## 1.2 Medical imaging

Medical imaging is a specific field of Computer Vision that deals with medical images. During the last decades many tools have been developed for the imaging of the body, first producing 2D representations of the body (X-Ray, Ultra-Sound) and more recently 3D volumes as sets of 2D images of the body. These imaging modalities, such as Computed Tomography (CT) or Magnetic Resonance Imaging (MRI), allow taking 3D images of the body in a non invasive way. Thus, medical images became increasingly important for medical practice. These images indeed contribute to many medical tasks such as diagnosis, follow-up of patients and evaluation of treatments. . . However, the amount of data produced by imaging modalities cannot be handled with manual approaches, especially since the introduction of 3D images. The current CT machines indeed allow acquiring 3D images as sets of 2D slices of the body, where each slice can measure less than a millimeter. This resolution implies that medical images acquired with recent clinical equipment will count lots of slices; the typical abdominal images will count around 500 slices, and full body images can count up to 2000 slices. Thus, techniques coming from computer vision have been introduced to extract information from the medical images. While many techniques from Computer Vision are used in medical imaging, medical imaging is also a domain on its own. Indeed, the methods and the problems differ significantly from usual Computer Vision, because of the specificity of the images and the particular nature of living organisms. In particular the 3D nature of the images requires specific algorithms in order to deal with the problems induced by 3D (e.g. partial volumes) and for taking advantage of the information brought by 3D; working on a 3D image is different from working independently on each slice that composes this volume.

In this study only CT images will be considered, thus the CT modality will be briefly introduced with a focus on the image acquisition and the tomographic reconstruction. CT scanners are relatively new machines. Their conception indeed began in the late 1960s before production in the 1970s. Then, scanners quickly showed their contribution in the medical domain. Therefore, their designer, Sir Godfrey

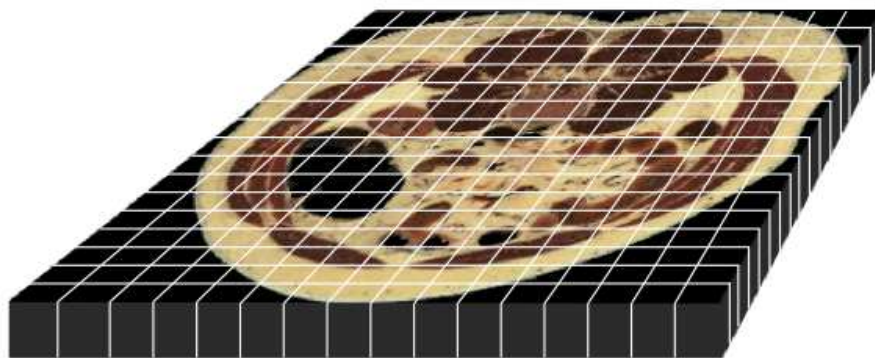
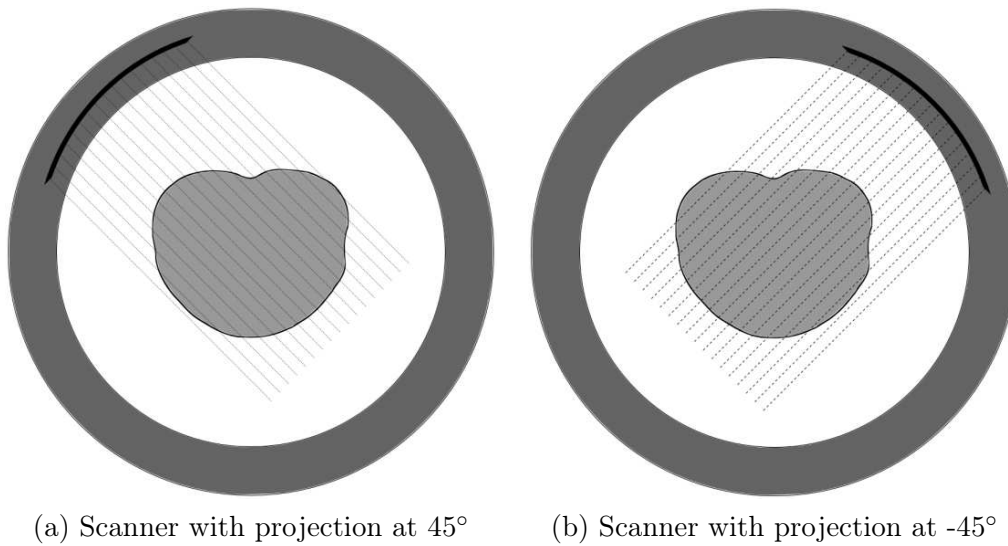


Figure 1.1: Picture of a scanner. There are two components, a table where the patient lies down and a hollow tube within which the CT images are acquired.

Hounsfield, received the Nobel Prize in Medicine in 1979. Concretely scanners are huge hollow tubes, where a patient lies down (fig. 1.1). Inside the inner surface of the tube moves a source of X-rays. Then the scanner registers the attenuation of the X-rays through the section of the patient that lies inside the tube. Finally, the registered attenuations are used to reconstruct the section of the patient inside the tube. The principle of the CT scanner will be first reviewed. Then, the Hounsfield Unit (HU) that measures the relative radiodensity of the voxels will be introduced.

A scanner measures the mean radiodensity for a set of rectangular cuboids that partition the scanned object. These rectangular cuboids are named voxels in 3D or pixels in 2D. To each voxel is given a mean radiodensity that is obtained by solving an inverse mathematical problem. The scanner indeed measures the cumulative attenuation of X-rays for diverse orientations of an X-ray source and for a same slice (fig. 1.2.a,b). Then, an image of this slice is obtained by retrieving the object that induced the observed cumulative attenuations. This task is named the tomographic reconstruction of the image. The tomographic reconstruction is an inverse mathematical problem where the observations are the attenuation data and the solution is the object that induced the observations. This problem is complex, even more for helical scanners where the observations are obtained for a helical trajectory and not all for a same slice. Thus, a quantization of the scanned object is introduced by dividing each slice in a number of voxels (fig. 1.2.c). The reconstruction is finally done by computing the mean radiodensity inside each voxel of this quantization.

The tomographic reconstruction provides a quantization of each slice of one



(c) Quantization of the abdomen that partitions a slice into a set of rectangular cuboid regions

Figure 1.2: Principle of the scanner. The absorption of X-rays is measured for a same slice of the body but for diverse directions (a,b). Then, the mean absorption inside a quantization of the scanned object (c) is computed by solving an inverse mathematical problem. The original image of the abdomen comes from the Visual Human Project<sup>®</sup> [NML 1989].

volume, where the mean radiodensity inside each voxel is known. Thus the reconstruction is expressed as an image volume where each voxel is given one intensity. This intensity is defined according to the mean radiodensity inside each voxel. However, a relative scale was introduced to replace the mean radiodensity as intensity value. Using this scale, each voxel has an intensity given in Hounsfield Unit (HU) that depends on the chemical composition inside each voxel (fig. 1.3). This scale is defined as a linear transformation of the measurement of the mean intensities that is defined by setting -1000 for the air and 0 for the water. These intensities then allow characterizing the tissues inside each voxel. However, these values are not absolute because of errors and differences between the patients (sect. 2.6).

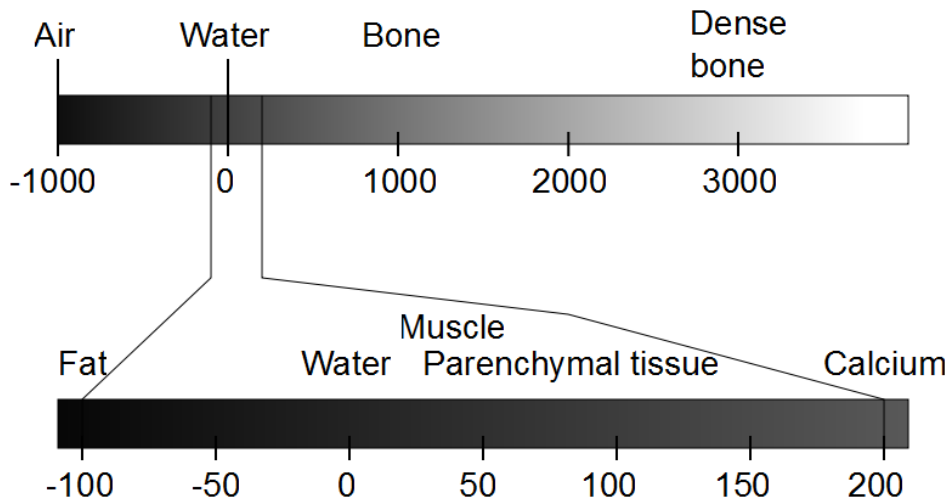


Figure 1.3: Hounsfield scale with some corresponding tissues. Values for air and water are the only stable values as they are used to set the Hounsfield scale. Other correspondences are given as rough estimates because they may vary between the exams.

### 1.3 Motivation

The segmentation of the liver tumors on CT images is achieved in a clinical prospect. Indeed, such segmentation has a significant value from a medical perspective. This work is nevertheless also relevant as a problem of Computer Vision. First, such segmentation is especially valuable when done for the CT modality. Then, this segmentation contributes to significant progresses for the physicians both for clinical routine and by opening new research possibilities. Finally, this segmentation is a complex problem of medical imaging that may benefit to other segmentation problems with texture objects in noisy images.

The segmentation of the liver tumors is achieved on CT images because this modality is the most common for the diagnosis of hepatic tumors [Hélénon 2003,



Choi 2006]. Indeed, while MRI is more relevant for some histological types, CT remains the best modality for the management of liver tumors in clinical routine. Thus, from a medical prospect, a segmentation of the liver tumors is more relevant when done on CT images. However, the proposed approaches are not limited to this specific modality. The algorithms developed during this thesis may indeed apply to any image modality as long as the image defines the intensity at each voxel with a single number.

This work is clinically motivated. Indeed, an automatic segmentation of the liver tumors would help physicians in clinical routine. Moreover, this work could open new research perspectives and could imply better therapeutic strategies and improved planning of several treatments. First, such segmentations would simplify and speed-up the work of physicians that cannot obtain segmentations of liver tumors without a long and tedious manual process. Then, an automatic segmentation method would allow for systematic delineation of the lesions. The ROIs for the lesions could affect many medical tasks next by offering new ways to manage these lesions. For example, these ROIs may be used to better evaluate the tumor lesions, or to plan several ablation techniques with the help of computers.

Segmentation of the liver tumors is not only valuable from a medical standpoint, but also for Computer Vision. Indeed, the proposed method applies to other segmentation problems in medical imaging. When segmenting the liver tumors, one has to deal with structures of high variability in a noisy medium where the structures differ by complex texture changes. Thus, the proposed method may contribute to other segmentation problems where the researched structures differ from other regions by complex texture changes. Moreover, due to the nature of the liver tumors, this segmentation should be robust to image noise and handle structures with complex and highly variable features.

## 1.4 Challenges

The segmentation of the liver tumors is a challenging task that is even complicated by the clinical prospect of this work. Indeed, the structures to segment vary widely and often a striking resemblance exists between tumors and other tissues. Moreover, the CT images are very noisy and also highly variable because enhancement is required to distinguish the lesions. Then, the clinical prospect of this study adds several constraints on the inputs and on the available resources for the segmentation process. Indeed, images from clinical routine should be handled by the segmentation method, and both the processing time and the available resources are limited.

From a Computer Vision standpoint, this segmentation is difficult. Indeed, segmentation has to be done in images with highly variable features, and for objects with miscellaneous appearances, and everything in a medium with a low signal noise ratio. First, the images to segment vary widely despite coming from a same modality. These variations may come from the characteristics that vary in clinical routine, meaning images from diverse CT machines, with various levels of resolution and for

both overlapping as well as non overlapping slices. The changes may also come from the enhancement of the images. Indeed, a contrast product is often injected to allow the detection of the tumor lesions. However, this injection impacts the acquired images that display diverse appearances and intensity ranges. Then, various tumors have to be segmented, with miscellaneous appearances within and between histological types, as well as with various appearances functions of the chosen enhancement. Moreover, often a striking resemblance exists between tumoral and healthy tissues. Finally, another difficulty comes from the liver itself. Being an organ with a high degree of vascularity, the images obtained are really noisy, a property that is amplified due to the enhancements.

The clinical prospect of the segmentation also imposes additional constraints on the resources available during the segmentation. First, the whole segmentation process should take no more than a few minutes in order to satisfy the speed standards of physicians. Indeed, speed is a critical parameter for the use in clinical routine. A segmentation that cannot be done in a reasonable time will be useless. However, the segmentation should remain accurate enough in order to be clinically relevant. Then, the segmentation process should not require unrealistic resources. Indeed, grid computing or computation on supercomputers cannot be considered for use inside hospitals. However, the available resources will not be precisely given. The whole process should nevertheless run on high-end personal computers.

## 1.5 Overview

This thesis will be divided in three chapters. A first chapter will precisely set the problem and explain the division of the segmentation of the liver tumors in two successive problems. A second chapter will then deal with the segmentation of a liver envelope. And the segmentation of the liver tumors will be finally presented in a last chapter.

The first chapter (sect. 2) introduces some medical background, and also defines and delimits the subject of this study. First, a medical introduction of the liver organ and the lesions to segment will be given, as well as some concise guidelines that are followed by radiologists. These brief reviews aim at a better understanding of the contribution of such segmentation, as well as an understanding of the environment where segmentation is done. Then, the framework of the study will be specified, before presenting the values of such segmentation. Indeed, the scope of the study will be stated; in particular the medical inputs will be detailed along with the constraints on the segmentation approach. Previous works will be reviewed next, followed by an overview of the segmentation process. In particular, the division in two steps will be justified. Indeed, the liver will be segmented first, and then the tumor lesions will be segmented within this envelope. Finally, the pre-process of the segmentation algorithm will be detailed.

The second chapter (sect. 3) deals with the segmentation of the liver on enhanced CT images of non-healthy patients. Three methods will be presented for

this problem, beginning with a simple morphological process, before introducing segmentations with prior knowledge. First, segmentation as an improvement of a partial segmentation will be reviewed. This first approach is a simple morphological treatment of a binary ROI that amounts to the computation of a smooth envelope that wraps around a mask. Then, a statistical atlas will be introduced as prior model of the liver. Using a statistical atlas, the segmentation will be achieved through the registration of this atlas and the deformation of a reference volume. Finally, additional priors will be added to the statistical atlas. An appearance prior and a registered spatial prior will indeed be introduced for the segmentation that will be obtained as the minimization of a MRF energy.

The last chapter (sect. 4) deals with the segmentation of the hepatic tumors within a liver envelope with normalized intensity. The proposed methods rely on advanced machine learning techniques for the separation of the tissues on a high-dimensional feature space. First, texture features will be combined to construct a classification function that distinguishes between the diverse types of tissues within the liver envelope. Then, a first segmentation of the liver tumors will be proposed simply by applying this classification function to one image. Finally, spatial constraints will be added during the segmentation and the reliability of the classification will be taken in account. Segmentation will indeed be achieved as a MRF minimization problem based on the probability of belonging to each class of tissue according to this previous classification function.

# Liver Tumor segmentation

---

*This chapter defines and delimits the subject of this study, namely the segmentation of the liver tumors, and introduces some medical background for a better understanding of the framework and the contribution of such segmentation. First, some medical background is given to introduce the liver organ and the lesions to segment, as well as some concise guidelines that are followed by radiologists. Then, the framework of the study will be specified, both for the medical inputs and the constraints on the segmentation approach, before presenting the values of such segmentation. Previous works will be reviewed next, followed by the presentation of the organization of the thesis. Finally, the pre-process of the segmentation algorithm will be detailed.*

---

**Contents**

<b>2.1</b>	<b>Intro</b>	<b>15</b>
2.1.1	Intro	15
2.1.2	Overview	15
<b>2.2</b>	<b>Medical background</b>	<b>16</b>
2.2.1	The liver	16
2.2.1.1	Anatomy	16
2.2.1.2	Role and regeneration	18
2.2.1.3	Pathologies	20
2.2.2	Injection protocols	21
2.2.3	Type of lesions	23
2.2.3.1	Incidence of liver cancers	23
2.2.3.2	Benign tumors	24
2.2.3.3	Primary malignant tumors	26
2.2.3.4	Metastases	29
2.2.3.5	Treatments	31
<b>2.3</b>	<b>Liver Tumor Segmentation, framework and purpose</b>	<b>34</b>
2.3.1	Setting the problem	34
2.3.1.1	Scope of the problem	35
2.3.1.2	Constraints	36
2.3.1.3	Myrian	36
2.3.2	Medical value	38
2.3.2.1	Immediate clinical value	38
2.3.2.2	New possibilities opened	39
<b>2.4</b>	<b>State of the art</b>	<b>40</b>
2.4.1	Review	40
2.4.2	Extracting some knowledge	42
<b>2.5</b>	<b>Overview of the proposed method</b>	<b>44</b>
<b>2.6</b>	<b>Pre-processing</b>	<b>45</b>
2.6.1	Normalization, working on a common basis	46
2.6.1.1	Intensity, a value not reliable enough	46
2.6.1.2	Existing normalization methods	46
2.6.1.3	Need of a more complex method	48
2.6.2	Retained method	48
2.6.2.1	Dynamic wrapping	49
2.6.2.2	Non linear histogram matching	50
2.6.2.3	Creation of reference	50

---

## 2.1 Intro

### 2.1.1 Intro

Liver cancers are deadly diseases with an important frequency in the world. Surgical resection is the best treatment available, but may apply only when some conditions on tumor sizes are met. Therefore early diagnosis and accurate appraisal of tumors are critical. The exceptional resolution of CT images allows good detection rates for most tumor types. However, the detection of liver tumors is challenging due to the small observable changes between healthy tissues and tumoral ones. Such a task is even challenging for clinical experts, where one can refer to an important volume variation (15-25%) between experts. Thus a good and repeatable method to detect the tumors would be an advantage. Being clinically motivated, such an approach should work for diverse tumor types at the same time, in order to avoid multiple and successive segmentations.

Such a task is quite difficult. First, various tumors have to be segmented, with miscellaneous appearances. Then, often a striking resemblance exists between tumoral and healthy tissues. Moreover, some tumors are only visible, or more distinguishable, when an appropriate phase has been considered. In general, these phases are roughly four, which correspond to diverse delays between the injection of a contrast product and the image acquisition. At the image level, the outcomes of these phases are diverse appearances and intensity ranges, both for the healthy tissues and the tumoral ones, along with shifting appearances for a same tumor from one phase to another. Another challenging problem is due to the liver itself. Being an organ with a high degree of vascularity, the images obtained are really noisy, a property that is amplified due to the enhancements. Finally, because of the focus on clinical use, the method should work for images in clinical routine, meaning images from diverse CT machines, with various levels of resolution and for both overlapping as well as non overlapping slices.

The segmentation of diverse tumor types will be achieved within the liver while satisfying clinical constraints. Indeed, this detection has to be done towards satisfying several constraints on the images handled by the process and on the resources required during the segmentation. First, images with diverse characteristics should be supported, like different resolution levels, various enhancement phases and protocols, and noisy anisotropic CT images. Then, the whole segmentation process should take no more than a few minutes in order to satisfy the speed standards of physicians.

### 2.1.2 Overview

This chapter will be organized as follows. First, anatomical, pathological and clinical data will be given. This information aims at offering a better understanding of the framework of this study and will be used for future references (sect. 2.2). Then, the scope of this study will be clarified as well as its contribution. A particular focus will be placed on the constraints due to the clinical purposes as well as on the

contributions, immediate or research motivated, of an automatic segmentation of liver tumors (sect. 2.3). Next, previous works for the segmentation of liver tumors will be reviewed (sect. 2.4). The segmentation process will be later introduced and justified. In particular, the segmentation of the liver before the segmentation of the lesions within this liver will be explained (sect. 2.5). Finally, the pre-process of the images will be presented, as a normalization of the images toward increasing the informative value of intensity (sect. 2.6).

## 2.2 Medical background

Some medical background will be introduced with a focus on the current facts. Future prospects or the contribution of segmentation of liver tumors for clinical use will not be mentioned in this section, but will be the topic of a subsequent section (sect. 2.3). First, the liver organ will be introduced. This review will focus on pathologies that are related with tumors or may impact the CT images, as well as on constraints coming from the CT modality. Then, injection protocols that allow enhancing lesions will be presented. Finally, the most common tumor types will be reviewed with a focus on etiology, characteristics and appearance on CT images.

### 2.2.1 The liver

The liver is the most voluminous solid organ inside human body (3% of body mass). This organ lays on the upper right part of the abdominal cavity, below the lungs and the heart, and to the right of the stomach, intestine and spleen. The liver anatomy will first be introduced. Then the numerous roles of the organ will be presented. Next a valuable characteristic of the liver will be detailed, regeneration. Finally, several liver pathologies will be reviewed.

#### 2.2.1.1 Anatomy

The liver is a highly variable organ that contains many vascular networks. The liver also has a highly variable shape that consists of several anatomical segments (8 or 9). As a soft organ, the liver shape indeed depends on the interactions with the other abdominal organs. Thus the organ variability is high. Moreover, this variability even increases with many types of pathologies.

The liver contains many vascular networks partially because of its specific location at the interface between the circulatory and the digestive systems (fig. 2.1). These numerous networks make liver a challenging organ for surgery and thus pre-operative planning is required before any liver operation. The liver indeed contains five vessel networks, three of which are blood networks. First, the liver receives blood from two blood networks that are very close or even interwoven, the portal vein and the hepatic artery. The former supplies the liver with nutrients coming from the intestine (shown in purple), and the latter provides the liver with oxygen (shown in red) (fig. 2.1). Then, the blood is drained from the liver by the hepatic

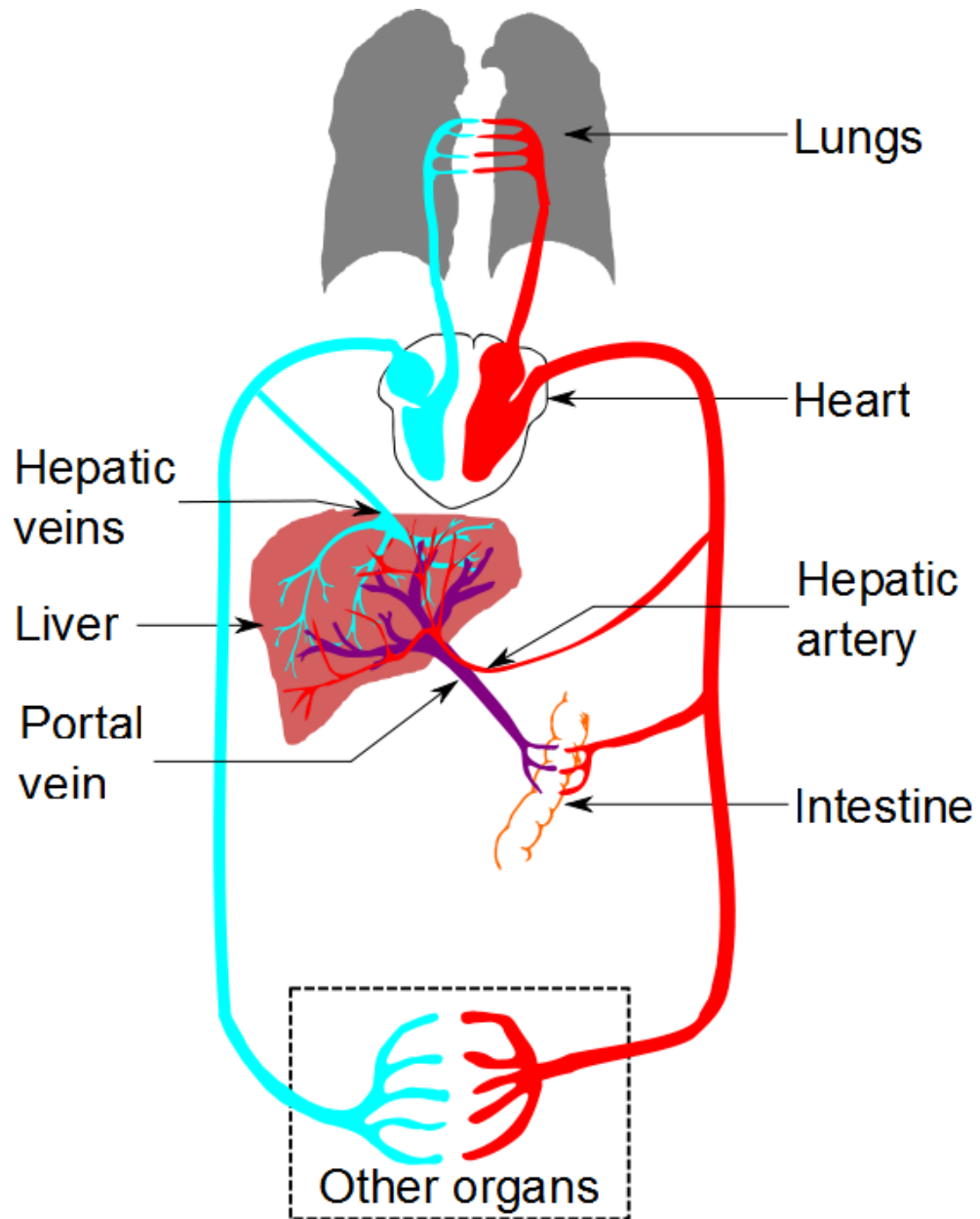


Figure 2.1: Human circulatory system with focus on the liver. The liver receives blood from two networks; the portal vein brings nutrients from the intestine and the hepatic artery supplies the liver with oxygen. Then, the hepatic veins drain de-oxygenated blood from the liver. Relative dimensions of the organs are not kept.



vein (light blue). These networks may be seen on a segmented example for a healthy liver (fig. 2.2). One may note the tree-like shape of the vessels, and their interwoven organization. The vessels are indeed trees as there are no loops inside each single network. Moreover, these networks are also interwoven, in particular the hepatic artery and the portal vein (fig. 2.2.a). Finally, a lymphatic network and a biliary network are also present inside the liver. The former does not impact the study as the lymphatic vessels remain invisible on CT images. On the opposite the biliary ducts are sometimes visible as hypodense regions. In particular, a dilatation of the intrahepatic bile ducts happens for several pathologies. This dilatation appears on the CT images because ducts become bigger than usual and thus more visible.

The liver is often divided in several (8 or 9) anatomical segments that amount to independent vascular regions. This anatomical segmentation was proposed by Couinaud in 1957 and still remains the medical standard [Couinaud 1957]. These anatomical segments are especially relevant for surgical resection in case of tumor lesions. These anatomical segments are indeed crucial when deciding for a cut liver while trying to limit the risk of cancer resurgence. In the subsequent parts these segments will be referred as anatomical segments in order to prevent any mix-up with segmentation as detection and labeling.

### 2.2.1.2 Role and regeneration

In addition or because of its specific location and anatomy, the liver is one of the organs that handle the highest number of functions. The liver has indeed three main types of roles for synthesis, breakdown of molecules and storage. First, the liver handles the synthesis of many molecules either by the metabolism of some or by the secretion of others. The liver indeed plays an important role in the transformation of the nutrients supplied by the digestive system. In particular, the liver performs several tasks for the carbohydrate metabolism and to a lesser extent for lipids. The liver also secretes some proteins such as coagulation factors, insulin growth factor, and bile. This latter is later stored inside the gallbladder through the biliary ducts. Then, the liver handles the breakdown of some molecules, in particular some drugs, hormones and toxins. For example liver plays a crucial role in the reduction of the blood alcohol level. Finally, the liver has also a storage role. Liver indeed stores many vitamins, glucose and iron. . .

Due to the diversity and the importance of the tasks handled by the liver, this organ is required for life; there are indeed neither natural nor artificial replacements such as dialysis for the kidneys. However, liver has a valuable characteristic that offsets this critical constraint for survival. Liver is indeed able to regenerate, which offers many possibilities of treatment. In particular the surgical resection of huge lesions is possible because the liver will grow again to replace the removed tissues. Furthermore living donor transplant is possible, i.e. a healthy person donates a part of his liver that is then transplanted to an ill recipient. Concerning the regeneration ability of the liver, a healthy volume of 20-30% is deemed sufficient to insure patient survival by keeping the hepatic function during the complete regeneration of the

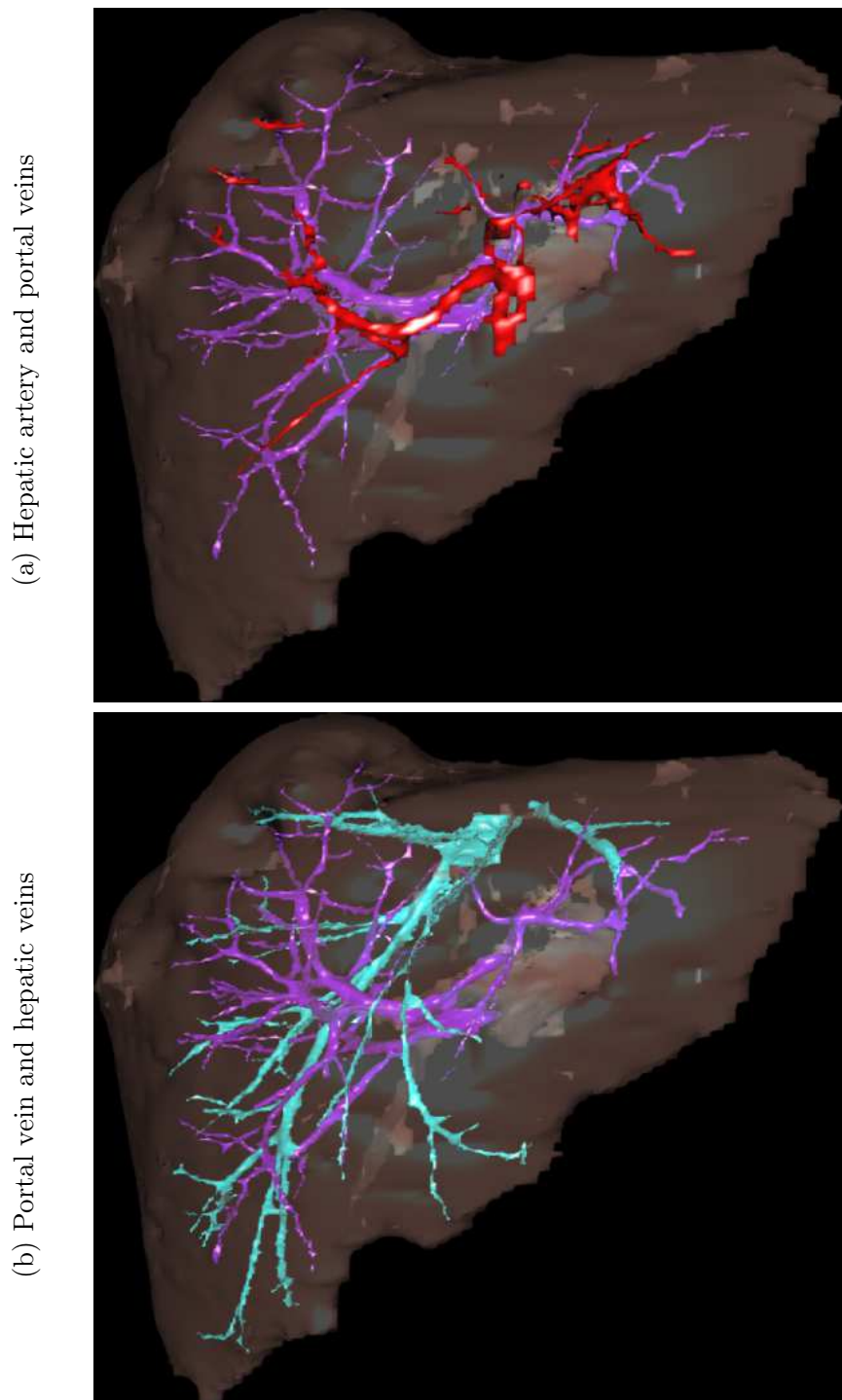


Figure 2.2: Anterior view of the blood networks inside the liver. The three blood networks are shown inside the liver. The portal veins are shown in purple (a,b), the hepatic artery in red (a) and the hepatic vein is displayed in light blue (b). Images are provided by CHU St Eloi. Segmentation and rendering were achieved with Myrian<sup>®</sup>.

liver [Fritsch 2003, Vilgrain 2002, H el enon 2003]. However, this threshold value is not yet accurately defined.

### 2.2.1.3 Pathologies

Several liver pathologies are found with diverse symptoms. In particular, these symptoms often include a jaundice caused by a dysfunction of the hepatic function, and an increase of the liver volume named hepatomegaly. Several of these diseases will be presented when they play a role related with the liver tumors. More precisely, pathologies with an impact on the CT images will be reviewed as well as diseases that may induce an oncogenesis. Several diseases indeed induce a malignant transformation, causing higher chances of tumor emergence. First, pathologies with an impact on the CT images will be presented, namely steatosis and hemochromatosis. Then, a short introduction of the liver tumors will be given, as this subject will be the topic of a subsequent section (sect. 2.2.3). Finally, cirrhosis and its causes will be reviewed because this chronic disease leads to several cancers. One may note that only CT modality is mentioned. This choice of modality is consistent with clinical routine, as the CT modality remains the best modality for the follow-up and the characterization of most lesions.

Hepatic steatosis or fatty liver is a chronic disease of the liver that modifies the liver appearance on CT images and may sometimes give tumors. First, fatty liver is a reversible pathology caused by an excess of triglyceride fat inside the liver cells. This pathology is the most common chronic disease of the liver; 31% of adults in USA indeed suffers from hepatic steatosis [Adams 2005]. Then, fatty liver induces important changes of the appearance of the liver on CT images. Indeed, the density of the liver on CT images decreases for fatty livers and heterogeneities may appear [Ernst 2005]. Moreover, some atypical appearances may be obtained for some lesions such as hemangiomas [Vilgrain 2000]. Finally, without treatment the hepatic steatosis may induce the appearance of cirrhosis in around 10% of cases that may later evolve with hepatocellular carcinomas [Sch utte 2009].

Hemochromatosis or iron overload is a disease caused by an excess of iron inside the liver. This pathology has mostly a genetic cause and induces a hyperdensity of the liver on CT images. Consequently many tumors will appear as hypodense on CT images and hypervascular lesions may disappear after injection [Ernst 2005]. This pathology might also evolve with cancers.

This thesis focuses on liver tumors, which will be detailed in a subsequent section (sect. 2.2.3). Various types of tumors may be found inside liver, either benign or malign. First, benign liver tumors are quite common; such lesions are indeed found in 1% of population [H el enon 2003]. The most common benign tumors are the hemangioma and cysts. Then, many malign lesions may be found inside the liver. These tumors are deadly; liver cancer is indeed the 4<sup>th</sup> deadliest cancer worldwide after lungs, stomach and not far from colorectal cancers [WHO 2009]. Several primary cancers may be found such as Hepatocellular Carcinoma (HCC) or Cholangiocarcinoma (CCC). Then, metastases are often found in the liver, because of its

location and its dual blood supply. The blood flow inside the liver is indeed second only to the lungs and its location favors the apparition of metastases from digestive cancers.

Several chronic liver diseases induce cirrhosis that causes important changes of appearance on CT images. Besides, cirrhosis may evolve in tumors. First, cirrhosis induces important changes of appearance and shape of the liver. Cirrhosis is indeed characterized by a fibrosis of the liver tissues and the apparition of scar tissues and regenerative nodules where healthy tissues are regenerated. Moreover, cirrhosis modifies the shape of the anatomical segments and impacts the CT images. Cirrhotic livers indeed appear with heterogeneities on CT images and with abnormal appearances for every phase. Then, cirrhosis is mostly caused by chronic diseases in particular the Hepatitis B and C (HBV and HCV) as well as chronic alcoholism [Ernst 2005, Gomaa 2008, Schütte 2009]. Finally, patients with cirrhosis are in a high risk population for HCC. Thus, cirrhotic livers are followed up by M.D. because 1-7% of cirrhosis cases leads to HCC each year [Merle 2005]. Moreover 70-90% of HCCs develop on cirrhotic livers, thus a follow-up of these patients contributes to better survival because lesions will be detected at an early stage [Blum 2005, Schütte 2009, Merle 2005].

Some infections such as hepatitis B and C viruses play a particular role for the liver, as they often induce an oncogenesis [Schütte 2009, Gomaa 2008]. HBV and HCV indeed induce chronic hepatitis, which first evolves into hepatitis-related cirrhosis and may then give cancer lesions. These two infections increase greatly the risk of liver cancer, and more specifically of HCC. Moreover, these viruses have a cumulative effect. First, HBV-related cirrhosis has a 2.7% probability of evolving into HCC per year [Gomaa 2008], which corresponds to an increase of the risk of HCC by a factor of 25-37[Lupberger 2007]. Then HCV has a similar effect. HCV evolves into chronic hepatitis after approximately 10-13 years, and then HCC appears 28-29 years later on average [Gomaa 2008]. Finally, cirrhosis induced by both HBV and HCV leads to an even greater increase of the risk of developing a HCC, which was computed at 45% after 10 years [Gomaa 2008].

### 2.2.2 Injection protocols

A contrast product is injected before the image acquisition for the diagnosis of most liver pathologies and in particular for the tumors. After injection, CT images are acquired after diverse time lengths in order to visualize the dynamic of the blood supply inside the tissues. These various images permit the detection of the tumors and their characterization. The value of these enhancements will be first illustrated with an example. Then, the reasons behind these injections will be explained. And finally, the injection protocols will be presented.

Enhancement is required for a good detection of the lesions first, and then for a correct diagnosis of the type of pathology. However, the most contributive enhancement depends on the type of lesion, and often several images are required for the characterization of one lesion. These enhancements are roughly four, no in-

jection, arterial phase, portal phase and delayed phase ordered by the time length between the injection of a contrast product and the image acquisition. The contribution and the dynamic of the enhancement are shown for a Hepatocellular Carcinoma (HCC) (fig. 2.3). The lesion remains invisible without the injection of a contrast product (fig. 2.3.a). Then, the center of the lesion slightly appears in arterial phase (fig. 2.3.b). Finally, marked boundaries appear in portal phase (fig. 2.3.c). For this example, segmentation and follow-up should be done in portal phase, because neither no injection nor arterial phase provide sufficient contrast for a good delineation between tumors and healthy tissues.

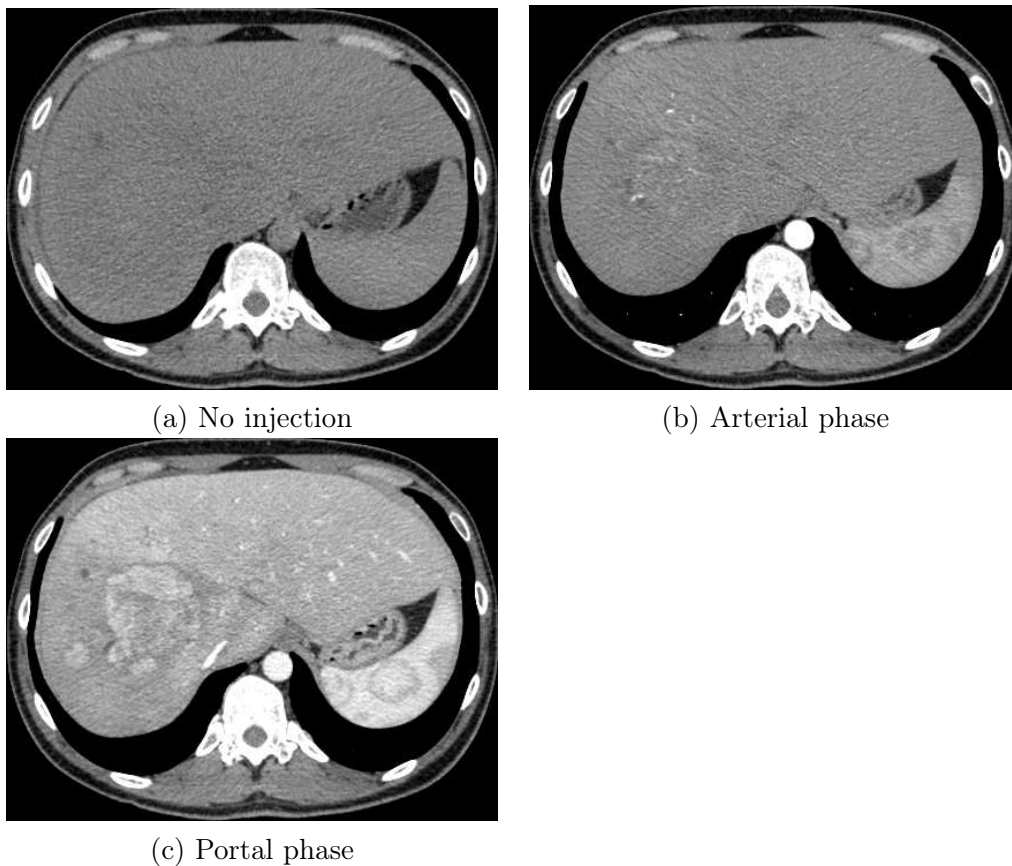


Figure 2.3: Contribution of enhancement for a HCC. A same axial slice is shown before the injection of a contrast product (a) and then for diverse enhancements, arterial phase (b) and portal phase (c). Images are courtesy from Beaujon's hospital.

The contribution of the injection of a contrast product comes from the difference of the blood supply between healthy and non healthy tissues. Tissues are indeed supplied in oxygen by both portal vein and hepatic artery. However, the balance of the supply between the two networks varies between healthy and tumoral tissues. Thus, the origin of the oxygen used by the tissues is a good characterization of the tissues. For example Matsui *et al.* showed that diagnosis of HCC against Adenomatous Hyperplasia (AH) could be done using the blood supply, as portal blood supply

was seen in 96% of cases of HA and only 6% of HCC [Matsui 1991]. More generally, HCCs are mostly dependent from the hepatic artery for blood supply, while healthy tissues depend mostly on the portal vein [Clark 2005, Bruix 2005, Ernst 2005].

Several enhancement protocols have been developed in order to enhance the tissues, functions of the origin of their blood supply. A single injection of a contrast product is done on the inner arm and images are acquired after diverse time lengths. These acquisitions allow enhancing tissues supplied by different blood networks because there is a time gap between the arrival of the contrast product inside the arterial and the portal networks [Ernst 2005]. Thus the contrast product passes through the various networks after different time lengths. The aorta and the tissues supplied by the hepatic artery will indeed be enhanced first, then tissues supplied by the portal vein and finally the liver will progressively enhance (fig. 2.4).

Four main enhancement phases are commonly mentioned, which are related to diverse time lengths after the injection of a contrast product. These phases are, in chronological order, no injection, hepatic arterial phase (HAP), portal venous phase (PVP), and delayed phase. However, these phases do not match with well defined protocols as there are neither worldwide nor national injection protocols for the liver. These four phases allow visualizing the dynamic of the blood supply inside the tissues. Without injection only abnormal structures appear such as calcifications or necrotic regions. Then, the arterial and the portal phases provide information on the origin of the blood supply. Finally, the delayed phase informs about the persistence of the enhancement. All phases are not useful in each case, but on the whole these phases are sufficient for the detection of lesions and their characterization when the CT images are combined with other medical analyses.

### 2.2.3 Type of lesions

#### 2.2.3.1 Incidence of liver cancers

Liver cancers are deadly, with a heterogeneous distribution of types and incidences worldwide. Moreover, these cancers have a high incidence that is even expected to increase. First, liver cancers are the third or the fourth deadliest cancers worldwide after lungs, stomach and close from colorectal cancers [WHO 2009, Merle 2005]. Moreover, liver cancers are of bad prognosis. For example HCC has less than 4-6% 5-year survival, and this survival rate did not improve significantly during the last decades [Clark 2005, Tang 2001]. Then, both the incidence and the type of liver cancer vary through the world. Africa and Asia are indeed regions of high incidence, while Western countries have lower incidence rates (fig. 2.5). Furthermore, liver cancers are mainly primary ones in the former countries, whereas they come mostly from metastases in the latter. Finally, the incidence of these cancers is high, is increasing and is expected to continue to increase. Liver cancers have a high incidence and a high mortality; in China HCC is for example the second cancer killer with 34.7/100000 deaths and the fifth worldwide for incidence [Tang 2001, Merle 2005]. Moreover, the incidence of HCC increased almost twofold for many countries dur-

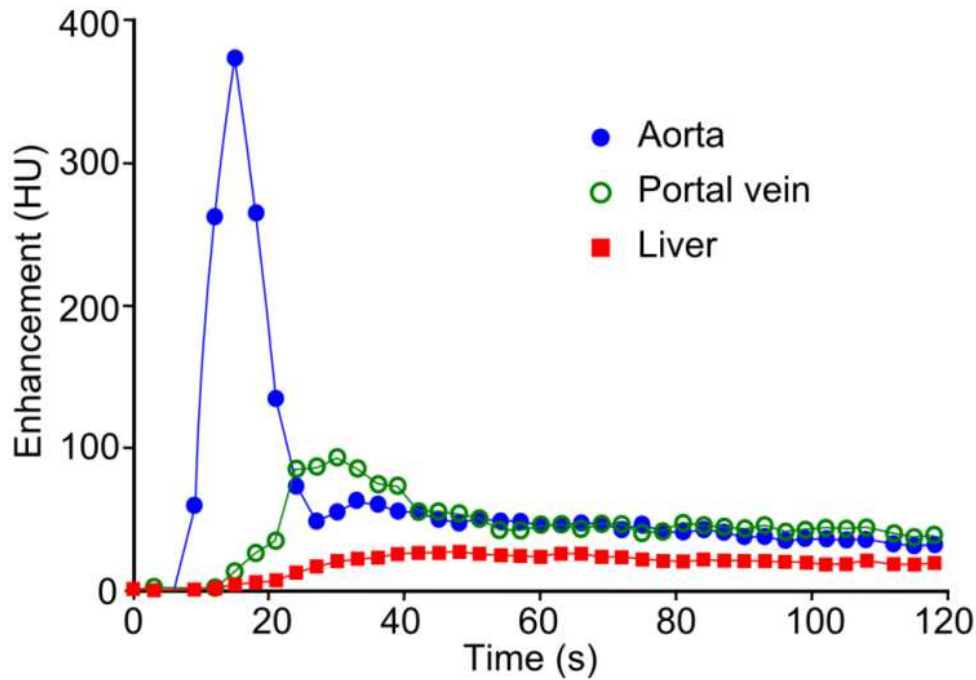


Figure 2.4: Enhancement of some anatomical structures, functions of the time length after injection. The enhancement is given in Hounsfield Unit (HU) for the aorta, the portal vein and the liver. The graph is taken from Blum’s book [Blum 2002].

ing the last two decades [Clark 2005, Merle 2005]. Besides, this incidence is also expected to increase due to the high number of persons with either HBV or HCV (resp. 300 and 200 millions); in particular an increase of the incidence of HCC of 150% for males and 200% for females is forecast in France [Merle 2005].

The liver tumors will now be presented as well as existing treatments for liver cancers. This presentation will not aim at providing complete guidelines for the detection and the characterization of the lesions. Indeed, this subject could and is already the topic of whole books. The most common lesions will be introduced by giving the basic epidemiology of each type as well as a brief presentation of their appearances on CT images. First, benign lesions will be briefly reviewed. Then, malignant tumors will be detailed beginning with the primary ones, and followed by the metastases. Finally, several treatments will be briefly presented with a focus on the planning of these treatments, which could benefit from an automatic segmentation of the lesions.

### 2.2.3.2 Benign tumors

Benign tumors will now be presented. These tumors are not directly valuable for this study, as their segmentation is not valuable from a medical perspective. Indeed, no treatment is required for most of them, apart from some exceptional cases. Thus, the segmentation of these lesions offers no medical benefits. The characteristics of

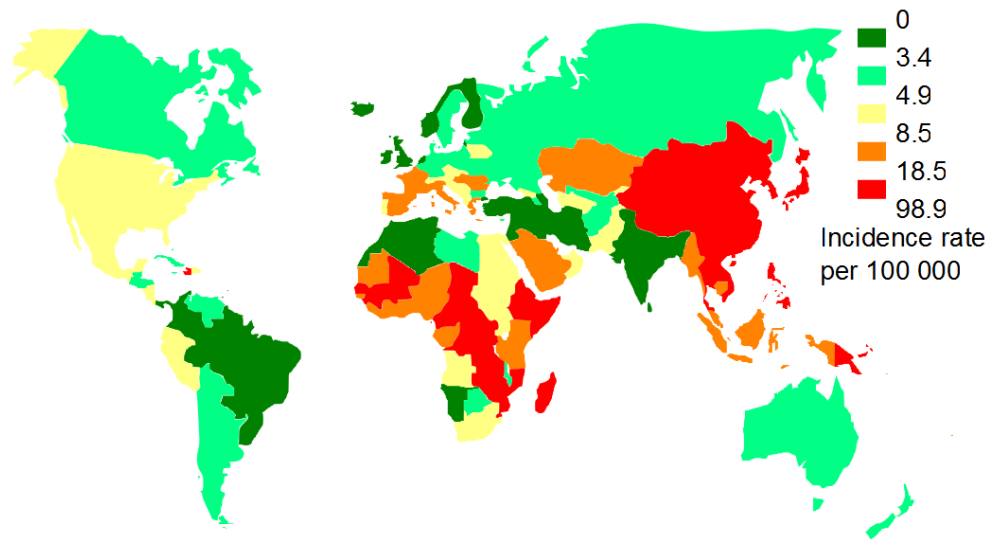


Figure 2.5: Worldwide incidence of liver cancers for males in 2002. Data was provided by CANCERmondial (IARC-WHO) [GLOBOCAN 2002].

these lesions are nevertheless useful to distinguish them from malignant lesions. Many types of benign liver tumors exist but only the three most common will be presented, Hemangioma, biliary Cysts and focal nodular hyperplasia (FNH).

### Hemangioma

Hemangioma is the most common benign tumor that affect 1-20% or 3-10% of the population depending on the authors [Vilgrain 2000, H el enon 2003]. This lesion is most of the time asymptomatic and no treatment is required. However, some huge lesions may induce symptoms and in these cases surgical resection of the lesions should be done. Diagnosis of these lesions is done on CT images for lesions bigger than 2cm; otherwise MRI is the best modality. When this condition on size is met, CT provides good sensitivity and specificity, 67% sensitivity and 99% specificity according to Nino-Murcia *et al.* [Nino-Murcia 2000, H el enon 2003]. The lesions first appear hypodense on non-enhanced CT images. Then a peripheral enhancement appears on HAP scans and a uniform enhancement of the lesion is obtained with a centripetal filling in PVP. This enhancement finally persists on delayed phase [Vilgrain 2000, H el enon 2003].

### Biliary Cysts

Biliary Cysts are also common lesions that affect 2-4% of population. These lesions are benign, asymptomatic, and with no possible evolution as malignant lesions. Thus, these tumors are often excluded from studies because of their lack of medical value. On CT images, biliary Cysts appear as spherical homogeneous and marked lesions that do not enhance whatever the enhancement phase [H el enon 2003,



Ernst 2005].

### Focal Nodular Hyperplasia

Focal Nodular Hyperplasia (FNH) is a rare liver tumor that affect 0.3-1% of population [Hél non 2003]. This type of tumor is asymptomatic and with no complications [Ernst 2005]. CT imaging is nevertheless valuable for the characterization of the type of the lesions. CT images indeed contribute to the characterization of FNH versus Adenoma, the latter being a malignant tumor [Van Hoe 1997]. On CT images the lesions are hypodense or isodense homogeneous without enhancement and only huge lesions are visible. Then, the lesions become quickly hyperdense on HAP scans, but this enhancement does not persist in PVP and delayed phase, where the lesions become isodense homogeneous [Ernst 2005, H l non 2003].

#### 2.2.3.3 Primary malignant tumors

The three main types of liver primary cancer will be reviewed with a focus on Hepatocellular Carcinoma. HCC is indeed the most common primary cancer of the liver by a wide margin. However, two other types of tumors are worth mentioning, Intrahepatic Cholangiocarcinoma (ICC) and Angiosarcoma. Several other types of liver cancers also exist, but are very rare.

### Hepatocellular Carcinoma

Hepatocellular Carcinoma (HCC) is the most common primary liver cancer, with a poor prognosis, as well as an important and increasing incidence. However, this cancer often develops on livers with existing pathologies, which allows following up high risk patients for an early detection of the lesions. The epidemiology of this cancer will be given first. Then, the oncogenesis of this cancer will be partially reviewed as well as its medical consequences. Very simple guidelines for the detection of HCC on CT images will be given next. And finally the current treatments and prognosis of HCC will be reviewed.

HCC is a deadly cancer with a high incidence that is increasing and is expected to continue to increase. First, HCC has a high incidence throughout the world, but with wide discrepancies between the regions. HCC is indeed the fifth most common cancer worldwide. However, the incidence varies of several magnitudes between different countries. In China HCC is for example the second cancer killer with 34.7/100000 deaths, while HCC is only the eighth one in France with an incidence of around 6.2/100000 [Tang 2001, Merle 2005, Binder-Foucard 2007]. Then, the incidence of HCC increased almost twofold in the last two decades in many countries [Clark 2005, Merle 2005, Clark 2005, Binder-Foucard 2007]. And this incidence is also expected to increase again due to the high number of persons with either HBV or HCV (resp. 300 and 200 millions); in particular an increase of the incidence of HCCs of 150% for males and 200% for females is forecast in France [Merle 2005, Clark 2005, Bruix 2005, El Serag 1999].

HCC often develops on a liver with an existing pathology, which allows following up high risk patients for an early detection of HCC. Yuki *et al.* indeed showed that 92.5% of patients with HCC had an existing pathology, either cirrhosis or chronic liver diseases and that only 7.5% had no major changes of the liver parenchyma [Yuki 1990]. Thus the high risk population is followed up at regular intervals to detect the appearance of HCC at an early stage. This follow-up is commonly done with 6 months intervals, but some physicians also propose 4 or 12 months [Bruix 2005, H el enon 2003]. Several pathologies induce an oncogenesis inside liver, in particular chronic alcoholism and infections such as HBV or HCV [Sch utte 2009, Goma a 2008]. HBC and HCV indeed induce a chronic hepatitis that gives hepatitis-related cirrhosis, which multiplies the risk of HCC by a factor of 25-37 [Lupberger 2007]. HBV-related cirrhosis indeed has a 2.7% probability of evolving into HCC per year, and HCC appears on average 28-29 years after HCV-related cirrhosis [Goma a 2008]. Furthermore, these two viruses have a cumulative effect and the risk of developing a HCC when having both HBV and HCV was evaluated at 45% after 10 years [Goma a 2008].

Enhancement pattern		Cases (%)
Arterial phase	Portal/Delayed phase	
Hyperattenuation	Hypoattenuation	26
Hyperattenuation	Isoattenuation	6
Peripheral hyperattenuation	Hypoattenuation	1
Peripheral hyperattenuation	Isoattenuation	1
Mixed attenuation	Hypoattenuation	13
Isoattenuation	Hypoattenuation	7
Isoattenuation	Isoattenuation	9
Hypoattenuation	Hypoattenuation	32
Hypoattenuation	Isoattenuation	3

Figure 2.6: Appearance of HCC on CT images for arterial and portal/delayed phases. Data is taken from Honda’s study [Honda 1993].

First detection of HCC is often done with ultrasounds, but CT imaging offers the best sensitivity for the detection and the characterization of HCC [H el enon 2003]. The main arterial supply of HCC indeed allows good detection and characterization of HCC on enhanced images; for example a lesion >2cm with characteristic arterial vascularization is a HCC at 95% [Bruix 2005]. First, without enhancement HCC appears hypodense most of the time, but may also sometimes appears as isodense with rare calcifications. Then, the lesions become heterogeneous in 40-60% cases with numerous possible features after injection of a contrast product. However, the lesions are hyperdense for 90% cases on HAP scans. A washout happens next during PVP and delayed phase for the lesions bigger >2cm, i.e. a marked decrease of the enhancement of the lesion. However, 25% of smaller lesions (<2cm) remain enhanced [Bruix 2005, H el enon 2003]. The actual enhancement dynamic is never-

theless slightly more complex. The dynamic and the appearance for each phase indeed vary widely (fig. 2.6).

HCC has a poor prognosis, with few treatments offering long term survival. HCC has indeed a poor prognosis with only 5% 5-year survival [Clark 2005, H el enon 2003]. However, when detected at an early stage and under some conditions, long term survival is possible. The choice of a treatment is nevertheless complex, requiring many physicians from various domains. Moreover the optimal treatment strategy is still under debate. The only treatments with a curative value are surgical resection, liver transplantation and percutaneous ablation. First, surgical resection is the best treatment for long term survival, but only 20% patients are candidates. This treatment offers a better outcome with 50% 5-year survival, however the recurrence rate is high (more than 70%). Then, liver transplantation is the second choice when surgical resection is not possible. Because the underlying disease is removed in the process, transplantation gives better results. A 5-year survival of 70% is indeed obtained for well chosen patients and with a smaller recurrence rate than for simple resection (15%). Moreover, the regenerative ability of the liver allows for transplantation from living donors to compensate for the lack of available organs. However, a small mortality exists for the donor (0.3-0.5%). Finally, several methods for percutaneous ablation have been developed for small lesions with no surgical option. Ordered by use, one may mention Percutaneous Ethanol Injection (PEI) and Radiofrequency Thermal Ablation (RFTA). When no curative option is available, palliative approaches are chosen. In particular, Transcatheter Arterial Embolization (TAE) and Transarterial Chemoembolization (TACE) induce a necrosis of lesions for more than 50% patients. The usual approaches to treat cancer are nevertheless almost useless. Chemotherapy and radiotherapy are indeed not proven to improve survival [Clark 2005, Bruix 2005, Blum 2005].

### Other types

Intrahepatic Cholangiocarcinoma (ICC) is the second most common primary liver cancer, which accounts for 5-30% of the primary ones. ICC is deadly as surgery is the only curative treatment. Moreover, surgery is available for few patients, and with poor prognosis; the 5-year survival rate is indeed of 0-39%. On CT images, diagnosis and detection are mostly done on the delayed phase. First, the lesions appear hypodense without enhancement. Then, the appearance of the ICC becomes quite random after injection. Indeed, a slight and incomplete enhancement of the lesion periphery is often seen on HAP or PVP scans. However, this enhancement is also sometimes marked and continuous or homogeneous. There may even be no enhancement or an enhancement of the entire lesion. On the opposite, the lesions are often well marked in delayed phase, where a medium to high enhancement is obtained for 74% cases after 10-20 minutes [H el enon 2003, Yalcin 2004].

Angiosarcomas are rare primary tumors of the liver, with a low incidence and a poor prognosis. Angiosarcomas indeed have a low incidence  $0.14-0.25/10^6$  and a short mean survival of 6 months. First, lesions appear hypodense without en-

hancement, with possible hemorrhages. Then, the lesions quickly become hyperdense after enhancement, with either a heterogeneous or a homogeneous enhancement [Hélénon 2003].

#### 2.2.3.4 Metastases

##### Epidemiology

The relative incidences of liver metastases compared to primary tumors vary widely throughout the world. However, these lesions remain quite common whatever the region due to the particular location of the liver. First, liver metastases are the most common types of liver tumors in Europe and USA, while primary ones are in China and Africa [Hélénon 2003]. Then, liver is the second biggest metastatic site after the lungs [Hélénon 2003]. Liver has indeed a double blood supply that drains a high blood flow, second only to the lungs. Moreover, the liver is the first organ affected by digestive metastases, because liver is the first organ after the intestine. These two facts contribute to making the liver an organ with many metastases. Actually most solid tumors may give hepatic metastases, in particular colorectal cancers, adenocarcinoma from pancreas and stomach, bronchial cancer, endocrine and breast cancers [Ernst 2005].

##### CT imaging

CT imaging remains the best modality for liver metastases, even if MRI offers better results in some cases. CT scans indeed offer exceptional resolution and good accuracy for the liver metastases. For example, sensitivities of respectively 69-70% and 73% have been shown for metastases and colorectal ones, as well as specificities of 86-91% and 96.5% [Choi 2006, Hélénon 2003]. The imaging of hepatic metastases answers several roles: the evaluation of suspected lesions, preoperative planning, the monitoring of treatment, and post-treatment follow-up [Choi 2006]. For this review, only the detection of the lesions will be presented as this step is common to all tasks. However, the characterization of one lesion will not be detailed. The differential diagnosis against several other types of tumor is indeed difficult for liver metastases. This difficulty is nevertheless irrelevant, as this study does not deal with the characterization of the tumors but their segmentation. Thus this difficulty will not be further developed [Ernst 2005, Hélénon 2003].

The primary site of the cancer impacts the appearance of the liver metastases. The appearance patterns indeed vary functions of the histological type of the lesion (fig. 2.7). The liver metastases keep several features from the primary cancer, which induce various appearance patterns. However, these appearance patterns are not sufficient to define the histological type of one metastasis [Hélénon 2003].

CT images for metastases are mostly acquired without delayed phase. Some authors even deny the contribution of CT scans without enhancement. The metastases will be divided in two classes, hypervascular and hypovascular, depending on their vascularization. Without enhancement, the lesions are almost never visible,

Enhancement pattern	Colon	Pancreas	Carcinoid	Breast	Others
Homogeneous	0	0	20	20	13
Abnormal internal vessels	0	0	20	0	0
Peripheral puddles	0	0	0	0	7
Complete ring	82	82	60	40	80
Incomplete ring	0	0	0	20	0
No enhancement	18	18	0	20	0

Figure 2.7: Appearance of several metastases in arterial phase on CT images, functions of the primary site. The possible appearances of four lesions are given along with the repartition of these appearances as percentages. Data was extracted from a paper by Nino-Murcia *et al.* [Nino-Murcia 2000].

but this phase allows the detection of some lesions that disappear after injection. First, hypovascular lesions are iso- to hypodense without enhancement, but remain often indistinguishable. This phase is nevertheless useful for the detection of calcifications, necrosis or hemorrhages. Then, some rare histological types of hypervascular lesions may appear hypodense. HAP is the optimal detection phase for hypervascular lesions. First, hypovascular lesions are hypodense, sometimes with a hyperdense rim, but the appearance persists in PVP and the contrast even increases. Then, hypervascular lesions are enhanced homogeneously with sometimes a hypodense boundary. HAP is the best phase for their detection, because the enhancement then decreases quickly until the lesions become isodense. PVP is the phase with the best sensitivity for the hypovascular tumors. First, hypovascular lesions appear as hypodense with sometimes a hyperdense rim. For this phase the sensitivity of detection of these hypovascular lesions is high; according to Kuszyk this sensitivity is 91% for lesions  $>1\text{cm}$ , but this measure decreases to 0-56% for smaller lesions. Then, often PVP does not contribute to the evaluation of hypervascular lesions, apart from a few ones that are isodense in HAP and visible only in PVP [Kuszyk 1996, Ernst 2005, H el enon 2003].

## Treatment

The treatment of metastases is highly dependent from the primary site and from the global state of the patient. The choice of a treatment is indeed dependent from the presence or the absence of other metastases in the body. Furthermore, the stage of the cancer in the primary site has a strong impact on the choice of a therapeutic strategy. Depending on the global health of the patient several therapeutic strategies may be considered. First, curative treatments may be chosen when available and if they offer some benefits for the survival of the patient. Then, many local treatments have been developed to treat lesions that are not candidates for surgery. Finally, the physician may choose to avoid treatment or to treat only the symptoms, when the patient will not or cannot benefit from the treatment of the metastases. For example, this choice might be done when the primary cancer is too advanced for

being cured, or when the number of metastatic sites is too high.

Surgical resection is the only curative treatment for metastases. However, only a fraction of the patients are candidates for surgery and even surgery offers low survival rates. Surgical resection is indeed available for less than 10% of patients. Moreover, survival after the resection is lower than for HCCs; depending on the histology and the extension of the lesions, the 5-year survival rate is only 20-40% [Hélénon 2003].

Several local methods have been developed for treatment when surgical resection is not available. Most are still evaluated and offer benefits only when the metastatic activity is limited to the liver and the lesions are small. The three main approaches for liver metastases are cryosurgery, Transarterial Chemoembolization (TACE) and Percutaneous Ethanol Injection (PEI). First, cryosurgery is an approach that showed some benefits for colorectal lesions. This treatment induces a necrosis of the lesions through cryotherapy, without destroying the healthy parenchyma or the vessels. Then, Transarterial Chemoembolization (TACE) showed similar results for colorectal metastases; the 2-year survival rate indeed increased. This treatment combines a localized embolization of the arteries with a local concentration of chemotherapy that leads to an anoxia of the lesion and an improved efficiency of the drugs. Finally, Percutaneous Ethanol Injection (PEI) is sometimes used, but is less efficient than for HCC. This treatment has nevertheless showed some effect on small endocrine lesions [Hélénon 2003, Choi 2006].

### 2.2.3.5 Treatments

The main available treatments will now be quickly reviewed. This review aims at providing some background information for a better understanding of the possible contributions of a segmentation of liver tumors, as well as the value of computer science from a medical prospect. The possible therapeutic strategies will be summarized by presenting several techniques, their efficiency and in particular whether the treatments are curative or palliative. The treatments will be divided in three main classes, depending on their therapeutic contribution. First, surgical approaches will be presented. Surgery is the only approach that offers long term survival. Then, minimally invasive approaches will be reviewed. These treatments offer lesser results than surgery, but are still sometimes efficient. Finally, diverse methods will be reviewed with mostly palliative effects.

#### Surgery

Surgical treatments are the only curative approaches. However, many patients are not candidates for such approaches. Moreover, surgery of the liver is a complex operation with significant mortality and morbidity. Thus, accurate selection of patients and preoperative planning are crucial [Hélénon 2003]. Surgical resection will be presented first. Then transplantation will be reviewed, from either a dead donor or a living one.

When available, surgical resection is the best treatment for most liver tumors.

However, surgical resection is a complex operation with high morbidity and mortality (1-3% mortality) [Bruix 2005]. Thus, the selection of the candidates, as well as preoperative planning are two crucial steps to avoid unnecessary operations and to decrease the risk of the operations. First, the selection of the candidates for surgical resection is based on many parameters, including several that are evaluated using imaging modalities. The lesions should indeed be detected and characterized in order to define the sizes and the number of lesions as well as their locations. Furthermore, the remaining volume after the resection should be evaluated in order to insure a sufficient hepatic activity during the regeneration of the liver. This minimal volume required after resection is not precisely defined, but is commonly set around 20-30%. Then, preoperative planning also benefits from the imaging. Images indeed allow defining in advance the cut liver and to prepare the operation. Preoperative planning also relies on two other segmentations. First, the blood vessels should be segmented to reduce the hemorrhages during the operation. Then, the anatomical segments of the liver should be known to define the regions to remove. Tumors are indeed removed by the ablation of complete anatomical segments to prevent cancer recurrence. The removal of whole anatomical segments indeed removes tissues that are supplied with blood in contact with one lesion. Thus the chance of missing tumoral cells is reduced, because the blood drains these cells inside the same anatomical segment [Hélénon 2003, Couinaud 1999].

Transplantation is the second choice of therapeutic strategy. However, many patients are not candidates for transplantation, and there is a shortage of organs. Fortunately, the regenerative ability of the liver allows for transplantation from a living donor. First, transplantation offers better survival rates than surgical resection for primary cancers, in particular when an underlying pathology exists. Moreover, good chances of long term survival are also expected for secondary cancers, in particular when there are no others tumors outside liver. Then, imaging plays an important role for the transplantation, by helping planning the operation and verifying the compatibility of the vascular networks between the donor and the recipient [Couinaud 1999]. Selection and planning are even more crucial when using a living donor. Indeed, the donor mortality cannot be ignored (0.3-0.5%) and complications happen for 20-40% of the donors. Thus, the compatibility of the graft with the anatomy of the recipient should be verified beforehand [Bruix 2005, Hélénon 2003].

### **Percutaneous ablation**

Several treatments by percutaneous ablation have been developed for patients that are not candidates for surgical resection, in particular for patients with multiple lesions of small size. These approaches are minimally invasive and are based on the local destruction of the lesions. The ablation of one lesion is indeed achieved without open surgery by guiding some tools through the skin with ultrasounds or CT until they are in contact with the lesions. These percutaneous approaches are not yet well evaluated, but offer lower results than surgical treatments. For these therapies, imaging plays an important role during the treatment and also after. Indeed, the le-

sions should be evaluated after treatment with CT or MRI [Sitruk 2004, Choi 2006]. Five therapeutic strategies will be reviewed, one chemical first (PEI), and then four thermal ones. The review of the thermal techniques will begin with the three based on heat, RFTA, MWA and LiTT. These techniques induce an immediate necrosis of lesions with a high local temperature (60°C). Then, the last thermal technique will be presented, cryosurgery. This last technique kills the cells by the cold (-20 - 30°C) [Sitruk 2004].

Percutaneous Ethanol Injection (PEI) uses ethanol to destroy the tumors. This technique is used mostly for cirrhotic liver with HCC. PEI is indeed the best percutaneous approach for small HCC lesions (<2cm) [Bruix 2005]. For PEI, a needle is first inserted up to a lesion. Then, ethanol is dropped through the needle, which induces a necrosis of 2-3cm around the needle. However, PEI is limited to only one injection at a time.

Radiofrequency Thermal Ablation (RFTA) relies on a single or multiple electrodes to deliver heat locally inside a lesion. RFTA is the most common percutaneous thermal technique and offers promising results. Similar effects on HCC are indeed obtained for PEI and RFTA, but RFTA is more efficient for lesions >2cm. This technique induces the necrosis of a small region around each electrode, but larger lesions may be treated by using multiple deliveries with overlaps, which was not possible with PEI. However, RFTA is inefficient in the neighborhood of vessels. The heat source in RFTA is indeed very small, thus most of the necrosis is due to the conductivity of heat. Consequently a heatsink effect appears around the vessels that act as coolant, and necrosis becomes incomplete [Bruix 2005, Sitruk 2004, Simon 2005, Clark 2005].

Two newer techniques of thermal ablation have been developed that offer similar results to RFTA: Microwave Ablation (MWA) and Laser-induced Interstitial Thermotherapy (LiTT). First, MWA replaces the heating source by microwaves. This change allows for better ablation around the vessels compared to RFTA. However, MWA is rarely used because of its additional price [Simon 2005]. Then, LiTT replaces the electrodes of RFTA by optic fibers that emit a powerful beam. LiTT offers two main improvements compared to RFTA; multiple optic fibers may be used at same time and MRI becomes available as a guide. The use of multiple optic fibers allows necrosis of regions of 6-7cm, and the use of MRI offers better resolution for the guidance and the location of the fibers, without the medical consequences of X-rays [Sitruk 2004].

Cryosurgery is similar to the other thermal techniques, with constraints similar to that of RFTA; only the cause of ablation changes. Cryosurgery indeed allows destructing small lesions by inducing a necrosis of the lesions through cryotherapy instead of heating. This technique showed some benefits for colorectal lesions [Hélénon 2003].



### Other treatments

The remaining therapeutic choices are palliative and aim at decreasing the symptoms or increasing the survival of patients. First, the usual therapeutic approaches will be reviewed, radiotherapy and chemotherapy. Then, two techniques based on arterial embolization will be reviewed, TAE and TACE.

Radiotherapy and systemic chemotherapy do not show any substantial success. First, radiotherapy is rarely used for liver tumors and even seldom mentioned. However, some authors believe that 3D conformal radiotherapy might have some value [Tang 2001]. Then, systemic chemotherapy has only a palliative effect, but also allows minimizing the symptoms [Clark 2005, Tang 2001].

Transcatheter Arterial Embolization (TAE) and Transarterial Chemoembolization (TACE) are two techniques based on the embolization of the arterial supply to induce an anoxia of the tumoral cells. However, these approaches apply only to lesions with a main arterial supply. For both treatments, the arterial blood supply is blocked as close from the lesion as possible. This blocking is achieved through a minimally invasive approach by passing a catheter inside the vessels. Then, the removal of the arterial supply has very different effects on healthy and tumoral tissues due to the difference between their blood supplies. An anoxia of the lesions should indeed appear when the main supplier of oxygen disappears, which might lead to a necrosis of the lesions. On the opposite, this technique should have no effect on healthy tissues, as they still receive oxygen through the portal vessels. First, TAE is only the embolization previously described. This technique does not offer significant results in average [Bruix 2005]. Then, TAE was improved by the addition of a local chemotherapy to the embolization. This evolution of the technique, named TACE, is still mostly palliative, but showed some effects for not far advanced HCCs. The addition of a local chemotherapy is an improvement compared to systemic chemotherapy, because the proximity with the lesion allows for higher concentration of chemicals [Vogl 2006, Clark 2005, Tang 2001].

## 2.3 Liver Tumor Segmentation, framework and purpose

The scope of this thesis will now be precisely defined, as well as the purpose of this study. First, problematic of this thesis will be explicitly set, by giving the scope of this problem and the constraints to respect. Then, the contribution of this study will be given, both for immediate clinical use and possible later improvements.

### 2.3.1 Setting the problem

Automatic segmentation of the liver tumors on CT images is achieved in a clinical prospect and as a plug-in for a medical imaging program. The clinical prospect imposes several constraints on the segmentation method. First, the inputs of the segmentation methods are constrained. Because of the routine clinical use, the method should indeed handle images with various features. However, the clinical

use also offers some information due to the role of the physician in the segmentation process. Then, the results of the method are constrained and hardware requirements are added by the clinical purpose of the study. Finally, this study extends the possibilities of an existing program for medical imaging, Myrian<sup>®</sup> that will be briefly introduced.

### 2.3.1.1 Scope of the problem

The segmentation of the liver tumors is achieved on enhanced CT images in a clinical prospect. This prospect imposes several constraints but also offers some information that defines the scope of the problem. The segmentation should indeed work on CT images coming from the clinical world, but for which the optimal enhancement phase is known.

Liver tumors should be segmented on CT images acquired in clinical conditions, meaning with diverse scanning machines, various sizes of voxel, diverse enhancement protocols, and for the main tumor types. To apply in a clinical environment the segmentation technique should not be limited to specific image characteristics, but should apply to varied image characteristics. First, the method should handle various scanning machines. Different machines indeed provide images with different features due to the tomographic reconstruction and the machine itself. Secondly, images with different sizes of voxel exhibit different features that should be dealt with. In particular, the level of noise and partial volumes effects vary significantly between 5mm and inframillimetric slice thicknesses. A maximal slice thickness of 5mm is a reasonable value, as it begins to provide sufficient resolution for the segmentation to be significant and is still widely used in clinical routine. Then, the segmentation technique should handle various enhancement protocols, either as different phases or as the differences of enhancement for a same phase between different hospitals. The tumors may indeed be segmented on several enhancement phases. Besides, no standard enhancement protocol exists. Thus, a same phase induces different enhancement levels and patterns for different hospitals. Finally, the segmentation should deal with several types of tumor at once in order to avoid multiple specialized segmentations. The segmentation of uncommon tumor types may nevertheless be ignored in order to avoid a gain for rare tumors at the cost of a decrease of the overall quality of segmentation.

The segmentation will be done on a single CT image while assuming that an adequate enhancement phase is selected. First, a single CT image for one phase will be used for the segmentation. Segmentation could take advantage of multiple enhancement phases, but multi phase segmentation will not be considered in this study. Then, segmentation will be assumed to be done on an adequate phase. Indeed, often a phase exists, where the tumors are more distinguishable. The segmentation will be assumed to be done for this phase. This assumption is reasonable, because physicians can select the best phase and sometimes already know this phase before the CT acquisition. The selection of an optimal phase is indeed nonsignificant compared to the time length required for segmentation. Besides, the optimal phase

is sometimes known prior to the CT acquisition. In particular, for follow-up the tumor type does not change and thus neither does the optimal enhancement phase.

### 2.3.1.2 Constraints

The clinical prospect of this study implies several constraints on the quality and the speed of the segmentation as well as on hardware requirements. First, the quality of the segmentation should be good for lesions  $>1\text{cm}$ . In order to be medically relevant, the segmentation technique should indeed provide sufficient results for lesions  $>1\text{cm}$ . Smaller lesions do not present a clinical value, and their segmentation is thus less valuable. Lesions  $<1\text{cm}$  are indeed excluded from medical tasks that require measurement of lesions, whether it be RECIST or WHO response criteria [Therasse 2000, Eisenhauer 2009, Miller 1981]. From a medical perspective, the detection of the lesions is also particularly significant. A lesion imperfectly segmented may indeed be manually corrected, whereas a lesion missed during the segmentation could have more harmful consequences. Then, the segmentation should be done in a few minutes. Being clinically motivated, the segmentation technique should indeed give results in a time frame compatible with the clinical use of the segmentation tool. Finally, the segmentation should run on reasonable CPU and memory resources. Because, the technique is intended for clinical use, grid computing and excessive memory requirements are prohibited. No explicit definition of the available resources will be given, but the whole technique should run on high-end personal computers.

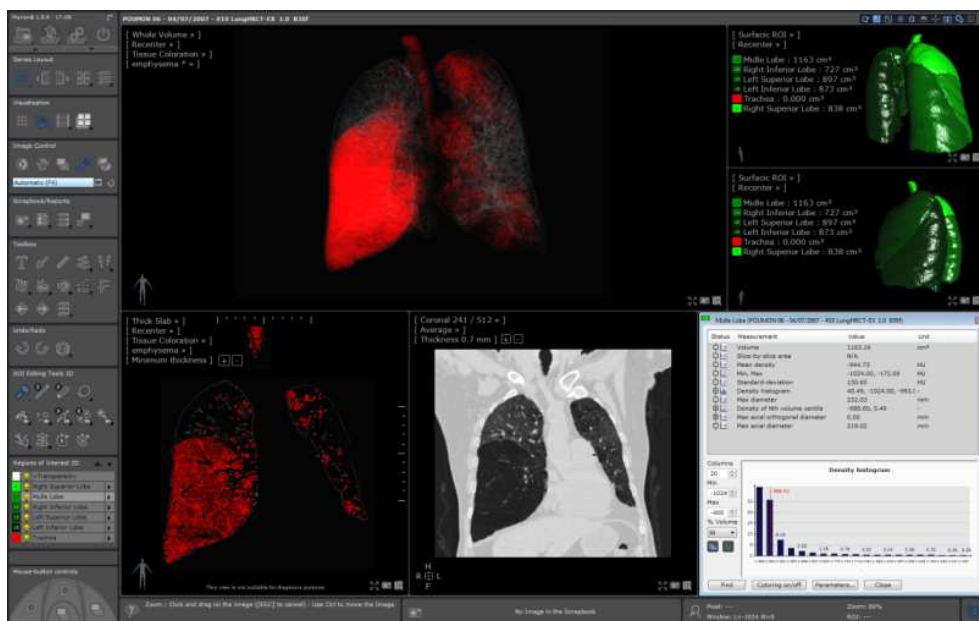
### 2.3.1.3 Myrian

Segmentation of the liver tumors is developed within a program for review of medical images and aided diagnosis, Myrian<sup>®</sup>. This software is dedicated to the clinical domain, either for radiologists, surgeons or clinicians. Myrian<sup>®</sup> offers several tools to help the physicians. In particular, two tools are especially relevant for this study, the visualization abilities and the segmentation engines. First, Myrian<sup>®</sup> displays the three medical planes (axial, sagittal, and coronal) and offers several 3D rendering views (fig. 2.8). Then, several tools are proposed for segmentation, either generic or specialized. Several generic tools are indeed available to speed-up manual segmentations or for correction purposes. Then, several expert engines already exist that handle specific problems such as the segmentation of the lungs prior to the volumetric evaluation of emphysema (fig. 2.8.b).

The value of Myrian<sup>®</sup> will be illustrated with a virtual hepatectomy, i.e. the planning of a surgical resection due to a liver tumor (fig. 2.8.b). The relevant structures for this operation are first segmented, namely hepatic and portals veins as well as the lesion and the liver. Then, the surgeon simulates the cut of the liver required for the removal of the diseased anatomical segments. The physician decides next if the operation is feasible by considering the resection volume and the remaining volume after resection. Finally, a 3D display of the liver with the



(a) Virtual Hepatectomy



(b) Lungs with emphysema

Figure 2.8: Clinical use of Myrian<sup>®</sup> for lungs and liver. A virtual hepatectomy is first shown, i.e. the planning of the surgical resection of a part of the liver (a). Then, the evaluation of emphysema inside lungs is presented (b).

intrahepatic vessels helps the surgeon to prepare his operation. Such a view indeed allows him to prepare for the difficulties brought by the vessels to cut.

### 2.3.2 Medical value

The contribution of an automatic segmentation of liver tumors from a medical perspective will now be detailed. The contributions will be reviewed while distinguishing the expected contributions from possible values at a longer term. The contributions will also be divided depending on the clinical task they are useful to. As said before, the imaging of liver lesions answers several roles: the evaluation of suspected lesions, the choice of a therapeutic strategy, the planning of treatment, and post-treatment follow-up (sect. 2.2.3.4). However, the evaluation of a lesion and post-treatment follow-up involve similar tasks. Thus, both tasks will be treated at same time. Finally, the contributions of an automatic segmentation of liver tumors will be detailed as follows. First, clinical as well as research possibilities opened by such segmentations will be detailed. Then, the expected contributions for each clinical task will be reviewed.

#### 2.3.2.1 Immediate clinical value

The contribution of an automatic segmentation of the liver tumors for the evaluation of a lesion before or after treatment is obvious. First, an automatic detection will help physicians to detect the lesions, in particular when the M.D. is not an expert radiologist of the liver. However, the study does not aim at replacing the M.D. but instead aims at helping him during the detection process. Then, a segmentation of the lesions directly defines the size of each lesion. In particular, such segmentations will provide directly the highest diameters of each lesion for the WHO and RECIST metrics. Moreover, the reliability of these measures will improve as the measures will be repeatable and should even be more precise than manual ones. Finally, a segmentation of the lesions offers a better repeatability than manual segmentations that may vary widely between M.D.

An automatic segmentation of the liver tumors also simplifies the choice of a therapeutic strategy. A main constraint for the availability of surgical resection indeed comes from the ratio tumor/liver after surgical resection (sect. 2.2.3.5). With a segmentation of the liver, the segmentation of the tumors allows easier computation of this ratio, and thus simplifies the choice of surgical resection as therapeutic strategy.

Segmented liver tumors have also several applications for treatment planning, in particular for surgical resection and arterial embolization. First, the segmentation contributes to preoperative planning. Because lesions are well defined, the definition of the regions to remove becomes easier. In particular, the segmentation provides a precise location of the lesions inside the anatomical segments of the liver, which will simplify the definition of the cut liver. Moreover, a 3D view of the liver with segmented tumors is a precious help for the surgeon, in particular when the blood

networks are also segmented. Then, TAE and TACE also benefit from this segmentation. Combined with a segmentation of the hepatic vessels, a segmentation of the liver tumors indeed allows planning the path to follow to reach the lesion.

### 2.3.2.2 New possibilities opened

A segmentation of the liver tumor opens new possibilities for the evaluation of the lesions. The solid tumors are indeed still evaluated using the one or two highest dimensions of a lesion in a single slice. However, these evaluation criteria are not sufficient according to several physicians, and are kept only because of the lack of methods to segment lesions in 3D. Moreover, these criteria do not take in account the necrosis of the lesions, while many treatments aim at necrosis (sect. 2.2.3.5) [Eisenhauer 2009, Miller 1981]. An automatic segmentation of liver tumor allows solving these two problems. Once a ROI is obtained, the evaluation of the lesions with other measures indeed becomes easy; in particular volumes and density histograms are directly available. The evaluation of the necrotic regions should also be simple because necrosis is often well marked. Thus the active region of each lesion becomes available, which offers several new possibilities to evaluate the lesions and the response to a treatment.

The choice of a therapeutic strategy could also gain from a segmentation of the liver tumors. First, the choice of surgical resection is dependent from the volume of healthy tissues after resection. However the minimal volume is not well defined. An automatic segmentation of the lesions might allow a better definition of this minimal volume. Moreover, one may imagine defining this volume, functions of the lesion, the volume of the liver. . . Then, such segmentations might improve the choice between percutaneous ablation techniques. Most techniques indeed have an optimal zone of destruction, with sometimes constraints around vessels. Thus, the ablation zones might be modeled for each technique and virtually applied to each patient in order to define the optimal technique depending on the shape and the location of each lesion.

The planning of treatment could gain a lot from a segmentation of liver tumors. Many therapeutic strategies could indeed gain for such segmentation, PEI, thermal percutaneous ablation, radiotherapy. First, an optimal position of the needle might be computed along with an optimal path of insertion for PEI. A position for optimal coverage of one lesion might indeed be computed, when knowing the necrosis pattern induced around the needle and a ROI for the lesion. Then, a ROI of the lesions offers many possibilities for percutaneous thermal ablation using heat (RFTA, LiTT, MWA). With these techniques, many lesions are indeed treated using multiple locations of the heating sources. Using a ROI and heating models, an optimal configuration of the heating sources might be automatically computed. Finally, 3D conformal radiotherapy might be evaluated for liver tumors once a ROI of the lesions is known.

## 2.4 State of the art

The segmentation of the liver tumors on CT images is a complex task due to the variability of the appearances of the lesions either between various tumor types or within a single type. Moreover, the segmentation has to be done on images with a low signal noise ratio. Therefore this task is still an open problem that various authors attempted to solve. Several approaches have been proposed for this segmentation and will be reviewed in a first section. Then, some conclusions will be extracted from these works.

### 2.4.1 Review

While recent methods often rely on multiple techniques to achieve an accurate segmentation of the tumors, only four different approaches are sufficient to define the main part of most proposed methods. First, the segmentation may be done using a threshold. Then, a region growing approach may be retained or active contours next. Finally, segmentation may take advantage of machine learning techniques. Historically, the first methods were often generic methods for segmentation in images with a low signal noise ratio. Then, methods have tended to add more and more prior knowledge on the lesions, for example as intensity distributions or appearance models. Many of these more evolved methods were proposed for the Liver Tumor Segmentation challenge 2008, which was a part of a workshop with MICCAI 2008 [Xiang 2008]. Thus, a short introduction to this challenge will be given.

Liver Tumor Segmentation challenge 2008 is a competition aiming at comparing the accuracy of segmentation of several algorithms for liver tumors [Xiang 2008]. Two datasets were provided for this competition, a training test with reference segmentations to tune the algorithm, and a test set without references (initially a third set was also used for qualification purposes, but is no longer available). The quality of each algorithm was finally assessed using a score defined as a weighting of several metrics, including the overlap error. The two data sets consist of 10 tumors from 4 patients for training, and 10 tumors from 5 patients for test. CT images were acquired on three different scanners, one 64-slice and two 40-slice using a standard four-phase contrast enhanced imaging protocol. The axial in-plane resolution of the CT images is 0.6-0.9mm and their slice thickness is 1mm or 1.5mm. All reference segmentations were manually done by an experienced radiologist, and confirmed by another radiologist. The comparison of the most interesting algorithms proposed during this context is given in the next section 2.9.

Historically, segmentation of the liver tumors was first achieved by thresholding. While basic, thresholding is still used with additional processes. However, such an approach is limited to simple metastatic lesions, and even sometimes only to hypodense ones. The first method of segmentation of the liver lesions was proposed by Soler *et al.* with a threshold value as part of an entire process for hepatic surgery [Soler 2001]. Then, the detection of tumors was handled using thresholding by several authors. Seo used simple thresholding [Seo 2005]. Park defined an op-

timal threshold [Park 2005]. Ciecholewski later introduced histogram equalization in order to simplify the choice of a threshold value [Ciecholewski 2007]. Indeed, the research of an optimal threshold is not trivial, thus several approaches have been proposed. Nugroho chooses a threshold that maximizes the between class variance [Nugroho 2008]. Choudhary relies on cross entropy for thresholding, before applying level set smoothing [Choudhary 2008]. Moltz defines a threshold using the quantiles of the intensity distribution inside a region given by a user, before applying adaptive thresholding to remove vessels connected to the lesion [Moltz 2008].

Region growing approaches have also been proposed in several semi automatic methods. These methods rely on an initial input by a user, who initializes a seed region for a region growing algorithm. These approaches have two main drawbacks. The proposed methods are strongly dependent from the initial seed, and do not apply to lesions that are not well marked. For example Wong proceeds to segmentation with a region growing algorithm based on intensity, for which exists a strong constraint on the relative sizes of the seed region and the whole lesion [Wong 2008]. Qi also proposed a region growing algorithm, but based on Bayesian decision rule with pdf models for the tumors that are updated during the algorithm [Qi 2008].

Active surfaces were also proposed to segment liver lesions that show a marked difference of intensity range compared to the liver parenchyma. Ben-Dan and Fernandez both propose a segmentation through active surfaces, with a Mumford-Shah energy expressed as an optimal separation between two regions with different mean intensities [Ben-Dan 2008, Fernández-de Manuel 2009, Mumford 1989]. However, both methods differ by the solving strategies and the space where the active surfaces are defined. Ben-Dan indeed proceeds to segmentation in a probability space, where probabilities are defined by a joint-likelihood ratio that models the tumor intensities [Ben-Dan 2008]. On the opposite, Fernandez expresses the energy with a level-set and carries out the segmentation on the CT image [Fernández-de Manuel 2009].

Due to the variability of the tumors and the difficulty to define thresholds, machine learning techniques have been proposed to achieve accurate segmentations of the lesions. These machine learning techniques have been used for two main purposes, either to cluster similar voxels inside an image, or to define prior models for the segmentation. First, machine learning techniques were retained to cluster similar voxels in an image. Massoptier uses k-means to classify the lesions against the parenchyma inside a liver envelope [Massoptier 2008]. Taieb classifies the voxels inside a ROI through Bayesian classification before finishing the segmentation with active contour [Taieb 2008]. Häme relies on fuzzy clustering with a spatial smoothing term, before a segmentation with deformable models [Häme 2008]. Then several approaches learn models that are then applied for segmentation, either online for Zhou, or offline for Li and Shimizu [Zhou 2008, Li 2006, Shimizu 2008]. First, Zhou proposed a semi automatic technique that propagates an initial segmentation on one slice using SVM that are updated after each segmented slice [Zhou 2008]. Li trained AdaBoost classifiers to locate boundaries on linear intensity profiles coming from a seed inside a lesion, before segmenting the lesions as a maximal score path on a polar representation of the image [Li 2006]. Shimizu proposed a segmentation



of the lesions as a labeling of voxels with an AdaBoost classifier that discriminates through a number of features including textural ones [Shimizu 2008]. This last approach is the most similar to the one developed in this study.

Two proposed methods rely on other semi automatic approaches. First, Smeets proceeds to segmentation in two steps. Smeets begins with the research of a minimal cost path on a spiral representation of the image from a seed inside a lesion. The segmentation is then achieved using level-sets with a speed dependent from a fuzzy classification [Smeets 2008]. Second, Jolly proposes a three steps segmentation that begins with an estimation of gray level distributions from a user input. Then, fuzzy connectedness segmentation is achieved in 2D, which relates to the research of paths with minimal intensity ranges. Finally, a 3D segmentation is obtained through random walkers [Jolly 2008].

### 2.4.2 Extracting some knowledge

Several conclusions may be extracted from these previous works. First, the choice of an automatic method should not damage the quality of segmentation compared to semi-automatic methods. Then, intensity alone does not seem sufficient for the segmentation of lesions other than metastases. Next, a liver envelope seems required for the segmentation of the liver tumors, in particular for automatic approaches. Finally, machine learning techniques fit well in the liver case.

Semi-automatic methods do not bring any improvement compared to automatic methods. On the opposite, the best method is an automatic one. The best results are indeed obtained using an automatic method (fig. 2.9). Then, the five best methods are equally divided between automatic and semi-automatic approaches. Furthermore, these five methods offer very similar results (fig. 2.9). Thus an automatic method presents two main advantages compared to a semi automatic one. On the one hand, such an approach does not require any user interaction and thus satisfies the speed constraint, as the segmentation may be done while the physician is not in front of the computer. On the other hand, this choice should not impair the quality of the segmentation.

Intensity alone is not sufficient to distinguish every type of lesion; texture features are required. Indeed, while many metastases and some primary tumors present distinct intensity ranges compared to healthy liver, many primary tumors may be distinguished only by a change of texture. This difference is found in the scientific literature as most methods dealing with more than liver metastases rely on texture instead of intensity as discrimination method (fig. 2.9)(fig. 2.10). For example, Smutek shows that texture features allow distinguishing between HCC and Cysts, and Smeets notes that texture features would offer more than intensity alone [Smutek 2006, Smeets 2008]. Moreover, the use of texture features seems to improve the overall segmentation (fig. 2.9)[Shimizu 2008].

According to previous works, the segmentation is best achieved when limited inside the liver, in particular for automatic segmentation. First, segmentation is often achieved inside a liver envelope (fig. 2.9)(fig. 2.10). Then, the best results are

	Automatic method	Segmentation inside liver	Machine Learning	Tumor types	Features for segmentation	Overlap Error (%)
[Shimizu 2008]	X	X	X	meta (other claimed)	texture	28.98
[Zhou 2008]		X	X	meta	intensity	30.02
[Moltz 2008]				meta	intensity	30.55
[Nugroho 2008]	X	X		meta	intensity	31.21
[Choudhary 2008]		X	X	hypo-dense	intensity	32.14
[Smeets 2008]		X		meta	intensity	34.6
[Wong 2008]				meta	intensity	39.4
[Taieb 2008]				meta	intensity	39.48
[Qi 2008]				meta	intensity	42.1
[Häme 2008]	X	X	X	meta	intensity	47.33
[Ben-Dan 2008]	X	X		meta	intensity	49.12
[Kubota 2008]	X			meta	intensity	53.72
[Soler 2001]	X	X		hypo-dense	intensity	-
[Jolly 2008]				-	intensity	-
[Li 2006]			X	-	intensity	-
[Massoptier 2008]	X	X	X	hypo-dense	intensity	-

Figure 2.9: Segmentation methods for the liver tumors, ordered by increasing overlap error. The overlap errors are given for a same set of images that was used for MICCAI segmentation challenge [Xiang 2008]. Some information was not available in the papers and will be shown by a dash.

	Detection inside liver	Machine Learning	Tumor types	Features for segmentation	Detection accuracy (%)
[Park 2005]	X		-	intensity	0.912
[Seo 2005]	X		-	intensity	0.858
[Ciecholewski 2007]	X		Hemangioma, Hepatoma	intensity	0.783
[Shimizu 2005]	X	X	HCC, Cyst	texture	seed detection
[Smutek 2006]		X	HCC, Cyst	texture	characteriza- tion

Figure 2.10: Detection and characterization methods for the liver tumors. Some information was not available in the papers and will be shown by a dash.

obtained when working inside a liver envelope. Among the five best segmentation methods, four are indeed limited to the liver, and the last one is limited to a ROI centered on the lesion that is given by the user (fig. 2.9). Finally, the only automatic technique that is not limited to the liver is also the worst one (fig. 2.9).

Machine learning contributes significantly to the segmentation of the liver tumors, in particular when lesions are distinguished thanks to texture changes. First, machine learning techniques are often used to segment the liver tumors. In particular, three of the five best methods use machine learning techniques, including the two best ones (fig. 2.9). Then, machine learning is especially relevant when using texture features, because the selection and the combination of these features are not straightforward. Machine learning techniques are indeed used in all methods that rely on texture features (fig. 2.9)(fig. 2.10).

## 2.5 Overview of the proposed method

A method of segmentation for the liver tumors on CT images is proposed. This approach consists of two successive segmentations, first the liver and then the tumors inside the liver. Both segmentation steps begin with a normalization that aims at defining a common intensity basis for the segmentation. First, a liver envelope is segmented after normalization of the entire image (sect. 3). Then, the liver tumors are segmented inside the liver envelope after normalization of the intensities within the liver envelope (sect. 4).

The intensities inside a CT image are normalized twice in order to define a relevant intensity basis at two different points of the segmentation algorithm. First, these normalizations aim at defining a common intensity basis that will simplify the automatic segmentations by increasing the relevance of the intensity value. The raw intensities inside one CT image are indeed of small value because the variations

of intensity due to external constraints may be bigger than the differences between the tissues to segment (sect. 2.6). Then, normalization is applied twice in order to better fit to each specific segmentation problem. The first normalization is indeed applied to the entire image in order to roughly obtain a stable intensity basis. A precise intensity basis is difficult to obtain at this point because the CT images may contain various anatomical regions such as entire body, abdomen/pelvis or thorax/abdomen. On the opposite, the second normalization is applied inside a well defined organ, thus the intensity normalization might be better done. Moreover, this second normalization is more crucial to avoid complicating an already complex problem.

The segmentation of the liver tumors begins with the segmentation of a liver envelope. This preliminary segmentation has two main roles; it improves the segmentation of the tumors and contributes to the diagnosis. First, automatic segmentation of the liver tumors was shown to be more accurate when limited inside the liver (sect. 2.4.2). This restriction indeed simplifies the segmentation problem by decreasing the number of possible tissues during the segmentation process. Then, a segmentation of the liver has a medical value for the treatment of tumor lesions (sect. 2.2.3.5). Both liver and lesions are indeed relevant for the choice of therapeutic strategies and preoperative planning...

The liver tumors are finally segmented inside the liver envelope using texture features and machine learning. The proposed segmentation approach deals with the segmentation of anatomical structures that differ by their texture appearance, inside images with a low signal noise ratio. Thus, the proposed approach may apply to other organs or lesions. First, the segmentation process is limited to a liver envelope in order to increase the accuracy of the segmentation. Then, intensity alone is not sufficient for the segmentation of any type of tumor lesions, thus texture features are retained to detect tumoral tissues (sect. 2.4.2). Finally, a significant contribution of machine learning was pointed by several authors, in particular when using texture features (sect. 2.4.2). Thus, state of the art machine learning techniques were retained to distinguish tumor tissues from other tissues inside the liver with an optimal speed.

## 2.6 Pre-processing

As a first step intensities inside the images are normalized using nonlinear histogram matching. This normalization allows obtaining a good match between intensities inside different images. Thus the intensity becomes a more reliable measure for the characterization of the tissues, as similar tissues will have similar intensities. This normalization was achieved by modifying and applying a work done by Turetsky and Ellis for the alignment of audio streams [Turetsky 2003].

## 2.6.1 Normalization, working on a common basis

### 2.6.1.1 Intensity, a value not reliable enough

CT images give a single intensity value at each voxel of the image, which characterizes the tissue inside a rectangular cuboid. However, the absolute CT intensities should not be used for the characterization of the tissues, as intensities for a same tissue vary functions of the patients and the machines. To better understand the value of intensity, its physical meaning will be first reviewed. Then, the causes of the unreliability of intensity will be detailed.

CT scanners reconstruct images as matrices of voxels, where each voxel measures the radiodensity inside the body, i.e. the mean attenuation of the X-rays inside a rectangular cuboid within the body. These mean attenuations are mainly dependent on the chemical composition of these rectangular cuboids as well as the spectrum of the X-ray source and the patient. Thus, these attenuations reflect the type of the tissues, which are given as Hounsfield Unit (HU) for the CT modality.

For CT images, intensity is not reliable enough to directly characterize a tissue, as a significant variability exists between machines and between patients, which may exceed the difference between the tissues [Levi 1982, Birnbaum 2007]. Moreover, the lack of a standard injection protocol introduces an additional variability due to the changes between the enhancements. First, the intensity varies between manufacturers' CT scanners and even between generations of a same CT scanner [Birnbaum 2007]. For example this last paper showed that the variability of the attenuation for a liver phantom could go up to 63.4 HU for a same scanner. Then, the orientation and the position of the patient inside the scanner impact the intensities inside the final image. Thus intensity will vary with time and between patients as identical positions cannot be obtained [Levi 1982]. Moreover, the size of the patient impacts the results as the volume of tissues the X-ray have to go through will vary. Finally, a contrast product is often injected to enhance the tumors inside liver. This enhancement will vary between the patients and the hospitals as there is no standard protocol. Thus a same tissue will take different HU depending on the characteristics of the injection. This difference is encountered for a same phase between different hospitals, and is even more important when different phases are considered. Consequently, the absolute intensity cannot be used for direct characterization of the tissues. Thus correction methods should be introduced to reduce this intensity variability. In particular, the variability of the intensity between the images should be decreased in order to remain inferior to the change of intensity between difference tissues.

### 2.6.1.2 Existing normalization methods

Several normalization methods have been proposed, but for this study only histogram matching was considered. Histogram matching relates to the research of one function that transforms a source histogram into a reference one. Once such a function is known, the source image may then be transformed into a new image where

the intensity distribution will become similar to that of the reference histogram. Thus, the link between intensities and tissues will become similar to that inside the reference, provided that the histogram distributions correspond to similar tissues.

### Histogram matching

Histogram matching relates with the research of a look up table (LUT) that transforms the intensities of a source image toward matching the histogram of this source image with a target histogram. Given a source image  $\mathcal{V}_{src}$  and a target histogram  $\mathcal{H}_{trg}$ , the process aims at finding a LUT  $f$  that modifies the intensities inside the source image such that the histogram for this transformed image  $\mathcal{H}_{f(\mathcal{V}_{src})}$  matches with the target histogram.

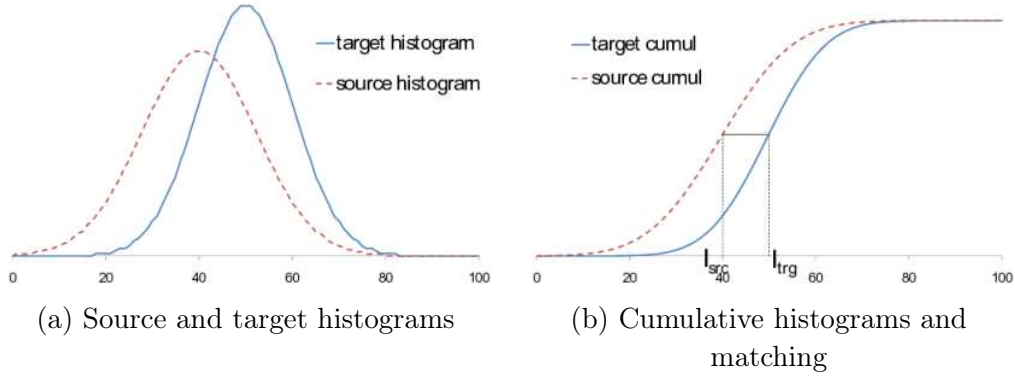


Figure 2.11: Histogram matching for two Gaussian distributions. Histograms (a) and cumulative histograms (b) are shown for two Gaussian distributions. A target is shown as a solid blue line, and the source is displayed as a red dashed one. The matching process for a single gray value  $I_{src}$  is then shown as the research of a correspondence between the two cumulative distributions (b).

Histogram matching is often achieved through cumulative histograms, by researching points with identical ordinate values inside the cumulative histograms. First, cumulative histograms are computed for the source image  $\mathcal{CH}_{src}$  and the reference histogram  $\mathcal{CH}_{trg}$ . These cumulative histograms  $\mathcal{CH}$  are defined on the range of intensity inside an image (or a ROI) as the sum of the probabilities for the inferior intensities (eq. 2.1). Then for every intensity  $i$  inside the source image a new intensity  $j$  is researched such that  $\mathcal{CH}_{src}(i) = \mathcal{CH}_{trg}(j)$ . Given an intensity  $I_{src}$ , the point inside the target cumulative histogram  $\mathcal{CH}_{trg}(j)$  with same ordinate is found, which provides a new intensity for the LUT (fig. 2.11.b). The LUT is finally given by the target intensity for each intensity value inside the source image.

$$\mathcal{CH}(i) = \sum_{j=0}^i \mathcal{H}(j) \quad (2.1)$$

### Histogram equalization

Histogram equalization is a particular case of histogram matching that allows improving the contrast inside an image by histogram matching with a constant target. Thus, the match is achieved with a linear cumulative histogram. This operation aims at better distributing the intensities within the range of possible values in order to improve the contrast. As a result, the most frequent intensity ranges are spread, which improves the contrast inside these regions. However, equalization may also increase the noise inside the image, in particular inside the background.

The effect of histogram equalization is shown for the peppers image (fig. 2.12). The dynamic of the image improves (fig. 2.12.a,b). Some details on the three central peppers are indeed more visible after equalization and the whole image seems less foggy. The effect of the process may be seen on the histogram too (fig. 2.12.c,d). The histogram indeed spreads on the entire range of intensity after equalization. Moreover, this equalization is done without losing the main shape of the histogram apart from intensity ranges with no or few pixels. The main effect may also be seen on the cumulative histogram that becomes close to linear as was requested by the approach (fig. 2.12.d).

#### 2.6.1.3 Need of a more complex method

Previous methods were supposed to be insufficient to match the histograms as a pre-process before segmentation in this work. However, usual histogram matching might have been better than the retained method *a posteriori*.

In this particular context the intensity distributions are highly variable. Indeed, the anatomical structures that contribute to the intensity distributions vary widely depending on the protocol of the exam; the organs and the tissues vary widely between acquisitions of the whole body, abdomen/pelvis or thorax/abdomen. Moreover, the pathologies may induce important changes of the intensity distributions. For example, the intensity distribution inside the liver will vary a lot, functions of the volume of the lesions and the vascularization of the lesions that may appear hypodense or hyperdense. Thus, histogram equalization as proposed by Ciecholewski will not apply for the segmentation of liver tumors as equalization would not give similar intensity values to similar tissues [Ciecholewski 2007]. Consequently, a method is required that forces the match between two distributions, while imposing a fit between peaks of the distributions that correspond to similar tissues.

#### 2.6.2 Retained method

The retained method is based on a work by Turetsky and Ellis on the force alignment of music signals, which was modified to apply to the matching of discrete intensity distributions [Turetsky 2003]. This method was retained in several tests inside this thesis. However, *a posteriori* simpler histogram matching technique might have been better.

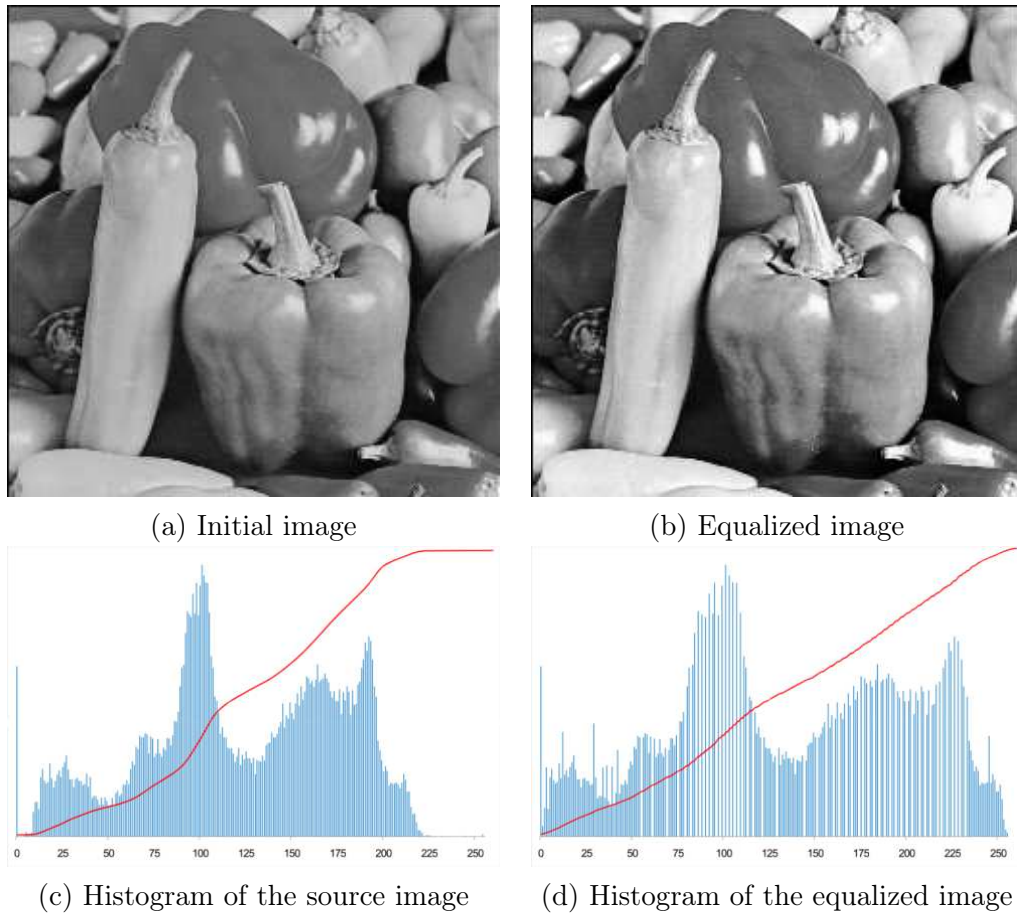


Figure 2.12: Histogram equalization for the peppers image. The histogram (c) of the initial image (a) is equalized (d), which gives an image with more contrast (b). Histograms are shown as blue bar graphs and cumulative histograms as red solid lines.

### 2.6.2.1 Dynamic wrapping

Turetsky and Ellis introduced a method to force the alignment of discrete audio signals [Turetsky 2003]. This technique is based on the research for a path that maximizes the local match between two aligned audio streams. First, a cost matrix is defined that gives the cost of the match for each pair of points. Then, an optimal path is found, which provides a transformation between the two signals.

Given two audio streams, a cost matrix is first defined between the two streams that are modeled by their discrete counterparts, using a set time step. This cost matrix is a  $m \times n$  matrix, whose dimensions  $m$  and  $n$  are given by the lengths of the two discrete audio streams. Each entry of the matrix gives the local match between two points of the discrete audio strings, defined as the distance between these two points. Once constructed, two special points are considered, the seed and the sink. These points respectively correspond to the pair of the two first points



of the discrete streams and the pair of the two last points. Seed and sink have a particular importance because they should always be aligned when forcing the alignment between the two streams.

The transformation between the two streams is given next by the optimal path between the seed and the sink. This optimal path is defined as the path that minimizes the sum of the local costs of match between the seed and the sink. This path is computed in a two-step process that uses methods related to the computation of distance maps. First, a cost map is computed that gives the minimal cost from the seed to any point of the cost matrix. This cost map is easily computed by front propagation from the cost matrix. Then, the optimal path is found by traceback from the sink. Beginning from the sink, the antecedent with minimal cost is recursively researched until the seed is reached. This traceback defines a minimal path from the seed to the sink, and thus a LUT to match one stream on the other.

### 2.6.2.2 Non linear histogram matching

Dynamic wrapping is applied for the first time for histogram matching. Dynamic wrapping applies to histogram matching by modifying the discrete series and the local cost function. First, the discrete time series are replaced by histogram distributions. Then, the local cost function is replaced by a weighted distance between the probabilities and the derivatives of the distributions. This definition aims at forcing identical location of Gaussian like parts. Formally, the local match between two points  $p$  and  $q$  in two different distributions  $h, k$  is defined as a weighted sum of two terms balanced by  $\lambda$ . The first term insures that the two points have similar probabilities, whereas the second insures that the global shapes of the histogram match.

$$d(p, q) = |h(p) - k(q)| + \lambda |h'(p) - k'(q)| \quad (2.2)$$

### 2.6.2.3 Creation of reference

A same reference distribution was used for matching each histogram in order to obtain a same intensity basis for the following processes. This reference distribution was obtained by computing an average histogram on a small set of images, from which abnormal distributions were absent.

# Segmenting the liver

---

*This chapter presents techniques for the segmentation of the liver on enhanced CT images of non-healthy patients. Three methods will be presented and ordered by increasing degree of complexity. First, segmentation as an improvement of a partial segmentation will be reviewed. This approach amounts to the computation of a smooth envelope that wraps around a binary mask. Then, segmentation by registration of an atlas and deformation of a reference volume will be proposed. And finally segmentation will be introduced as the minimization of MRF energy with an appearance prior and a registered spatial prior.*

---

**Contents**


---

<b>3.1</b>	<b>Intro</b>	<b>54</b>
3.1.1	Liver problematic	54
3.1.2	Context	55
3.1.3	Worth	58
3.1.4	Overview	58
<b>3.2</b>	<b>Previous work</b>	<b>59</b>
3.2.1	Region growing strategies	60
3.2.2	Probabilistic Atlases	60
3.2.3	Statistical Shape Model	61
<b>3.3</b>	<b>A simple approach: improving existing segmentation</b>	<b>62</b>
3.3.1	Intro	62
3.3.1.1	Context	62
3.3.1.2	Improving an existing segmentation	63
3.3.2	Transformation into a correct envelope	64
3.3.2.1	Filling surface lesion as holes inside a mesh	64
3.3.2.2	Introducing hole filling for meshes	67
3.3.3	Filling surface cavities	68
3.3.3.1	Principle	68
3.3.3.2	Obtaining a simple contour from a ROI	69
3.3.3.3	Detection of hole boundaries	70
3.3.3.4	Matching hole boundaries	71
3.3.3.5	Filling holes	73
3.3.4	Test protocol	73
3.3.5	Discussion, results	74
<b>3.4</b>	<b>Statistical atlas to represent image variability</b>	<b>74</b>
3.4.1	Motivation	74
3.4.1.1	Introducing statistical atlases	74
3.4.1.2	Atlas' worth	75
3.4.2	Theoretical background	76
3.4.2.1	Image registration, working on a same basis	76
3.4.2.2	Statistical atlas as one Gaussian per pixel	80
3.4.2.3	Gaussian Mixture Models	81
3.4.2.4	K-means	82
3.4.2.5	Expectation Maximization algorithm	84
3.4.2.6	Minimum Description Length	86
3.4.3	Creation of a statistical atlas	87
3.4.3.1	A simple example	87
3.4.3.2	Definition	89

---

3.4.3.3	Construction . . . . .	92
3.4.3.4	Implementation issues . . . . .	92
3.4.3.5	Method . . . . .	94
<b>3.5</b>	<b>Segmentation, atlas based . . . . .</b>	<b>95</b>
3.5.1	Intro . . . . .	95
3.5.2	MRF for image segmentation . . . . .	96
3.5.2.1	Definition . . . . .	96
3.5.2.2	Solving MRF . . . . .	98
3.5.3	Atlas Registration . . . . .	101
3.5.3.1	Dense Image Registration . . . . .	101
3.5.3.2	Iterative Multiscale Registration . . . . .	103
3.5.3.3	Atlas Registration . . . . .	106
3.5.4	Segmentation by registering a statistical atlas . . . . .	107
3.5.4.1	Principle . . . . .	107
3.5.4.2	Method . . . . .	108
3.5.5	Test protocol . . . . .	109
3.5.5.1	Comparison metric . . . . .	109
3.5.5.2	Segmentation priors . . . . .	110
3.5.5.3	Protocol . . . . .	110
3.5.6	Results, discussion . . . . .	111
<b>3.6</b>	<b>Combined segmentation . . . . .</b>	<b>113</b>
3.6.1	Intro . . . . .	113
3.6.2	Introducing prior models . . . . .	114
3.6.2.1	Definition, motivation . . . . .	114
3.6.2.2	Choosing prior models . . . . .	114
3.6.2.3	Spatial probability map . . . . .	115
3.6.2.4	Liver appearance model . . . . .	116
3.6.3	Segmentation using prior models . . . . .	117
3.6.3.1	Principle . . . . .	117
3.6.3.2	Segmentation energy . . . . .	117
3.6.3.3	Solving the problem . . . . .	119
3.6.4	Protocol, method . . . . .	121
3.6.4.1	Segmentation priors . . . . .	121
3.6.4.2	Test protocol . . . . .	121
3.6.5	Discussion, Results . . . . .	121
<b>3.7</b>	<b>Conclusion, summary . . . . .</b>	<b>125</b>

---

## 3.1 Intro

Segmentation of the liver from Computed Tomography (CT) volumes plays an important role during the choice of therapeutic strategies for hepatic diseases. Moreover, this segmentation is a prerequisite for many clinical and research applications. Despite a lot of attention this segmentation remains a challenging task due to the lack of visible edges on most boundaries of the organ. The high variability of both intensity patterns and anatomical appearances makes segmentation even more difficult. Moreover, all these variations are increased by the pathologies.

Due to the challenges previously stated, liver segmentation is often processed using prior information, in the manner of probabilistic atlas or statistical models [Okada 2007, Zhou 2005]. However, these methods often rely on a specific voxel size and protocol, which do not match our clinical prospect. In this chapter, the problem of liver segmentation in a clinical prospect is addressed, meaning with diverse enhancement protocols, diverse voxel dimensions, and various pathologies. To our further knowledge this problem was addressed only three times, but using very different approaches based on shape models or probabilistic atlases [Ling 2008, Linguraru 2010, Okada 2007]. Three methods will be proposed to solve this problem. First, segmentation will be done by computing a smooth envelope around a binary ROI that segments the healthy parenchyma. Then a statistical atlas will be introduced and used for segmentation through a nonrigid registration process. Finally, this second method will be improved by the introduction of priors on the spatial location of the liver and its appearance pattern.

### 3.1.1 Liver problematic

Segmentation of the liver is a complex problem because liver combines two difficulties for segmentation; the liver is both of highly variable shape and displays low difference of appearance compared to the neighbor organs. First, the liver is a soft organ, whose shape is highly dependent on the other organs inside the abdomen. Moreover, many pathologies have a strong effect on the appearance and the shape of the liver. Then, no edges are visible on many sides of the liver. In particular the differences between the liver and the diaphragm or the stomach or the spleen are very small [Masutani 2006].

The liver has a highly variable shape. First, the liver volume varies a lot; the volume of the liver envelope indeed varies by a magnitude of more than 3 for the shown examples (fig. 3.1). However, formulas exist that estimate the hepatic volume, functions of the age and the weight of the patient and sometimes using the height too, but these formulas cannot extend to the characterization of the hepatic lobes or anatomical segments and are even incorrect for pathological livers [Schiano 2000, Lin 1998]. Joyeux *et al.* indeed showed that the correlations between the liver volume and the volumes of the liver lobes or anatomical segments were very low at most and often null [Joyeux 2003]. The variability of the liver shape is also illustrated by six examples of liver segmentations displayed with 3D surface rendering (fig. 3.1). The

variation of volume and the change of shape cannot be missed. Indeed, the longest dimension, the location and the relative dimensions of the anatomical segments of the liver differ between the examples.

The characterization of the liver by its appearance alone is also difficult, due to its nature and the clinical use of the images. On some boundaries the difference between the liver and the other organs is easily seen, for example for the hepatic dome (fig. 3.2.a). However, this difference decreases at other locations. The difference between the liver and the intestine or the stomach is already more complex (fig. 3.2.b). Then, this complexity increases again with the spleen and the diaphragm that can be distinguished with even less ease (fig. 3.2.c). Moreover, the overall appearance of the liver changes in presence of pathologies such as tumors, cirrhosis, haemochromatosis. . . Furthermore, CT images are often acquired following an injection protocol in order to enhance some useful structure for diagnosis. However, this enhancement introduces two additional difficulties. First, the appearance changes with the injection phase, and then the enhancement increases the noise inside the images that are already very noisy for the liver without any enhancement.

### 3.1.2 Context

Segmentation of the liver is achieved on CT images in a clinical prospect with potential applications for diagnosis and as a first step for the segmentation of the liver tumors. Because of these potential applications, several constraints are introduced on the segmentation method. First, the method should work with diverse enhancement phases. Then, the approach should apply to healthy and non healthy livers. Finally, the approach should be of sufficient quality and fast enough.

Segmentation should apply to CT images acquired with diverse injection protocols. In order to get a correct diagnosis, a contrast product is often injected prior to the image acquisition. This injection indeed enhances several anatomical structures. In particular this enhancement allows distinguishing lesions from healthy tissues, and characterizing the type of one tumor, as the type of blood flow varies between the tumor types and compared to healthy tissues. Thus, the liver segmentation should apply to any common enhancement phase in order to be clinically useful, and to be usable for the subsequent segmentation of hepatic lesions.

Segmentation should handle healthy as well as pathological livers. Segmentation of the liver is already a complex problem. However, this simple case is non relevant for clinical use, as a segmentation of the liver is more contributive in case of pathology. The segmentation technique should thus deal with pathological livers. This extension of the scope of the segmentation comes with an increase of the difficulty of the problem. Pathologies indeed induce both anatomical changes of the liver and modifications of the appearance of the tissues within the liver envelope on CT images.

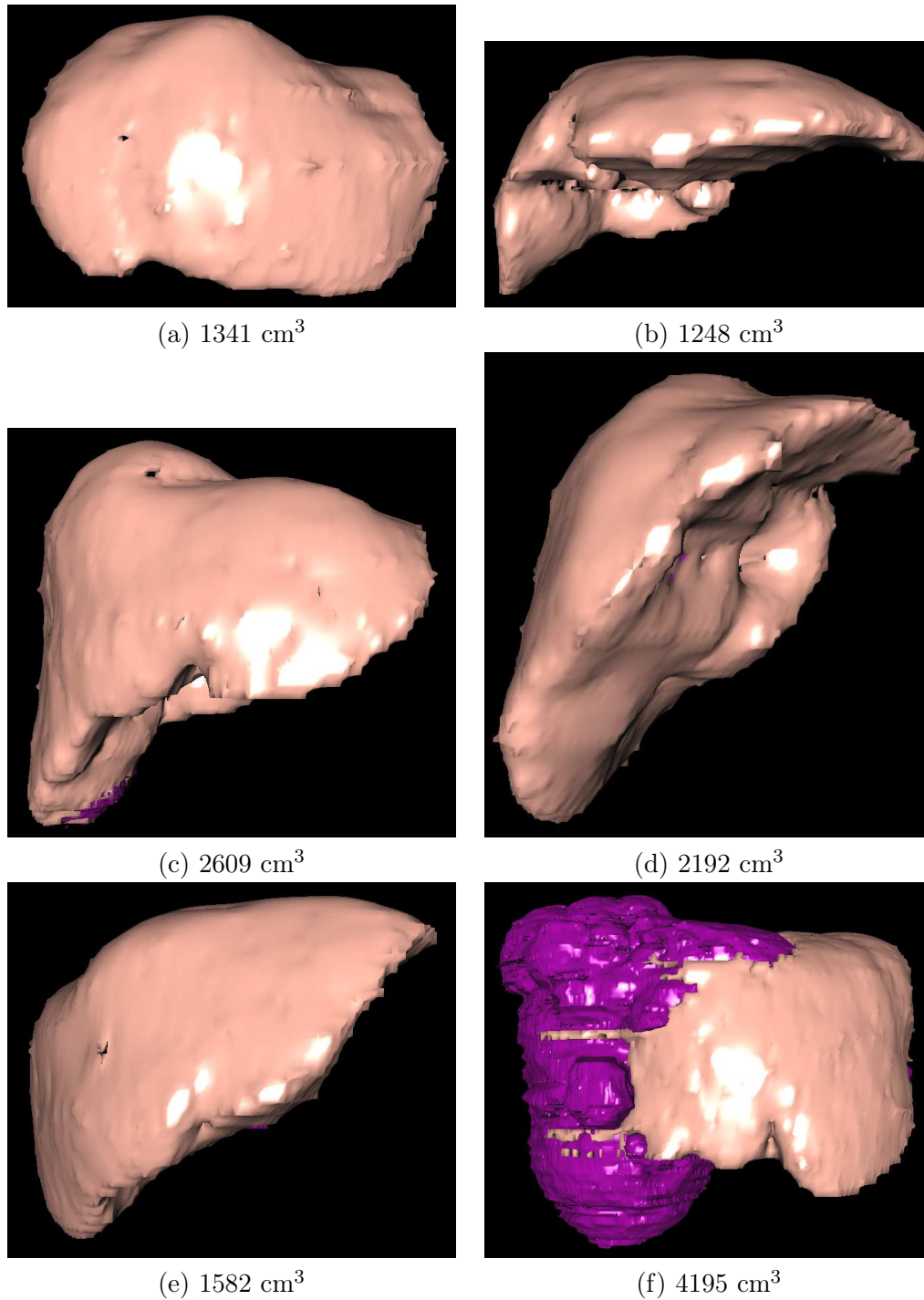


Figure 3.1: Variability of shape and volume for several livers. 6 reference segmentations for tumoral liver are shown using Myrian<sup>®</sup> with surface rendering in frontal view. The volume of the hepatic envelope is given for each image. The envelope is displayed on each image in light pink and lesions in purple. Images are courtesy from Beaujon's hospital.

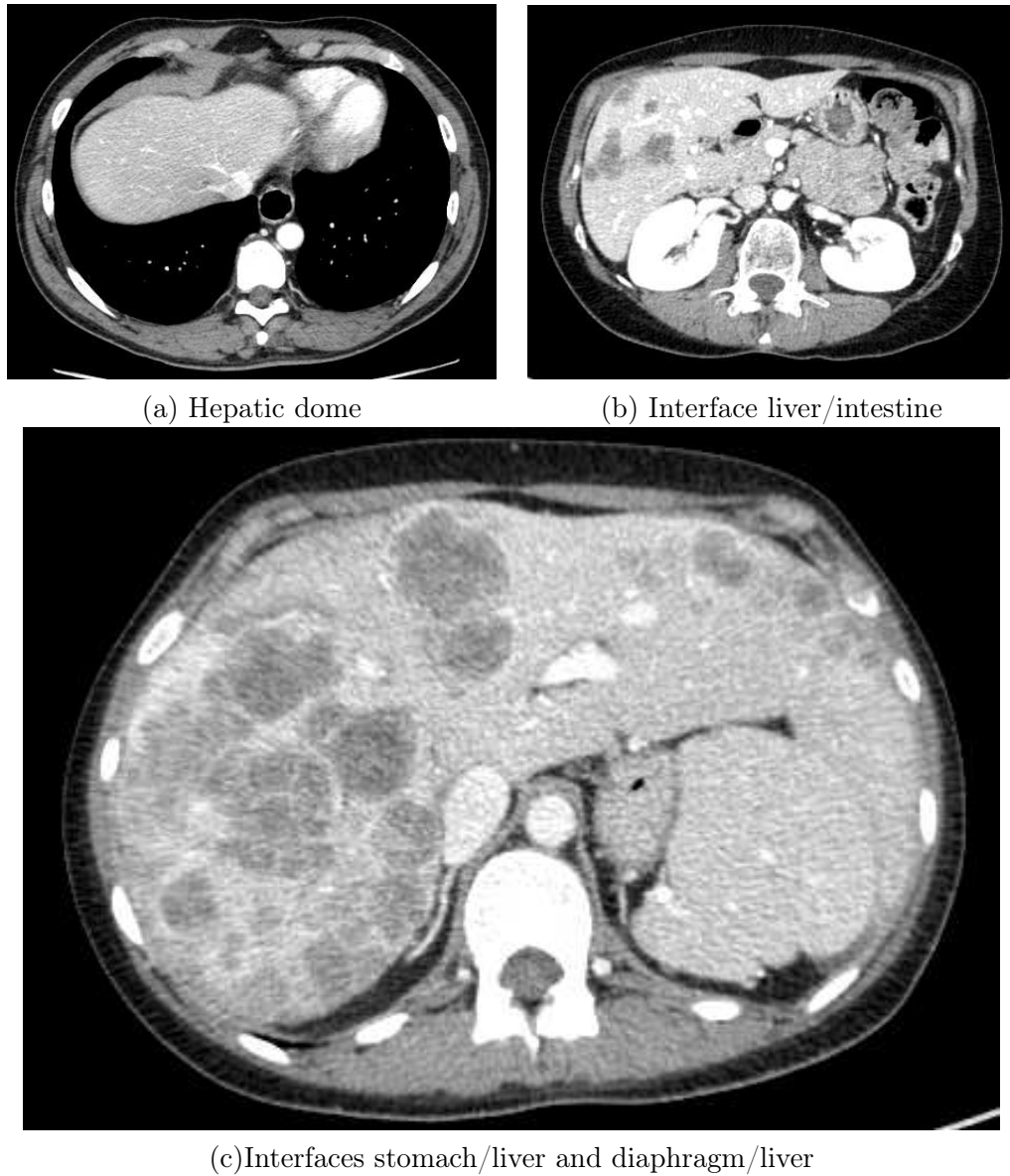


Figure 3.2: Variability of appearance for three liver examples with colorectal metastases. Images are courtesy from the CRLC Val d'Aurelle.



### 3.1.3 Worth

The segmentation of the liver is valuable from a medical perspective first, as an intermediate step in the overall process of this study next, and finally from the point of view of Computer Vision. First, the segmentation of the liver contributes to diagnosis and therapeutic decisions. Then, this segmentation is a required step for the segmentation of liver tumors. Finally, the proposed methods are not liver specific, but may also apply to a wider range of problems in medical imaging and image processing.

A precise segmentation of the liver is a precious help for diagnosis (cirrhosis, tumors. . .) and for the choice of a therapeutic strategy [Joyeux 2003]. For example Lin *et al.* showed that the volume of the liver correlates with the severity and the etiology of some chronic liver diseases [Lin 1998]. The liver plays an even more important role for the choice of a therapeutic strategy in case of cancer. For example the remaining volume of healthy liver after surgical resection is a crucial measure for the choice of surgical resection as treatment [Kubota 1997]. This measure is also important for living-donor liver transplantation [Strong 2006, Schiano 2000].

A liver segmentation also contributes to the main goal of this study. First, segmentation of the liver is the first step of the segmentation of the liver tumors. Then, from a medical prospect segmentation of both liver and lesions is more valuable than segmentation of tumors only. In particular the choice of surgical resection is dependent on the ratio between healthy and tumoral liver [Hélénon 2003]. Furthermore, both lesions and liver are useful for preoperative planning [Choi 2006]. Thus, segmenting the liver along with the tumors is more coherent with the clinical prospect of this study.

From a Computer Vision perspective, the proposed approaches offer three main contributions, a wrapping morphological tool, a new statistical atlas and the non rigid registration of this atlas. First, a method to compute smooth envelopes that wrap around a binary object is introduced. Then, a new statistical atlas is proposed that gives an optimal intensity distribution at each point of the space. A method of construction and nonrigid registration of such an atlas are finally proposed.

### 3.1.4 Overview

Three methods will be proposed for the segmentation of the liver envelope. First, segmentation will be achieved through the computation of a smooth envelope around a binary ROI that segments the healthy parenchyma. Then a statistical atlas will be introduced with a segmentation achieved through a nonrigid registration process. Finally, prior information on spatial location and appearance of the liver will be added to the segmentation process in order to improve the previous approach.

The computation of a smooth envelope around a binary ROI will be proposed in a first section, and will be applied to the segmentation of a liver envelope (sect. 3.3). This approach aims at taking advantage of an existing segmentation of the healthy parenchyma for the segmentation of a liver envelope. The proposed technique applies

existing methods from the field of digital reconstruction as morphological operations to fill non smooth cavities at the surface of a binary object.

A new definition of statistical atlas is introduced next and applied to the segmentation of a liver envelope through the nonrigid registration of this atlas and the transformation of a reference segmentation (sect. 3.4) (sect. 3.5). A new definition of statistical atlas is introduced that gives one optimal model of intensity distribution for each point of the space. Then, a state of the art method for nonrigid registration is modified to apply to atlases. Finally, such an atlas is created for the liver and applied to the segmentation of a liver envelope through the deformation of a segmentation reference for the atlas.

Spatial and appearance priors are finally added to the statistical atlas during the segmentation process (sect. 3.6). A multiscale segmentation process is introduced that registers a statistical atlas before proceeding to segmentation as a balance between the adequacies with an appearance prior and a spatial prior.

## 3.2 Previous work

Despite lots of attention, fully automatic liver segmentation from a 3D CT volume remains a challenge, mainly because of the variability of the liver shape and of the intensity patterns inside and in the neighborhood of the liver. Besides, segmentation of the liver with pathologies and on enhanced CT images is an even more complex problem that was less often dealt with. Three main approaches have been considered for the segmentation of the liver. First, several approaches follow a region growing strategy to segment the liver. Then, many approaches have been proposed that impose strong *a priori* on the liver. The liver is indeed often modeled with an appearance prior combined with probabilistic atlases (PA) or statistical shape models (SSM).

The three approaches exhibit different strong and weak points. First, while many authors stress the value of prior information, region growing strategies nevertheless offer good results because of their ability to adapt to the specific characteristics of each image. However, region growing strategies have the disadvantage of badly handling lesions. Secondly, PA and SSM both introduce prior information on the liver with more success with SSM. PAs indeed offer lesser results, but their accuracy could improve when using better standardization techniques. Moreover, PAs not only model the liver, but also often model other abdominal organs. Then, the nature of the images that are handled also varies. Whereas PAs handle mainly non contrast images, SSMs show better abilities when working on enhanced images. Finally, despite the advantages previously mentioned, SSMs involve two major difficulties. The initialization of a SSM is indeed complex and crucial, and the possible deformations of the liver modeled by a SSM are limited, which prevents capturing the high variability of the liver shape. Thus SSMs approaches often rely on other methods for initialization and as final step for obtaining a finer segmentation.

### 3.2.1 Region growing strategies

Region growing strategies are data driven approaches with no or few *a priori* that iteratively construct a ROI defined at pixel level beginning with an initial set of seeds. Such approaches iteratively cluster neighbor voxels by deciding whether these voxels are close enough to be added to the ROI at a given step. Because of this definition, region growing strategies do not rely on a specific prior model, but adapt to each image depending on the seeds given by the user. Despite their simplicity, region growing approaches still provide satisfactory results compared to more evolved approaches. Ruskó's method indeed offered the second best results at MICCAI segmentation challenge 2007 [Ruskó 2007]. However, the nature of the method prevents getting good segmentations when huge lesions are found in the CT images.

A first segmentation of the liver was proposed by Chemouny, who segments the liver in PVP through a region growing approach after denoising the CT images with the Nagao's filter [Chemouny 1999, Chemouny 2001]. This approach begins with a filter of the CT images with Nagao's (sect. 4.3.2.4). A segmentation of the healthy parenchyma is obtained next through an optimal region growing strategy with coarse-to-fine merging conditions. Candidate regions are indeed merged with the growing ROI using conditions of increasing complexities. Then, Pham segments the liver on a probability image with a split and merge approach followed by a region growing algorithm [Pham 2007]. The specificity of this approach comes from the probability image that gives the probability of being inside the liver at each voxel of the image for a texture model of the liver. Finally, Ruskó proposed a technique to segment the liver in PVP, where the merging condition is dependent from an online intensity distribution and a wide neighborhood system [Ruskó 2007]. This region growing approach indeed relies on the intensity distribution inside the segmented ROI, and uses huge spherical neighborhood to decide on or against the merging of new voxels. This latter approach provides accurate segmentations. This method indeed ranked second at MICCAI segmentation challenge 2007. However, large lesions are not handled by this technique.

### 3.2.2 Probabilistic Atlases

Several approaches introduced priors on the spatial location of the liver in order to cope with the unmarked boundaries of the liver. Region growing approaches indeed tend to oversegment the liver and in particular often include the stomach, the spleen or the pancreas. Thus, probabilistic atlases (PA) have been introduced to provide the probability of being inside the liver at each location inside one image. Such PA of course require a spatial standardization of the images in order to decrease the spatial changes between the images. This standardization is crucial and its accuracy impacts the final segmentation. Concerning their accuracy, PA offer lesser results than SSM. However, this accuracy might improve when applying a better standardization of the images. While PA offer spatial information that region growing approaches cannot accede, PA have two disadvantages compared to region growing

techniques. Indeed, PA do not handle well the enhancements and most approaches apply only to non contrast images.

Park introduced the first abdominal atlas for four abdominal organs on non enhanced CT images [Park 2003]. Using a PA, he segments the liver by optimizing a MRF formula that depends on intensity distributions computed through the registration of the PA by thin plates wrapping. Then Zhou proposed a segmentation of the liver in non-enhanced images as a threshold of the probabilities of being inside liver, where the probability is defined at voxel level through the PA and an intensity model [Zhou 2005, Zhou 2006]. However, Okada criticizes this last approach because Zhou *et al.* verified the performance of this technique only on normal and few mildly diseased livers and not on severely diseased ones [Okada 2007]. Shimizu later proposed a segmentation technique for 12 abdominal organs inside the abdominal cavity, still on non contrast CT images [Shimizu 2007]. This technique begins with a rough segmentation obtained through the PA and priors on intensity distributions inside each organ. The rough segmentation is refined with level-sets next. Finally Linguraru proposed a segmentation of pathologic liver and spleen on contrast enhanced images [Linguraru 2010]. An initial segmentation is first obtained through series of rigid, affine and non linear registrations of one PA. This segmentation is then refined using geodesic active contour and an estimation of the distribution parameters inside each organ.

### 3.2.3 Statistical Shape Model

Statistical Shape Models (SSMs) have often been proposed for the segmentation of the liver. These SSMs define a mesh for the liver envelope along with possible deformations of the nodes describing the liver boundary. Compared to deformable models, SSMs do not only store the shape of the liver, but also the main modes of variation. Thus, the SSM imposes a strong *a priori* on the liver shape that should improve the segmentation, as *segmentation methods without prior information are prone to fail* according to Heimann [Heimann 2006]. However, the high variability of the liver is highly challenging because of the difficulty to define a SSM that captures the huge variance of the liver shape. The limited transformations of SSM indeed impede the fit with structures of large natural variability [Heimann 2007, Heimann 2009]. Consequently, later authors using SSMs also introduced an additional step in order to free them from the limited transformations of SSMs. However, the contribution of SSMs still remains significant as a recent review showed that SSMs with a subsequent free deformation step offered the best results during automatic liver segmentation [Heimann 2009]. Indeed, among ten automatic methods that were compared on same data, the three best rated were all based on SSM followed by an additional deformation.

### SSM alone

Lamecker constructed a SSM of the liver that was used for segmentation of livers with lesions in PVP, using models of gray levels profiles along surface normals to fit the SSM on new images [Lamecker 2002, Lamecker 2004]. Kainmüller later extended this method by using free forms to define the SSM and by adding a specific step to handle the lesions during the segmentation [Kainmüller 2007]. This approach proved to be the best one during MICCAI segmentation challenge 2007. Heimann then introduced a SSM to segment the liver without major pathologies in PVP [Heimann 2006]. He mainly follows Lamecker's approach, to which he adds a multi-resolution algorithm and active shape models. The multi-resolution approach allows improving the robustness of the method, while the active shapes are used to authorize small deformations that are prevented by the SSM. Heimann later improved his approach with an automatic initialization of the shape model and by replacing the active shape models by a more complex technique [Heimann 2007]. In a last step, Heimann indeed defines a deformable model that is fitted on the image while keeping the global shape of the fitted SSM.

### Combined approaches

Okada introduced a segmentation approach that combines PA and multi-level SSM to segment the liver [Okada 2007]. An initial segmentation is done using PA and voxel similarity measures. The segmentation is achieved next with multi-level SSMs. Multi-level SSMs are coarse-to-fine models that divide a SSM in multiple patches in order to attempt to limit the modifications induced by lesions. These multi-level SSMs indeed allow decreasing the impact of lesions by fitting multiple SSMs in diverse parts of the liver. This definition also offers more variability than a single SSM as many smaller SSMs are used. Moreover, the increased variability will not affect the overall shape of the liver because of continuity conditions between the neighbor shapes. Ling further continued with the division of the liver in distinct objects by proposing a segmentation using a hierarchical shape model, where liver boundaries are detected using learning based approaches [Ling 2008]. Because gradient or intensity distributions are not sufficient, Ling introduced a shape model divided in five regions with various prior information learned for the detection of the boundary.

## 3.3 A simple approach: improving existing segmentation

### 3.3.1 Intro

#### 3.3.1.1 Context

An extension of an existing tool was considered as first approach. Indeed, the segmentation of the healthy liver parenchyma is already available in a segmentation engine that follows the approach proposed by Chemouny (sect. 3.2.1). However, this segmentation is imperfect and cannot be used as a liver envelope. Instead of

beginning from scratch, the improvement of this segmentation in order to obtain a segmentation of the liver envelope was considered as a first approach. This problem was further restricted by considering only the ROI provided by the existing tool. Thus the computation of a liver envelope becomes the research of a smooth envelope that encompasses a binary ROI.

A tool already exists that segments the healthy liver parenchyma with more or less success for CT images in portal phase. This tool segments the healthy liver parenchyma, but is often imperfect on some boundaries and the compartment around the other structures inside the liver envelope is quite variable. Sometimes lesions and vessels will be excluded from the segmentation, sometimes they will be included, and sometimes they will be partially segmented. In the whole and despite its imperfections, this tool provides a correct segmentation of the tissues inside the liver, which may then be extended to the segmentation of the entire liver envelope.

The proposed method computes a liver envelope from a segmentation of the healthy hepatic parenchyma that is assumed to be of sufficient quality. This initial segmentation is not assumed to be perfect, because the method should nevertheless handle some imperfections of the initial ROI for robustness purposes. However, this first approach will not deal with the improvement of segmentation of the healthy parenchyma. Instead, the proposed approach will proceed to the computation of an envelope in order to segment the tissues of the liver that were not segmented.

This entire method lies on the assumption that the healthy parenchyma is the most voluminous part of the liver and is well distributed inside the liver envelope. Computation of the liver envelope is indeed done by computing a smooth ROI that encompasses the segmentation of the healthy parenchyma. When this initial ROI is well distributed inside the liver, the smoothness constraint on the envelope will drive the segmentation to include the missing parts of the liver. Otherwise, the proposed approach will not recover a correct envelope. This constraint is not too restrictive because the healthy parenchyma most of the time accounts for the majority of the liver volume. Besides, when the healthy parenchyma is not predominant, segmentation of lesions do not have any clinical value as the lesions will be too voluminous compared to the healthy parenchyma. In fact, the only problematic case involves huge lesion located on a boundary of the liver with high curvature. For this configuration the smoothness constraint will not be able to include the missing lesion, but only a part of this missing region. This constraint is illustrated by an artificial example on a true liver segmentation (fig. 3.3). While the smoothness constraint allows defining a correct envelope when the initial segmentation is well distributed inside the liver (fig. 3.3.a), a missing part inside the initial ROI leads to poor results when located on a region with high curvature (fig. 3.3.b).

### 3.3.1.2 Improving an existing segmentation

If the healthy parenchyma is correctly segmented, the computation of the envelope amounts to filling the holes due to the missing parts of the liver, in particular the tumors and the vessels (fig. 3.4.b). These missing parts will be divided in two

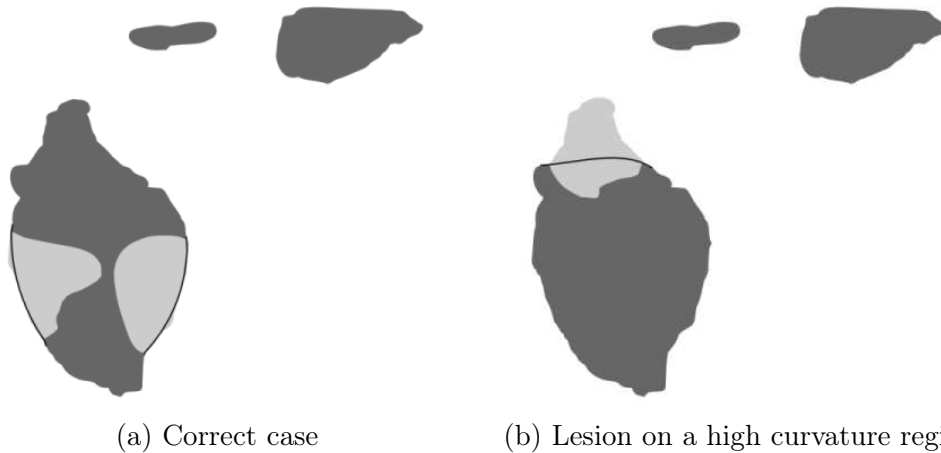


Figure 3.3: Limits of the computation of a smooth envelope from a liver ROI. Reference ROI is shown in dark gray, missing parts in light gray and the correction with the envelope is displayed as a black line.

types depending on the difficulty they induce, which is directly linked with their locations. These holes will be referred as interior holes or surface cavities. For the former case, the missing parts are located inside the liver, which leads to ROIs with holes inside (fig. 3.4.c.1). For the latter case, cavities are located at the surface of the liver, in particular due to abut tumors (fig. 3.4.c.2).

Interior holes and surface cavities cannot be filled using the same approach. Interior holes do not present any difficulty as they may be easily filled using standard morphological operations with a connexity map. However, surface cavities are more challenging and will be the main difficulty this approach has to deal with. Indeed, the difference between a normal variation of the liver and a cavity on the surface of the envelope is not straightforward.

Approaches proposed for filling holes on meshes obtained while scanning 3D objects are retained for the computation of a smooth envelope that encompasses a binary ROI. The proposed approach is divided in three steps. First, the surface cavities are detected and characterized by their boundaries. Then, the boundaries that enclose a cavity are matched. Finally, each surface cavity is filled smoothly with a condition on the continuity of the outline. This approach explicitly deals only with surface cavities; however, the interior holes are also filled as the result of the technique is a contour.

### 3.3.2 Transformation into a correct envelope

#### 3.3.2.1 Filling surface lesion as holes inside a mesh

The computation of a smooth envelope that encompasses a binary ROI is achieved using a method related to the filling of holes inside 3D meshes. The idea is to create a mesh from the initial binary ROI, before cutting this mesh in order to introduce

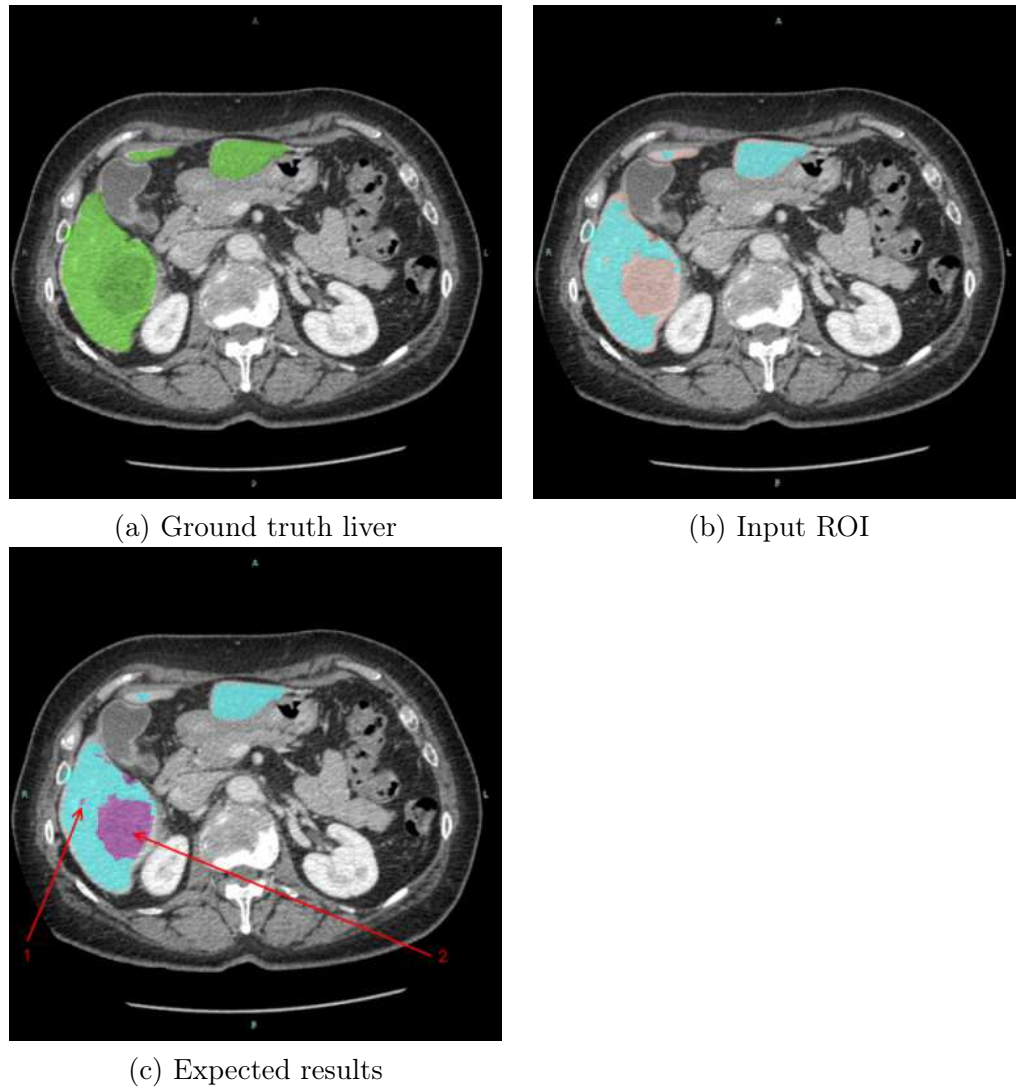


Figure 3.4: Segmentation of the hepatic envelope from the healthy parenchyma. Ground truth liver is shown in green, the healthy hepatic parenchyma given as input is blue, and the result of the hole filling process is displayed as purple. The hole filling process should fill interior holes (c.1) and cavities on surface (c.2).



holes on the mesh surface at same locations than the cavities on liver surface. These holes may then be filled with existing methods to fill mesh holes. One should note that this problem is simpler than the usual problem in digital reconstruction, because the envelope to compute is quite continuous and regular. Thus the main difficulty becomes the detection of the cavities on the surface of the initial ROI.

Several other approaches were considered, model registration, fitting B-Splines on the binary mask, and deformation of a mesh. However, all these other approaches were not retained because they were not efficient on preliminary tests or introduced difficulties with no obvious solution.

Model registration aims at matching a model of the liver on the binary ROI, where the model is a representative liver volume. Such method was assumed to correct some of the errors from the initial ROI, because of the introduction of prior knowledge. However, this approach did not provide good results on preliminary results, as the matching of binary volumes is not well done. Moreover, a single representative volume as model does not allow capturing the high variability of the liver. Furthermore, using an atlas instead of a volume as model introduces an additional difficulty. The model will indeed not be binary, thus there is still some way before defining a final ROI from a probability map.

Because of the smoothness constraint on the envelope, one may imagine fitting B-Splines or NURBS on the boundaries of the initial ROI in order to obtain the envelope. Identical approaches have been retained to reconstruct volumes after digitalization, by fitting splines on a set of points [Kumar 2007, Yin 2004, Daniels 2008]. With these approaches the cavities on liver surface could be avoided by setting a number of control points in order to capture the true shape of the object without going inside the smaller cavities on liver surface. Moreover, the constraint set by the number of control points may be able to cope with some initial errors. However, the choice of a number of control points seems unsolvable. This number indeed depends on the shape of the liver; for compact livers only a small number of control points would be required, and on the opposite, livers with distinguishable lobes would require many points. Therefore this approach cannot be used to compute the envelope, because the choice of a number of control points depends on the envelope.

Deformation of a mesh aims at wrapping a mesh around the binary ROI and the possible cavities on surface in order to obtain the envelope. However, this approach cannot provide the expected results. This impossibility comes from two opposite and incompatible constraints. First the mesh should be close to the initial ROI. Then the mesh should not wrap inside the surface cavities. This second constraint is obtained by penalizing high curvatures and concave parts. Because of the size of the surface cavities, these two constraints are impossible to met, even for simple artificial phantoms. Indeed, meeting the first constraint requires giving weight to the data term, and on the opposite the weight of the regularization term has to be greatly increased in order to avoid wrapping inside the surface cavities.

### 3.3.2.2 Introducing hole filling for meshes

In the field of digital reconstruction meshes represent volumes. These meshes are prone to holes, thus methods to infer unknown parts of objects have been developed. These methods deal with the computation a point set surface, meaning a surface that better match a set of points, while using the known parts of the objects to infer the missing ones [Zhao 2007, Wu 2008, Casciola 2005, Pérez 2008].

The digitalization and reconstruction of 3D objects as meshes is often used in computer graphics and computer-aided design. . . Whatever the origin, 3D-scanners, design software or others, such meshes are often incomplete. In particular these meshes often have holes on their surface, which is a problem because these holes prevent the use of some algorithms and introduce compatibility problems between software. Thus, methods have been developed to patch these holes in a likely manner.

Many methods have been proposed, however only methods based on patches by continuity have been retained, as they may apply to the transformation of a binary ROI into a smooth envelope. Such methods begin by identifying the holes on the mesh. Then, they define an initial patch, with diverse methods, but always while using the neighborhood of the hole to estimate the orientations of the mesh faces inside the hole. Finally, this patch is regularized using various methods.

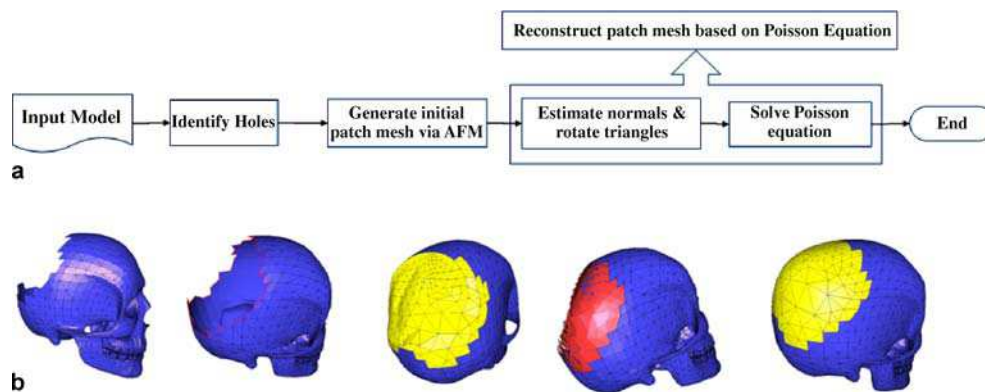


Figure 3.5: Filling holes on meshes as proposed by Zhao [Zhao 2007]. A flowchart of the proposed method is given (a), along with the corresponding states for a skull (b). The figure is extracted from Zhao’s paper.

The principle of these techniques will be illustrated using Zhao’s method for an example given in one of his paper (fig. 3.5) [Zhao 2007]. Given an initial mesh, the hole is first detected. Then an initial patch is computed using the Advancing Front Mesh (AFM). The idea behind this technique is to iteratively construct the patch, rim by rim around the boundary of the hole, by adding faces following the average direction of the neighbor faces. Finally, the Poisson equation is used to impose a smooth connection between the faces of the patch mesh in order to approximate the missing region more accurately.

### 3.3.3 Filling surface cavities

#### 3.3.3.1 Principle

The retained approach mimics the hole filling process of meshes in digital reconstruction (sect. 3.3.2.2). This approach draws a parallel between the holes on a mesh in digital reconstruction and the cavities on the surface of a ROI. From an initial binary ROI  $\mathcal{S}_{bin}$  is first created a surface mesh  $\mathcal{Q}$ . Then, holes are artificially cut inside this mesh, where the cavities on the surface of the binary ROI are located. Finally, these holes on the mesh are patched with existing methods to fill mesh holes. For this approach the main difficulty comes from the detection of the cavities on the boundary of the binary ROI  $\mathcal{S}_{bin}$ , because the location of the holes on the mesh will always be known and because the envelope to compute is quite continuous and regular. One should note that interior holes will be dealt with by this approach as the mesh will define only the outer boundary of the binary ROI.

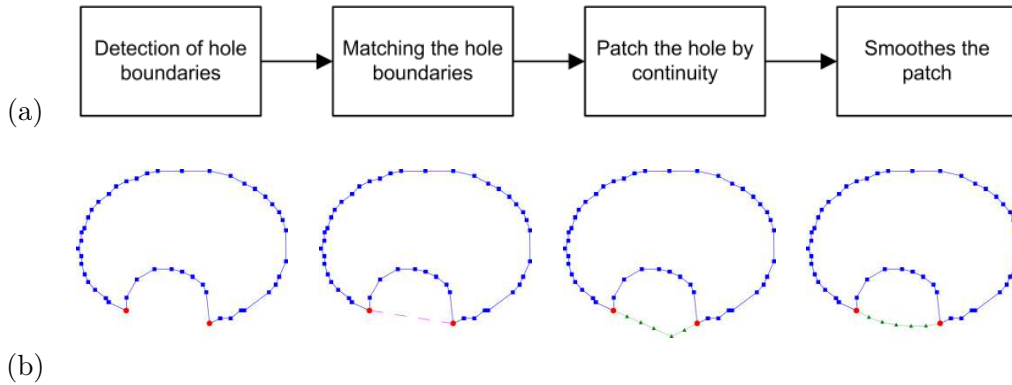


Figure 3.6: Principle of the hole filling process for an artificial example. The process flowchart (a) for the hole filling process is presented and illustrated for an artificial example created by the exclusion of an ellipsis and a circle (b.)

The entire process will now be presented through a simple example defined as an ellipsis with a missing circular part (fig. 3.6). First a mesh  $\mathcal{Q}$  is computed for the binary ROI  $\mathcal{S}_{bin}$ . Then, the holes are detected and defined by their boundaries on the mesh. Each hole is indeed delimited by a pair of nodes of the mesh. Finally, each hole is patched using a heuristic value to begin with the most probable holes and to prevent filling normal variations of the binary object.

The holes are detected in two steps; first are detected candidate nodes and then these candidates are paired in order to define the boundaries of each hole. Candidate nodes on the mesh, namely nodes that could be on the boundary of a hole, are detected by looking for nodes where the curvature is varying fast and not as a one-off phenomenon. This constraint indeed allows the detection of hole boundaries, because the researched envelope is assumed to be smooth, thus a breakdown of the smoothness of the surface should come from a cavity. The candidates as hole boundaries are then matched in order to define the holes.

Given a set of holes that are each defined by a set of nodes, the holes are then patched. These patches are ordered by a heuristic value that aims at measuring the breakdown of the smoothness of the surface of the mesh for the path between a pair of candidate boundaries. This heuristic allows beginning to patch the holes that have the highest probability of being a hole on the surface of the mesh. This value also allows stopping when the candidate holes have low chance of being irregularities on surface, but could rather be normal variations of the object. Each hole is then patched following the idea of Zhao [Zhao 2007]. First all nodes inside the boundary defined by the set of nodes are removed. Then, the AFM is used to define an initial patch. Finally, the patch is smoothed using a simplification of Zhao's approach; all faces inside the patch are modified in order to minimize the variance of the angles between successive neighbor face.

In this study, the hole filling process was achieved with a 2D approach, where each connex component is processed independently. The 2D approach may indeed be sufficient as the initial binary ROI is already smooth in the third dimension. Moreover, the definition of the boundary of each hole is tricky in 3D, because the constraint on curvature may be not strong enough to detect an entire boundary. Thus, detection of the border of each hole would require to well detect some nodes on this border and then to complete the boundary with well chosen nodes. Moreover, with this construction of the hole boundaries, the distinction between normal variations of the object and missing parts would become a more complex problem.

### 3.3.3.2 Obtaining a simple contour from a ROI

The computation of the initial mesh is done while satisfying two constraints, the mesh should be precise enough for the detection of the holes, and it should not be too detailed in order to keep only the relevant features of the contour. Obtaining the initial contour will be introduced with an example (fig. 3.7). First, an initial contour is obtained using the Freeman chain code histogram [Freeman 1974]. This treatment defines a contour with one node per pixel on the boundary of the object (fig. 3.7.b). Then, this initial contour is pruned in three steps. Non contributive nodes are first removed (fig. 3.7.c), then nodes that may be removed with few change are deleted (fig. 3.7.b), and finally too close nodes are merged.

An initial contour is computed from a binary ROI by following the boundary of the object, pixel after pixel. This contour is obtained by moving along the boundary of the ROI while following the directions introduced by Freeman. Freeman indeed described a coding chain that defines a line by a set of successive directions to follow in order to move around this line while going through each pixel of the line [Freeman 1974]. By computing this coding chain from a point on the boundary of the object, the contour of the ROI may be obtained when the boundary of the object is used as a line. Then, obtaining the contour from the coding chain is straightforward. And getting the initial point on boundary is also very easy; one only has to go through the image and take the first point inside the object.

The contour is then simplified in three steps, removal of non contributive nodes

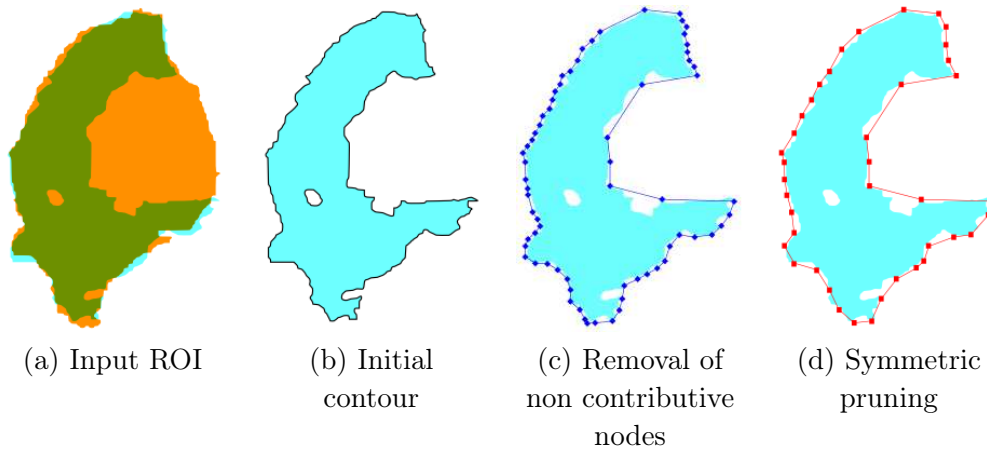


Figure 3.7: Computation of a contour from a ROI, example for a connex component of the liver. An initial segmentation is shown in light blue, along with the reference for the liver in orange (a). A missing part of the liver on the right is due to a tumor lesion. The contour for the ROI is first computed using Freeman coding chain (b). Then, non contributive nodes are removed from this contour (c). Finally, this contour is pruned in order to simplify its representation (d).

first, symmetric pruning then and finally a merging of close nodes. First, are removed the nodes that can be removed without any change of the contour. Consequently, straight line segments of the contour are defined by only their boundaries. Then, a symmetric pruning is done by removing the extreme two nodes of three, each time three successive nodes are inside a same circle of a set size. This step allows removing small variations of the contour without modifying the contour too much as what would happen when considering only a pair of nodes. Finally, nodes that are within a small distance from each other are merged in a new node on the middle of both.

### 3.3.3.3 Detection of hole boundaries

The detection of possible hole boundaries is a critical step for the detection of the holes. This detection is done in a coarse-to-fine manner in order to progressively find candidates by excluding nodes that cannot be on the boundary of a hole. First, the nodes that are almost aligned are removed. Then, the curvature in a neighborhood is computed for the remaining nodes. This curvature is retained next to exclude nodes that are in a region that evolves smoothly. Finally, only one node is retained when many candidates are found in a same part.

First, only nodes that are not aligned with their neighbors are marked as possible candidates. Aligned nodes are indeed not relevant for the problem of hole detection. This first selection is achieved with a threshold for an obtuse angle that is chosen to prevent false negatives. In this study, nodes were retained as candidates when the absolute change of angle between two successive edges was smaller than  $\frac{3\pi}{4}$ .

Candidate nodes are then pruned by searching if the curvature is local or not. This second step is achieved by computing the curvature using second order neighbors (neighbors of neighbors). The same threshold on angle is then used to exclude some candidates.

At this point remain some parts of the contour with many consecutive candidate nodes. Only the most significant node is kept inside each set of consecutive nodes; the node with the highest curvature. The neighborhood for merging nodes was set to 10 pixels (2 times the pruning radius).

### 3.3.3.4 Matching hole boundaries

The candidate boundaries are then matched in order to define each hole. This match should answer to three problems. First, the method should be able to match candidates that define a hole. Then, false positive should be dealt with. Finally, surface cavities should be distinguished from normal anatomical variations. This match is achieved using a constraint of direct line of view to define possible pairs and a heuristic value to distinguish holes from anatomical variations.

The constraint of direct line of view allows excluding pairs that cannot define a hole, by searching whether or not a straight line between the two nodes intersects with the outline. When two nodes are on the boundaries of a same hole, they may indeed see each other, which means that the line segment between them do not intersect the outline. This definition handles the matching of candidates separated by a hole, but does not distinguish between two nodes on the boundary of a same hole and two nodes inside a hole. However, this step allows dealing with the natural boundaries of the liver and excluding some false positive candidates. The contribution of the direct line of view is shown for a liver slice (fig. 3.8.a). This constraint excludes pairs that do not relate with a trough of the contour and defines possible pairs. However, this step does not define neither the nature of the holes nor their relevance.

A heuristic value is then introduced to estimate the relevance of a hole and to define the order to follow for filling the holes. This heuristic value orders the possible pairs of candidates depending on their probability of being a hole that is assumed to be related to the deepness of the hole between two matched nodes. This value is defined as the ratio between geodesic and square Euclidean distances between two matched nodes. This heuristic value favors pairs with a small neck compared to the contour between them. Thus, non smooth parts will be filled first. Moreover, a threshold may be set to stop filling holes that have low probability of being surface cavities. The contribution of the heuristic values is shown on an example (fig. 3.8.b). One may see that the surface cavities are filled first, before progressively filling parts that modify more and more the global shape of the object and hence have less probability of being surface cavities.

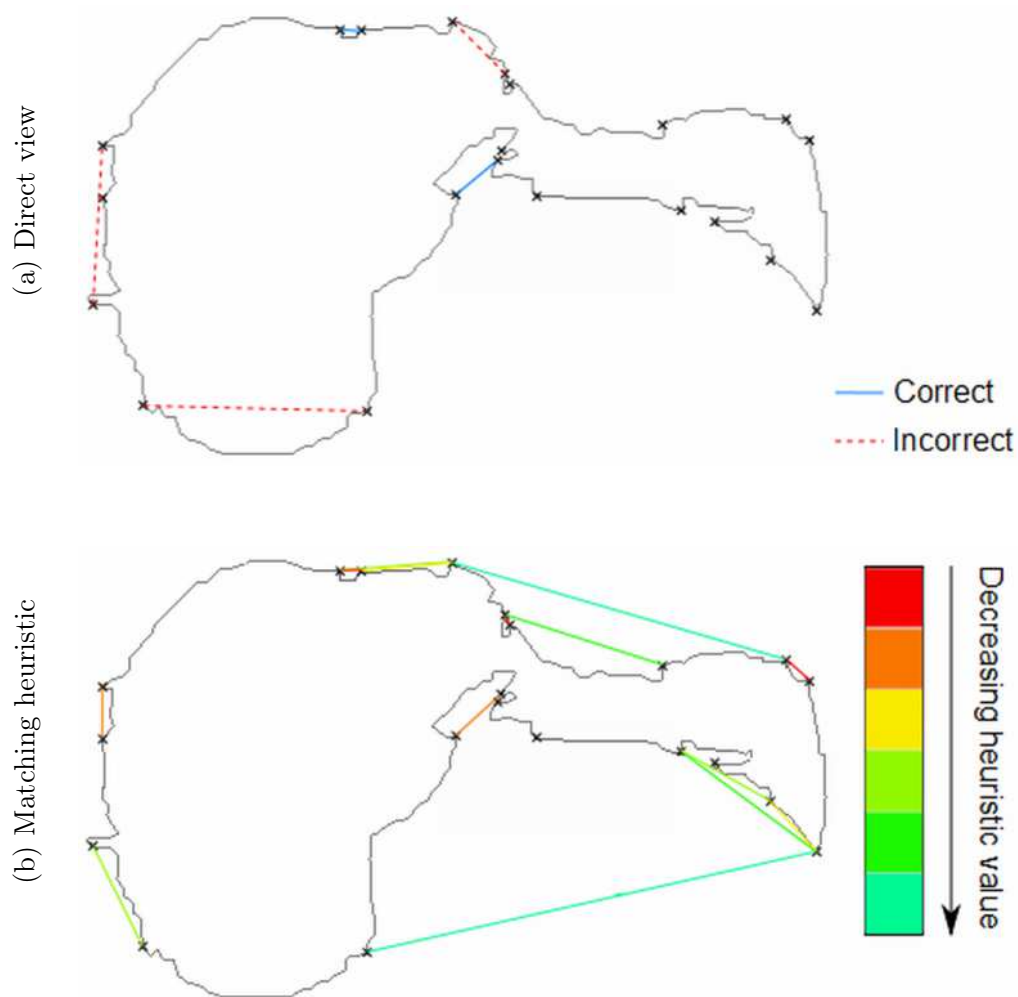


Figure 3.8: Matching candidate boundaries. The matching process is illustrated for a liver outline shown in medium gray, where candidate boundaries are shown as black x. The direct view constraint is presented on a first example (a), with solid blue line for correct pairs and dashed red ones for the incorrect matches. The contribution of the heuristic value is shown in a second example that describes the heuristic value of the match with colored lines (b).

### 3.3.3.5 Filling holes

The filling process is applied to a pair of boundary nodes obtained with the previous steps. This process is divided in two steps. First, the part of the contour between the two nodes is replaced by extending the contour while following the direction on either side of the hole. Then, this patch is modified by minimizing the sum of the absolute difference of angles between successive line segments inside the patch. This part of the approach was only done on phantoms and not applied in the main method, in particular for the subsequent tests.

An initial patch is first done by continuity on either side of the hole. Given two matched nodes around a supposed hole, the contour is first opened by removing the edges between these two nodes. Then, the contour is closed by adding a patch between the two nodes by continuity on either side of the hole. The direction of the edges on each side of the hole is retained to extend the outline from both boundary nodes until the intersection of both extensions. This approach may yet be insufficient as it may deal poorly with noise. The direction on either side of the hole is indeed crucial for the smoothing step. Thus, average directions on the neighborhood of each hole boundary might be more robust.

The initial patch is finally smoothed by minimizing the variance of the angles between successive edges within the patch. This smoothing is achieved by minimizing an energy using a gradient descent, where the energy is the angular variances within the patch. One should note that the angles between the outline and the first segments on either side of the hole are included for the computation of the mean angular variation and also for the variance. However, these two segments will remain unchanged during the minimization process in order to insure the continuity with the outline on either side of the hole.

### 3.3.4 Test protocol

Slices extracted from 2 volumes were retained to define the parameters of the approach. In a first step an initial contour is created using symmetric pruning and merging of nodes, with a pruning radius of 5 pixels and a merging radius of 10. Then, candidate boundaries are selected in multiple steps. The angle definition for characterization of boundary nodes was set to  $\frac{3\pi}{4}$ , and a minimal heuristic value of 0.05 was chosen to distinguish relevant holes from normal variations. Finally, the filling process was replaced by a straight edge between the selected pairs of nodes.

The method was then applied to 8 new volumes that were chosen because they contain tumors at diverse locations and of diverse sizes. A binary ROI for the liver was first computed using an existing segmentation engine based on the work by Chemouny (sect. 3.2.1). Then the hole filling process was applied slice by slice and connex component by connex component on each slice of the binary ROI.

Results were evaluated by comparing the obtained envelope with a correct reference. In particular, excesses and defaults were counted, namely parts filled while not being inside the liver and surface cavities that were missed. This quantification



of the results was done by counting the number of correct and incorrect detections of holes. Comparison of volumes was not retained because the last filling step was simplified by a straight line.

### 3.3.5 Discussion, results

Most lesions are filled when the initial binary ROI is of sufficient quality. However this comes at a price; many parts of the images are indeed wrongly filled. The detection of the holes due to surface lesions is correct; 91.7% of lesions are filled (780/851) and this rate increases to 96.5% when partial fills are also counted. However, some regions are mistakenly filled because they look like surface cavities. In particular, the aorta is filled for 80% of tests, and folds in the liver are almost always filled. These overfilling are not too problematic. However, incorrect fillings of the kidneys are more annoying and still common; they are indeed partially filled in 65% of cases.

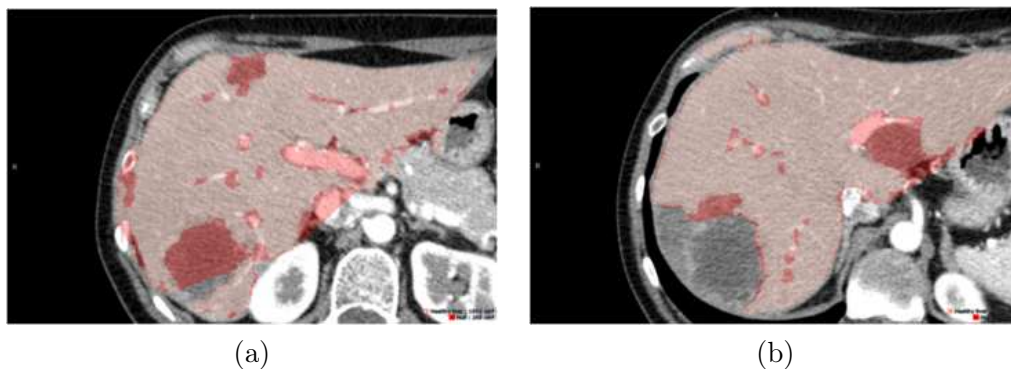


Figure 3.9: Examples of results for the hole filling process. The initial binary ROI and the filled parts are shown as masks on CT images. The former is shown in light pink and the latter in red.

The proposed approach provides correct results in many cases. However this approach suffers from many lacks. First, the approach assumes that holes boundaries are marked. When this is not the case, the hole is missed. Then, the matching sometimes fails, because the condition of direct line of view is not met or because a boundary point is missing as at bottom left of (fig. 3.9.b). Finally, the treatment of separate connex components often prevents a better filling of the ROI, as in (fig. 3.9.a). Thus other approaches should be considered.

## 3.4 Statistical atlas to represent image variability

### 3.4.1 Motivation

#### 3.4.1.1 Introducing statistical atlases

Atlases combine a set of volumes in a single representative object that is named atlas. The simplest possible atlas is an average volume as it represents all volumes

inside a training set. However, such definition is too simplistic to be truly useful. Thus, statistical atlases were introduced to model the statistical variations of the volumes inside a training set.

Statistical atlases may be used to model the intensity distribution at each point of the atlas domain, or to model the variations of shapes of some objects inside a volume. The former was proposed by Glocker *et al.* for a Gaussian model of intensity distribution [Glocker 2007a], and the latter was proposed by many authors either as probabilistic atlases (PAs) or as statistical shape models (SSMs). First, PAs were introduced by many authors in the liver case in order to model the spatial probability of belonging to the liver, functions of the spatial location [Park 2003, Shimizu 2007, Okada 2007, Zhou 2005, Lamecker 2002]. Then, SSMs were introduced to define a representative mesh for the liver boundaries along with its possible variations [Lamecker 2002, Okada 2007].

Every atlas definition requires the knowledge of the spatial location inside a volume. Thus a common spatial basis should be defined to make sense. Several approaches have been proposed that will be ordered by increasing precision. Shimizu and Zhou retained landmark standardization [Shimizu 2007, Zhou 2005], Park chose affine volume mapping [Park 2003], and Okada and Glocker selected nonrigid registration [Okada 2007, Glocker 2007a]. First, landmark standardization aims at mapping a number of characteristics points together in order to align some visible structures. Then, the other methods research a transformation that aligns the images, using only affine transformations in the affine case and transformations with more degrees of freedom in the case of nonrigid registration. Nevertheless, each approach aims at decreasing the spatial variability between the volumes. Thus, only the relevant variations of the structures inside the images are captured and not the spatial variations between the volumes.

As part of this study, the retained statistical atlases will only model the intensity distributions, functions of the spatial location in the atlas. Then, only nonrigid registration will be considered for the spatial normalization as this technique offers the best mapping abilities.

#### 3.4.1.2 Atlas' worth

Because of their ability to model intra-patient variability, atlases are often retained as *a priori* for the segmentation of objects inside images. Statistical atlas are indeed an improvement compared to simple volumes, as they allow capturing the variations within a training set, and thus are more precise than a single volume. This precision contributes to the relevance of an atlas for segmentation, in particular when the region to model is highly variable such as inside the abdomen.

Statistical atlases have already been applied for segmentation. First, Glocker *et al.* constructed a statistical atlas that was applied for cartilage segmentation using nonrigid registration [Glocker 2007a]. Then, Shimizu and Zhou segment the liver inside CT images using a PA for the spatial location of liver along with a pdf for the liver intensity [Shimizu 2007, Zhou 2005]. . .

A new atlas will be introduced that relies on the state of the art nonrigid registration technique retained for Glocker’s atlas [Glocker 2007a], but models the intensity distributions with a more complex statistical model, GMM. First, this atlas models the intensity variations at voxel level, because a single distribution for an entire organ is *a priori* not sufficient for the liver. Then, a state of the art nonrigid registration technique is retained for this atlas, which allows for accurate and quick fusions of images or atlases. Finally, statistical modeling is not done with simple Gaussian distributions, but with Gaussian Mixture Models. Indeed, the different phases of enhancement induce radical changes of intensity ranges that cannot be well described with a Gaussian pdf.

### 3.4.2 Theoretical background

Some background tools and theories will be introduced for subsequent sections. First registration will be presented; registration is at the heart of atlas creation and of many possible clinical applications. Then, a simple example of statistical atlas will be introduced. Finally, theoretical notions required for the definition of the new statistical atlas will be given, namely Gaussian mixtures, k-means and the EM algorithm.

#### 3.4.2.1 Image registration, working on a same basis

##### Definition

Registration is the task of aligning two images on a same spatial basis. Indeed, registration aims at finding a correspondence between two images in order to locate identical structures in both images. The registration can then be used for fusion of both images, namely the simultaneous visualization of both images one atop the other. This fusion has many applications either from a medical standpoint, or in Computer Vision. Fusion and registration will be first introduced through a 2D example. And then, the physical factors that impose registration will be detailed.

Two CT-slices taken from two different patients are registered (fig. 3.10.a,b). Both images are around the same location in the body. However, the images are quite different; liver, stomach and even skin do not have the same shapes. Thus a fusion of both images cannot be directly used for comparison; registration is required. Registration defines a deformation field (fig. 3.10.c) that allows transforming the source image into a deformed one (fig. 3.10.d) that better matches with the target image. A fusion of this deformed image with the target image is then more meaningful as they are more similar than before. However, one may note some artifacts on the left part of the liver and on the stomach. The fusion indeed distorts the image, which might be detrimental to the initial structure of objects (e.g. transformation of one circle onto a square). To prevent such destructive distortions, registration should be done as a balance between similarity with a target image and preservation of the shapes of the structures.

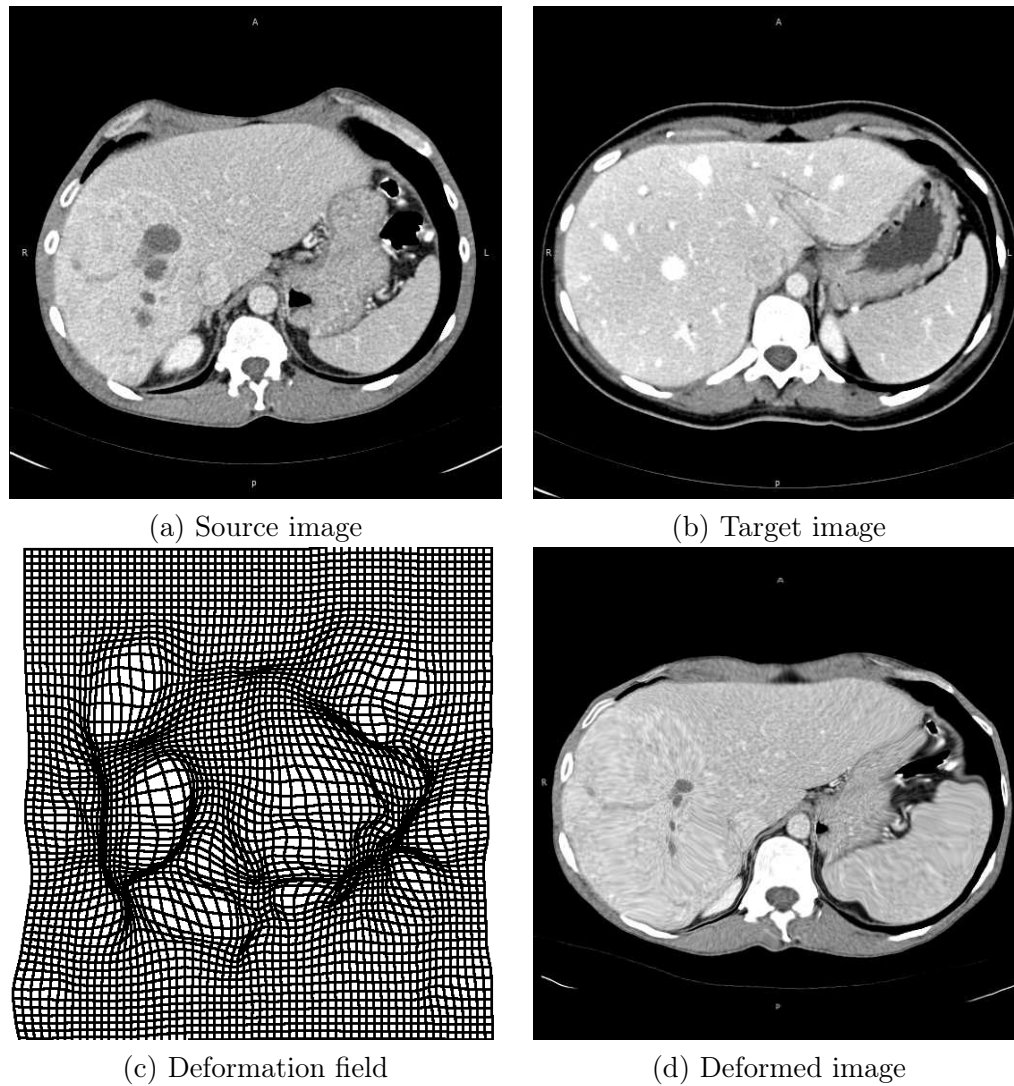


Figure 3.10: Nonrigid registration of two slices of the abdomen. The source image (a) is registered on the target image (b). This registration is done through the computation of a deformation field (c). This deformation field is then applied to the source image in order to define a deformed image (d) that better matches with the target image. For this example, registration was achieved with the drop2D software using SAD as similarity measure [Glocker 2009, Komodakis 2009b].

The spatial coordinates of one voxel cannot be directly used to find the correspondence between two images, both for registration of images from a same patient and for registration of images from two different persons. First, the position of one patient changes between two image acquisitions, because his position inside the imaging machine will never be exactly the same. Then, internal organs move or even change. Because patients breathe, internal organs move. In particular, the movements inside the abdomen are especially large as the abdomen is both close from the lungs and composed mainly of soft tissues. Moreover, organs evolve due to pathologies; for example lesions may grow in the context of this study. Finally, the anatomical structures vary widely because of the anatomical variability between patients.

### Value of registration

Fusion of images has many applications either from a medical or a Computer Vision perspective. First, image fusion eases and improves the follow-up of lesions. Indeed, fusion allows displaying a same lesion for various exams spread over time one atop the other. Thus, the evolution of one lesion is more easily seen. Fusion is also useful when numerous lesions have to be followed up. For example, searching a same lesion inside various images might be difficult for pulmonary nodules. In this case, a fusion brings a time gain as it provides the correspondences between the lesions. Then, registration enables the fusion of images from diverse imaging modalities, which may contribute to better diagnosis. However, multi-modality registration will not be further developed as this study focuses on CT only. Finally, registration provides a way to obtain many images on a same basis, which makes the creation of atlases possible. Atlases are collections of maps that represent an object, either an organ or a volume. These atlases account for the anatomical variability of tissues, while excluding their spatial movements. Thus, registration is required for the creation of these atlases. Moreover, registration is also a crucial step when segmenting the anatomical structures that are modeled by one atlas. Indeed, the image where segmentation is done should be spatially aligned with the atlas for the atlas to have some value.

In this study, the registration domain will be limited to intrinsic dense nonrigid registration methods [Andronache 2006]. First, extrinsic methods rely on artificial markers placed before the image acquisition. These approaches are not relevant for this study. Indeed, obtaining these markers would require a change of the acquisition protocols and would not account for the wide anatomical variations inside the abdomen. Thus, only intrinsic methods are retained, i.e. approaches that use the entire image volume for registration. Then, methods based on voxel similarity are retained. Methods based on landmarks are indeed both less precise and hard to apply to the liver. Moreover, using entire images is more informative than using a few landmarks. Furthermore, the automatic detection of landmarks is difficult for the liver as some boundaries are not well marked and many landmarks disappear between various injection phases. Finally, rigid registration methods are excluded

from this study because they cannot account for the internal movements of organs and the anatomical variations. Indeed, rigid registration consists in finding the 6 degrees of freedom that better match one image onto another (3 translations and 3 rotations). Thus, rigid registration cannot account for the internal movements nor for the anatomical variability.

### Nonrigid registration

While rigid registration is sufficient for many medical applications, nonrigid registration is required for intersubject registration and atlas matching. Indeed, the rigid constraint cannot account for the nonlinear variations between patients. The anatomical variability and the movements due to breathing induce a high variability of the structures between different subjects that cannot be explained through a rigid transformation. Thus nonrigid registration methods are required to create and use a statistical atlas. These methods can be described by three components: a transformation model, a similarity measure and an optimization technique [Wang 2007].

As mentioned before, registration aims to find a transformation  $\mathcal{T}$  that better match a source image  $\mathcal{V}_{src}$  onto a target one  $\mathcal{V}_{trg}$  for the chosen similarity measure. In the field of nonrigid registration, often Free Form Deformations (FFD) based on B-Splines have been introduced as deformation model. The idea is to embed one image into a solid that is then deformed to fit onto another image. One of the main contributions of this approach is the ability to describe complex deformations with only a small set of displacements for a set of control points. Thus, this method has been widely used since its introduction by Rueckert *et al.* [Rueckert 1998].

The similarity measures the adequacy between two voxels, which provides a way to find the optimal transformation between two images. This optimal transformation is found when distance between the target and the deformed image is minimal. While for landmark registration this distance may be simply defined as the Euclidean distance between the landmarks, dense registration requires more complex distances defined between voxels. Thus, similarity measures are introduced for dense registration as distances between two voxels. Two different measures are considered, SAD (Sum of Absolute Differences) and MI (Mutual Information). SAD is the simplest similarity measure, defined for each pair of voxels as the absolute difference of intensity between the two voxels. For this measure to remain relevant, the intensity information should remain similar between the two images. Thus this distance is relevant for registration of images of same modality. On the opposite, MI is a similarity measure designed for registration of multimodal images that was introduced almost simultaneously by Viola and Maes [Viola 1995, Maes 1996]. Mutual Information comes from the information theory and measures in this particular context the statistical dependences between two voxels. As no assumption is done for this measure, MI applies well whatever the characteristics of both images.

The optimization technique aims to compute the optimal transformation following the chosen transformation model, by minimizing the chosen similarity measure. This step is crucial, because the quality of the optimization impacts the overall

quality of the transformation. Moreover, from the speed of the optimization depends the clinical prospect of the registration. A recent improvement by Glocker *et al.* using a multiscale approach along with graph-cuts resolution attracted lots of interest because of its ability to provide good registration accuracy along with small computation times [Glocker 2008, Glocker 2007b]. In this approach the registration problem is expressed as a MRF energy that is then optimized using a precise and fast algorithm. This method will be detailed in a subsequent section (sect. 3.5.3.1).

### 3.4.2.2 Statistical atlas as one Gaussian per pixel

A statistical atlas aims to model the statistical variations of intensities inside a volume, in order to represent as well as possible the structures inside this volume. The choice of a statistical atlas against a simple mean volume is justified by the improvement brought. The statistical atlas indeed allows capturing the variations within a set of images, when an average volume would be too simplistic.

In order to model only the anatomical variations to the exclusion of spatial variations, statistical atlases have to be computed on a set of registered volumes. As seen in previous section (sect. 3.4.2.1) registration defines a transformation that matches one image onto another. Such transformation allows removing spatial variations between the images. Moreover the registration allows keeping only the relevant anatomical variations; the small anatomical variations and the changes of locations due to the breathing should indeed be removed or lessened by the registration. Thus a set of registered volume  $\mathcal{W} = \{\mathcal{V}_1, \dots, \mathcal{V}_n\}$  is introduced to create an atlas by modeling the intensity variation within the training set and for each location inside the volumes.

In a previous paper Glocker *et al.* constructed a statistical atlas on such a set of registered volumes, or training set, using a normal distribution of intensity as model, before carrying out cartilage segmentation [Glocker 2007a]. For the given set of registered volume, Glocker *et al.* construct an atlas that gives a Probability Density Function (pdf) for each voxel  $\mathbf{x}$  of the volume. This atlas is composed of an optimal representative volume and a variance map, which defines a statistical atlas with one Gaussian pdf per voxel. The optimal representative volume  $\mathcal{V}_{\mathcal{M}} : \Omega \rightarrow \mathbb{R}$  defines the mean value for each point of the atlas. And the variance map  $\sigma_{\mathcal{M}} : \Omega \rightarrow \mathbb{R}$  gives the deviation between the optimal volume and the training set. These two volumes contain the parameters for the distribution models on each point of the space. For each voxel  $\mathbf{x} \in \Omega$  the atlas defines a statistical distribution for this location as a normal distribution with a mean  $\mathcal{V}_{\mathcal{M}}(\mathbf{x})$  and a standard deviation  $\sigma_{\mathcal{M}}(\mathbf{x})$ .

$$\begin{aligned} p_{\mathbf{x}}(i) &= \mathcal{N}(\mathcal{V}_{\mathcal{M}}(\mathbf{x}), \sigma_{\mathcal{M}}(\mathbf{x})) \\ &= \frac{1}{\sqrt{2\pi}\sigma_{\mathcal{M}}(\mathbf{x})} e^{-\frac{(i-\mathcal{V}_{\mathcal{M}}(\mathbf{x}))^2}{2(\sigma_{\mathcal{M}}(\mathbf{x}))^2}} \end{aligned} \quad (3.1)$$

### 3.4.2.3 Gaussian Mixture Models

Gaussian Mixtures Models (GMMs) are models expressed as weighted sums of Gaussian distributions. They may be regarded either as statistical models for distributions created by clustered data, or as statistical clustering methods. From a mathematical point of view these Gaussian Mixture Models (eq. 3.2) are defined as weighted sums of Gaussian (or normal) distributions (eq. 3.3).

$$f(x) = \sum_{i=1}^l \pi_i \mathcal{N}(\mu_i, \sigma_i) \quad (3.2)$$

$$\mathcal{N}(\mu, \sigma) : x \rightarrow \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (3.3)$$

In medical imaging, many tissues with very different intensity ranges are encountered, thus GMMs are often used to model intensity distributions. In such cases, the normal distribution  $\mathcal{N}(\mu_i, \sigma_i)$  defines the intensity distribution for one cluster, and the weight  $\pi_i$  gives the probability of being inside this cluster. For example, the distribution shown in (fig. 3.11) is composed of three clusters with diverse weights. The first cluster (dashed blue) has a higher probability of happening than the red dotted one.

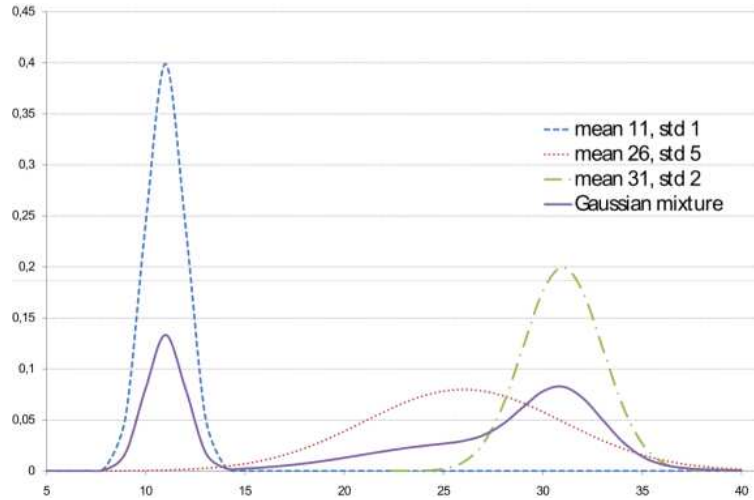


Figure 3.11: Intensity distribution of one GMM along with its three components. The global distribution (solid purple) is defined as a weighted sum of three normal distributions with diverse parameters. Each basis distribution is related to a different type of object (or tissue).

GMMs may be used in the opposite way too, namely given a data sample one wishes to find the GMM that best fit with this sample. This approach will model the data as coming from diverse modes, while defining the importance of each mode. This clustering ability is illustrated for a sample distribution composed of two different clusters (fig. 3.12.a). Fitting a GMM on such sample means defining each cluster



by its distribution  $\mathcal{N}(\mu_i, \sigma_i)$  as well as by its relative probability  $\pi_i$ . The difficulty lies with the computation of the parameters of the GMM (fig. 3.12.b). Indeed, the number of components for the GMM and the parameters of each component are not easily found. These parameters are nevertheless obtained for a set number of components thanks to the Expectation Maximization algorithm (EM) (sect. 3.4.2.5). Moreover, a measure will be introduced that allows defining the optimal number of components for a specific data sample (sect. 3.4.2.6).

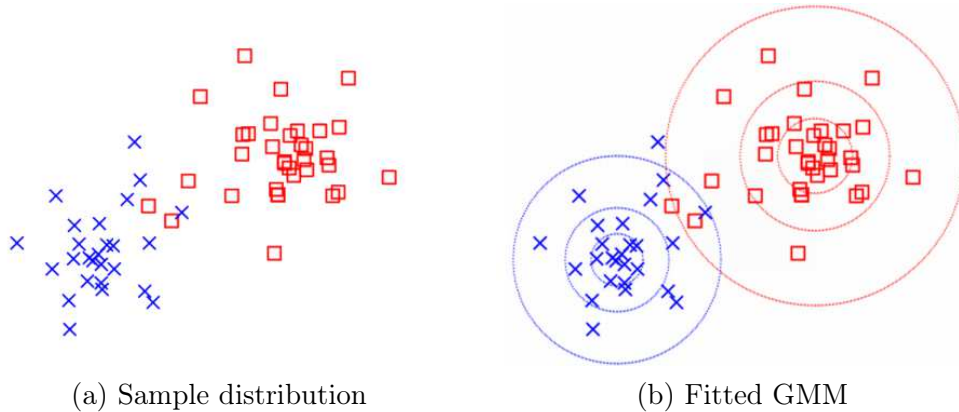


Figure 3.12: GMM for clustering. An artificial sample distribution contains spatial locations of two distinct modes shown as blue crosses and red squares (a). By fitting a GMM with 2 components on this data, the two modes may be retrieved (b).

#### 3.4.2.4 K-means

K-means is an algorithm that constructs a partition of a population into  $k$  clusters by minimizing the sum of squares within clusters. This simple algorithm was introduced by MacQueen in 1967 and is still widely used because of its good clustering ability and its simplicity [MacQueen 1967]. In particular k-means is often used as initialization for other clustering algorithms like the EM one. The main idea is to place  $k$  centroids and to iteratively alternate two steps. First, all points of the population are assigned to the closest centroid, which defines a partition for each centroid. Then, the centroids are updated to take in account the new partitions.

Given a set of  $n$  observations  $x = \{x_1, \dots, x_n\}$ , k-means clustering aims to partition the observations  $x$  into  $k$  sets ( $k < n$ ). This partition  $C = \{C_1, \dots, C_k\}$  is obtained by minimizing an objective function defined as the sum of squares within clusters (eq. 3.4). This minimization aims to define  $k$  clusters  $C_i$  where sample points remain close from their centroids  $\mu_i$  for the distance function  $\|\cdot\|$ .

$$\operatorname{argmin}_C \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2 \quad (3.4)$$

The optimal partition  $C$  is often a NP-hard problem, thus a heuristic approach was developed to find an approximate partition through an iterative refinement

approach [Mahajan 2009, Dasgupta 2008]. This approach iteratively consists of one assignment step followed by an update of centroids. The first step aims at defining each cluster  $C_i$  of the partition by assigning every observation to the cluster with the closest centroid (fig. 3.13.b,c). The second part aims at updating centroids according to the new partitions (fig. 3.13.c,d).

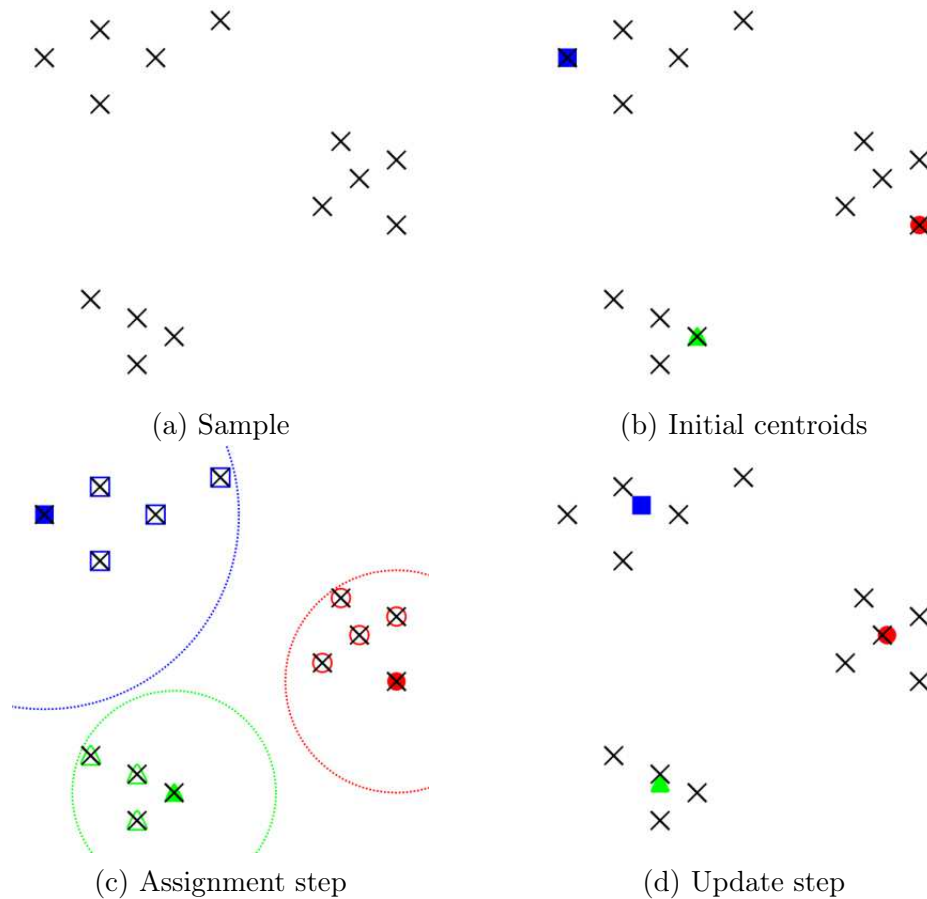


Figure 3.13: One iteration for the k-means algorithm. Observations are shown as crosses. Clusters are shown as circles, squares and triangles. The initial observations are shown in (a). First an initial pick of centroids is defined by taking some observations as cluster centroids (b). Then, each observation is assigned to the cluster with the closest centroid (c). Finally, new centroids are computed in (d), which may be different from the existing observations.

Cluster centroids  $\{\mu_1^{(t)}, \dots, \mu_k^{(t)}\}$  being known for step  $t$ , the new clusters  $C_i^{(t+1)}$  are first defined by assigning each observation to the cluster with the closest centroid (eq. 3.5).

$$C_i^{(t+1)} = \left\{ x_j \in x : \left\| x_j - \mu_i^{(t)} \right\| < \min_{i \neq k} \left\| x_j - \mu_k^{(t)} \right\| \right\} \quad (3.5)$$

Then, the centroids are updated with the components of the new clusters (eq. 3.6).

And finally, the optimal partition for this heuristic is obtained when assignments no longer change.

$$\mu_i^{(t+1)} = \frac{1}{|C_i^{(t+1)}|} \sum_{x_j \in C_i^{(t+1)}} x_j \quad (3.6)$$

This method has two main drawbacks. First, the heuristic approach does not guarantee the optimality of the solution. Thus, the obtained partition may be sub-optimal. Moreover, the solution is dependent of the initialization. Then, a wrong choice of  $k$  may lead to incorrect results. The choice of the initial centroids is not detailed, but is important, because the entire technique depends on this initialization; different initializations may indeed give different partitions. Thus the algorithm is often applied more than once, with different random picks inside observations as initialization. The centroids may also be initialized by taking observations far from each other. The number of clusters is also a source of errors as a wrong choice of  $k$  may lead to poor results. When the number of modes inside the sample differs from  $k$ , clusters will indeed not match with the true modes.

### 3.4.2.5 Expectation Maximization algorithm

#### Expectation Maximization

The Expectation Maximization algorithm (EM) is an efficient algorithm to estimate the parameters of a model for which some observations are most likely. The EM algorithm was introduced in 1977 by Dempster to estimate the parameters that maximize the likelihood on incomplete data, and has been widely used since [Dempster 1977]. This algorithm is composed of two steps, an E-step and an M-step. First the E-step aims to estimate the missing data using the observations and the current estimates for the model parameters. Then the M-step relies on these estimates of missing data to maximize the likelihood function.

The EM algorithm seems similar to the k-means algorithm. Indeed, they both apply to the problem of clustering data. However, there are several major differences. First, with k-means, data is clustered without prior knowledge, whereas the model to optimize introduces a major *a priori* for the EM algorithm. One may argue that the distance function inside k-means is indeed a prior knowledge, but this distance is a weak *a priori*. Then, with k-means every observations is explicitly assigned to a single cluster, while EM deals with probabilistic assignments of both observations and hidden observations. Finally, the quality of the final results differs. With k-means a heuristic approach is followed with no guarantee on the final solution. On the opposite the EM approach guarantees the local optimality of the solution.

Given a set of  $n$  observations  $x = \{x_1, \dots, x_n\}$  and a proposed model  $p(x | \Theta)$ , where  $\Theta$  are the model parameters, the EM algorithm aims to find a set of parameters  $\Theta$  that maximizes the likelihood of the model  $p(x | \Theta)$  for the observations  $x$ . The specificity of the EM approach comes from the addition of hidden observations. Indeed, a small set of observations cannot describe entirely one model; some missing

observations may be important. The EM algorithm aims to take these missing observations in account by adding some hidden observations  $z = \{z_1, \dots, z_m\}$  that are estimated through the algorithm. The problem then becomes the maximization of the incomplete-data log likelihood (eq. 3.7), i.e. the maximization of the likelihood of the observations and the hidden ones.

$$\operatorname{argmax}_{\Theta} p(x | \Theta) = \operatorname{argmax}_{\Theta} \left[ \log \sum_z p(x, z | \Theta) \right] \quad (3.7)$$

The maximization of the log likelihood is done by a coordinate ascent by maximizing the expected complete log likelihood  $Q(\Theta | \Theta^{(t)})$ . This maximization is done iteratively by alternating an E-step and an M-step. First, the E-step aims to estimate the conditional likelihood of the hidden data given the observations  $x$  and the estimation of parameters at step  $t$   $\Theta^{(t)}$  (eq. 3.8). Then, the M-step aims to update the model parameters, by maximizing the expected complete likelihood (eq. 3.9).

$$Q(\Theta | \Theta^{(t)}) = E_{p(z | x, \Theta^{(t)})} [\log p(x, z | \Theta)] \quad (3.8)$$

$$\Theta^{(t+1)} = \operatorname{argmax}_{\Theta} Q(\Theta | \Theta^{(t)}) \quad (3.9)$$

The EM algorithm requires some initial estimates for the parameters that will impact the final result. Such estimation is often done using the k-means algorithm because of its speed and robustness.

### Expectation Maximization for Gaussian Mixture Models

The EM algorithm applies to GMMs, for which an interesting property appears. For GMMs the two steps of the EM algorithm may indeed be done at once because an analytical solution exists, where all model parameters can be directly computed from previous estimated parameters using the posterior distributions.

Given a set of  $n$  observations  $x = \{x_1, \dots, x_n\}$  the EM algorithm is applied to a Gaussian Mixture Model  $p(x | \Theta) = \sum_{i=1}^l \pi_i \mathcal{N}(\mu_i, \sigma_i)$ , where the model parameters  $\Theta$  are defined by  $\Theta = (\pi_i, \mu_i, \sigma_i)_{i \in [1, l]}$ . The parameters of this model are iteratively computed using only the previous parameters and the *a posteriori* distributions  $p(i | x_j, \Theta^{(t)})$  for each mixture component, given the observations and the previous model parameters (eq. 3.10).

$$\begin{aligned}
\pi_i^{(t+1)} &= \frac{1}{n} \sum_{j=1}^n p_{i,j}^{(t)} \\
\mu_i^{(t+1)} &= \frac{\sum_{j=1}^n x_j p_{i,j}^{(t)}}{\sum_{j=1}^n p_{i,j}^{(t)}} \\
\sigma_i^{(t+1)} &= \frac{\sum_{j=1}^n p_{i,j}^{(t)} (x_j - \mu_i^{(t+1)})^2}{\sum_{j=1}^n p_{i,j}^{(t)}}
\end{aligned} \tag{3.10}$$

where  $\forall (i, j) \in [1, l] \times [1, n]$   $p_{i,j}^{(t)} = p(i | x_j, \Theta^{(t)})$

### 3.4.2.6 Minimum Description Length

The Minimum Description Length (MDL) introduces a way to find the optimal number of clusters that represents some data. In particular, for GMMs, MDL relates with the computation of the optimal number of components to represent some data.

Given a set number of mixture components, one may easily fit a Gaussian mixture on a sample distribution  $x$ , using the Expectation Maximum algorithm. However, the choice of an optimal number of Gaussian components  $l_x$  for a sample  $x$  is not straightforward. The model should indeed satisfy two constraints. The model for the sample  $x$  should reflect as much as possible the intrinsic properties of the sample  $x$ , while being the shortest possible one. These constraints mean that the model should represent the sample as well as possible, while remaining robust to the noise inside the sample. The best model should also use the least possible parameters to code the data, which allows decreasing computation time and memory consumption, as well as preventing overfitting.

Rissanen solved this problem by the introduction of the Minimum Description Length (MDL) [Barron 1998]. He described a way to find the optimal number of clusters that represents some data, meaning the shortest description that represents some samples, while taking in account both the model and its error. The MDL is defined as follows and allows defining the best model for one sample, as the one with the smallest MDL.

$$MDL = \min_{k, \Theta} -\log(p(X | \Theta)) + \frac{1}{2}k \times \log(n) \tag{3.11}$$

where the first component is the log likelihood of the model given a set of parameters  $\Theta$ , and the second one a penalty proportional to the number of parameters for the model ( $k$ ) multiplied by the log size of the sample ( $n$ ). This definition was applied to GMMs by Kyrgyzov *et al.*, who proposed an analytic formula for the MDL value when using Gaussian Mixture Models [Kyrgyzov 2007].

$$\Lambda(l) = -\frac{1}{2} \sum_{j=1}^l n_j \log\left(\frac{n_j^2}{|\sigma_j|}\right) + \frac{2n + 3l - 1}{2} \log(n) - \frac{n}{2} \log(2\pi) - \frac{n^2}{2} \tag{3.12}$$

where  $l$  is the length of the Gaussian mixture,  $n_j$  is the number of samples belonging to the  $j^{\text{th}}$  Gaussian component of the mixture. This last number ( $n_j$ ) is the posteriori computed while fitting the mixture on the sample  $x$  with the EM algorithm.

The research of an optimal model for a same sample  $x$  and using only GMMs may then be restated as a simpler minimization problem. The number of observations indeed becomes a constant because the sample distribution does not change.

$$\Lambda'(l) = - \sum_{j=1}^l n_j \log\left(\frac{n_j^2}{|\sigma_j|}\right) + 3l \cdot \log(n) \quad (3.13)$$

### 3.4.3 Creation of a statistical atlas

#### 3.4.3.1 A simple example

Atlases aim at defining a representative model for a set of images. This modeling should be done on registered images in order to remove the variations due the spatial changes between the images. First, a mean image would be the simplest model for the variations within an image set. However, such representation cannot account for complex images. Thus, statistical atlases were introduced to improve the modeling of images, first using Gaussian models and then using GMMs. Statistical atlases will now be introduced through a simple example, while explaining some choices and pointing some flaws of the models.

Atlases should be computed on registered sets of images. Otherwise the models become more dependent on the spatial variations between the images than on the variations of the structures inside the images. This constraint is shown for an artificial data set (fig. 3.14). A small data set was created using ellipses on images with two gray levels. A registered set of images was then created by registering each image onto a same target. Two representative images were computed next as mean images, the first one on the raw images (fig. 3.14.a), and the second one on the registered images (fig. 3.14.b). The mean image computed on the source images displays a high variability around the location of the ellipses due to the diverse shapes of the ellipses. On the contrary, the mean image computed on registered images shows little variability due to the possible shapes of the ellipsis. This feature explains why atlases are computed on registered sets. This mean image also shows that mean images cannot well model a set of images, as the intensity difference between the ellipsis and the background is null inside the mean image computed from the registered set (fig. 3.14.b). Thus, statistical atlases were introduced to improve the modeling of intensities inside the image set.

Statistical atlases were introduced to improve the modeling of intensity distributions compared to a mean image. Two types of statistical atlases were considered and applied to the previous set of registered images (fig. 3.15). The first atlas follows the definition proposed by Glocker *et al.* (sect. 3.4.2.2) [Glocker 2007a]. This atlas indeed models the intensity distributions using Gaussian pdf. Then, the second one is an improvement and a generalization of this last method, which models the

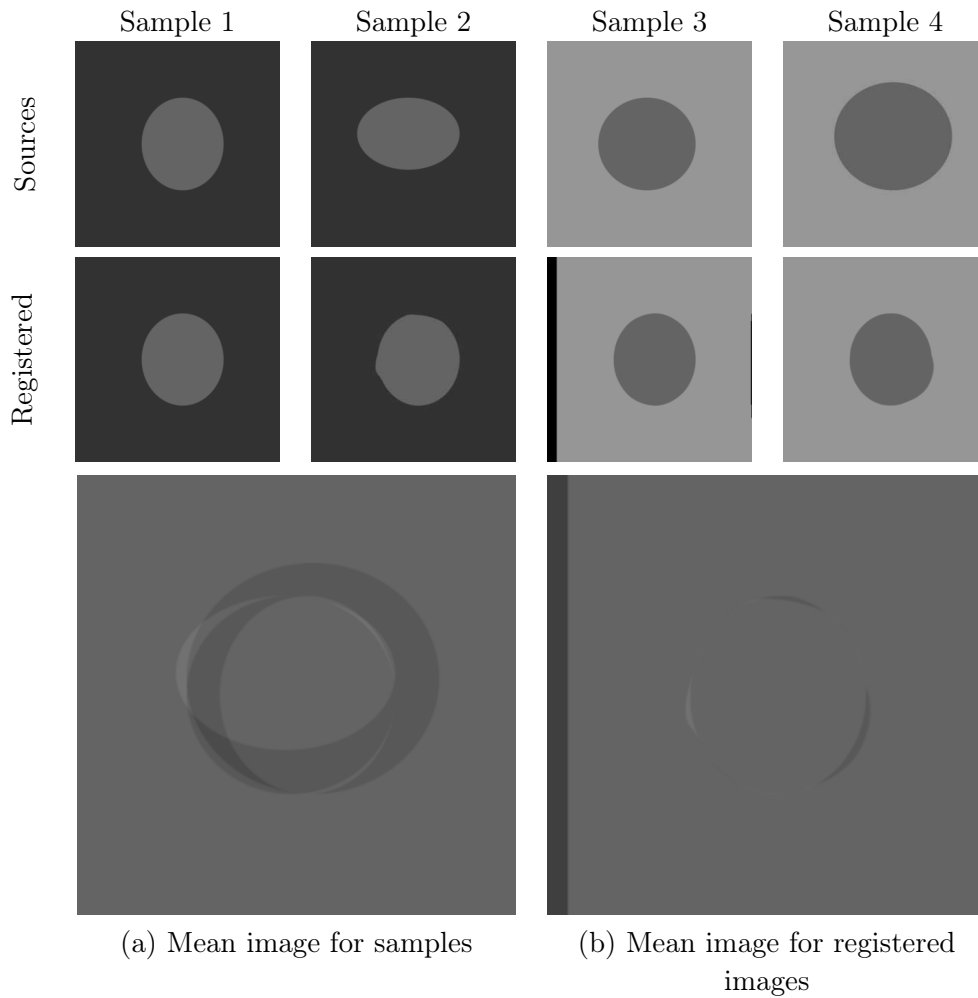


Figure 3.14: Atlases as mean images created on raw and registered image sets. Elliptic phantoms are considered to create an atlas as a representative image. A mean image computed on the raw samples is first shown (a). Then a mean image is computed on a registered set (b). The registered set was created using the first sample as target image during the registration using drop2D with NMI or SAD as similarity measure [Glocker 2009].

intensities using GMMs. These atlas definitions are finally compared for the previous set of registered images. First, the imperfections of registration give atlases that are composed of five regions with diverse intensity distributions, including two significant parts and three smaller regions. Then, statistical atlases offer better differentiation abilities than mean images. Indeed, with statistical atlases, the intensity distributions differ between the ellipsis and the background. However, a Gaussian model for the distribution is not sufficient to get a good difference between background and object, as both distributions cover a same intensity range (fig. 3.15.a). On the opposite, GMMs allow handling cases with multiple modes. Thus the difference between background and object becomes more marked (fig. 3.15.b). However, using GMMs introduce an additional risk because non-significant modes might be modeled.

### 3.4.3.2 Definition

A statistical atlas is introduced as one optimal Gaussian mixture per voxel. This construction improves a statistical atlas defined as one Gaussian distribution per voxel (sect. 3.4.2.2). Such an atlas indeed provides a better modeling of intensity distributions without overfitting. This definition is also a generalization of the statistical atlas with one Gaussian pdf per voxel, as a Gaussian distribution is simply a GMM with a single component.

Modeling the intensity distributions as GMMs instead of Gaussians allows better modeling. Moreover, the condition of optimality prevents overfitting. First, Gaussian distributions cannot well describe every intensity distribution inside the abdomen. For example, intensities vary widely inside blood vessels, functions of the injection phase. Thus, a normal distribution cannot account for the various cases. On the opposite GMMs can. They indeed allow setting different components corresponding to the different phases. Then, GMMs are added with low risk of overfitting because of the optimality constraint. Indeed, while GMM offers better modeling abilities, this could come to the cost of overfitting, namely modeling even the random variations of intensity. Thus only optimal GMMs are retained to model the intensity distributions. Optimality is defined following Occam's razor, as the simplest model that matches well with the data in order to minimize the risk of overfitting. The optimality is thus defined by the MDL (sect. 3.4.2.6).

From a mathematical perspective, the atlas  $\mathcal{A}$  is defined as a function from a volume  $\Omega$  to a space of Gaussian Mixtures  $\Xi$  that maps voxels  $\mathbf{x}$  to GMMs of length  $l_{\mathbf{x}}$ . A GMM is defined for each voxel of the volume  $\mathbf{x} \in \Omega$ , which models the intensity distribution at the location of the voxel  $\mathbf{x}$  inside space. Each GMM is guaranteed to be optimal by the selection of the best model as defined by the MDL.

$$\mathcal{A} : \begin{cases} \Omega & \longrightarrow \Xi \\ \mathbf{x} & \longrightarrow p_{\mathbf{x}} = \sum_{i=1}^{l_{\mathbf{x}}} \pi_{\mathbf{x},i} \mathcal{N}(\mu_{\mathbf{x},i}, \sigma_{\mathbf{x},i}) \end{cases} \quad (3.14)$$



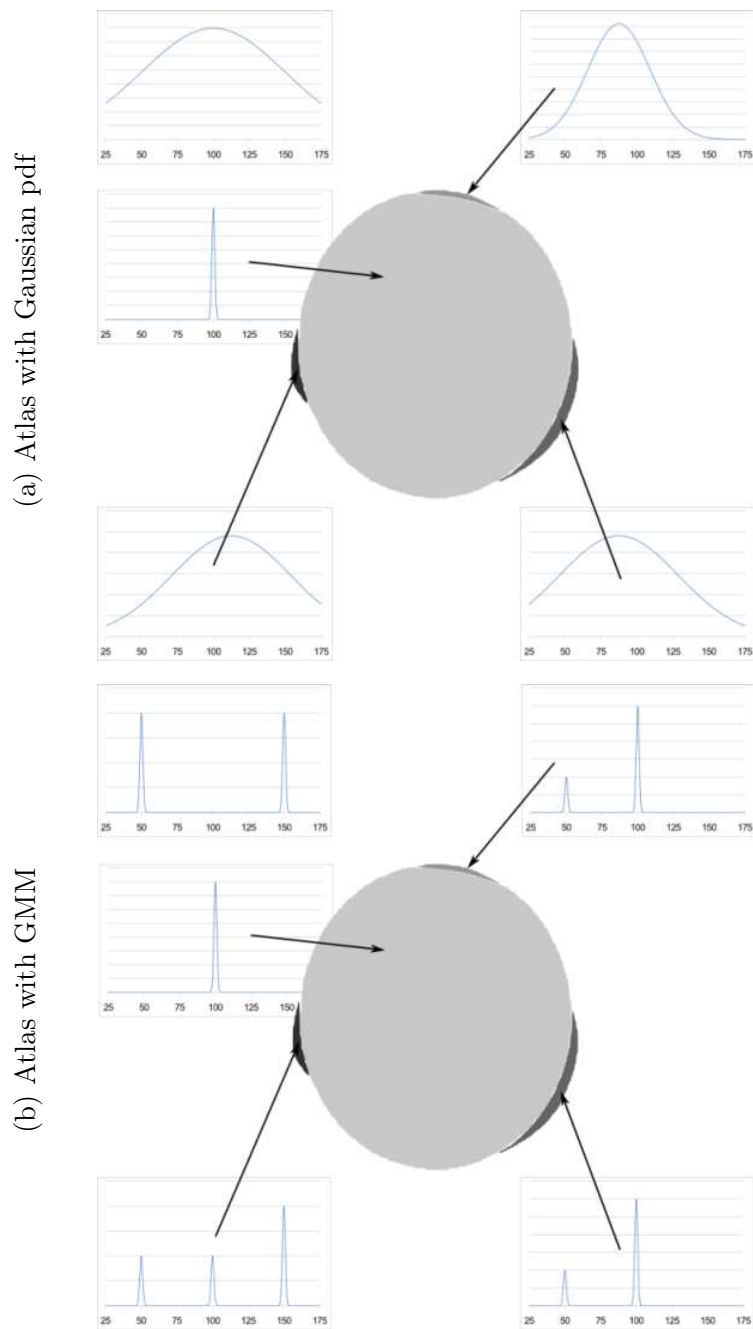


Figure 3.15: Comparison of two definitions for statistical atlases. Two atlases are computed for the previous set of registered ellipses (fig. 3.14). First a statistical atlas defined with one Gaussian pdf per pixel is shown (a). Then an atlas defined with GMMs is displayed (b). For both atlases the intensity distribution is shown for each part of the image.

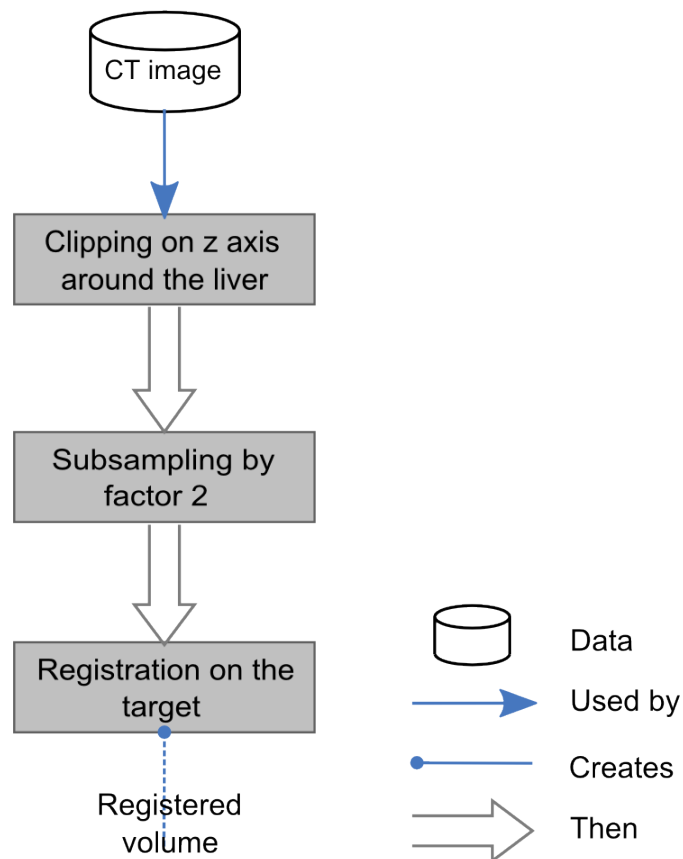


Figure 3.16: Creation of registered volumes for use during atlas construction. Several volumes are registered on a same target. All these volumes are first clipped around the liver region in order to simplify and speed-up the process. Then, they are subsampled by a factor 2 for performance reasons (both because of reduced memory consumption and additional speed). Finally, the subsampled volumes are registered on a same target and in a soft manner.

### 3.4.3.3 Construction

The atlas is defined at each voxel of one volume by computing the optimal GMM that models an intensity sample extracted from a set of registered images. A set of  $n$  registered volumes  $\mathcal{W} = \{\mathcal{V}_1, \dots, \mathcal{V}_n\}$  is first created. Registered volumes are indeed required in order to remove the spatial variability inside the CT images. This set defines an intensity sample for each voxel of the volume that is then used to compute a good intensity model at each voxel.

First, a set of registered volumes is created by registering volumes on a same target (fig. 3.16). This registration may be done using any method as long as the registration is nonrigid and keeps the relative positions of organs. First, registration should be nonrigid because the atlas aims at modeling the anatomical variations between medical images. Indeed, rigid or no registration would not remove the spatial variations between the images. Thus, atlases computed on such sets would not have any use for the characterization of anatomical variations. Then, the registration should be done while insuring that the deformation field remains smooth and without folds. Otherwise the boundaries between organs would be modified, which would create artifacts in the training set and possibly inside the atlas.

Then, a pdf is defined at each voxel  $\mathbf{x}$  of the volume as an optimal Gaussian mixture (fig. 3.17). At each voxel  $\mathbf{x}$ , the set of registered volumes  $\mathcal{W}$  defines an intensity sample that is used to compute the optimal GMM. The EM algorithm is first used to fit GMMs with various sizes on the sample. Then, the GMM with the smallest MDL value is retained to model the intensity distribution at voxel  $\mathbf{x}$ .

### 3.4.3.4 Implementation issues

The computation of the optimal GMM was simplified by the introduction of a maximum length  $l_{max}$  for GMMs. This parameter simplifies the research of the optimal MDL by setting a finite number of possible candidates. Given a sample, GMMs are fitted for every length between 1 and  $l_{max}$  using the EM algorithm after initialization with the k-means algorithm. Then, the MDL value is computed for each GMM, which provides the optimal GMM.

Degenerate distributions may sometimes be introduced through the fitting process of GMMs. Thus, some post-treatments are added to remove Dirac's components of GMMs (components with 0 standard deviation). Moreover, a default model is introduced to model the intensity distribution inside background in order to keep a constant background. First, fitting GMM may sometimes introduce components with 0 standard deviation. These components are modified with a set standard deviation in order to keep continuous distributions. Then, the voxels in background are given a default GMM. The intensities outside the body are indeed small, but also very noisy. Thus, the optimal GMM within this region often contains many components that are not relevant. Therefore, these GMMs are replaced by a default GMM in order to decrease memory consumption, speed-up the evaluation of GMMs and improve the robustness to changes of intensity inside the background.

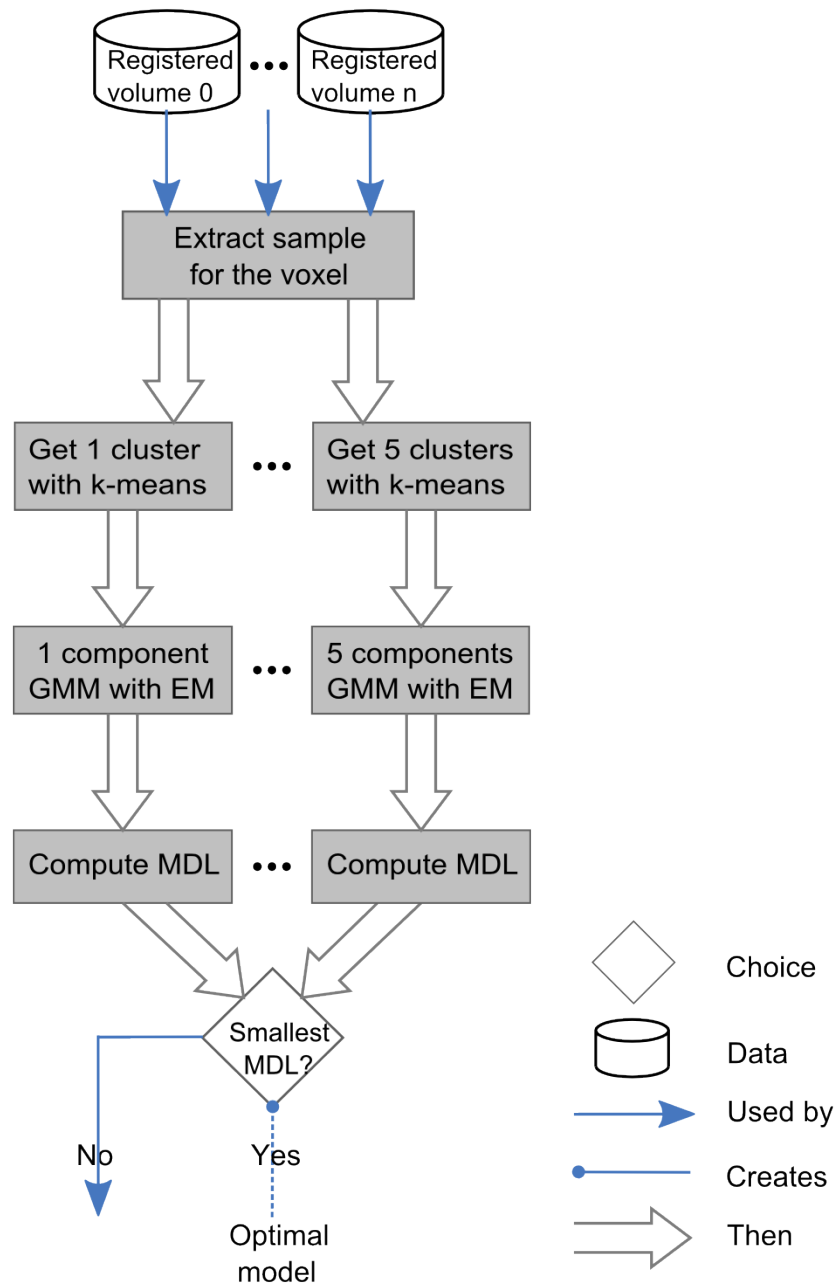


Figure 3.17: Creation of an atlas from a set of registered volumes. Given a set of  $n$  volumes registered on a same target, an intensity sample is first extracted for each voxel of the space. Then, each sample is modeled by GMM with exactly 1, 2, 3, 4 or 5 components. The optimal model at each spatial location is finally decided by the MDL measure.

### 3.4.3.5 Method

A set of registered volumes was created by registering a set of clipped and subsampled volumes on a same target volume. First, images were clipped to contain only a region around the liver. Then, images were subsampled by a factor 2 on each axis. Finally, registration was achieved with the drop software developed by Glocker [Glocker 2009].

In a first step CT volumes are clipped to contain only a region around the liver. This step aims at simplifying the problem by working on a smaller volume by excluding non relevant parts for this study (higher part of lungs, legs...). This process is not too restrictive and may be automated. First, this clip is done without trying to obtain a precise region in order to avoid introducing a bias in the registration due to repetitive locations. Then, this clip is indeed one of the standard views for this kind of medical image. For follow-up or detection of liver tumors, CT scans of the abdomen are the standard proceeding, along with thorax or pelvis in some cases. Finally, transforming any problem into this simpler one is not difficult. The upper part of the liver may be easily found by searching for the lungs and the heart, and the bottom part may be found using the hips. Indeed, several authors described methods to go back to this smaller volume [Soler 2001, Zhou 2005, Shimizu 2007, Choudhary 2008].

Original images were subsampled by a factor 2 on each axis. Original images were of various sizes, typically 512x512x., and subsampled to 256x256x. The contribution of this subsampling is twofold. First, it simplifies the creation of the initial set of registered volumes. Then, it decreases the memory required to load an atlas. Indeed, this subsampling decreases the memory requirements of the atlas by a magnitude of 8. This extra memory will later benefit to segmentation with an atlas.

The set of registered images is finally obtained by registering a set of raw images on a well chosen image. The choice of the target image introduces a bias on the set of registered volumes. Thus, this target volume should be a well chosen representative volume in order to minimize the bias. An uncommon target image would indeed introduce a huge bias, because registrations would become less precise. Consequently, many errors would be introduced in the atlas.

The computation of each optimal GMM is achieved through the computation of the GMM with the smallest MDL for a sample extracted from the set of registered images. First, an intensity sample is created for each voxel of the target image used to create the set of registered images. This sample is extracted from the registered set of images by taking all intensities for the same voxel within the set. Then, a GMM is fitted on each sample by the EM algorithm with an initialization with the k-means algorithm. Fitting is done for every possible number of components between 1 and  $l_{max}$ . Next, MDL is computed for each GMM, and the GMM with the smallest MDL value is chosen as optimal GMM. Finally, the optimal GMM is parsed to replace Dirac's components by normal distributions with a small standard deviation and to replace GMMs in the background of the image by a default distribution.

## 3.5 Segmentation, atlas based

### 3.5.1 Intro

Atlases may be used for the segmentation of images. Given a reference segmentation  $\mathcal{A}_{seg}$  for an atlas  $\mathcal{A}$ , any new image  $\mathcal{V}_i$  may be segmented by applying a deformation field to the reference segmentation  $\mathcal{A}_{seg}$ , where the deformation field is obtained while registering the atlas onto the new image  $\mathcal{V}_i$ . The atlas  $\mathcal{A}$  is first registered on a volume  $\mathcal{V}_i$ . This registration defines a deformation field that matches the atlas on the new volume. Then, this deformation field is applied to the reference segmentation, which gives the location of the same segmented structure inside the new volume  $\mathcal{V}_i$ . Thus, the deformed reference segmentation provides the segmentation inside the new image provided that the registration is correct.

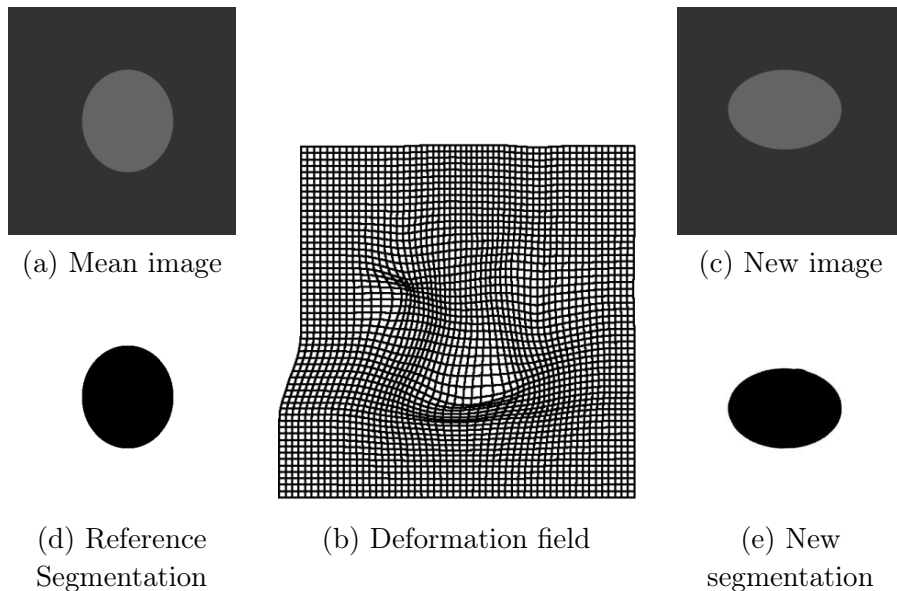


Figure 3.18: Segmentation through registration of a representative image. A representative image (a) is registered onto a new image (c) through a transformation defined by a deformation field (b). The segmentation of the new image (e) is then obtained by applying this deformation field on a reference segmentation for the representative image (d).

Segmentation through atlas registration will be first illustrated for a simplistic representative volume on artificial images (fig. 3.18). A mean volume  $\mathcal{V}_M$  and a reference segmentation  $\mathcal{V}_{seg}$  for the liver are introduced for the artificial set used as example at the beginning of the previous section (fig. 3.14). First, the mean volume (fig. 3.18.a) is registered onto a new image (fig. 3.18.c) through a deformation field (fig. 3.18.b). Then, this deformation field is applied to the segmentation reference (fig. 3.18.d), which defines a segmentation for the new volume (fig. 3.18.e). However, a mean volume as atlas is not very promising, because it uses only one reference, and does not take in account the variability of the object to segment.

Even for simple cases mean images do not well model the variability of one volume (fig. 3.14). As an organ with a very high degree of variability, even more in the presence of tumors, one should not wait too much from mean volumes in the liver case.

Segmentation through registration of statistical atlases is introduced to correct the lacks of mean volumes as atlases. Indeed, mean volumes do not take in account the variations of the structures, nor the information given by the training set used to create the atlas. In fact this previous approach amounts to a simple registration between two images; only one of the images is an artificial one in an attempt to model the diverse appearances of one object. Thus, the registration of a statistical atlas on an image volume is introduced to take in account the variations of the images modeled by the atlas. This registration is done following the approach developed by Glocker *et al.* and adapted to the proposed atlas model [Glocker 2007a, Glocker 2008]. First, the MRFs that are used to achieve registration will be introduced. Then, the registration approach proposed by Glocker *et al.* will be reviewed, before being adapted to the proposed atlas. Finally, the approach will be evaluated with some tests.

### 3.5.2 MRF for image segmentation

A Markov Random Field (MRF) is a graphical model with many applications, in particular in Computer Vision. According to Kindermann, MRFs were introduced by works of Preston and Spitzer as a generalization of an older model, the Ising's model [Kindermann 1980, Preston 1974, Spitzer 1971, Ising 1925]. MRFs aim to model a set of variables that take discrete values, where the values are dependent only from the neighbors variables (two variables are independent when they are not neighbors). MRFs will be first formally introduced. Then, some examples of use will be given. Finally, optimization techniques on MRFs will be reviewed with a focus on the method by Komodakis *et al.* that is the current state of the art [Komodakis 2008].

#### 3.5.2.1 Definition

##### MRF

A Markov Random Field models a set of variables  $V$  that take discrete values inside a set of labels  $\mathcal{L}$ , and where the dependences between the variables are defined by a neighborhood system  $\mathcal{N}$ . This MRF is defined by an undirected graph  $\mathcal{G} = (V, \mathcal{N})$ , where the variables  $v \in V$  are the nodes and the neighborhood system  $\mathcal{N}$  defines the edges of the graph. The neighborhood system defines the dependences between the variables of the model. These dependences are given by cliques, where cliques are undirected complete graph  $\mathcal{G}_{clique}$ , for which each node is connected with every other node inside the graph clique  $\mathcal{G}_{clique}$ .

Pairwise MRFs are a subset of the MRFs, where the neighborhood system  $\mathcal{N}$  contains only cliques with two nodes. This simpler definition implies that only two variables have to be examined at same time because they will be independent from

the other ones. This definition is an important restriction from the general case that offers fewer possibilities. However, pairwise MRFs remain widely used because optimization on these models is more developed and easier, whereas the cost of optimization for higher-order MRFs cost is prohibitive. Higher-order MRFs should nevertheless become more common as a new optimization technique was recently proposed by Komodakis *et al.* [Komodakis 2009a]. In the subsequent paragraphs, only pairwise MRFs will be considered.

### MRF optimization problem

The MRF optimization problem is stated as the research of an optimal labeling  $\mathcal{C}^*$  for which a cost is minimal. This cost is composed of two terms. The first one is often called data term and measures the adequacy between nodes and labels. The second one is called smoothness or regularization term and often aims at imposing a continuity of the labels.

Given an undirected graph  $\mathcal{G} = (V, \mathcal{N})$  and a discrete set of labels  $\mathcal{L}$  the optimization problem aims at assigning a label  $u_v$  to each node  $v \in V$  ( $\mathcal{C} = \{u_1, \dots, u_k\}$ ). However, assigning a label has a cost that comes from both the singular potential  $V_p(u_p)$  at the level of the node and from a pairwise potential  $V_{p,q}(u_p, u_q)$  due to the regularization term. Both potentials are problem specific, which offers a wide range of possible applications. However, such energy is difficult to minimize, because this energy function is a highly non-convex function on a high dimension space.

$$\mathcal{C}^* = \underset{\mathcal{C}}{\operatorname{argmin}} \sum_{p \in V} V_p(u_p) + \sum_{p,q \in \mathcal{N}} V_{p,q}(u_p, u_q) \quad (3.15)$$

### Field of use

MRF allows expressing a wide range of problems, in many fields and in particular for Computer Vision. Indeed, MRFs were first introduced for statistical physics and later applied to many other fields including economics, sociology, and machine learning [Kindermann 1980, Komodakis 2008]. In particular, MRFs were applied to many problems of Computer Vision, such as 3D reconstruction, image denoising, optical flow estimation, stereo, extraction of texture information... [Komodakis 2008, Kolmogorov 2004, Pham 2007].

MRFs have often been applied to the segmentation of objects inside images [Boykov 2006, Kolmogorov 2004]. MRFs may indeed express many segmentation problems due to the flexibility of the potentials. Moreover, the absence of constraints regarding topology or dimensions makes MRFs especially valuable. Besides, optimization methods have been developed that provide global optimality for these MRF problems. However, these optimization methods often add constraints on the potentials that can be used, and on the usability for other applications than academic research. The first methods, based on Graph Cuts, were indeed slow and could hardly be used in a clinical prospect [Boykov 2001b, Boykov 2006, Kolmogorov 2004]. However, Komodakis *et al.* recently proposed a new optimization approach that deals



with more types of potentials and provides at least as good results as previous Graph-Cuts approaches but significantly faster.

### 3.5.2.2 Solving MRF

#### Graph-Cuts, optimal optimizers for metric pairwise potentials

Discrete MRF problems have been first solved by general-purpose optimization techniques such as simulated annealing. These techniques did not provide a good minimum and were very slow in practice, thus MRF problems were of small value. However, efficient methods have been developed to solve these problems during the last decade. These efficient techniques are commonly separated in two families: approaches based on graph cuts, or Loopy Belief Propagation (LBP) [Szeliski 2006, Komodakis 2008].

In this study, only methods based on graphs cuts have been considered, with a particular focus on a recent improvement of the expansion approach proposed by Komodakis *et al.* [Komodakis 2007a]. First, older methods cannot be retained in this study due to their poor efficiency. Methods such as simulated annealing provide neither a good solution, nor a fast result. For example, Boykov showed that simulated annealing provides results 500 times slower and with an error twice as big compared to the expansion algorithm for a segmentation problem [Boykov 2001a]. Then, LBP methods were not retained because they are less theoretically based and are less efficient than graph cuts in experimental tests. First, the optimality of LBP is not proved and the convergence is not well defined in the general case [Komodakis 2007a]. Then, Szeliski *et al.* compared LBP and graph cuts methods for several optimization problems; graph cuts approaches are always more efficient when the conditions of use are met [Szeliski 2006, Szeliski 2008]. Indeed, only when non-metrics pairwise potentials are used do LBP give better results. And for all other cases graph cuts are more efficient. Moreover, the final errors are almost equals in all benchmark tests (less than 0.5% difference), but the optimization is always significantly faster with graph cuts; 50 times for stereo matching, 100 times for photo montage and 10 times for segmentation.

#### Graph Cuts approaches

Boykov and Jolly introduced graph cuts to solve pairwise MRFs using metric pairwise potentials [Boykov 2001b]. The initial paper dealt only with binary segmentation of images, but the proposed approach generalizes to any number of labels. However, optimization with graph cuts for the multi-label case becomes NP-hard [Kolmogorov 2004].

The main idea behind graph cut approaches is to take advantage of the graph expression of the MRF optimization problem and to define the potentials of the energy as costs on the edges of a new graph (eq. 3.15). The approach contains three steps. First, a new graph is created that contains the initial graph  $\mathcal{G}$ . Then, costs are defined for all edges of the new graph  $\mathcal{G}'$ . Finally, the MRF optimization

problem is solved by computing a cut on the new graph  $\mathcal{G}'$ . The approach will be illustrated with a simple example before a more formal introduction of the method.

The segmentation of a 3x3 image is achieved through the optimization of a MRF problem, whose singular potentials are defined through the adequacy with an intensity model and pairwise potentials aim at penalizing changes of labels between neighbor pixels (fig. 3.19). A graph  $\mathcal{G}$  is first constructed, whose nodes  $V$  are defined as one node per pixel and edges  $\mathcal{N}$  by the 4-connexity inside the image. Then, costs are set for all the edges  $\mathcal{N}$  of the initial graph, which are bigger when the pixels are similar. These costs come from the pairwise potentials that aim to give a same label to spatially close pixels with similar intensities. Thus, cutting an edge between two nodes will cost more when pixels are similar, which will favor giving a same label to similar pixels. Singular potentials for the first class are then introduced by defining a source  $S$  (light gray) with an edge to every initial node, where the cost on each edge is the singular potential for the first class and for the linked node. The same is done for the singular potentials of the second class with a sink  $T$ . At this point the edges of the constructed graph define all the potentials of the MRF problem (fig. 3.19.b). As may be seen, edges have very various costs; configurations (node+label) with a high confidence are shown as larger edges, similar nodes are linked by large edges, while different nodes are linked by thin ones. . . Finally, the optimization is achieved with a minimal cut. This cut creates a partition of the graph in two parts, one that contains the source and the other with the sink. The graph with the source defines all nodes that should be labeled as in the first class, and resp. the graph with the sink provides the labeling of nodes inside the second class.

Formally, a new graph  $\mathcal{G}' = (V', \mathcal{N}')$  is created as a supergraph of the initial graph  $\mathcal{G}$ . First, a new set of nodes  $V' = V \cup \{S, T\}$  is defined by adding a source  $S$  and a sink  $T$  to the initial set of nodes  $V$ . Then new edges are added to the initial edges  $\mathcal{N}$ , which link each initial node in  $V$  to either the source  $S$  or the sink  $T$ . Thus, the new edges are defined as  $\mathcal{N}' = \mathcal{N} \cup \mathcal{N}_S \cup \mathcal{N}_T$ , where  $\mathcal{N}_S$  are edges between the source  $S$  and each initial node in  $V$  and  $\mathcal{N}_T$  the edges between the sink  $T$  and the nodes in  $V$ .

The potentials of the energy are added to the graph  $\mathcal{G}'$  as costs on the edges. The pairwise potentials are added by setting costs for the edges of the initial graph  $\mathcal{G}$  and the singular potentials are added as costs between the initial nodes  $V$  and the two new nodes ( $S$  and  $T$ ). The singular potential  $V_p(1)$  for the first class at node  $p$  is set on the edge between the node  $p$  and the source  $S$ . The same is done for the second class, where the singular potentials are set for the edges  $\mathcal{N}_T$  going to the sink  $T$ .

Finally, the optimization problem is solved by computing a  $S - T$ -cut, namely a partition of the nodes  $V'$  into two disjoint sets that minimizes the cost of the cut edges. The optimal labeling is then given by the remaining links to either the source or the sink. All nodes that are finally labeled in the first class will remain linked with the source, while links to the source will be cut for nodes in the second class. Two main methods have been introduced to compute this cut for any finite number of labels, algorithms based on swap moves or on expansion moves. Swap algorithms

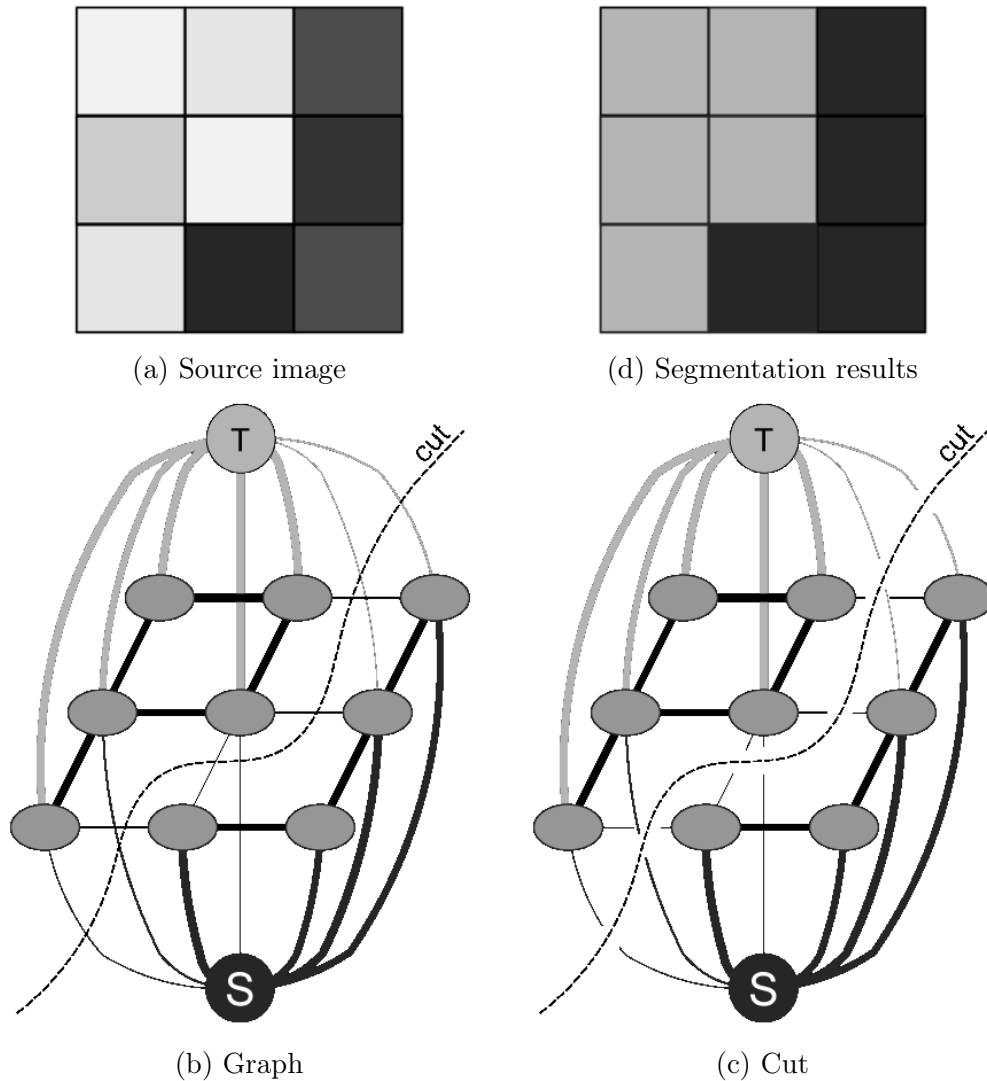


Figure 3.19: Graph cut segmentation of a small image. A graph is defined (b), where each node is one point of the source image (a) and where two terminal nodes are added (S for object and T for background). The edges of the graph are defined either by the connectivity inside the image (black) or by additional edges between each node of the image and both terminal nodes (light and medium gray). To each edge corresponds a cost shown by the width of the edge, either as a continuity condition (black edges) or as a similarity measure for a class. Finally a cut is done while minimizing the cost of the cut edges (c). This cut defines two distinct graphs that define the segmentation results. The example is taken from [Boykov 2001b]

handle more general energy functions with non-metrics potentials, while expansion algorithms deal only with metric or semi-metric pairwise potentials. However, the expansion algorithm provides an upper bound for the error and give experimentally better results than swap algorithms and also faster [Boykov 2001a, Szeliski 2008].

### Solving MRF with linear programming

The  $\alpha$ -expansion algorithm proposed by Boykov and Jolly has long been the reference for MRF optimization, until a new method was introduced by Komodakis *et al.*, who solve the MRF optimization problem using primal/dual strategies and linear programming [Komodakis 2007b, Boykov 2001b]. This new method has the advantage of being very fast (around a magnitude ten speedup compared to the expansion algorithm), but its drawback is high memory consumption. In term of quality, the computed solution remains the same than with the expansion algorithm for metric pairwise potentials. The new algorithm can even handle some non metric pairwise potentials, as long as pairwise potentials remain positive and are null only when labels are equals  $V_{p,q}(u_p, u_q) \geq 0$  and  $V_{p,q}(u_p, u_q) = 0 \Rightarrow u_p = u_q$ .

### 3.5.3 Atlas Registration

In this section, the registration approach proposed by Glocker *et al.* will be first reviewed and then applied to atlas registration [Glocker 2007b, Glocker 2008]. Glocker proposed a nonrigid registration method using FFD as deformation model. This method attracted lots of interest, because of its ability to provide good registration accuracy along with small computation times for diverse similarity measures. This method will be first reviewed for the case of dense image registration. Then an adaptation for the proposed atlas will be introduced.

#### 3.5.3.1 Dense Image Registration

As mentioned before, a registration method is defined by three components: a transformation model, a similarity measure and an optimization method (sect. 3.4.2.1). Glocker's technique relies on Free Form Deformation (FFD) as transformation model and may use any similarity measure, which explains the high genericity of the approach. The registration problem is expressed as a MRF problem, where only the control points of the deformation field are taken in account.

The optimal transformation takes advantage of the recent improvements for the optimization of MRFs brought by Komodakis *et al.* In particular, the optimal transformation may be obtained in a short amount of time and with confidence regarding the quality of the solution due to the characteristics of the optimization technique.

The registration method will now be reviewed for point-wise similarity measures. This approach extends to more complex similarity and statistical measures with the introduction of local image patches centered on every controls points [Glocker 2008].

However such measures are not relevant for this study, because of the chosen similarity measure for the atlas.

### Theory

Registration aims at finding an optimal transformation  $\mathcal{T}^*$  that better matches a source image  $f : \Omega \rightarrow \mathbb{R}$  on a target image  $g : \Omega' \rightarrow \mathbb{R}$  for a similarity measure  $\rho$ . The research of the optimal transformation  $\mathcal{T}^*$  is done by minimizing an energy  $E_{reg}$  defined as the sum of two energy components. The first one,  $E_{data}$  aims to minimize the difference between the target image and the transformed source. This difference is defined as the sum of the distances between each transformed voxels of the source image and the voxels of the target image for the chosen similarity measure  $\rho$ . The second energy component introduces a smoothness energy  $E_{smooth}$  that aims to impose a constraint on the regularity of the transformation.

$$\mathcal{T}^* = \underset{\mathcal{T}}{\operatorname{argmin}} E_{reg}(\mathcal{T}) \quad (3.16)$$

$$E_{reg}(\mathcal{T}) = E_{data}(\mathcal{T}) + E_{smooth}(\mathcal{T})$$

$$E_{data}(\mathcal{T}) = \int_{\mathbf{x} \in \Omega} \rho(g(\mathbf{x}), (f \circ \mathcal{T})(\mathbf{x})) d\mathbf{x} \quad (3.17)$$

Glocker *et al.* introduced FFD based on B-Splines as model of transformation. This transformation is given by a deformation field that is defined by a set of control points located on each intersection of a uniform grid  $\mathcal{G}_U : [1, M] \times [1, N] \times [1, P]$ . Using this grid, the transformation of each voxel  $\mathbf{x}$  can be expressed as a weighted combination of the displacements  $\mathbf{d}_g$  of the grid points  $g$ , with positions  $\mathbf{g}$ . The weight for each grid point  $\eta(\cdot)$  is defined as a function that gives the contribution of a control point  $g$  to the displacement field  $\mathcal{D}$ .

$$\begin{aligned} \mathcal{T}(\mathbf{x}) &= \mathbf{x} + \mathcal{D}(\mathbf{x}) \\ \mathcal{D}(\mathbf{x}) &= \sum_{g \in \mathcal{G}} \eta(|\mathbf{x} - \mathbf{g}|) \mathbf{d}_g \end{aligned} \quad (3.18)$$

Using this deformation model the previous registration energy (eq. 3.17) can be stated on this deformation grid, functions of the influence  $\eta^{-1}(\cdot)$  of an image voxel  $\mathbf{x}$  to a control point  $g$ .

$$E_{data}(\mathcal{T}) = \frac{1}{|\mathcal{G}_U|} \sum_{g \in \mathcal{G}_U} \int_{\mathbf{x} \in \Omega} \eta^{-1}(|\mathbf{x} - \mathbf{g}|) \rho(g(\mathbf{x}), f \circ \mathcal{T}(\mathbf{x})) d\mathbf{x} \quad (3.19)$$

$$\eta^{-1}(|\mathbf{x} - \mathbf{g}|) = \frac{\eta(|\mathbf{x} - \mathbf{g}|)}{\int_{\Omega} \eta(|\mathbf{y} - \mathbf{g}|) d\mathbf{y}} \quad (3.20)$$

Then, instead of using continuous displacement vectors, a discrete set of  $k$  displacement vectors is defined  $\{d^1, \dots, d^k\}$ , that corresponds to a discrete set of labels  $\mathcal{L} = \{u^1, \dots, u^k\}$ . With each node of the grid  $g$ , is associated a label  $u_g$  that gives the displacement vector for the node  $g$ . This discrete framework allows approximating the data term of the energy (eq. 3.19) as the singular term of a MRF.

$$E_{data}(\mathcal{T}) \approx \sum_{g \in \mathcal{G}_U} V_g(u_g) \quad (3.21)$$

$$V_g(u_g) = \int_{\mathbf{x} \in \Omega} \eta^{-1}(|\mathbf{x} - \mathbf{g}|) \rho(g(\mathbf{x}), f \circ \mathcal{T}(\mathbf{x})) d\mathbf{x}$$

The smoothness energy is then defined to impose a smooth deformation field by insuring that the direction of displacement of the control points does not change too fast. This constraint is chosen as the Euclidean distance  $\|\cdot\|_e$  between two displacement vectors for all available neighbors pairs between the control points  $g$ . It should be noted that this pairwise potential satisfies the conditions of use for the MRF optimization method developed by Komodakis *et al.*

$$E_{smooth}(\mathcal{T}) = \sum_{(g,h) \in \mathcal{N}_g} V_{g,h}(u_g, u_h) \quad (3.22)$$

$$V_{g,h}(u_g, u_h) = \lambda_{gh} \|d^{u_g} - d^{u_h}\|_e$$

Finally the registration problem is expressed as a MRF in a discrete domain as a balance by  $\lambda_{smoo}$  between the two term of the registration energy. This MRF energy is then optimized using Komodakis's algorithm, as the energy satisfies its conditions of use.

$$E_{reg}(\mathcal{T}) = \sum_{g \in \mathcal{G}_U} V_g(u_g) + \lambda_{smoo} \sum_{(g,h) \in \mathcal{N}_g} V_{g,h}(u_g, u_h) \quad (3.23)$$

### 3.5.3.2 Iterative Multiscale Registration

#### Principle

The transformation is iteratively computed in a coarse-to-fine manner. First, the transformation is computed following a multiscale approach in a coarse-to-fine manner. A rough estimate of the transformation is first computed on subsampled images at low resolution, before refining the transformation using higher resolutions of image. Then the transformation is iteratively constructed at each scale in order to use only a small number of possible displacements. The transformation is indeed computed by combining basic possible transformation in order to progressively construct a complex transformation, while using only a small set of basic displacements.

This approach is illustrated on an example (fig. 3.20). As may be seen the transformation is progressively constructed for one scale. Then, the deformation progressively becomes more and more complex with the finer scales, while keeping the same rough aspect.

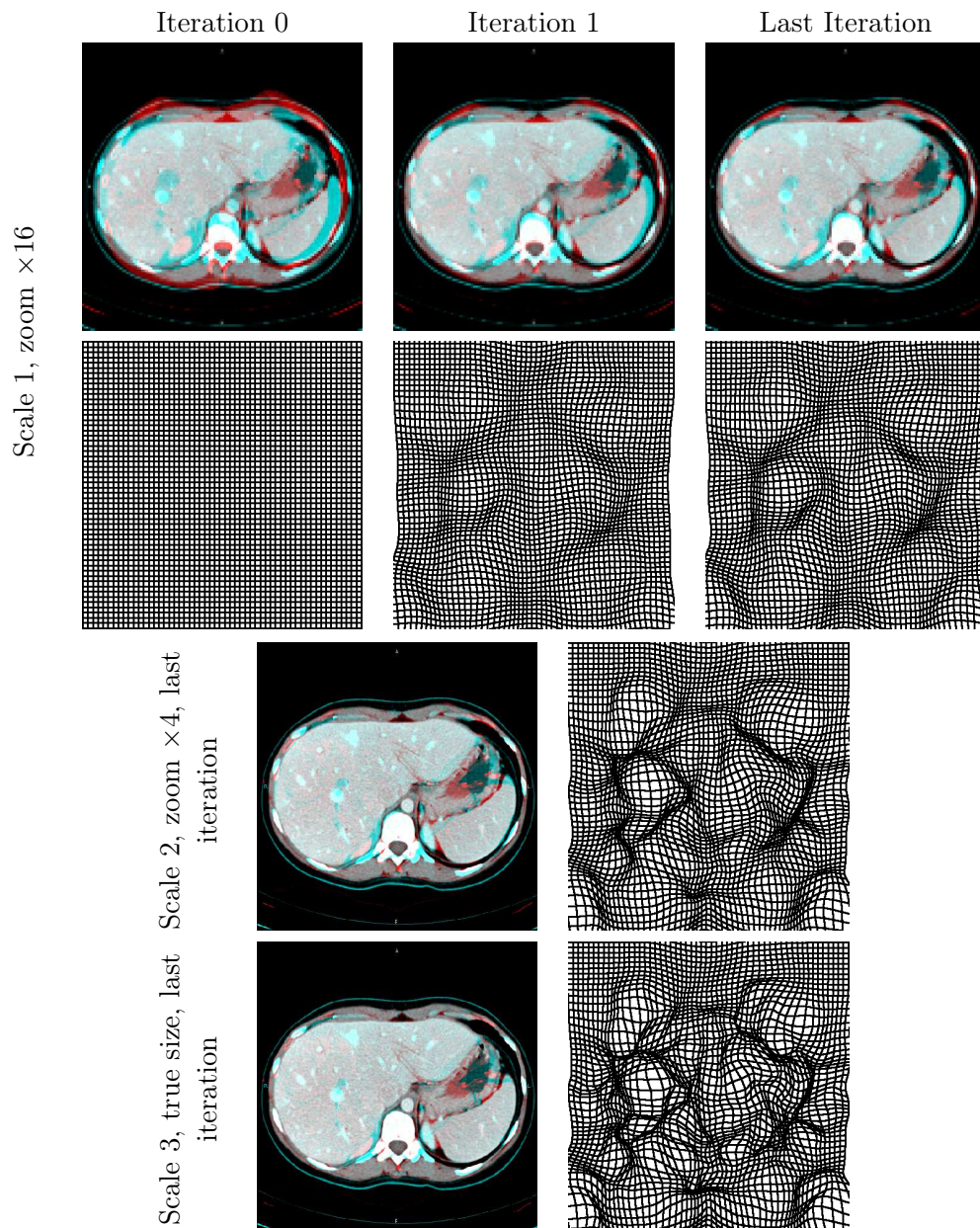


Figure 3.20: Registration process for the fusion of two liver slices. The deformed source and target are shown along with related deformation fields for an example previously introduced (fig. 3.10). The deformed source is shown in red and the target image in blue on blended images. Intermediate deformations and blended images are shown for diverse scales. The incremental construction of the deformation is also shown for the coarser scale.

### Iterative multiscale scheme

The registration is done following a multiscale approach in a coarse-to-fine manner in order to improve the robustness and the quality of the registration. Indeed, the quality of the deformation field is constrained by both the quantization of the displacement vectors (defined by the labels) and the number of control points (the nodes). A high number of control points along with a good quantization of the displacement vectors cannot be met at the same time. Thus, a global to local approach is followed. The idea is to begin with a small number of nodes with a good quantization of displacements, and refine later by using more control points, but with less labels. The decrease of the number of labels will not matter, because an approximate direction is already known, thus some displacement vectors can be suppressed, as useless. In practice, registration begins with subsampled images and aim only at defining a rough transformation. The registration is then gradually refined using increasingly detailed images but modifying less and less the deformation field.

This multiscale approach allows speeding up the registration, because only the last steps of the registration technique will require the entire image. Combined with the speed and the quality of the retained optimization technique, this implementation contributes to the overall speed of the method.

The quality and the characteristics of the registration are directly driven by the choice of the available displacement vectors. Choosing more displacement vectors allows for finer registration, but longer computation times due to the increased number of possible labels. Moreover, the problem may become unsolvable for all practical purposes for a high number of displacement vectors due to hardware constraints. Thus, the registration is done in several steps, with only a small number of available displacement vectors. The final deformation is gradually constructed by the addition of all successive displacement vectors that are computed. Formally, the transformation is iteratively constructed while taking in account the transformation at previous step. At step  $t$ , the transformation is defined as the deformation field at previous step plus an unknown displacement that should be optimized (eq. 3.24).

$$V_g^{(t)}(u_g) = \int_{\mathbf{x} \in \Omega} \eta^{-1}(|\mathbf{x} - \mathbf{g}|) \rho(g(\mathbf{x}), f(\mathbf{d}_g + \mathcal{T}^{(t-1)}(\mathbf{x}))) d\mathbf{x} \quad (3.24)$$

### Quantization of displacement vectors and diffeomorphism

The retained displacement vectors are chosen only on axes directions and with a maximal displacement of 0.4 times the spacing of control points. First, displacement vectors on axes are sufficient to define any vector of the space through the iterative scheme. Then, the distance constraint insures that the final registration is a diffeomorphism. Obtaining a diffeomorphic transformation is important because it insures that the transformation is both invertible and structure preserving. The former may be valuable for some applications, while the latter is crucial for either medical application or the creation of atlases. The preservation of structures indeed insures that no information is lost during registration. Otherwise it may happen



that a lesion at the boundary of a structure is removed during registration, which would prevent correct diagnosis. Moreover, registered images that are folded would lead to incorrect samples for atlas creation. This diffeomorphic constraint is insured by restricting the set of available displacements to be 0.4 times the control point spacing at most. This restriction is indeed sufficient to get a diffeomorphic transformation. First, Choi *et al.* proved that a deformation field produced by 3D B-Splines with a maximal displacement of 0.4 times the spacing of control point is a diffeomorphism [Choi 2000]. Then, the composition of diffeomorphisms is a diffeomorphism (multiscale iterative construction).

### 3.5.3.3 Atlas Registration

Glocker's approach for image registration (sect. 3.5.3.1) was applied to atlas registration. However, registration of an atlas implies some changes because the similarity measure is not defined between two images but between one atlas and one image.

#### Theory

The registration of an atlas aims to find an optimal transformation  $\mathcal{T}^*$  that better matches an atlas  $\mathcal{A} : \Omega \rightarrow \Xi$  on a target image  $g : \Omega' \rightarrow \mathbb{R}$ . The research of the optimal transformation  $\mathcal{T}^*$  is done as for images by minimizing an energy  $E_{reg} = E_{data} + E_{smooth}$ . The smoothness component does not change compared to dense image registration. However, the data term of the energy  $E_{data}$  is modified in order to maximize the adequacy between the voxels of the image  $g$  and the model for the transformed atlas.

$$E_{data}(\mathcal{T}) = \int_{\mathbf{x} \in \Omega} \rho_{\mathcal{A}}(g(x), (\mathcal{A} \circ \mathcal{T})(\mathbf{x})) dx \quad (3.25)$$

The similarity measure  $\rho_{\mathcal{A}}$  aims at maximizing the probability of a match between a voxel in  $g$  and the intensity model defined by the transformed atlas for this location  $p_{\mathcal{A} \circ \mathcal{T}(\mathbf{x})}(g(x))$ . Such a probability exists and is well defined. Indeed, for each voxel  $\mathbf{x}$  in the atlas  $\mathcal{A}$  is known a pdf, and the corresponding intensity in  $g$  is defined in a unique manner by the transformation  $\mathcal{T}$ . The similarity measure is finally defined as the log likelihood of this probability of match in order to go back to a minimization problem.

$$\rho_{\mathcal{A}}(g(x), (\mathcal{A} \circ \mathcal{T})(\mathbf{x})) = -\log(p_{\mathcal{A} \circ \mathcal{T}(\mathbf{x})}(g(x))) \quad (3.26)$$

The registration problem is finally expressed and solved as a MRF optimization problem, as for image registration (sect. 3.5.3.1). The only difference comes from the choice of a new similarity measure (sect. 3.26). It should also be noted that while image registration was symmetric, the proposed method for atlas registration deals only with the registration of one atlas on an image. The opposite registration is another problem.

### Implementation

Atlas registration is done as for dense image registration, meaning in a multiscale approach while insuring that the successive transformations remain diffeomorphic. A difficulty is nevertheless posed by the subsampling of the atlas during the multiscale process. Subsampling atlases is indeed difficult as it amounts to averaging GMMs. The simpler approach would be to sum GMMs. However, no improvement would come from the multiscale approach in this case. The total number of Gaussian components inside the GMMs would indeed remain the same for the entire atlas whatever the scale factor. Then, simplifying a weighted sum of GMMs by another GMM would be another approach. However, this method opposes the clinical prospect of this study as this simplification would require complex computations during the registration, which would also offset the contribution of the multiscale approach. Therefore, the registration is achieved using one atlas per scale, with atlases that are computed off line.

Atlas registration is achieved using a set of atlases for different scales that are all computed on a same set of registered images. Instead of computing a single atlas, one atlas is created for each scale of the registration algorithm. This approach removes the problems that come from atlas subsampling and allows keeping the improvement brought by the multiscale approach. However, this approach introduces an additional constraint; the registration should always be done with the resolution of the atlas, as the atlas cannot be subsampled nor upsampled to match the dimensions of a target image.

This multi-scale atlas is computed by constructing an atlas for each scale of the registration algorithm, while extracting intensity samples from the same set of registered images, but with diverse subsampling factors. Indeed, whatever the scale the atlas should be computed on the same set. Otherwise, the coarse-to-fine approach might fail as changes of scale might require bigger deformations than available. The length of the available displacements indeed decreases with each scale of the registration algorithm. Thus, a change of atlas may induce a displacement of the structures that is bigger than those available.

### 3.5.4 Segmentation by registering a statistical atlas

#### 3.5.4.1 Principle

Atlases may be used to segment images through the transformation of a reference with a deformation field obtained by registering an atlas on an image. The principle does not change compared to the case of segmentation through registration of a representative volume (fig. 3.18). However, use of an atlas is an improvement compared to representative volumes as it accounts for the variations of appearances.

The contribution of atlases for segmentation through registration will be illustrated with the previously introduced example (fig. 3.14). Mean images could not account for the variability of appearances inside the retained artificial sample (fig. 3.14). Moreover, a representative volume could not be chosen to handle

all images inside the set of images. On the contrary, atlases (fig. 3.21.a) take in account this variability. Thus, registration of an atlas should provide a deformation field (fig. 3.21.b) that may be used to segment a new image, whereas the registration of an image would not provide a correct transformation.

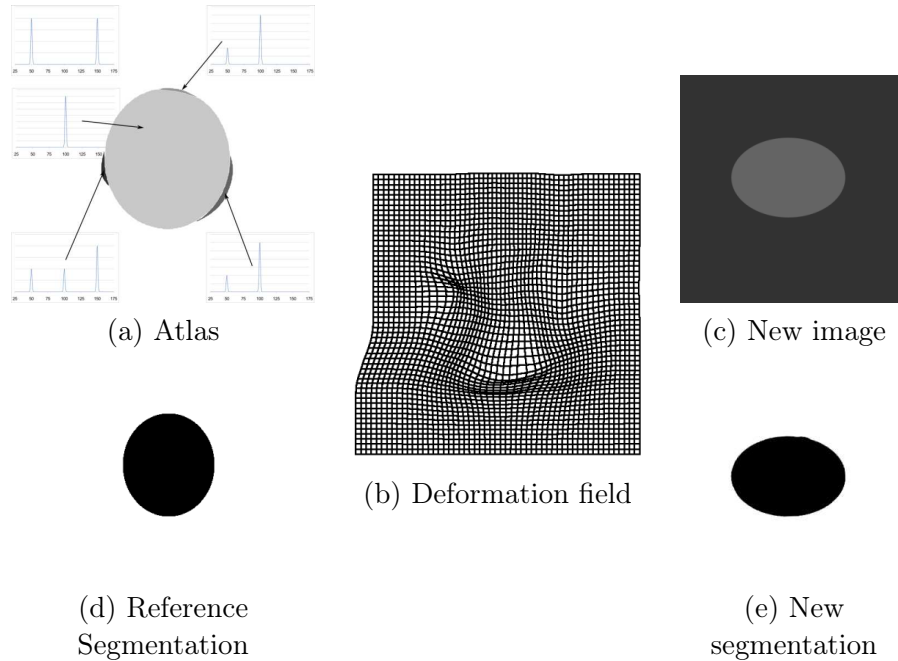


Figure 3.21: Segmentation through the registration of an atlas. An atlas (a) is registered onto a new image (c) through a transformation defined by a deformation field (b). The segmentation of the new image (e) is then obtained by applying this deformation field on a reference segmentation for the atlas (d).

#### 3.5.4.2 Method

Segmentation through atlas registration is achieved through three main steps. First a multiscale atlas is created. Then this atlas is registered on the image to segment in order to obtain a deformation field. This deformation field is finally applied to a reference segmentation for the atlas in order to segment the new image.

Before beginning any segmentation two tasks should be done once for all, the creation of the atlas  $\mathcal{A}$ , and the definition of a reference segmentation  $\mathcal{A}_{seg}$ . Both tasks are done off line and only once. First, the multiscale atlas  $\mathcal{A}$  is computed on a same set of registered images. This atlas is composed of one atlas per subsampling factor required by the scales of the registration process (sect. 3.5.3.3). Then, a reference segmentation  $\mathcal{A}_{seg}$  is given by the manual segmentation for the target image used to create the set of registered images.

The segmentation of a new image  $\mathcal{V}_i$  begins with the registration of the atlas  $\mathcal{A}$  on this new image  $\mathcal{V}_i$ , which defines a transformation  $\mathcal{T}_{\mathcal{A}}$ . The atlas  $\mathcal{A}$  is

registered on the new image  $\mathcal{V}_i$  following a multiscale approach using MRF optimization (sect. 3.5.3.3). This registration defines a transformation  $\mathcal{T}_{\mathcal{A}}$  that matches the atlas on the new volume.

Finally, segmentation is achieved by applying the computed transformation  $\mathcal{T}_{\mathcal{A}}$  on the reference segmentation  $\mathcal{A}_{seg}$  for the atlas  $\mathcal{A}$ . The transformation  $\mathcal{T}_{\mathcal{A}}$  matches the atlas  $\mathcal{A}$  onto the new image  $\mathcal{V}_i$ . Therefore, this transformation should match the reference segmentation on the same structures inside the new volume  $\mathcal{V}_i$ . Consequently the deformed reference segmentation  $\mathcal{T}_{\mathcal{A}}(\mathcal{A}_{seg})$  provides the segmentation inside a new image provided that three conditions are met. The registration should indeed be correct, the reference segmentation should be relevant and the object to segment should remain relatively similar.

### 3.5.5 Test protocol

#### 3.5.5.1 Comparison metric

Several metrics have been introduced to quantify the quality of segmentation. For this study, three common metrics were retained to quantify the quality of segmentation; the sensitivity, the specificity and Dice Similarity Coefficient (DSC), as well as the Jaccard index in some rare case. All these metrics are defined functions of the number of voxels that are correctly or incorrectly classified as well as their expected classes. First, the sensitivity gives the percentage of tumor that is correctly classified. Then, the specificity quantifies the quality of the segmentation for healthy tissues. Finally, DSC and Jaccard index measure how well the object is segmented.

The sensitivity (eq. 3.27) defines the fraction of correctly segmented voxels inside tumors, or number of true positives (TP) divided by the real number of voxels inside tumors. This total number of voxels inside tumors is defined by the sum of TP with the number of false negatives (FN), where FN defines voxels wrongly set as outside tumors.

$$\text{sensitivity} = \frac{TP}{TP + FN} \quad (3.27)$$

The specificity (eq. 3.28) defines the fraction of correctly segmented voxels inside healthy tissues. This percentage is defined as the number of well classified voxels inside healthy tissues, or number of true negative (TN) divided by the real number of healthy voxels. As for sensitivity, the number of voxels inside healthy tissues is defined as the sum of well classified voxel inside healthy tissues (TN) plus the number of voxels wrongly set as tumoral or false positives (FP).

$$\text{specificity} = \frac{TN}{TN + FP} \quad (3.28)$$

The Dice Similarity Coefficient (DSC) measures how well the object is segmented, while taking in account both the defaults and excesses of the segmentation (eq. 3.29). This metric is defined as twice the intersection of reference and segmented volumes over their addition.

$$DSC = \frac{2TP}{2TP + FP + FN} \quad (3.29)$$

The Jaccard index also measures how well an object is segmented, but its value is stricter than DSC. This metric is indeed defined as the intersection of reference and segmented volumes over their union (eq. 3.30). Thus, the incorrectly classified voxels have more weight than for DSC. It should be noted that the Jaccard index equals 1 minus the overlap error.

$$DSC = \frac{TP}{TP + FP + FN} \quad (3.30)$$

### 3.5.5.2 Segmentation priors

First, a set of 65 volumes  $\mathcal{I}$  was chosen for training. All images within this set were anisotropic axial CT images with liver tumors that were subsampled by a factor two. The images within this set were chosen to well sample the histological types of the tumors, the types of the machines, and the enhancement phases. The size of voxel inside this image set  $\mathcal{I}$  spans between 1.12mm and 1.86mm for the axial slices and between 1.6 and 10mm for the slice thicknesses. All volumes were manually centered around the liver region, but without a precise location. This last step may also be automatically done, by computing the position of the bottom of the lungs and the upper part of the hips. Indeed, Zhou *et al.* successfully did it this way, but it may also be done by registration at low resolution [Zhou 2005].

Then, a set of 65 registered volumes  $\mathcal{W}$  was created. These registered volumes were computed by the registration of every image within the initial set  $\mathcal{I}$  on a same target  $\mathcal{V}_{trg}$ , whose voxel size was 1.37x1.37x3.77 mm. This registration was achieved using drop software, with 5 pyramid levels, 5 iterations per level, a sparse sampling of labels, with SAD as similarity measure, and for an initial grid spacing of 180mm [Glocker 2009]. However, the weight of the regularization term was not constant, as the deformation of the liver varies significantly between the CT images. Typical values retained for the weight of the smoothness term were chosen as power of 10 between  $10^{-3}$  and 10.

Finally, the statistical atlas was constructed using 5 resolution levels. Five atlases were computed for different resolutions in order to define the multiscale atlas. For each scale, intensity samples were first extracted from the set of registered images  $\mathcal{W}$  for a chosen level of resolution. Then, pdf models were computed at each voxel with the EM algorithm after initialization with the k-means, and for a maximum number of Gaussian components of  $l_{max} = 5$ . Finally, the atlas was cleaned using the process described in (sect. 3.4.3.4). Dirac's components were given a standard deviation of 1 and a background pdf model was chosen as  $\mathcal{N}(-950, 50)$ .

### 3.5.5.3 Protocol

17 new CT volumes with manual segmentations were used to assess the quality of segmentation. These 17 volumes were chosen in order to be representative from

the variability of the possible cases. Indeed, the images were chosen with diverse sizes of voxel, various enhancement phases, and with diverse liver pathologies. For example, the voxel size within this set spans between 2.5 and 10mm, and volumes with diverse pathologies were chosen, including 6 cases of HCC, 1 of adenoma, 1 with cirrhosis and 9 with metastases from diverse primary sites.

The quality of segmentation was then assessed using the three metrics previously introduced (sect. 3.5.5.1). First, intensities were normalized for every image of the test set. Then, an estimate of the liver segmentation was computed for each of these images. The segmentation was achieved using the proposed approach that segments an image through atlas registration (sect. 3.5). For this step, registration was achieved using the same parameters than during atlas creation (sect. 3.5.5.2). However, during testing the weight of the smoothness term was constant and set to 0.25. Finally, the segmented regions were compared to reference segmentations.

### 3.5.6 Results, discussion

The accuracy of the segmentation is poor with segmentation through atlas registration. Indeed, the sensitivity for the segmentation of the liver is only 73% and the DSC only 70% (fig. 3.22). Moreover, these poor characteristics are not due to a high variability regarding the quality of the results, but remain stable on the test set. Indeed, the standard deviation of these measures is small, and the quality of the segmentation remains similar for all images but one.

	Sensitivity	Specificity	DSC
Atlas only	$0.73 \pm 0.02$	$0.98 \pm 5 \cdot 10^{-5}$	$0.70 \pm 0.01$

Figure 3.22: Segmentation through atlas registration, quantitative evaluation.

The reasons behind these poor results will now be explained through some examples for two livers with HCC and metastases (fig. 3.23). First, one may note that the segmentation is more accurate when the liver is large and has smooth boundaries (fig. 3.23.b,d). On the opposite, the quality of the segmentation is poorer when the liver is more carved (fig. 3.23.a,c). This behavior is explained by the influence of the shape of the liver. Indeed, when the liver is more carved, the correspondence between the atlas and an image is less precise because the smoothness constraint of the registration prevents sharp changes of the deformation field. Then, the segmentation is often incorrect for the bottom of the liver (fig. 3.23.a). This part of the liver is highly variable and thin compared to the upper part of the liver. Thus, this behavior could have two explanations. Either, the atlas is not reliable enough because of the variability of the intensity distribution at this location. Or, the liver is neglected due to its small size and because of the similarity of the liver with the surrounding tissues. Finally, the segmentation is sometimes close from the true boundary of the liver but still incorrect, while the liver can easily be distinguished from the surrounding tissues. For example for the bottom of the HCC liver (fig. 3.23.c), darker tissues are wrongly segmented on the upper right part of

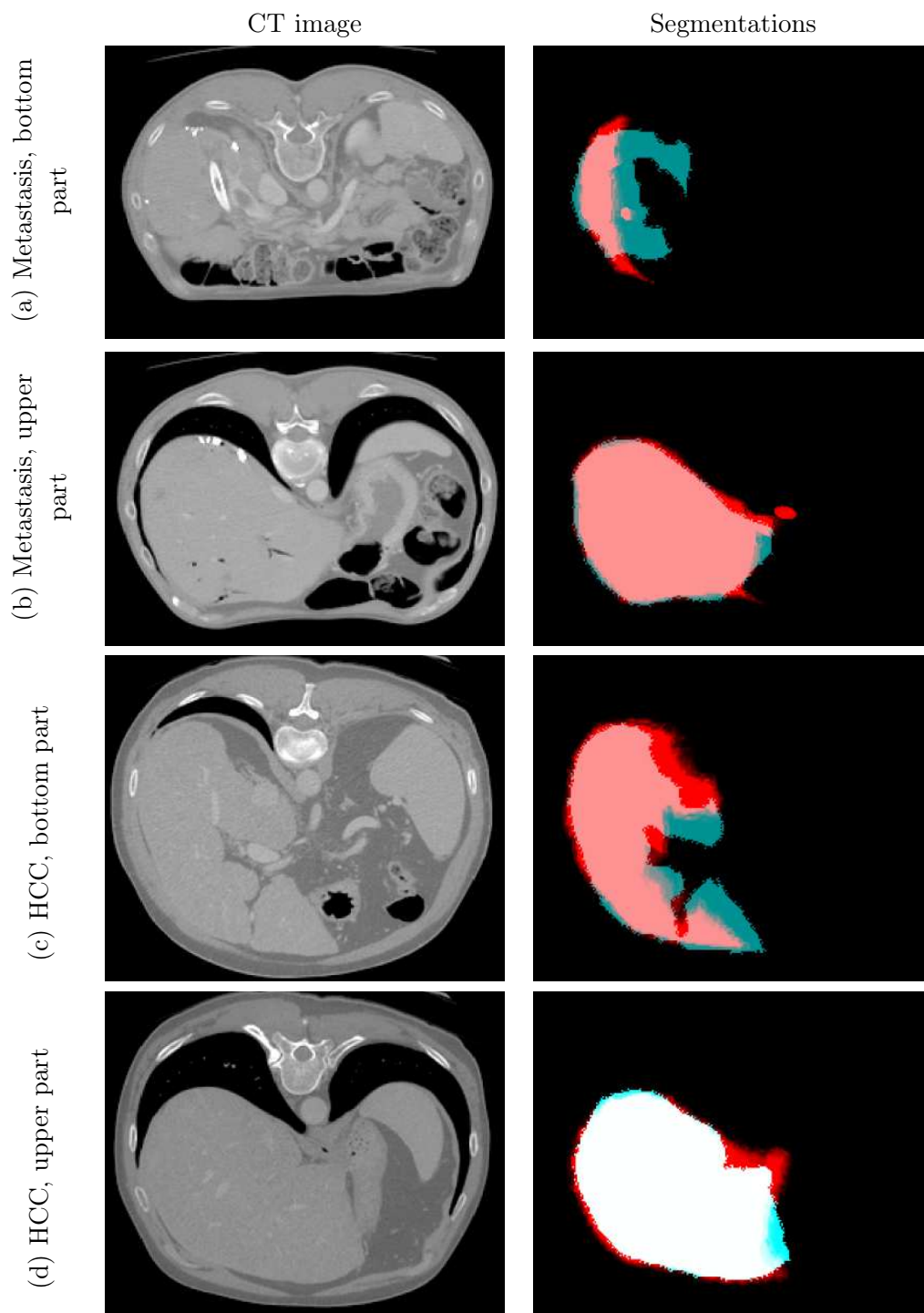


Figure 3.23: Segmentation through atlas registration, examples for two livers. Several results of segmentation are given for one metastatic liver and one with HCC. The reference segmentation is shown in blue and the automatic segmentation in red.

the liver, while the true boundary of the liver is easily seen.

To conclude, the segmentation of the liver through atlas registration does not offer accurate segmentations. Indeed, the quality of the segmentation is poor because of several lacks of the approach. First, the registration of the atlas cannot always provide perfect correspondences between the anatomical structures inside the image and the atlas. These imperfections sometimes come from lacks of the atlas, but are more often caused by the inability of obtaining a perfect registration due to the high variability of the liver and its sharpness. Therefore, the segmentation should be done while offering some slack from the correspondences given by the registration. Furthermore, the importance of the slack should not be uniform. Indeed, while some slack would be relevant when the registration is imprecise, it would become detrimental when the registration is accurate. Thus, a map should be used that defines the authorized slack and its importance, functions of the spatial location in the image. Then, the segmentation is sometimes imprecise while a marked liver boundary exists. Consequently, appearance patterns should also be used to take advantage of regions where the liver can be distinguished from the surrounding organs.

## 3.6 Combined segmentation

### 3.6.1 Intro

Simple atlas registration is limited by both the performance of the registration, and the relevance of a single reference segmentation to model a highly variable structure. Indeed, segmentation through atlas registration was shown to lack spatial information regarding the accuracy of registration and the locations with high liver variability. Moreover, not using information on the appearance of the liver was also shown to be detrimental (sect. 3.5.6). Thus a new segmentation method is introduced that relies on the registration to proceed to segmentation using both spatial and appearance priors.

The previous approach has several lacks. First, even for a highly efficient method like the retained registration technique, perfect match cannot be met. Then, the liver is highly variable, thus often a perfect registration cannot be obtained. Moreover, a single reference segmentation cannot account for every possible shapes of the liver, even after transformation. For example, the 8 or 9 anatomical segments of the liver have very diverse and non smooth shapes, thus registration cannot define a transformation to register any liver on any other one. Consequently, no correct segmentation can be obtained using only the transformation. To conclude, the objects to segment are not perfectly aligned in the fusion image and consequently the segmentation is imperfect.

Registration alone cannot provide a good segmentation. In particular the computed segmentation may sometimes be visibly incorrect. Moreover, the correct boundary of the object to segment may be visible, e.g. the boundary liver/lungs. Therefore, the visual information should be added in order to improve the final



segmentation, but should not be used alone. Registration indeed offers spatial information about the location of the liver that is very useful to distinguish the liver from neighbor organs with similar appearances such as the spleen. Thus, a new approach is introduced that segments the image with a balance between the spatial location and the appearance of the voxels. This approach should improve the segmentation, as appearance is added to the spatial information given by the registration, and because spatial information is no longer defined as binary but as a probability.

The combined segmentation begins like the previous approach, but the transformation is then used to spatially align a spatial probability map instead of directly proceeding to segmentation. This aligned spatial probability map is used next for segmentation in combination with an appearance prior. First the atlas is registered on a new image, with no change compared to the previous approach. Then, the obtained transformation is used to deform a spatial probability map in order to obtain the spatial probabilities on the new image. Finally, the segmentation is achieved as a balance between the spatial probabilities given by the transformed probability map and an appearance prior.

### 3.6.2 Introducing prior models

#### 3.6.2.1 Definition, motivation

Prior models define known features of one object, which contributes to more robust segmentations. Indeed, this additional knowledge offers better ways to discriminate between the objects than what generic techniques allow. For example, segmentation becomes complex when no boundaries are visible or when the difference between two objects is slight. In these cases prior information adds some knowledge that may offset the lack of visible landmarks in the image.

Prior knowledge may be introduced in many ways. For example several models have been introduced in the liver case; shape models have been introduced by Okada and Lamecker [Okada 2007, Lamecker 2002], Tesar introduced a feature model for the tissues that was applied to segmentation of abdominal organs [Tesar 2008], and the statistical atlas previously proposed is also a statistical model.

Due to its appearance and nature, the liver is a good candidate for the introduction of prior models. Indeed, the liver shows no visible boundaries on many parts. Moreover, the liver is truly similar to close organs such as the spleen. Thus, prior models are introduced to improve the segmentation. These priors should satisfy two constraints. First, the prior models should handle the parts of the liver where no change of appearance is visible. Then, the prior models should rely on the appearance when available.

#### 3.6.2.2 Choosing prior models

Two prior models are added to the statistical atlas, a spatial probability map and an appearance model. The two criteria previously stated cannot be met with a single

*a priori*. Thus two prior models are introduced, a spatial prior and an appearance prior that answer each to a specific constraint previously mentioned. These models are finally used in combination under the assumption that their combination will fit the aforementioned criteria.

A spatial probability map is introduced that replaces the reference segmentation as spatial prior. This map gives the possible liver locations in the space in order to model the anatomical variability of the liver. Such spatial map is available because the registration provides a transformation that allows adapting the spatial prior to any new image. In particular this model should drive the segmentation on parts where the change of organ is imperceptible. However, this spatial prior will have small value when boundaries are obvious, because the spatial prior might be slightly shifted.

An appearance prior is also introduced to improve the segmentation where change of appearance is marked between the object to segment and the background, e.g. the liver/lungs and liver/colon interfaces in the liver case. This appearance prior alone is not sufficient to proceed to the segmentation of the liver. Such prior would indeed have difficulty to distinguish between liver and diaphragm or between liver and spleen. However, this prior could improve segmentation when used with a spatial prior. Using both priors, one may hope for improvement when the contributions of both approaches are combined.

### 3.6.2.3 Spatial probability map

With a perfect registration and no anatomical variability, the registration of an atlas would define precise correspondences between similar structures. Thus, any structure could be segmented by the deformation of a reference segmentation inside the atlas. To develop this idea spatial probability maps are introduced for each class of tissue inside the atlas  $\mathcal{A}$ . These maps aim at capturing the variations of shape of the liver or other structures by giving the probability of being from one class for each voxel of the image. These spatial probability maps are computed using registered segmentation references for the same set of images that was used to create the atlas.

A spatial probability map  $\mathcal{M}_{c_i} : \Omega \rightarrow \mathbb{R}$  gives the probability of being of class  $c_i$  for each voxel of the space  $\mathbf{x} \in \Omega$ . In order to be relevant, this prior is defined on the atlas basis  $\mathcal{A} : \Omega \rightarrow \Xi$ . First, a common basis is mandatory to define identical spatial locations. Thus, spatial probability maps are created on a set of registered volumes. Then, the spatial probability map should be defined on the atlas basis. Therefore the image set  $\mathcal{W} = \{\mathcal{V}_1, \dots, \mathcal{V}_n\}$  used to construct the atlas is retained. Otherwise the spatial prior would become useless as there would be no way to match this prior with any new image. Finally, the probability for each class of tissues  $c_i$  is defined at each point of the space  $\mathbf{x} \in \Omega$  as the percentage of voxels at the same location inside the registered image set  $\mathcal{W}$  that are from the class  $c_i$ .

The contribution of this probability map is twofold. First, the spatial probability map captures the anatomical variability of all objects. Then, the map may provide information regarding the robustness of the registration. On some boundaries of the

objects, the registration indeed tends to be less precise. Thus, lower probabilities will be given on such parts. On the opposite, parts where one object is always present will be given higher probability. Thus the spatial probability maps provide information related to the accuracy of the registration, functions of the location in space.

A spatial probability map is useless without spatial standardization. Thus, the atlas registration is required in order to adapt the spatial prior to any new image. Otherwise the spatial prior would become non relevant. Given a transformation  $\mathcal{T}_{\mathcal{A}}$  that registers the atlas  $\mathcal{A}$  on a new image  $\mathcal{V}_i$ , the spatial map  $\mathcal{M}_{c_i}$  for the class  $c_i$  applies to a new image as the transformed spatial map  $\mathcal{T}_{\mathcal{A}}(\mathcal{M}_{c_i})$ . The validity of this transformed probability map  $\mathcal{T}_{\mathcal{A}}(\mathcal{M}_{c_i})$  is insured by the construction of the spatial probability map  $\mathcal{M}_{c_i}$  on the atlas basis.

### 3.6.2.4 Liver appearance model

The spatial location is not sufficient to obtain accurate segmentations. Thus an appearance prior is introduced to compute a more precise separation between the classes. This prior aims to model the appearances of the classes of tissue. One should note that the construction of this prior is done independently from the atlas, and applied independently from the registration process. Two appearance models have been introduced that provide the probability  $p_{app}(\mathbf{x} | i)$  of belonging to a class  $c_i$  at pixel  $\mathbf{x}$ . The intensity distribution within a class was retained as first appearance model, and a texture model was chosen as second possible one.

The intensity distribution inside a class of tissue  $c_i$  was retained as first appearance model. This appearance prior is given by one histogram distribution  $\mathcal{H}_{c_i}$  for each class  $c_i$ . The probability of belonging to a class is then directly obtained from this distribution. For any voxel  $\mathbf{x}$  of a volume  $\mathcal{V}_j$  the probability of belonging to the class  $c_i$  is simply defined by the value of the distribution  $\mathcal{H}_{c_i}$  for the intensity  $\mathcal{V}_j(\mathbf{x})$  (eq. 3.31). This prior distribution is computed as an average distribution for a set of images. There are no specific constraints on the set of images that should be used to create the histogram prior. However, using the image set that was used for the creation of the atlas seems the best choice. This choice indeed has several values. First, it prevents bias due to the use of same volumes for training and test, because volumes used to create the atlas cannot be used after this first step. Then, this choice allows avoiding segmenting additional references for training. Finally, using a same set for all prior knowledge is sound because the same images will be used for all training parts. Moreover, this choice offers more images for test.

$$p_{app}(\mathbf{x} | i) = p(\mathcal{V}_j(\mathbf{x}) | c_i) = \mathcal{H}_{c_i}(\mathcal{V}_j(\mathbf{x})) \quad (3.31)$$

A second prior was introduced for the case of segmentation with only two classes as a classification function based on texture features. For this prior a classification function is trained to distinguish between diverse tissues appearances using machine learning techniques. This prior is detailed in next chapter and gives probabilities

of belonging to a class tissue. However, this definition is limited to problems with only two classes (sect. 4.4.2.2).

### 3.6.3 Segmentation using prior models

#### 3.6.3.1 Principle

The new segmentation approach relies on a transformed spatial prior and an appearance prior to segment a new image, using atlas registration to define the transformation that matches the spatial probability map. The segmentation is done in two steps. First an atlas is registered, which defines a transformation that is used to match the spatial prior on any new image. Then segmentation is achieved at pixel level as a balance between the class probabilities given by both priors. Both steps are finally solved using MRF expressions.

#### 3.6.3.2 Segmentation energy

To simplify the notations, and because there is a direct link between the image and the graph that is constructed, notations inside the image and inside the graph will be identical. Each voxel is indeed related to a single node and each node to a single voxel. The same is true for edges and neighbors.

Given a new image  $\mathcal{V} : \Omega' \rightarrow \mathbb{R}$ , the atlas  $\mathcal{A}$  is first registered on this new image with a transformation  $\mathcal{T}_{\mathcal{A}}$  (sect. 3.5.3.3). The segmentation problem is then stated as the minimization of a MRF energy  $E_{seg}$  for a graph  $\mathcal{G}_{seg} = (\Omega', \mathcal{N}_n)$ , whose solution defines a labeling  $\mathcal{C}^*$  of the image and thus provides a segmentation of this image. This MRF energy is expressed as a MRF where the nodes are defined by the voxels inside the image  $\Omega'$ , and the edges are defined by the neighborhood system  $\mathcal{N}_n$  inside the image. Any labeling  $\mathcal{C} = \{u_{\mathbf{x}}, \mathbf{x} \in \Omega'\}$  assigns a label  $u_{\mathbf{x}}$  to each node  $\mathbf{x}$  of the image, which gives the class of the tissue  $c_{\mathbf{x}}$  at each voxel. The optimal labeling  $\mathcal{C}^*$  is one of the possible labeling, for which the segmentation energy is minimal and hence should provide an accurate segmentation of the image.

$$\begin{aligned} \mathcal{C}^* &= \underset{\mathcal{C}}{\operatorname{argmin}} E_{seg}(\mathcal{C}, \mathcal{T}_{\mathcal{A}}) \\ E_{seg}(\mathcal{C}, \mathcal{T}_{\mathcal{A}}) &= \sum_{\mathbf{x} \in \Omega'} V_{\mathbf{x}}(u_{\mathbf{x}}) + \lambda \sum_{\mathbf{x}, \mathbf{y} \in \mathcal{N}_n} V_{\mathbf{x}, \mathbf{y}}(u_{\mathbf{x}}, u_{\mathbf{y}}) \end{aligned} \quad (3.32)$$

This energy is composed of two terms balanced by  $\lambda$ . The first term is a data term that aims at maximizing the adequacy between each voxel and the possible classes. As for the second term, namely the regularization term, it aims at penalizing the change of class between neighbor pixels in order to remove parasite fluctuations of the labels.

### Data term

The data term aims at maximizing the adequacy between each voxel  $\mathbf{x}$  and its class  $c_{\mathbf{x}}$ , where the adequacy is defined as a balance of the adequacies with the spatial and the appearance priors. In order to come back to a minimization problem, the singular potentials are defined as a balance between the negative log likelihoods of the appearance prior and the spatial prior, where the balance is controlled by the positive weight  $\alpha \geq 0$ . Thus, the singular potential  $V_{\mathbf{x}}(u_{\mathbf{x}})$  measures how well each voxel  $\mathbf{x}$  fits into the known class models  $c_{\mathbf{x}}$  for a known transformation  $\mathcal{T}_{\mathcal{A}}$ .

$$V_{\mathbf{x}}(u_{\mathbf{x}}) = -\log(p_{app}(\mathbf{x} | c_{\mathbf{x}})) - \alpha \log(\mathcal{T}_{\mathcal{A}}(\mathcal{M}_{c_{\mathbf{x}}})(\mathbf{x})) \quad (3.33)$$

### Regularization term

The regularization term aims at penalizing changes of labels between neighbor pixels. This term aims at decreasing the influence of small variations between neighbor voxels. In particular this potential should correct erroneous labels due to incorrect singular potentials thanks to the neighbor voxels. Two definitions have been considered for this regularization term. First, the aforementioned aim was retained as definition. Then, the distance and the intensity of the voxels was also taken in account.

The regularization term was first defined as a penalty between neighbor voxels with different labels. This definition is the simplest one for the pairwise potential  $V_{\mathbf{x},\mathbf{y}}(u_{\mathbf{x}}, u_{\mathbf{y}})$ . The pairwise potentials are indeed defined by a positive constant value when labels are different and 0 when they are identical. The potential is thus directly given by the inverse Kronecker's delta  $\bar{\delta}_{x,y}$ .

$$V_{\mathbf{x},\mathbf{y}}(u_{\mathbf{x}}, u_{\mathbf{y}}) = \bar{\delta}_{u_{\mathbf{x}},u_{\mathbf{y}}} \quad (3.34)$$

$$\bar{\delta}_{x,y} = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{otherwise} \end{cases} \quad (3.35)$$

This first definition has two main flaws. First, this pairwise potential does not take in account the distance between two neighbor voxels, which is a lack for anisotropic images. Then, this potential does not take in account the relative class probability of both voxels. Thus a new potential  $V_{\mathbf{x},\mathbf{y}}(u_{\mathbf{x}}, u_{\mathbf{y}})$  is introduced that penalizes the difference of labels between two neighbor voxels as a function inversely proportional to the distance between the voxels and functions of the difference of intensity between them. An additional parameter  $\sigma$  is also introduced in order to characterize the image noise and thus the relevance of intensity differences.

$$V_{\mathbf{x},\mathbf{y}}(u_{\mathbf{x}}, u_{\mathbf{y}}) = \frac{1}{\|\mathbf{x} - \mathbf{y}\|_e} e^{-\frac{(\nu(\mathbf{x}) - \nu(\mathbf{y}))^2}{2\sigma^2}} \bar{\delta}_{u_{\mathbf{x}},u_{\mathbf{y}}} \quad (3.36)$$

### 3.6.3.3 Solving the problem

#### Principle

The segmentation problem is divided in two parts, registration and segmentation; however both steps cannot be done independently. The registration problem may be solved using a multi scale approach in a coarse-to-fine manner as described by Glocker *et al.* [Glocker 2008]. However, the optimization method for MRFs developed by Komodakis *et al.* cannot apply to the segmentation step due to its high memory consumption [Komodakis 2007b]. Segmentation of full images would indeed require either an unrealistic amount of memory, or to work on images with smaller sizes. Both approaches are not compatible with the clinical prospect of this study. Thus the multiscale approach for registration is extended to the segmentation in order to gradually refine a region of interest (ROI) that includes the structure to segment. Such an approach allows keeping a reasonable size of graph through the segmentation process, without compromising over its quality.

With a coarse-to-fine manner, the segmentation may begin with the entire image and progressively focus on a smaller part of the entire image while the resolution increases. At first segmentation is carried out on the entire image, but with a high subsampling factor. Thus, the segmentation will be feasible in practice as the number of nodes will remain small. Then, the segmentation is achieved at finer levels while using the ROI from the previous subsampling factor to restrict the segmentation to a smaller part of the initial image. This restriction allows keeping a graph with a reasonable size throughout the entire process while allowing segmentation at any chosen precision. However, the difficulty lies with the propagation of the ROI from a coarse level to a finer one. Indeed, one may not simply take the ROI and compute the same ROI with a smaller voxel size, because the entire set of possible displacements induced by the next iteration of the registration should also be taken in account. This set of displacement is nevertheless known, because the constraint for diffeomorphic deformations imposes maximal displacements for the finer level (sect. 3.5.3.2).

#### Implementation

The segmentation takes advantage of the multi resolution registration in order to gradually refine a region of interest for the structure to segment. Such an approach allows keeping a reasonable size of graph throughout the segmentation problem without compromising over its quality.

The segmentation of an image  $\mathcal{V}$  is achieved through an iterative process with diverse image scales. The segmentation process will be presented for a case with two classes but also extends to any number of classes (fig. 3.24). At each scale  $s$  of the approach are successively computed a transformation for the registration of the atlas  $\mathcal{T}_{\mathcal{A}}^{(s)}$ , a segmentation for the current scale  $\mathcal{V}_{seg}^{(s)}$  and an expected area for the next scale  $\mathcal{V}_{mask}^{(s+1)}$ . First, the research area for the first scale  $\mathcal{V}_{mask}^{(1)}$  is initialized as the entire volume. Then, the iterative process begins. For each scale of

image, the registration of the atlas  $\mathcal{A}$  on the volume for the current scale  $\mathcal{V}^{(s)}$  is first done using the proposed technique for atlas registration. This registration defines a transformation  $\mathcal{T}_{\mathcal{A}}^{(s)}$  for the current scale that is then used for the segmentation of the image  $\mathcal{V}^{(s)}$ . This segmentation is achieved using the optimization technique from Komodakis *et al.* and is limited to the segmentation mask previously computed  $\mathcal{V}_{mask}^{(s)}$  [Komodakis 2008]. Finally, the area of research for the next step  $\mathcal{V}_{mask}^{(s+1)}$  is computed from the current segmentation  $\mathcal{V}_{seg}^{(s)}$ . In the two class case, this segmentation  $\mathcal{V}_{seg}^{(s)}$  is the segmentation of the object against the background. This approach extends to the multiple class case by defining  $\mathcal{V}_{seg}^{(s)}$  as the union of all objects against the background. However, this extension will not be feasible when the union of the objects or the object contains too many voxels.

Set the initial mask  $\mathcal{V}_{mask}^{(1)}$  for segmentation as the entire image  $\mathcal{V}^{(1)}$ .  
 For each scale  $s = 1, \dots, S$

1. Compute the transformation  $\mathcal{T}_{\mathcal{A}}^{(s)}$  that matches the atlas  $\mathcal{A}$  on the subsampled image  $\mathcal{V}^{(s)}$ .
2. Segment the subsampled image for current scale  $\mathcal{V}^{(s)}$  inside the current mask  $\mathcal{V}_{mask}^{(s)}$ .
3. Using the current segmentation  $\mathcal{V}_{seg}^{(s)}$  compute a maximal area  $\mathcal{V}_{mask}^{(s+1)}$  where the object to segment should remain at next step.

Figure 3.24: Combined segmentation, solving process.

The difficulty lies with the propagation of the ROI from a coarse level to a finer one. Atlas registration was indeed treated before (sect. 3.5.3.3), and the segmentation problem meets the requirements of the MRF optimization technique developed by Komodakis *et al.* [Komodakis 2008, Komodakis 2009b]. Thus, only remains the computation of the segmentation mask  $\mathcal{V}_{mask}^{(s+1)}$  from the segmentation at one scale  $\mathcal{V}_{seg}^{(s)}$ . Indeed, the update of the segmentation mask cannot be done by simply up-sampling for a smaller size of voxel. In order to define this mask, one has to take in account the displacements induced by the registration at next step. However, the diffeomorphic constraint on the registration imposes a maximal displacement for each control point at the finer level (sect. 3.5.3.2). Thus, maximal authorized displacements of the points are known, and the segmentation  $\mathcal{V}_{seg}^{(s)}$  may be deformed to obtain a bounding mask for the next step  $\mathcal{V}_{mask}^{(s+1)}$ .

### 3.6.4 Protocol, method

#### 3.6.4.1 Segmentation priors

The set of images used for training and the atlas were identical to that of the segmentation with a single atlas (sect. 3.5.5.2). Only a spatial and an appearance prior were added.

The spatial probability maps were computed by deforming segmentation references on the initial image set  $\mathcal{I}$ . First, M.D. manually defined ground truth segmentations  $\mathcal{S}$  on the initial set of images  $\mathcal{I}$ . Then, every segmented volume was deformed following the deformation defined for the creation of the set of registered images  $\mathcal{W}$ . Each segmentation  $\mathcal{S}_i$  was transformed with the deformation field matching  $\mathcal{I}_i$  onto  $\mathcal{V}_{trg}$ . Thus, a liver segmentation was obtained for every registered volume. Finally, the spatial probability maps were computed as the average probability of being inside or outside liver for each voxel of the target volume.

The appearance priors were learned on the initial set of images  $\mathcal{I}$ . For every image inside this set, a reference segmentation is known for each class of tissue  $\mathcal{S}_i$ . These reference segmentations were retained to train the appearance priors, either as an average intensity distribution, or by training a classification function (sect. 4).

#### 3.6.4.2 Test protocol

The accuracy of segmentation was evaluated on the same 17 volumes than with the previous method (sect. 3.5.5.3). Then, the quality of segmentation was assessed using the three metrics previously introduced (sect. 3.5.5.1). First, intensities were normalized for every image of the test set. Then, an estimate of the liver segmentation was computed for each of these images. The segmentation was achieved using the combined segmentation (sect. 3.6.3) with an appearance prior defined either with a pdf model or by a classification function with texture features (sect. 3.6.2.4). Finally, the segmented regions were compared to the reference segmentations.

The combined segmentation divides into two parts, atlas registration, and segmentation with priors. First, atlas registration was achieved using the same parameters than in the previous approach (sect. 3.5.5.3). Then, the segmentation with the priors was achieved using the intensity dependent smoothness term (eq. 3.36) with a weight of 0.5 and an image noise set to 50. The singular potentials were defined next as a balance between the adequacies with the two priors. This balance was set to  $\alpha = 10$ , which gives ten times more weight to the spatial prior.

### 3.6.5 Discussion, Results

The addition of priors improves the quality of the segmentation of the liver. However, the resulting quality remains insufficient for use in clinical routine. First, the quality of segmentation improves when using combined segmentation (fig. 3.25). Indeed, the sensitivity improves of 7% and 10% compared to segmentation via atlas registration. Moreover, the improvement is even greater for DSC, where increases of 12% and



	Sensitivity	Specificity	DSC
Atlas only	$0.73 \pm 0.02$	$0.98 \pm 5 \cdot 10^{-5}$	$0.70 \pm 0.01$
Combined with pdf	$0.79 \pm 0.02$	$0.99 \pm 5 \cdot 10^{-5}$	$0.78 \pm 0.01$
Combined with texture	$0.80 \pm 0.01$	$0.99 \pm 3 \cdot 10^{-5}$	$0.83 \pm 4 \cdot 10^{-3}$

Figure 3.25: Combined segmentation, evaluation on several images. First, the results with atlas only are reminded. Then, the results of combined segmentation are given for two appearance priors, either an intensity distribution, or a classification function based on texture features.

19% are obtained. Then, the improvement of the segmentation is greater when more complex appearance models are retained. Indeed, the measures are all higher when using texture features instead of pdf as appearance model. Moreover, this improvement is significant for the DSC. Finally, an unexpected behavior is obtained when using a classification function as appearance prior; the variance of the results decreases by a magnitude between 2 or 3 compared to pdf as appearance prior. This gain probably comes from the better reliability of the second appearance prior. Indeed, a classification function is a more robust way to distinguish between the liver and the other tissues than a pdf.

The evolution of the segmentation with combined segmentation will now be qualitatively presented using the examples chosen for segmentation through atlas registration (sect. 3.5.6). First, an improvement may be seen when using the new priors. Moreover, this improvement is visible whatever the case (fig. 3.26)(fig. 3.27). Then, combined segmentation allows correcting cases with large difference between reference and segmentation through atlas registration. For example, on the bottom of the metastatic liver (fig. 3.26), combined segmentation allows getting a large part of the liver that was missed. Finally, combined segmentation deals better with sections of the liver with sharp boundaries. For example, on the bottom of the liver with HCC (fig. 3.27), the segmentations better follow the true shape of the liver. In particular, sharp contours are obtained. However, the segmentation remains often far from the true boundaries of the liver.

To conclude, while combined segmentation is an improvement compared to segmentation through atlas registration, the quality of the segmentation of the liver is not sufficient for clinical use. Indeed, the segmented livers are often incorrectly segmented at locations with high shape variability, for example at liver bottom and for some parts with sharp anatomical segments. Therefore, the precision of a combined segmentation of the liver is not sufficient to have a medical value. Indeed, the average DSC for the segmentations remains relatively small (83%) and far from state of the art techniques (95-96%) [Linguraru 2010, Massoptier 2008]. Moreover, the proposed approach has small value as initialization of a more precise segmentation technique. Indeed, the segmentation of the liver is often done in two steps, beginning with a rough initialization, before precisely segmenting the organ. However, the length of the combined segmentation prevents its use as an initialization step. Indeed, with an initialization of several minutes, it would become impossible to compete with

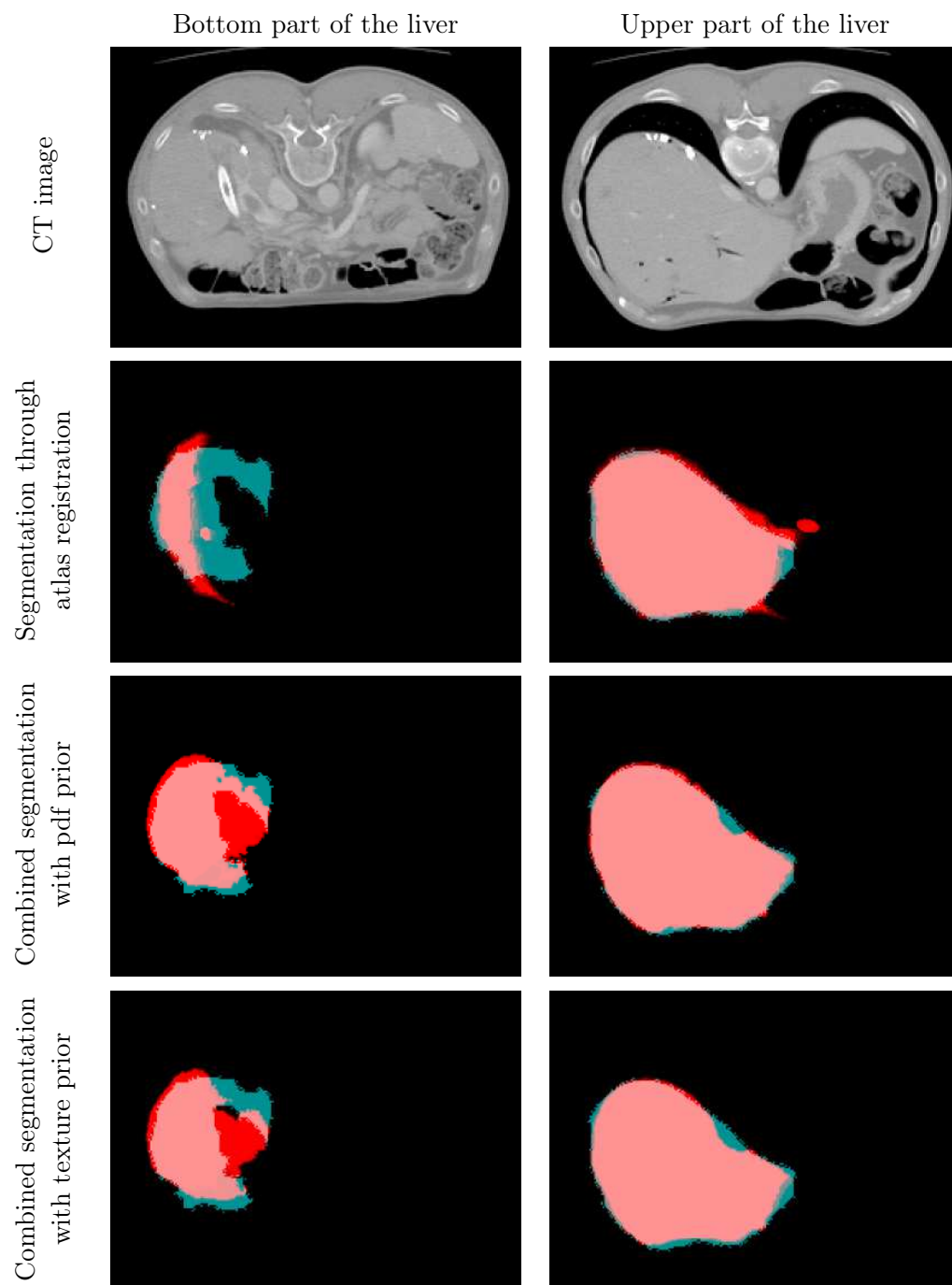


Figure 3.26: Combined segmentation, examples for a metastatic liver. Several results of segmentation are compared for segmentation through atlas registration, and combined segmentation with either a pdf or a texture prior. The reference segmentation is shown in blue and the automatic segmentation in red. The examples are given for a metastatic liver on a CT image with a size of voxel of  $1.6 \times 1.6 \times 10$ mm.

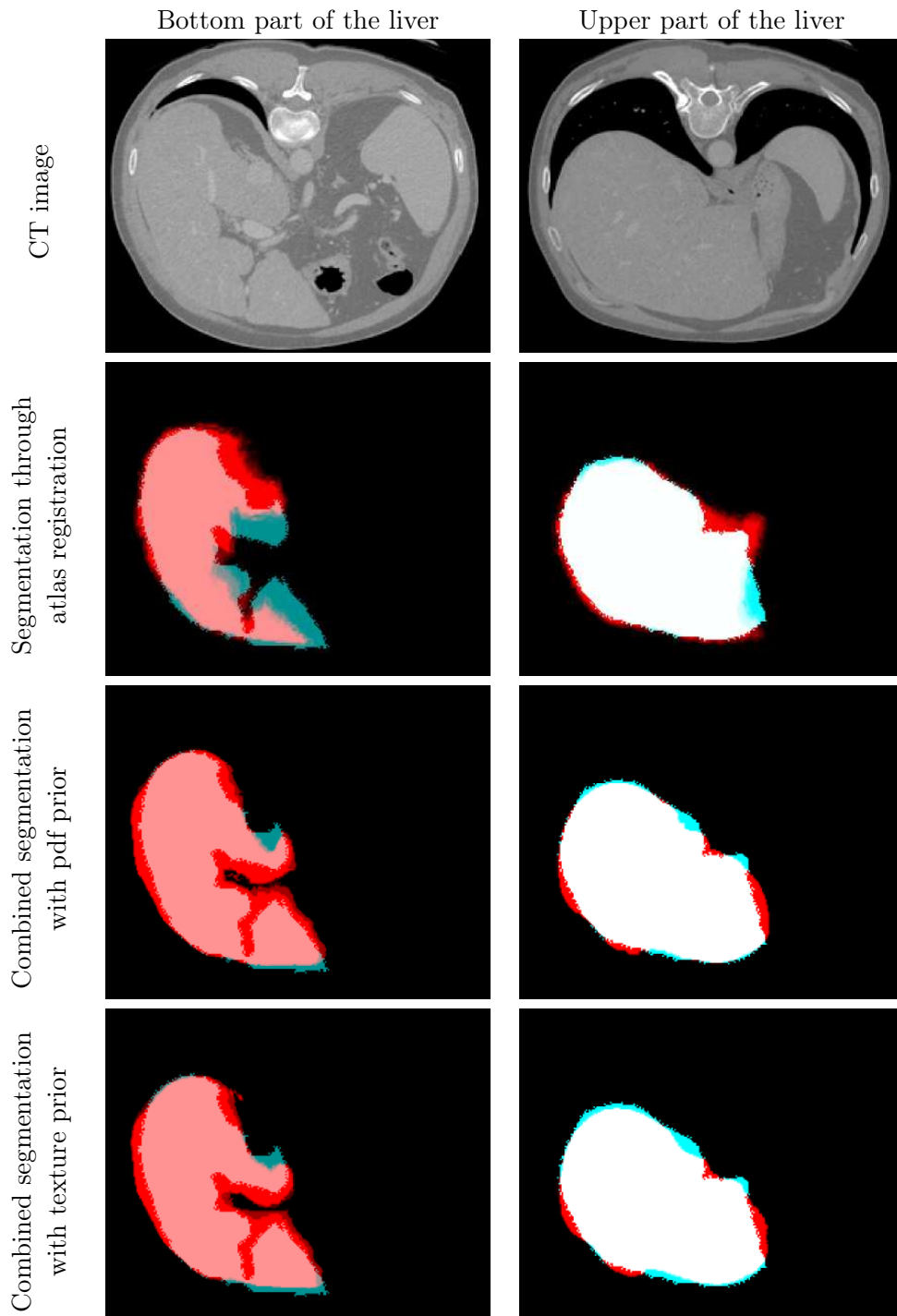


Figure 3.27: Combined segmentation, examples for a liver with HCC. Several examples of segmentation are compared for segmentation through atlas registration, and combined segmentation with either a pdf or a texture prior. The reference segmentation is shown in blue and the automatic segmentation in red. The examples are given for a liver with HCC on a CT image with a size of voxel of  $1.5 \times 1.5 \times 2.5$ mm.

methods that segment the liver in a matter of a dozen of seconds [Ling 2008].

### 3.7 Conclusion, summary

Three approaches have been proposed to segment the liver. First, a morphological technique was introduced to obtain an envelope from an approximate segmentation of the healthy parenchyma. Then, two methods based on the registration of an atlas were proposed. None of these methods offered sufficient results for the use of an automatic liver segmentation in clinical routine. Consequently, the segmentation of the liver tumors within a liver envelope will be done most of the time using a gold standard in the next chapter. However, the proposed methods still have some value.

A morphological tool was proposed that computes a smooth envelope that wraps around a binary ROI. First, this tool is generic and may apply to many other organs because many have smooth boundaries. Indeed, this tool allows segmenting an entire organ including the abnormal parts, when a segmentation of the healthy region is known. Then, the proposed approach might be improved. In particular, the computation of an initial outline should be further developed in order to obtain an approximate contour with well distributed nodes. This change should also improve the detection of the hole boundaries, because all measures inside the heuristic detection of these points would become more reliable.

Two methods of segmentation through atlas registration were proposed with poor success in the liver case. However, these methods might be valuable for other problems. First, a new statistical atlas was proposed that gives an optimal pdf at each location of the space. This atlas generalizes a previous atlas definition with Gaussian models. Moreover, the optimality condition chosen to construct the atlas extends to any type of distribution model. Indeed, while GMMs were retained in this work, the MDL condition is also valid for any other model. Thus, an atlas with non Gaussian models or with various models at same time may be constructed without any difficulty by following the proposed approach. Then, the proposed methods failed because the atlas could not capture the high variability of the liver and because the registration technique could not offer transformations precise enough to match the livers. However, many organs have lower shape variability than the liver. For such organs, the proposed segmentation techniques might be valuable. Finally, modeling the deformations induced by the lesions might have some value. Indeed, Ashraf proposed a tumor deformation model to deal with the registration of brain tumor images with an atlas [Ashraf 2006]. This approach induced a significant reduction of the registration error. Thus, a similar model in the case of the liver might be valuable to deal with the liver tumors. However, such model might be significantly more complex in the liver case, because of the impact of the lesions on the overall shape and size of the liver. Moreover, modeling tumor deformation will have no use when dealing with the normal anatomical variability of the liver.



# Tumor segmentation inside a liver envelope

---

*This chapter deals with the segmentation of the hepatic tumors within a liver envelope with normalized intensity. The proposed methods rely on the separation of the tissues on a high-dimensional feature space with advanced machine learning techniques. First, the construction of a classification function will be treated, which distinguishes between the diverse types of tissues inside the envelope based on their texture appearance. Then, this classification function will be used to segment the liver tumors by direct use of the learned separation of the tissues. And finally segmentation will be achieved as a MRF minimization problem based on the probability of belonging to each class of tissue according to this previous classification function.*

*Although the methods were chosen for the liver case, they may also be applied to any other problem where the objects to segment differ by their texture appearances.*

---

**Contents**


---

<b>4.1</b>	<b>Intro</b>	<b>130</b>
4.1.1	Context	130
4.1.1.1	Medical Background	130
4.1.1.2	Contribution of segmentation	132
4.1.2	Challenge	132
4.1.3	Role in the global process	133
4.1.4	Overview	134
<b>4.2</b>	<b>Machine learning</b>	<b>135</b>
4.2.1	Solving complex problems	135
4.2.2	State of the Art	136
4.2.2.1	Supervised methods	136
4.2.2.2	Unsupervised techniques	137
4.2.3	Importance of validation	138
4.2.4	AdaBoost, a relevant method for our problem	139
4.2.4.1	Characteristics	140
4.2.4.2	Theoretical background	141
4.2.4.3	Validation for AdaBoost	143
4.2.4.4	Adequacy to our problem	145
<b>4.3</b>	<b>Selection of texture features</b>	<b>146</b>
4.3.1	Texture metrics, quantifying local appearance	147
4.3.1.1	Texture, spatial arrangement of intensities	147
4.3.1.2	Quantifying intensity distribution	148
4.3.1.3	Co-occurrence matrix, tracking pixel pairs	149
4.3.1.4	Haralick's, quantifying pixel relations	150
4.3.2	Filtering, preparing images before treatment	152
4.3.2.1	Convolution to filter an image	152
4.3.2.2	Smoothing the image	154
4.3.2.3	Gabor's, enhancing texture characteristics	156
4.3.2.4	Relevant filters in the liver case	157
4.3.3	Defining texture features	162
4.3.3.1	Defining the features	162
4.3.3.2	Choosing a texture size	165
4.3.4	Feature selection	166
4.3.4.1	Worth of feature selection	166
4.3.4.2	Method	168
<b>4.4</b>	<b>Segmentation of tumoral tissues</b>	<b>171</b>
4.4.1	Creation of a classification function	171
4.4.1.1	Generation of a good training set	171

---

4.4.1.2	Definition of weak learners . . . . .	172
4.4.1.3	Learning the classification function . . . . .	172
4.4.1.4	Validation . . . . .	174
4.4.2	Segmentation based on pixel classification . . . . .	174
4.4.2.1	Straight segmentation . . . . .	174
4.4.2.2	MRF for better segmentation . . . . .	175
<b>4.5</b>	<b>Protocol . . . . .</b>	<b>177</b>
4.5.1	Defining the classification function . . . . .	177
4.5.1.1	Choice of image sets . . . . .	177
4.5.1.2	Normalization . . . . .	179
4.5.1.3	Feature Selection . . . . .	179
4.5.1.4	Learning . . . . .	180
4.5.1.5	Validation . . . . .	180
4.5.2	Implementation issues . . . . .	180
4.5.2.1	Classification . . . . .	181
4.5.2.2	Filtering . . . . .	181
4.5.2.3	Fast computation of first order features . . . . .	182
4.5.2.4	Fast computation of second order features . . . . .	184
4.5.2.5	Solving the MRF . . . . .	186
4.5.3	Test . . . . .	186
4.5.3.1	Straight segmentation of tumors . . . . .	186
4.5.3.2	Segmentation of tumors, MRF based . . . . .	187
<b>4.6</b>	<b>Discussion, Results . . . . .</b>	<b>188</b>
4.6.1	Straight segmentation of tumors . . . . .	188
4.6.1.1	Visual results . . . . .	188
4.6.1.2	Quantifying the results . . . . .	191
4.6.1.3	Impact of lesion size . . . . .	192
4.6.2	Segmentation of tumors using MRFs . . . . .	193
4.6.2.1	Visual observations . . . . .	193
4.6.2.2	Evaluation for several lesions . . . . .	195
<b>4.7</b>	<b>Conclusion, Summary . . . . .</b>	<b>196</b>

---



## 4.1 Intro

Liver cancers are deadly diseases with important frequencies in the world (4<sup>th</sup> most deadly cancer in the world and second in China [Tang 2001]). Surgical resection is the best treatment available, but may apply only when some conditions on tumor sizes are met. Therefore early diagnosis and accurate appraisal of tumors are critical. The exceptional resolution of CT images allows good detection rates for most tumor types. However, the detection of liver tumors is challenging due to the small observable changes between healthy tissues and tumoral ones. Such a task is even challenging for clinical experts. Thus a good and repeatable method to detect the tumors would be an advantage. Being clinically motivated, such an approach should work for diverse tumor types at the same time in order to avoid multiple and successive segmentations, as well as for several protocols and CT machines to prevent methods specific to one site. Besides these constraints, the problem is reduced to segmentation inside a liver envelope, where the segmentation is easier to achieve.

### 4.1.1 Context

First, the medical background of liver tumors will be quickly reviewed. Then, focus will be on the value of an automatic segmentation compared to a manual one. First, background information will be reviewed, but with less details than in previous chapter (sect. 2.2). The focus will be on relevant characteristics of the liver and tumors along with the role of segmentation for therapeutic strategy, while focusing on some common types of cancer. Then, the contribution of an automatic segmentation will be discussed from a medical perspective, while distinguishing improvements of existing strategies from possible novelties.

#### 4.1.1.1 Medical Background

The liver is a voluminous organ located inside the abdomen and at crossroads between many vascular networks. From this particular position come many critical functions and a surprising ability, regeneration. The liver is the biggest non-hollow organ inside the human body, where it amounts for around 3% of the human body mass. Liver has many functions related to digestion, namely transformation of some nutrients, storage of some components and barrier against some poisons. These functions come in part from the location of the liver at a junction between several networks. The liver indeed contains three blood-networks, portal vein, hepatic arteries and hepatic veins, to which are added a biliary tree and a lymphatic network. Besides these functions, liver has the surprising ability to regenerate. The liver is indeed able to grow again from a small initial volume, which provides many possibilities for treatment by surgical resection [Vilgrain 2002, Fritsch 2003]. However, a sufficient volume of healthy liver is required for the patient to survive. The hepatic function is indeed necessary for life, and no substitute exists like dialysis for kidneys. Some companies are admittedly endeavoring to create an artificial liver, but such

substitutes are still far from ready [Demetriou 2004, HepaLife Technologies 2009, McGuckin 2008].

The liver is an organ where many tumors can be found, either as a metastatic site or as a primary one. Liver is one of the most common locations for metastases, second only after the lungs. Many tumors can indeed create hepatic metastases; in particular Pickren showed that 41% of patients with cancer death had liver metastases [Pickren 1982]. Explanations for this high metastatic activity are twofold. First, liver is the first organ met when coming from the digestive system. Thus liver is the first location for metastases from digestive cancers. Then, the blood flow inside liver is second only to the lungs, which increases the probability of metastases. Concerning primary cancers, Hepatocellular carcinoma (HCC) is the most common primary type of liver cancer. This cancer not only has a poor prognosis (less than 5% 5-year survival rate [Clark 2005]), but the death toll is increasing and is forecast to continue to increase. Indeed, while many cancers are less and less deadly, deaths from HCC are increasing (1.4/100000 for 1976-1980 and 2.4/100000 for 1990-1995 in USA) [Obad 2007, Tang 2001, Taylor-Robinson 1997, El Serag 1999, Bruix 2005]. Moreover, the incidence of this cancer is forecasted to increase for western countries.

CT scanner is the best imaging modality for the detection of liver tumors and pre-operative planning, as well as for treatment monitoring [Hélénon 2003, Choi 2006]. However, a correct detection of lesions requires the choice of an adequate enhancement phase, as well as good reading abilities to detect small differences between tissues in a noisy environment. First, a striking resemblance often exists between tumoral and healthy tissues. However lesions become more visible after enhancement with a contrast product. These injection protocols, or phases, are roughly four, which correspond to diverse delays between the injection of a contrast product and the image acquisition. At the image level, an adequate phase allows enhancing the tumors, but this enhancement comes with some drawbacks. Appearances and intensity ranges indeed change both for the healthy tissues and the tumoral ones from one phase to another. Then, another challenge is due to the liver itself. Being an organ with a high degree of vascularity, the images obtained are really noisy, a property that is amplified due to the enhancements.

For most tumors, surgical resection is the only treatment that can lead to complete healing, but it cannot apply to all cases. For metastases, resection is the best available treatment, allowing good survival rates (20-40% 5-year survival rate on selected patients [Hélénon 2003]). Surgical resection is also the preferred treatment for HCC [Tang 1989], with a favorable prognostic (40-60% 5-year survival rate according to Tang [Tang 2001] and 50-70% for Blum [Blum 2005]). Other treatments are inefficient but not useless. The role of the other therapeutic strategies is twofold. First, the treatment may be palliative. Then, treatments may be used to make surgery possible. One of the requirements for surgical resection is indeed to preserve a sufficient percentage of healthy liver after resection (20-30%). Treatments may be used to modify this ratio and transform a non surgical case into a case where surgical resection becomes possible.

#### 4.1.1.2 Contribution of segmentation

An automatic segmentation has many applications both for routine clinical use and for research purposes. First, an automatic segmentation offers a time gain for the physician for some repetitive tasks such as follow-up and preoperative planning. Then, an automatic segmentation offers more information than actual segmentation techniques, which may be relevant for clinical assessment of treatments or for better characterization of lesions.

Determining the volume of a lesion is important for the selection of therapeutic strategies. Thus, physicians often have to segment lesions for follow-up of patients or to plan surgery. Some patients are indeed known to have a higher risk of developing a hepatic cancer, e.g. because of a pre-existent pathology in the HCC case. These patients are followed up regularly to detect the development of a malignant lesion at an early stage. Thus, physicians have to check images at regular intervals to detect these lesions (4-6 months [Bruix 2005, Okada 1993]). An automatic tool would mean a time gain for these physicians. This gain would be even more important in the case of surgery planning. Before deciding for surgical resection, physicians have first to check if the patient is eligible for resection by evaluating the remaining healthy liver after surgery. Then, they have to plan the surgery by defining where to cut. For both these tasks a 3D segmentation is an advantage.

The availability of a 3D segmentation offers interesting possibilities for the characterization of tumors. Indeed, the current methods have some lacks for the evaluation of the responses of lesions to treatment. Thus, physicians are currently trying to find better methods of characterization, but are limited by the lack of availability for 3D segmentation techniques [Eisenhauer 2009]. Current characterization methods are indeed still using only 2D information, with a single diameter to characterize a lesion (RECIST [Therasse 2000, Eisenhauer 2009]) or two perpendicular diameters (WHO response criteria [Miller 1981]), where 3D could offer more. However, 3D characterization requires a segmentation method, because assuming that physicians will do everything manually is not reasonable. 3D offers a better insight of lesions by providing the volume of a lesion, which is more informative than the diameter(s) of a section. Moreover, a ROI allows quantifying the pixel density inside a lesion, which permits in particular to characterize necrotic regions. This last prospect is a strong improvement for the assessment of therapeutic medicines. Many modern approaches aim indeed for necrosis of lesions instead of shrinking of tumors, while present response measures do not take this parameter in account.

#### 4.1.2 Challenge

The segmentation of liver tumors is a challenging task. First, the wide range of possible appearances for both tumoral and healthy tissues makes the distinction between tissues more complex. Then, segmentation is achieved on a noisy background. Finally, the clinical prospects introduce additional constraints in term of speed and robustness.

Healthy and tumoral tissues inside liver have various intensities and appearances arising from the various types of tumors, the changes induced by the injection phases and the technical differences between imaging machines. First, various tumors have to be segmented, with miscellaneous appearances and often a striking resemblance with healthy tissues. Moreover, some tumors are only visible, or more distinguishable, when an appropriate phase has been considered. Unfortunately this enhancement impacts all appearances and shifts the intensities of both tumoral and healthy tissues. Finally, segmentation is done for a multicenter study, which introduces variations caused by changes of injection protocols and imaging machines.

Besides this variability of appearances and intensity ranges, the segmentation should cope with the non-existent edges of some lesions as well as with intensity variations between tumors and healthy tissues that may be inferior to the background noise. First, some lesions do not contain visible edges (e.g. diffuse cancers or some HCCs), which prevents the use of approaches based on edges. Then, the nature of liver and the requirement of enhancement introduce a high degree of noise. The liver is indeed an organ with a high degree of vascularity and a contrast product is injected prior to image acquisition. Thus, the CT images are noisy, which introduces uncertainty for the link between the intensity and the type of tissue. This uncertainty implies that the intensity value at one pixel does not directly describe the type of tissue contained inside this pixel, as the difference of intensity between two tissues may be smaller than the amplitude of the noise.

The clinical prospect of this study introduces several other constraints, on the robustness of the method, on its speed and on quality of detection. First, the method should work for real images, meaning images from diverse CT machines, with various levels of resolution and for both overlapping as well as non overlapping slices. Indeed, the method should handle images, whose quality is sufficient for the diagnosis by a physician, and not only perfect images with low noise. Then, the total length of segmentation should remain reasonable. Segmentation should not take more than a few minutes from the opening of a medical image to obtaining the results of segmentation. Finally, segmentation should be as good as possible for lesions  $\geq 10\text{mm}$ . This requirement of quality comes from the need of a good segmentation for diagnosis, while the constraint on size comes from medical guidelines. Lesions smaller than 10 mm are indeed considered as non-measurable lesions (RECIST [Therasse 2000, Eisenhauer 2009]); thus, these lesions are excluded from protocols.

### 4.1.3 Role in the global process

Segmentation of the hepatic tumors inside an already known liver envelope is the last step of the proposed method to segment tumors inside CT images. Before the segmentation presented in this chapter, a liver envelope was computed (sect. 3), and the intensities were normalized within this envelope (sect. 2.6). The normalization of CT images is a crucial step, because it allows using a common intensity basis, where intensity variability between images remains small. Working inside a liver

envelope allows simplifying the problem and getting better results. Previous works indeed showed that segmentation of hepatic tumors was more accurate when done inside liver only, in particular for automatic methods (sect. 2.4.2).

The proposed segmentation approach deals with a more global problem than segmentation of liver tumors only. Indeed, this approach applies to any segmentation problem where the researched objects differ by their texture appearances. Moreover, the proposed approach was chosen for its robustness toward noise and its computational efficiency, thus it could be a relevant technique for many other segmentation problems with clinical applications. In particular, the approach was tested to define an appearance model for the segmentation of liver in a previous chapter (sect. 3.6.2.4).

#### 4.1.4 Overview

The proposed method addresses the problem of segmentation of diverse tumor types versus the healthy tissues within the liver. This detection is done while satisfying a number of constraints, like different resolution levels, various enhancement phases and protocols, in noisy anisotropic CT images and in a reasonable time. A learning based approach is chosen to address this task, which addresses features selection, tumor classification and the use of classification for segmentation. The intensities are assumed normalized, and from a predefined set of features are selected the ones for which an optimal classifier exists to separate healthy versus non-healthy tissues. This classifier is based on the AdaBoost method that decomposes the process into a number of weak classification tests [Freund 1997]. Once such a multi-level, multi-feature classifier has been determined, the task of segmentation consists of combining the results of classification applied to a new volume towards achieving a good segmentation.

A learning approach was followed because no edge method applies due to the high noise in the liver, and because there are different kinds of objects to segment (some tumors are so dissimilar that they have nothing in common), and because the appearance of each object vary widely. Moreover, the analysis of previous works shows the contribution of machine learning for the segmentation of liver tumors (sect. 2.4.2). Machine learning techniques search for complex patterns inside data, which is precisely the case here. Indeed, the tumor appearances vary widely and there are neither prior knowledge for every possible appearance of the lesions nor techniques to combine this prior information. Consequently, machine learning seems an appropriate approach to distinguish healthy from tumoral tissues.

Machine learning techniques determine a process that separates a set of observations. In the case of this study, observations consist of intensities and classes correspond to tumor versus non-infected tissues. Despite the resolution of CT images, one can imagine that the separation of healthy versus non-healthy samples in this space is almost impossible as intensity alone cannot distinguish between the diverse tissues. The use of filters and their responses, as well as texture metrics is a convenient way to take into account the relative context, and consider features

with better discrimination power along with more robustness towards noise. One can consider either the responses themselves or seek a separation on a subspace that encodes the dependencies at the local scale of these responses, which is precisely the contribution of machine learning.

The proposed approach can extend to any kind of segmentation problem, where objects differ by their texture patterns. While initially aimed at liver tumor segmentation, the entire technique does not rely on any information specific to the liver, meaning that the approach is applied to the liver only by setting some parameters. Thus, the proposed technique may be applied to other segmentation problems, while keeping the same advantages concerning noise management, clinical setting and speed.

## 4.2 Machine learning

### 4.2.1 Solving complex problems

Machine learning is a scientific domain concerned with the creation of algorithms able to modify their behaviors in function of data. On contrary of expert systems that rely on structured knowledge bases to treat new information, machine learning aims to construct algorithms able to deal with problems when no structured knowledge is available. This domain is strongly related with many other scientific domains such as statistic, Artificial Intelligence, and data mining. . . In particular, machine learning is often used for pattern recognition and medical imaging, because of the difficulty to translate such problems into intelligible ones for a computer.

In this study, machine learning will be reduced to the case of classification with inductive approaches. Given a set of samples, classification techniques aim at deducing rules allowing predicting the classes of new samples, where the classes are labels describing the type of the samples. Then, only inductive approaches are retained as possible machine learning techniques. Inductive approaches aim at learning a set of rules that remains unchanged once learned, as opposed to transductive approaches that update the set of rules while they are applied. Transductive techniques were excluded because they are not compatible with a clinical prospect. First, obtaining varying results over time is unacceptable for physicians, because measures obtained through classification would not be reliable. For example, the clinical assessment of treatments cannot be achieved with these approaches. The evolution of classification would indeed introduce a bias on all measures, which would prevent any conclusion. Finally, varying behaviors are often against the law in the medical domain. For example, the FDA does not give clearance for medical software, whose behavior evolves over time.

The sets of rules learned for classification using inductive methods will thereafter be called classification functions or classifiers.

Two main types of algorithms will be considered, supervised and non-supervised. Supervised methods deal with the training of a classifier for data sets where both samples and expected classes are known. Given a set of samples and a set of expected

classes, supervised approaches aim at maximizing the number of well classified samples, where the expected class of well classified samples is the one predicted by a classifier. On the opposite, the non-supervised approaches deal with the research of similar samples in a data set without knowing the expected class of any sample. For non-supervised methods, only samples are required. Expected classes are unnecessary, because such approaches aim at gathering similar samples in clusters. These clustering approaches qualify as classification methods because the cluster that better matches with a new sample can be used to define the class of this sample. Because clusters gather similar samples, one can reasonably assume that each cluster contains mainly elements of a same class. Thus a class can be assigned to each cluster, which allows classification.

## 4.2.2 State of the Art

### 4.2.2.1 Supervised methods

#### Support Vector Machines

Support Vector Machines (SVM) belong to supervised machine learning techniques and are used for classification problems. SVM combine linear separation in a feature space with a non-linear kernel in order to define a non-linear classification in the direct space. SVM apply to many problems by choosing an adequate kernel and define a measure of the robustness of classification. However, some constraints are introduced by the underlying theory. Indeed, the size of the training set is limited by the numerical solvers, and the classification of new samples may be slow.

SVM were introduced by Boser, Guyon and Vapnik as a classifier that optimizes the margin between training samples and the decision boundary in a feature space [Boser 1992]. Usually this feature space has a bigger dimension than the direct space, but it does not complicate the estimation problem, because simpler methods can be used (mostly linear methods in feature space) [Müller 2001]. The difficulty is the mapping between the two spaces, which can be difficult and complex to use and often impossible to get (you only have to find an image big enough and the mapping would be too big). However, thanks to the Mercer's trick, the mapping constraint disappears. Mercer's theorem indeed implies that *every linear algorithm that only uses scalar products can implicitly be executed in [the feature space] by using kernels, i.e. one can very elegantly construct a nonlinear version of a linear algorithm* [Schölkopf 1998b]. Thus a hyperplane that separates two classes may be found inside a feature space, using only kernels applied to samples from the training set. Implications of this approach are many; the decision boundary is constrained mostly by the closest samples from the boundary, called Support Vectors, the size of the margin mirrors the robustness of the separation (a large margin will imply more difference between the classes).

### AdaBoost

AdaBoost is a supervised machine learning technique introduced by Freund and Schapire, which constructs a strong classifier using only basic functionals with low discrimination ability [Freund 1997, Schapire 1999]. AdaBoost stands for adaptive Boosting and is one of the Boosting algorithms, which are meta algorithms used for supervised machine learning. AdaBoost relies on low level functionals, called weak learners, to construct a classifier in an adaptive evolution [Meir 2003].

AdaBoost has been applied in many domains including computer vision, because of its genericity, its simplicity of use, and its speed. First, the AdaBoost algorithm applies to many problems without lots of tuning. Only the number of rounds has to be chosen, as well as the definition of the weak learners. The main difficulty lies in the choice of adequate weak learners. These weak learners should indeed be chosen accordingly to the considered problem. As mentioned before, the AdaBoost approach applies to many domains by selecting various discrimination methods. However, the value of the approach depends entirely on the choice of adequate weak learners. Indeed, from these weak learners stems the customizability of the approach. Finally, the machine learning technique provides a classification function as a linear sum of weak learners. Thus, the classification is fast when the weak learners are fast. The AdaBoost technique will be further detailed in a later section (sect. 4.2.4).

#### 4.2.2.2 Unsupervised techniques

##### K-means

K-means is a non-supervised algorithm that allows partitioning a population into  $k$  clusters by minimizing the sum of squares within clusters. This algorithm was introduced by MacQueen in 1967 and is still widely used because of its good clustering ability [MacQueen 1967]. From a set number ( $k$ ) of centroids, the idea is to iteratively assign samples to the closest centroids before updating these centroids using the samples inside each cluster. This technique provides  $k$  clusters, where all samples are similar with respect to a chosen distance. This clustering technique was detailed in a previous chapter (sect. 3.4.2.4).

##### Kernel Principal Component Analysis

Kernel Principal Component Analysis or KPCA aims to apply Principal Components Analysis (PCA) in feature space in order to extract features that represent data with fewer uncorrelated variables. KPCA was presented by Schölkopf *et al.* as a non-linear generalization of PCA, where the non-linearity is introduced with non-linear kernels [Schölkopf 1998b].

The PCA is widely used in the statistic domain to represent data using a small set of uncorrelated variables that better explain the data distribution [Pearson 1901]. From a mathematical perspective, PCA relates to the research of an ordering of axes inside the feature space, for which the axes are sorted in descending order,



functions of the significance of each axis to explain the data distribution. It is thus possible to define subspaces embedded in the initial feature space, where data is approximated with a chosen precision. This reduction of variables has several applications; it can be used for visual representation of high dimensional data, to find linked component, to approximate data (not using features from the lesser relevant components), for denoising for example [Mika 1999]. PCA also impacts the classification methods. For example, the projection of samples in a subspace simplifies the subsequent application of other classification methods, and sometimes makes the method possible.

KPCA applies PCA in a feature space where features may be non-linear in the direct space. Hence KPCA can extract nonlinear structures in data. This approach offers several advantages; experimentally fewer components than with PCA are required to obtain similar classification results, and KPCA can be used for every case where PCA applies, while the choice of a kernel offers a wider range of possibilities [Schölkopf 1998b, Schölkopf 1999]. This approach may be seen as a non-supervised learning method, because combined with a clustering algorithm like k-means (sect. 4.2.2.2), it can define mean appearances in feature space for each class (possibly multiple appearances for each class). However, returning in direct space is still complex and not always possible, as the mapping from direct to feature space is not surjective. The research of an approximate point in direct space that represents a point in feature space is called the preimage problem and is still discussed [Schölkopf 1998a, Bakir 2004, Kwok 2004]. The method is also limited by the required computations. Indeed, the algorithm does not require any optimization technique, but requires diagonalization of a matrix. This diagonalization is an advantage for small feature spaces, but for higher dimensional spaces it often becomes an unsolvable problem.

### 4.2.3 Importance of validation

The validation is a crucial step between learning and use of a classifier. This validation aims to verify that the learned classifier is valid. This step is sometimes considered as the final step of the learning process, whose aim is to verify that the patterns found in the training set are still found in other data sets. The learning algorithm may indeed create a classifier that cannot be generalized to other data sets, either because the training set was not representative of the general case, or because the learning method did not apply to the problem. Moreover, this step also allows the detection of overfitting.

The validation consists in evaluating the performance of a classifier on a new test set. A validation set is first created and composed of samples and expected classes, where no sample was previously used for training. The learned classifier is then applied to every sample from the validation set and the performance of the classifier is evaluated by comparing predicted classes to expected values. The performance is finally assessed and a choice is made about the future of the classifier.

The performance allows deciding whether the classifier should be kept or rejected.

If the performance is sufficient for the considered problem, the classifier is ready for use. Otherwise the learned classifier should not be further used. Several causes may explain this bad performance. First, the retained machine learning technique may be unsuitable for the considered problem. Then, the training set may be non representative of the general case. Finally, the bad results may be due to overfitting.

A classifier with good generalization abilities cannot be trained on a non representative training set. As inductive machine learning techniques are used, the training set is used to infer rules applicable in the general case. Thus, rules learned on a biased set cannot provide good results in the general case.

Overfitting occurs when the learned classifier is more complex than it should be. Any training set will contain some noise. When a classifier gives too much importance to this noise, the classifier will tend to model random variations inside the training set, which will affect the predictive value of the classifier. An example of overfitting is given for the classification of two classes linearly separated in a 2D-space with some incorrect samples (fig. 4.1). Two methods are used to find a separation between the two classes. A linear separation provides a simple and good dividing line for the artificial sample (fig. 4.1.a), while a B-splines boundary better classifies all samples (fig. 4.1.b), but at the expense of the simplicity of the regression model. This second separation gives too much weight to two incorrect samples, which might induce errors of classification on other data sets. When overfitting, classifiers will tend to be too complex, which will in turn lessen their predictive value. As Occam's razor states, the simplest solution is usually the correct one.

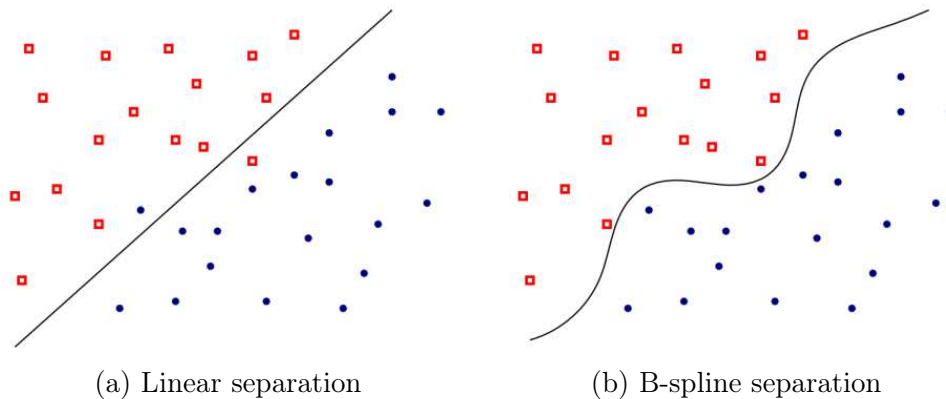


Figure 4.1: Separation of two noisy classes in the space, the problem of overfitting. An artificial sample is considered, with two classes shown as red squares and blue circles that are linearly separated. However, some noise is added to the sample. A separation is then researched to distinguish these two classes. First, a linear separation is computed (a), and then a separation based on B-splines is researched (b).

#### 4.2.4 AdaBoost, a relevant method for our problem

Freund and Schapire introduced a supervised learning method named AdaBoost that uses several weak learners to construct a strong classifier [Freund 1997]. This

method has been widely used, because it runs fast (when the weak learners are fast) and applies in many cases. However, the quality of the results is dependent on the selection of adapted weak learners.

Given a training set, composed of samples with expected results, AdaBoost provides a learning algorithm to construct a strong classifier by combining low level functionals (named weak learners) in a coarse-to-fine manner. This strong classifier may later be applied to treat new samples. Variations of AdaBoost apply to multi-class problems. However the context of this work requires only binary classification, therefore AdaBoost will be reviewed only for the binary case.

#### 4.2.4.1 Characteristics

The AdaBoost approach offers several advantages. In particular, the AdaBoost algorithm offers very fast results of classification, and applies easily to many problems. This ease to customize comes from the ability to obtain strong classifications from basic discrimination functionals. However, these advantages come at the price of incorrect results when the weak learners or the training set are inadequate. Thus, the validation step is particularly relevant for AdaBoost.

While the learning process may take a long time, applying a classifier is often fast. A slow learning process is not too restrictive because it is done off line and only once. However, fast application in everyday use is a strong incentive. The overall speed of the strong classifier is admittedly limited by the choice of the weak learners, but being low level functionals they should be fast.

AdaBoost is easy to use and to customize for many problems. The user only has to define two inputs for the learning algorithm, a maximum number of learning rounds and a set of weak learners. The maximal number of rounds is easy to set, however the definition of the weak learners is more complex and problem dependent. Indeed, the freedom of choice for the weak learners induces the good adaptability to any problem. However, an incorrect choice of weak learners will lead to poor results.

AdaBoost provides a strong learning method. The technique is able to construct good classification functions even for weak learners with small discrimination ability. Using weak learners that offer the best discrimination possible is nevertheless better. Freund and Schapire indeed showed that the training error decreases exponentially with the number of weak learners, and that the better the weak learners are, the better is the final classification function (sect. 4.2.4.2). These two facts imply that the quality of the weak learners impacts both the quality and the speed of the final classifier.

The choice of an adequate training set is crucial for the quality of this classification function. Indeed, the training set should represent as well as possible the variety of the problem that is aimed at. If there are more samples of one class, this class will be given more weight in the learning process. Thus the number of objects in each class inside the training set matters. Consequently, an unequal distribution of samples within the classes will impact the learning process. It may however be useful in order to favor one class.

The quality of classification on the training does not allow inferring the quality for a new set, even for a well chosen training set. The classifier should be validated and sometimes truncated whatever the quality of classification on the training set. The learning process follows indeed a coarse-to-fine approach, which implies that the later stages of the algorithm are more prone to overfit. Moreover Meir stated that AdaBoost is sensitive to noise, mostly during the late learning stage [Meir 2003]. Thus the contribution of the end of the classifier may be very small or even detrimental to the classification of new samples. A validation should therefore be done in order to keep only the relevant components of the classifier.

#### 4.2.4.2 Theoretical background

##### Introduction

Given a training set  $\chi$  composed of labeled pairs  $(x_i, y_i)$ , AdaBoost aims to construct a strong classifier  $H$  that provides rules to predict the class of a sample  $x_i$  with a good accuracy for the training set  $\chi$ . Given a set of simple functionals, named weak learners  $\{h_j\}_j$  and a maximal number of learning rounds  $T$ , the algorithm constructs this classifier  $H$  as a weighted sum of  $T$  weak learners balanced by their weights  $\alpha_t$  (eq. 4.2). It should be noted that a same weak learner may be used more than once, and that some available weak learners may remain unused.

$$\begin{aligned} \chi &= \{(x_1, y_1), \dots, (x_m, y_m)\} \\ x_i &\in X, \text{ with } X \text{ instance space} \\ y_i &\in Y \subset \mathbb{Z}, \text{ class of } x_i \end{aligned} \quad (4.1)$$

$$H(.) : x \rightarrow \text{sign} \left( \sum_{t=1}^T \alpha_t h_t(x) \right) \quad (4.2)$$

The learning problem consists in defining a classifier that minimizes the prediction error for the training set  $\chi$ . This minimization problem may be restated as the research of a set of weights  $\alpha_i$  along with a set of weak learners  $\{h_1, \dots, h_T\}$  that minimize the classification error on the training set. This error is well defined, because for each sample or set of features representing the sample  $x_i$ , the expected class  $y_i$  is known.

##### Learning algorithm

The main idea behind the learning process is to begin by classifying the easiest cases and then focus on the more difficult cases. This adaptation is done using a distribution  $D_t$  that assigns a weight to each training sample. During the learning process the weights of the well classified samples will decrease while the weight for misclassified samples will increase in order to focus on the cases that are still not well handled. These weights allow creating iteratively the classifier, beginning with the more general cases, before focusing on the most difficult ones.

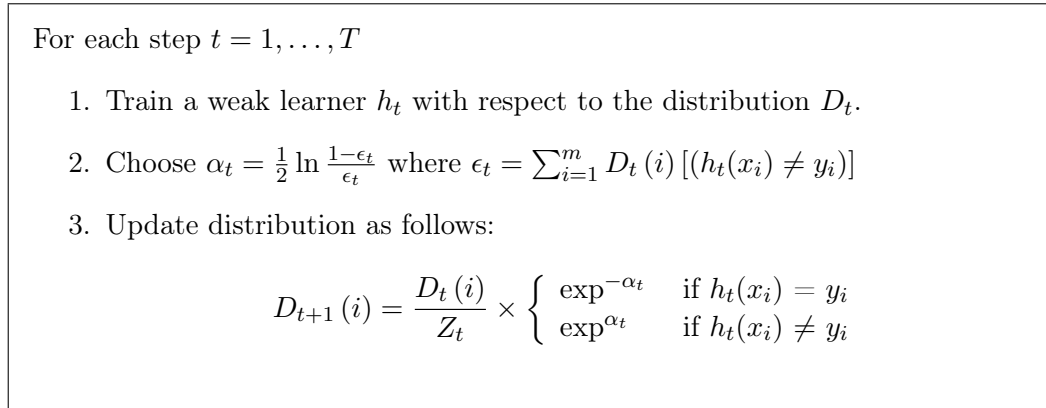


Figure 4.2: AdaBoost, algorithm of the learning process.

The learning process is iterative and contains three main steps (fig. 4.2). First a weak learner is chosen, then the associated error is computed to define the weight of the weak learner in the strong classifier, and finally the distribution is updated to take in account the evolution of the classification function.

First a good weak learner  $h_t$  has to be chosen. The only constraint is the selection a weak learner with an error smaller than 0.5 for the current distribution  $D_t$ . When there is no error or when no weak learner does better than the random classifier, the algorithm stops, because the accuracy of the classification cannot be improved on the training set. For the first case, classification is already perfect for the training set, thus no improvement is possible. And for the second case, the addition of another component will be detrimental to the classification. Thus, stopping prevents a reduction of the classification ability. As mentioned before, weak learners with small errors induce a faster decrease of the error in the strong classifier. Thus a common method of choice is the selection of the weak learner with the smallest error in order to obtain good results with fewer components. However other methods may be chosen, for example to maximize the robustness of the classifier.

Then the error  $\epsilon_t$  is computed with respect to the current distribution  $D_t$  and later used to define the weight  $\alpha_t$  for the weak classifier  $h_t$ .

Finally the new distribution  $D_{t+1}$  is computed in order to focus more on samples that were not well classified during the step t, and to reduce the importance of well handled samples. A normalization term  $Z_t$  is also introduced to keep a distribution.

The initial distribution often gives an equal weight to each sample hen no previous knowledge is known,  $\forall i D_1(i) = \frac{1}{m}$ . However the distribution may also be used to give more weights to some samples or to compensate for an unbalanced training set.

### Training error

Freund and Schapire provided an upper bound for the training error, and showed that this error decreases exponentially with the number of weak learners. Moreover, the authors showed that the better the weak classifiers are the better is the final classification function [Freund 1997].

Let us write the error  $\epsilon_t$  for the weak learner  $h_t$  as  $\epsilon_t = \frac{1}{2} - \gamma_t$ , where  $\gamma_t$  represents how much better than random classification is  $h_t$ . Freund and Schapire proved that an upper bound of the classification error is expressed as an inverse exponential of the sum of all successive square  $\gamma_t$  (eq. 4.3). Consequently, the error decreases exponentially and the better a weak learner is, the better is the induced improvement.

$$\prod_t \left[ 2\sqrt{\epsilon_t(1-\epsilon_t)} \right] = \prod_t \sqrt{1-4\gamma_t^2} \leq \exp\left(-2\sum_t \gamma_t^2\right) \quad (4.3)$$

### From score to probability

The learning process constructs a classifier that returns a binary result. However the AdaBoost score  $F$  computed before applying the sign function contains more information than a single binary result (eq. 4.4). Friedman *et al.* indeed showed that the AdaBoost score could be used to compute the class probability [Friedman 2000] (eq. 4.5).

$$F(x) = \sum_{t=1}^T \alpha_t h_t(x) \quad (4.4)$$

$$p(\mathbf{x}) = \frac{\exp^{F(\mathbf{x})}}{1 + \exp^{F(\mathbf{x})}} \quad (4.5)$$

#### 4.2.4.3 Validation for AdaBoost

As mentioned before, validation is a crucial step between learning of a classification function and use of this function on new data sets (sect. 4.2.3). This step aims generally at checking that the learned classifier is relevant for the problem. For the case of AdaBoost, validation also allows doing some final tuning before freezing the classifier for later use.

First, validation allows verifying that the classification function is relevant for the considered problem. It may happen that learning failed, or was applied to a problem different from the one that was considered first because of a bias in the training set. A new set of samples is thus created and used to evaluate the quality of classification on new samples. This evaluation attempts to quantify the generalization error of the classification function and its robustness when dealing with new data. It also allows rejecting incorrect classification functions.

Then, validation may be used to tune up the classification function, by truncation of the classifier aiming at keeping only the relevant components of the weighted

sum that defines this classifier. The contribution of this step will be illustrated with an example showing the quality of classification, functions of the number of components for two classifiers obtained while selecting features for the classification of liver colorectal metastases (fig. 4.3). For these two examples, the quality increases with the number of components at first, and later tends to stall (fig. 4.3.a,b) or even worsens (fig. 4.3.b) with additional components. In both cases, a truncation of the classifier is relevant. For the first example using 400 components instead of 210 does not bring any significant improvement of quality, while being more time consuming (fig. 4.3.a). The truncation is even more worthwhile in the second case, where using 400 components instead of 275 will not only require more time, but will also reduce the sensitivity with no increase of specificity (fig. 4.3.b).

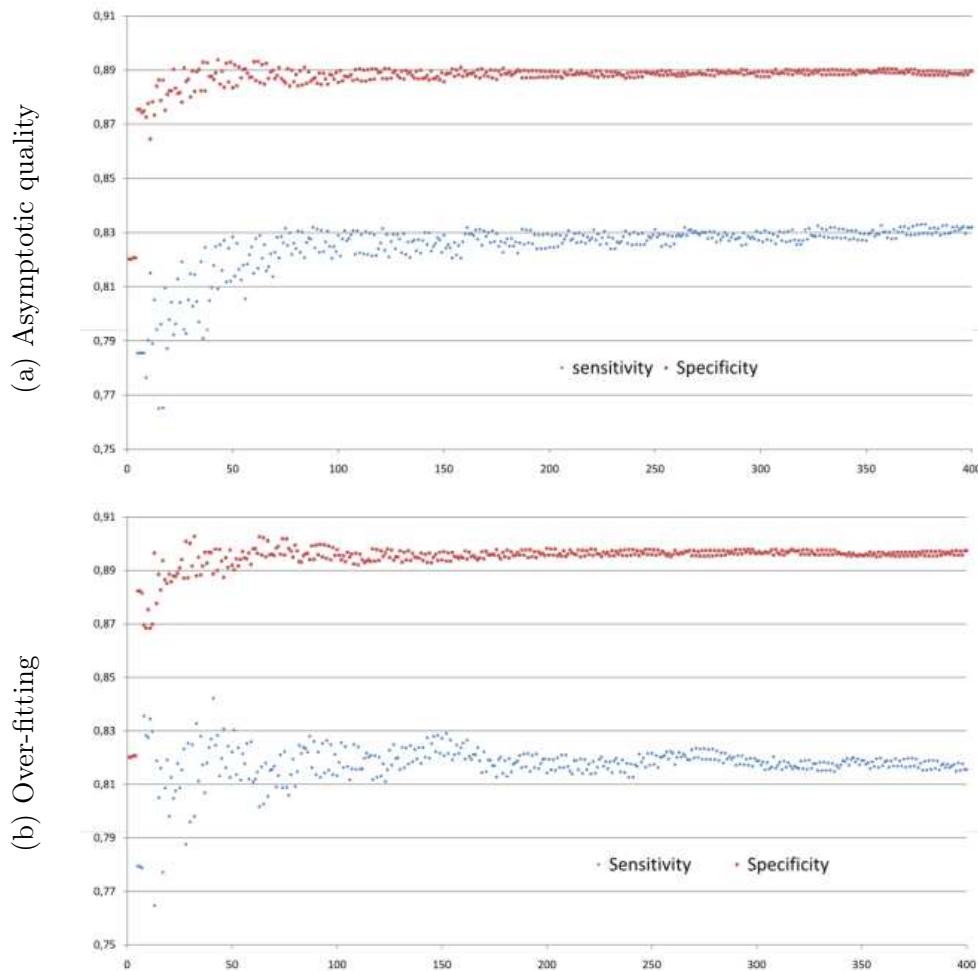


Figure 4.3: Contribution of validation for two classification functions obtained while selecting features for the detection of liver colorectal metastases. Sensitivity and specificity are displayed functions of the length of the classifier.

The length of a classification function is dependent on the maximum number of

rounds inside the learning process that is given as input by the user. Because of the coarse-to-fine approach of the AdaBoost algorithm, components added in the later stages of the learning process bring less gain to the overall quality of classification, while being more error prone. Thus, the validation should be used to determine an optimal classifier length by studying the quality of the classification, functions of the number of components. This optimal length is simply defined as the number of components at the beginning of the asymptotic part of the plot of quality, or as the last number of components before a steady worsening of the classification.

The contributions of this tuning are many. First, it allows making up for some noise inside the training set. Then, it improves the classification speed. Finally, it allows detecting non optimal use of the learning process. When there is some noise in the training set, assumed to be in minority compared to correct samples, these incorrect samples should be treated in the last stages of the learning process (because of the coarse-to-fine approach). Thus, removing the last components of a classification function will lessen the weight of these samples in the global classifier, which may prevent a worsening of the classification due to the last stages of the classification. This truncation allows in fact reducing the problem of overfitting. Then, removing components will decrease the number of calculi during the classification process, which will lessen the computation time. Finally, plots of quality allow detecting cases where the number of rounds for the learning process was not correctly set. If the quality of classification is still improving for the last components of a classification function, chances are that the quality may be further improved by doing more rounds during the learning process. Thus, these plots allow detecting when the user set a too small maximal number of rounds for training.

#### 4.2.4.4 Adequacy to our problem

AdaBoost meets the requirements of this study. Indeed, the technique is fast to apply, which answers our clinical prospects. Moreover, its flexibility should provide a good framework to distinguish tumoral from healthy tissues. Finally, this machine learning technique provides better results than other methods for a same training set.

AdaBoost is by definition fast to apply. In particular, other methods based on kernels (SVM, KPCA, ...) would not apply well to this problem. Such methods rely on matrices with dimensions linked with the size of the training set. Because this training set has to be big to represent the variability of the possible appearances, applying kernel methods would be slow and may even be impossible.

AdaBoost offers better experimental results than other machine learning techniques. Given a same training set where the samples are represented by a same set of 144 features, AdaBoost shows the best classification ability on a new test set (sect. 4.3). A training set of 1014 samples was used to train several classifiers using the machine learning software RapidMiner [rapid i 2008]. The quality of classification was later assessed on a set of 75987 samples. This test validates the choice of AdaBoost to discriminate healthy and tumoral tissues inside liver, be-



cause not only does AdaBoost provide the best accuracy but it does so with a 10% margin (fig. 4.4).

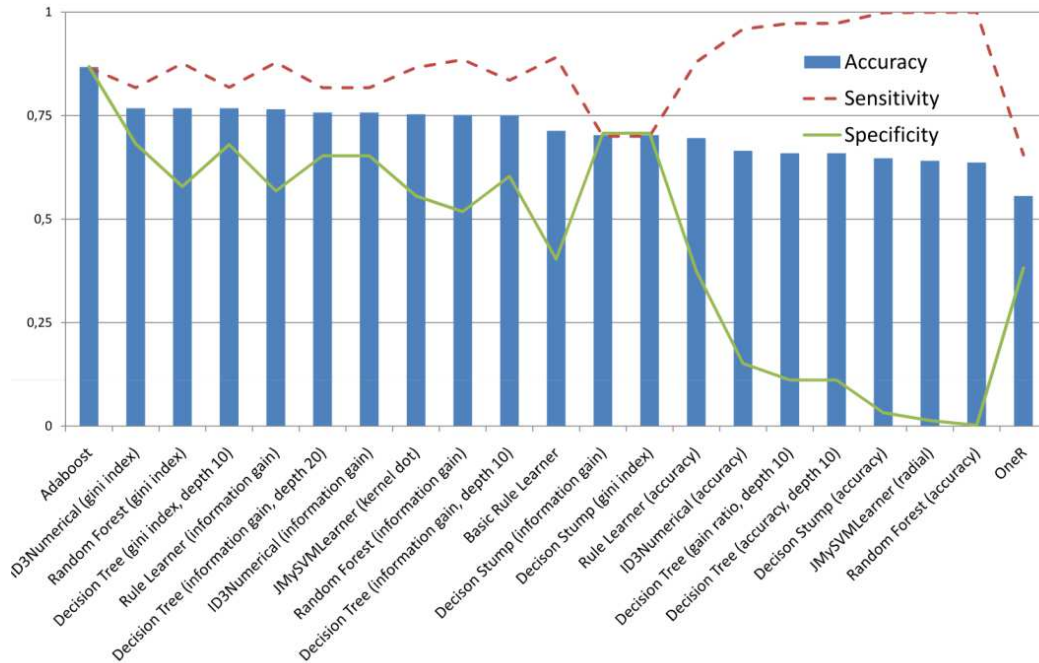


Figure 4.4: Comparison of diverse machine learning techniques for the classification of healthy vs. tumoral tissues. Training was done with a set containing 1014 samples described by 144 features. The learned classifier was then evaluated on a test set of 75987 samples.

### 4.3 Selection of texture features

Machine learning techniques allow distinguishing elements from different classes using features describing these elements. However, the learned discrimination method is dependent upon the features describing each element. Thus, features allowing a good discrimination between the classes should be selected.

First, intensity only cannot discriminate well enough between healthy and tumoral tissues inside the liver (sect. 2.4.2). Thus, not only information at pixel level will be used, but the neighborhood too. Then, raw images may be insufficient or inadequate for direct use, thus images will be treated in order to work on better conditions, either by correction or enhancement of the local features.

### 4.3.1 Texture metrics, quantifying local appearance

#### 4.3.1.1 Texture, spatial arrangement of intensities

Texture is an attribute representing the spatial arrangement of the gray levels of the pixels in a region [Herdilou 1999]

Images are made of pixels, where each pixel gives the intensity for a specific location. An exact location may be described by the intensity of the closest pixel. However, this location may be more precisely described by using more pixels nearby, because spatial information is added to the local intensity. Such a patch is called texture and may be characterized by the existence of repeated patterns, by the spatial distribution in a neighborhood. . . For example two different textures are shown in (fig. 4.5). While the central voxel has the same intensity, the two patches cannot be seen as similar, because the textures are different. While the first texture appears random (fig. 4.5.a), the second one has a cleaner look, with linear patterns (fig. 4.5.b).

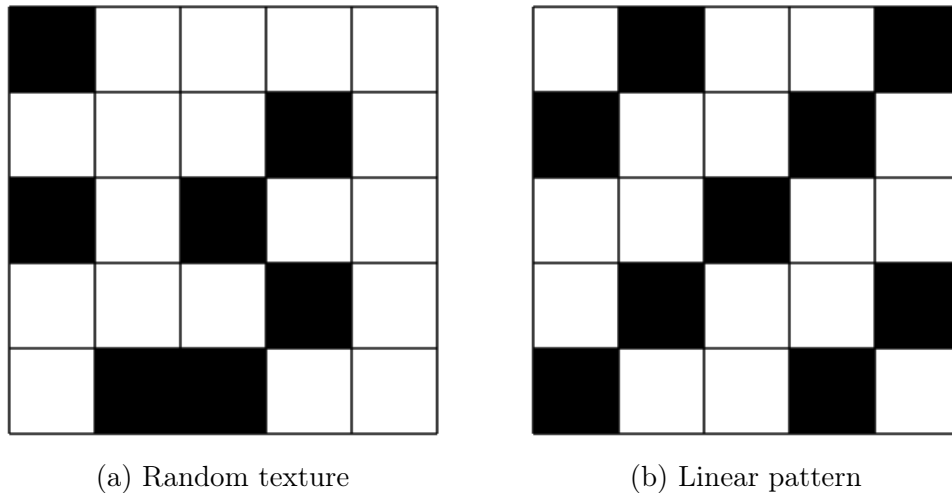


Figure 4.5: Two voxels with the same intensity, but with very different surrounding textures. The first texture (a) seems random, while the second one (b) contains linear patterns.

The human eye is able to distinguish similar textures from different ones. The difficulty lies in quantifying this difference from a mathematical point of view. Several approaches have been proposed to characterize a texture: spectral, structural or statistical analysis. First, structural approaches aim to build a grammar that explains how to build a texture using texture primitives. A new texture is then parsed to check if it matches with the created grammar. Then, spectral analysis deals with the characterization of textures in the frequency domain, using both frequency and orientation in the space. Finally, statistical approaches aim to characterize textures by their statistical properties, namely the distribution of intensities within a texture patch, sometimes combined with the spatial locations of the pixels.

Statistical approaches were retained to characterize the texture. First, structural approaches are not compatible with the retained approach. AdaBoost indeed requires some features to characterize a texture, what structural approaches do not provide. Then, statistical approaches provide metrics that can be used as features for the learning process. Finally, structural analysis might have been used too, but this domain is more relevant for periodic textures and measures are slower to compute than for statistical analysis [Srinivasan 2008, Chen 1998].

Statistical analysis provides a number of metrics that quantify diverse characteristics of the texture: contrast, randomness and linearity... These metrics are characterized by the number of pixels in relation that are used for computation. This number of pixels is named the order of the metric. A first order method will only take intensity in account. A second order method will use relations between pairs of pixels... It should be noted that humans can discriminate between textures with different second order metrics, but are bad with third order [Julesz 1981]. The study will be limited to first and second order metrics, because metrics with higher orders do not apply to this study. Higher order metrics indeed aim to distinguish textures with fibers and the liver does not contain fibers.

The texture features that were retained will now be presented, while representing texture patches by  $n \times n$  square matrices  $I$  of gray levels  $I_{i,j}$ . The size of the textures  $n$  will always be odd, in order to be able to center textures on one pixel.

#### 4.3.1.2 Quantifying intensity distribution

First order features rely on the distribution of intensities within each texture patch to quantify some characteristics of the texture. These metrics mostly characterize the histogram inside the texture patch. Given a texture patch described as a square matrix  $I$  with odd size  $n$ , the retained metrics are defined as follows.

Mean : Define the position of the histogram on the gray scale.

$$MEAN = \frac{1}{n^2} \sum_{(i,j) \in [1:n]^2} I_{i,j} \quad (4.6)$$

Variance : Measure the repartition of gray levels around the mean value.

$$VAR = \frac{1}{n^2} \sum_{(i,j) \in [1:n]^2} (I_{i,j} - MEAN)^2 \quad (4.7)$$

Skewness : Quantify the difference compared to a symmetric distribution (deviation to lower or higher intensities).

$$SKEW = \frac{1}{n^2} \sum_{(i,j) \in [1:n]^2} (I_{i,j} - MEAN)^3 \quad (4.8)$$

Kurtosis : Characterize the histogram peak (the smaller, the flatter).

$$KURT = \frac{1}{n^2} \sum_{(i,j) \in [1:n]^2} (I_{i,j} - MEAN)^4 \quad (4.9)$$

Signal-to-Noise Ratio : Evaluate the corruption of one signal by the background noise.

$$SNR = \frac{MEAN}{\sqrt{VAR}} \quad (4.10)$$

#### 4.3.1.3 Co-occurrence matrix, tracking pixel pairs

Co-occurrence matrices are matrices that keep track of the distribution of pixel pairs inside a texture patch. These matrices are useful for the computation of second order metrics, which are by definition computed using pixel pairs. Given a texture patch described as a square matrix  $I$  with odd size  $n$ , a co-occurrence matrix  $P_{\Delta_x, \Delta_y}$  defines the probabilities of pixel pairs separated by an offset  $(\Delta_x, \Delta_y)$ . This matrix has a size given by the range of intensity within the texture patch.

$$P_{\Delta_x, \Delta_y}(i, j) = \sum_{k=1}^n \sum_{l=1}^n \delta_{I(k,l), i} \delta_{I(k+\Delta_x, l+\Delta_y), j} \quad (4.11)$$

where  $\delta$  is Kronecker's delta  $\delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$

However, this definition has a main drawback; it is not invariant by rotation. Thus co-occurrence matrices are more often defined in a polar basis. These matrices  $P(d, \theta)$  are keeping track of the probabilities of pixel pairs for a direction  $\theta$  and a distance  $d$ . An example is given for the co-occurrence matrix  $P(1, 0)$  (fig. 4.6). Given a texture patch (fig. 4.6.a), a co-occurrence matrix is iteratively constructed by counting the number of occurrences for each pair of pixel intensities (fig. 4.6.b). The co-occurrence matrix is finally defined by normalizing the matrix that counts the occurrences of each pixel pair (fig. 4.6.c).

Only four directions  $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$  are considered. First, only directions as multiple of  $45^\circ$  are retained, because these directions allow direct use of intensities, without any interpolation. Then, the order of pixels inside a pair is not taken in account for the construction of co-occurrence matrices. Thus co-occurrence matrices are symmetric, and are invariant for textures transformed by rotation of  $180^\circ$ .

The dimension of a co-occurrence matrix is defined by the intensity range within a patch, which is a problem because of the large range of intensity inside medical images. Thus a mapping of intensities to an admitted number of gray levels is introduced. For large intensity ranges the co-occurrence matrices will be mostly empty. Moreover, a change of a few Hounsfield units is not truly relevant for the analysis because such variations are non-significant compared to the image noise. Consequently, a mapping of the intensities onto a smaller range of intensities is

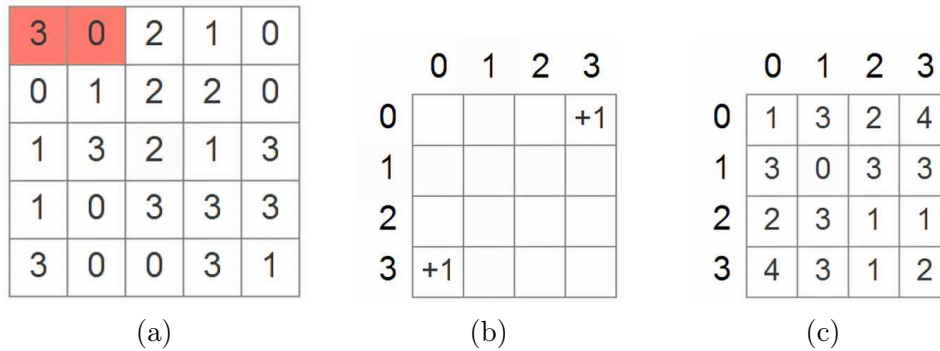


Figure 4.6: Creation of the co-occurrence matrix  $P(1,0)$ . The occurrences of each pair of pixels are counted (c) for a texture patch (a). The count is done iteratively by considering each pair of pixels at distance 1 and for angle 0. A pair of pixels (shown in light red) is considered and used to update the co-occurrence matrix, while ignoring the order of the intensities (b).

introduced. The target range of intensities is set to a number of admitted gray levels  $m$ . This mapping offers some advantages, both in term of computation and information gain. Indeed, using smaller matrices will simplify and speed up the computations, while decreasing the number of possible intensity values allow keeping only the relevant variations of intensity.

#### 4.3.1.4 Haralick's, quantifying pixel relations

Haralick's texture descriptors were retained as second order texture descriptors. These Haralick's descriptors are metrics upon pairs of pixels that were introduced by Haralick in 1973 and have been widely used since [Haralick 1973]. In particular Pham showed that these texture descriptors were informative for the liver case [Pham 2007].

Haralick introduced 14 descriptors that are all computed using any single co-occurrence matrix. As previously written, many co-occurrence matrices may be defined for a same texture patch, by choosing diverse distances  $d$  and angles  $\theta$ . Consequently, for a same texture patch, the Haralick's descriptors provide a high number of features to quantify the texture characteristics. For this work, only 9 descriptors were retained. Each descriptor is defined for a co-occurrence matrix  $P(d, \theta)$  of size  $m$  that will be noted as  $P$  to simplify the notations.

Entropy : Measure the randomness of the distribution of pixel pairs

$$-\sum_{i=1}^m \sum_{j=1}^m P_{i,j} \log P_{i,j} \quad (4.12)$$

Energy : Measure the homogeneity of the texture (the smoother the texture, the higher the homogeneity)

$$\sum_{i=1}^m \sum_{j=1}^m P_{i,j}^2 \quad (4.13)$$

Contrast : Measure the local contrast

$$\sum_{i=1}^m \sum_{j=1}^m (i-j)^2 P_{i,j} \quad (4.14)$$

Sum Average : Measure the average of pixel pairs

$$\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m i P_{i,j} + j P_{i,j} \quad (4.15)$$

Variance : Measure the variation of the gray level distribution

$$\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (i - \mu_r)^2 P_{i,j} + (j - \mu_c)^2 P_{i,j} \quad (4.16)$$

Correlation : Measure the linearity of the image (high when there are many linear structures)

$$\sum_{i=1}^m \sum_{j=1}^m \frac{(i - \mu_r)(j - \mu_c) P_{i,j}}{\sqrt{\sigma_r^2 \sigma_c^2}} \quad (4.17)$$

Maximum Probability : Give the probability of the most common pixel pair in texture

$$\max P_{i,j} \quad (4.18)$$

Inverse Difference Moment : Measure the smoothness of a texture

$$\sum_{i=1}^m \sum_{j=1}^m \frac{P_{i,j}}{1 + (i-j)^2} \quad (4.19)$$

Cluster Tendency : Measure the grouping of pixels with close intensities (granularity)

$$\sum_{i=1}^m \sum_{j=1}^m (i - \mu_r + j - \mu_c) P_{i,j} \quad (4.20)$$

where  $\mu_r$ ,  $\mu_c$ ,  $\sigma_r^2$  and  $\sigma_c^2$  are means and variance for rows and columns.

$$\mu_r = \sum_{i=1}^m \sum_{j=1}^m i P_{i,j} \quad \sigma_r^2 = \sum_{i=1}^m \sum_{j=1}^m (i - \mu_r)^2 P_{i,j} \quad (4.21)$$

$$\mu_c = \sum_{i=1}^m \sum_{j=1}^m j P_{i,j} \quad \sigma_c^2 = \sum_{i=1}^m \sum_{j=1}^m (j - \mu_c)^2 P_{i,j} \quad (4.22)$$

### 4.3.2 Filtering, preparing images before treatment

A filter is a process that treats an image to remove or enhance some features. Seeing an image as a signal, filtering is the processing of this signal. However, contrary to usual signal processing, image processing offers more ways to treat the signal by taking advantage of the local information (2D or 3D signal, instead of 1D signal for electricity). Filtering allows removing some unwanted components of the image like the noise. The process also provides ways to enhance some features in an image, either by flattening similar regions, thus making the difference between regions more visible, or by enhancing some structures or patterns.

Many filters have been developed, some generics, other specialized for specific tasks or domains. In this work only deterministic filters are retained because filters with varying behaviors or that have to be applied an unknown number of times do not match the clinical prospect of this study. Filters retained for this study include some classical filters for smoothing, filters with proven relevance in the liver case, and filters related to human vision (for the texture appearance).

#### 4.3.2.1 Convolution to filter an image

The convolution is a linear operator that creates a new image by linearly combining pixel intensities using a kernel to define the weight of each pixel. For each pixel of the image, convolution defines a linear operation that calculates a new intensity for this pixel. This new intensity is computed as a weighted sum of the neighbor pixels, with weights given by a kernel that may be seen as another smaller image. This kernel may define any kind of combination of pixels inside each local patch. Thus, the convolution is a powerful tool of image processing that is often used for filtering. Because of its flexibility, the convolution is relevant for many tasks such as edge detection, smoothing, denoising, and enhancement...

From a mathematical point of view, the convolution is a linear operator that combines two functions to create a third one. Convolution ( $f * g$ ) of two functions  $f$  and  $g$  is the integral of the product of the functions  $f$  and  $g$  with a shift.

$$(f * g)(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u, v)g(x - u, y - v)dudv \quad (4.23)$$

Convolution applies to image processing by using a discrete version of the convolution, where both image and kernels are seen as functions. Given an image  $J$  and a kernel  $K$  of size  $2m + 1 \times 2n + 1$ , where  $K$  is smaller than the image, the discrete convolution is defined as a sum between the products of both images with an offset. Intuitively the convolution is composed of 3 steps. First, a patch is selected inside an image with same dimensions than the chosen kernel (fig. 4.7.a). Then, the intensity after convolution is computed as the sum of intensities inside the patch multiplied by the weights defined inside the kernel (fig. 4.7.b). Finally, the new intensity is added into the filtered image and a new pixel can be treated.

$$(K * I)(i, j) = \sum_{k=-m}^m \sum_{l=-n}^n K(k, l) I(i - k, j - l) \quad (4.24)$$

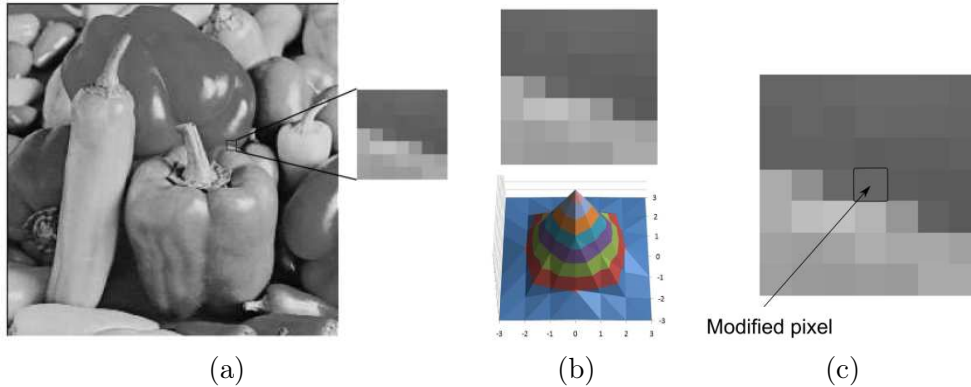


Figure 4.7: Convolution of an image, practical use. Given an image (a) convolution of the entire image is done pixel by pixel. Treatment of each pixel is done by taking a patch around this pixel first, whose size is that of the kernel, and then by computing the new intensity of the pixel with a weighted sum between patch intensities and weights inside the kernel (b, c).

The definition of the discrete convolution highlights a possible difficulty for the boundaries of the image  $I$ . The filtered value  $I(0, 0)$  indeed requires the value of intensities outside the image. Several approaches have been proposed to solve this problem. First, one may consider that intensity outside the image is null, but this choice induces artifacts on the boundaries of the image (intensities will tend to be lower). Then, a mirror approach may be used, by mirroring the boundaries of the image on the outside, such that  $I(-1, 0) = I(1, 0)$ . Skipping the boundaries of the image is nevertheless the easiest way. Avoiding treating the rim of the image indeed makes the introduction of special cases on boundaries non relevant. Moreover, relevant parts of the image are rarely on boundaries and apart from large kernels this does not leave numerous untreated pixels.

The convolution applies to functions or images of any dimension. This study is based on 3D images, thus the convolution will also be done in 3D. Doing otherwise would mean losing the information provided by the third dimension. Then, the filters will be applied while taking in account both the voxel anisotropy and the change of voxel sizes between images. Image voxels are indeed rarely be cubic. Moreover, the size of the voxels varies between images of two patients. In particular the voxel depth often varies between 0.8 and 5mm. Thus, most filters should be applied while taking in account the size of voxel in order to keep a similar response for any voxel size. Otherwise a set number of pixels would lead to the use of very different anatomical structures for a same location. In the given example, filtering the central lesion would depend on different anatomical structures when using a set radius of one pixel (fig. 4.8). While pixels used for filtering are mostly inside



the lesion for slices of 1mm (fig. 4.8.a), the upper vessel is also included for 5mm slices (fig. 4.8.b).

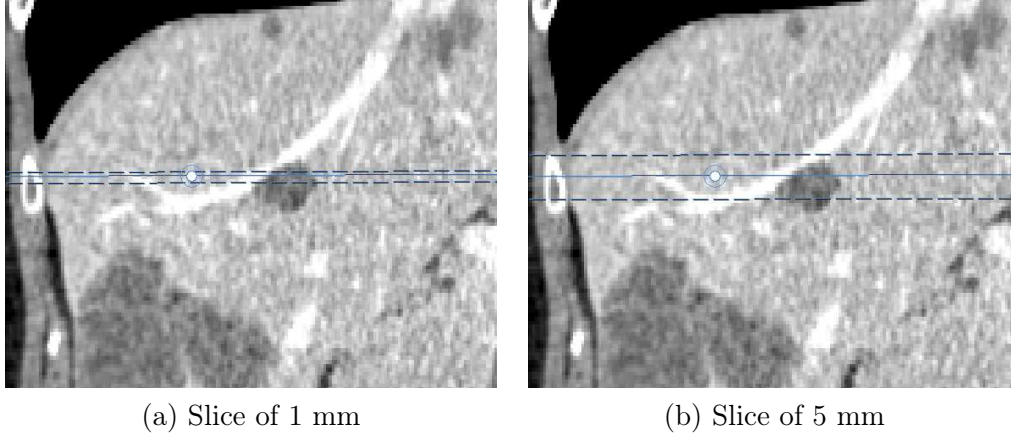


Figure 4.8: Influence of voxel sizes for filtering, sagittal view. The volume used by taking three slices is shown between two vertical lines. The anatomical structures contained inside the 3 slices are very different between images with a slice thickness of 1mm (a) and 5mm (b).

#### 4.3.2.2 Smoothing the image

Smoothing relates to the decrease of intensity variability in the image, by decreasing the difference in intensity between neighbor pixels. This task is related to denoising because the intensity of pixels will be more modified when pixels are more different from the neighborhood. Thus, when regions are assumed to be homogeneous, pixels with higher probability of being noisy will be more modified. However, this comes at the price of a blurring effect that makes edges less visible. Two linear filters based on convolution were considered, the mean and the Gaussian filter, with diverse sizes of kernel.

##### Mean Filter

The mean filter is the simplest way to smooth an image. Mean filtering averages the intensities using a local patch, whose dimensions are given by the dimensions of the kernel. Mean filtering relies on a kernel  $K_{mean}$  of size  $X \times Y \times Z$  with constant weights that are normalized in order to prevent a change of magnitude for the intensity (eq. 4.25). The effect of mean filtering is shown on an image with synthetic noise (fig. 4.9.c-e). The smoothing effect may be seen in any figure. Indeed, the noise tends to become less visible when the size of the kernel grows. However, this denoising effect comes with a blurring effect on the boundaries of the objects that compose the image.

$$(K_{mean})(i, j, k) = \frac{1}{XYZ} \quad (4.25)$$

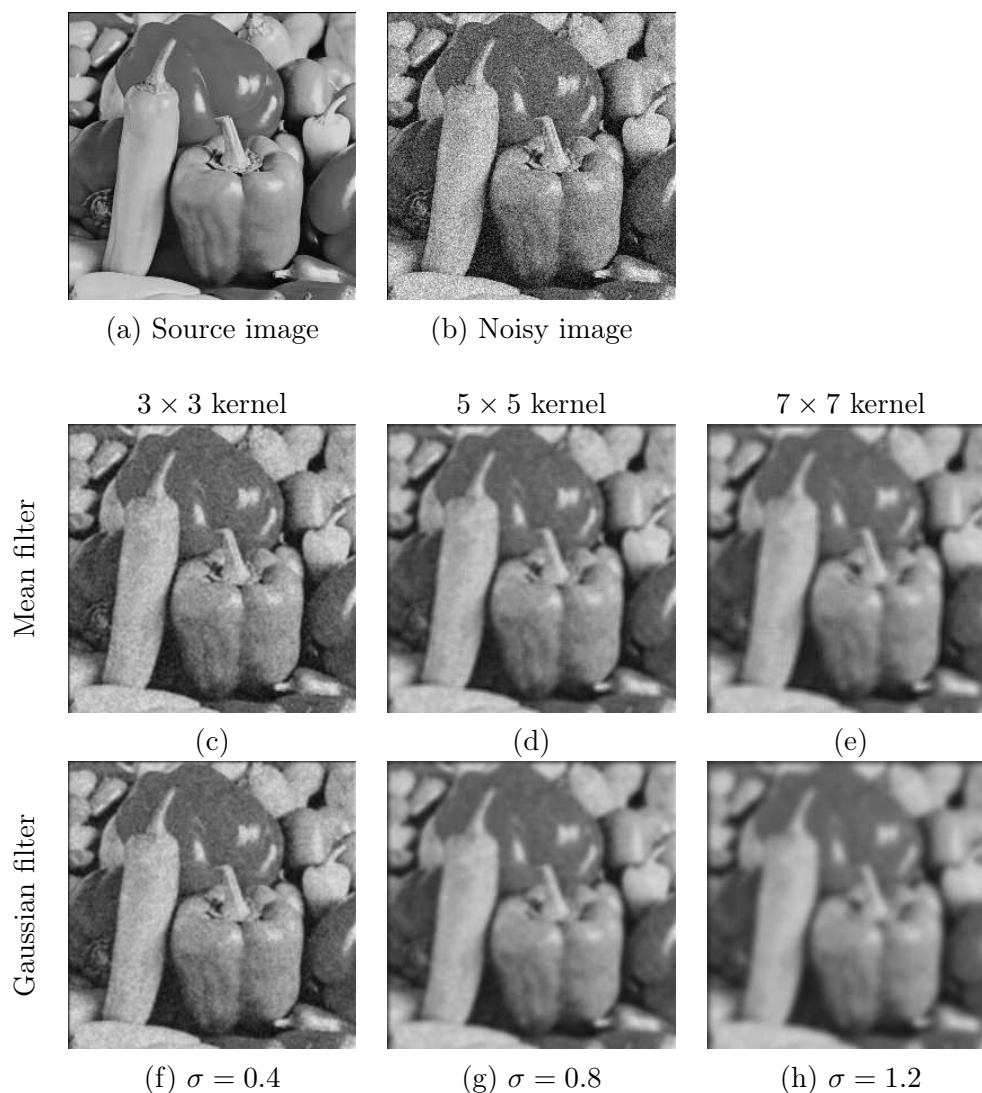


Figure 4.9: Impact of mean and Gaussian filters on a noisy image. The pepper image (a) was modified by adding a Gaussian noise (0 mean and deviation 25) (b). The noisy image is then smoothed for diverse sizes of kernel using a mean filter (c,d,e) and a Gaussian filter with diverse standard deviations (f,g,h).

### Gaussian Filter

The Gaussian filter is a smoothing filter that gives more weight to the spatial location of the pixels. While the mean filter gives the same weight to all pixels inside the kernel mask, the Gaussian filter aims to give more weight to pixels in the center of the patch. This change is driven by the idea that the farther from center is a pixel, the less relevant it is to compute the new intensity. The difficulty comes from the choice of good weights. A Gaussian distribution is a good compromise for the convolution kernel, when distance from kernel center is computed using the size

of the voxels (eq. 4.26). Such an approach should still give a smoothed image, but should also remain closer to the initial image than the mean filter. In particular, the edges should remain clearer than for the mean filter. Gaussian filtering is applied to the same image than mean filters (fig. 4.9). The increase of the blurring effect along with the size of the kernel is similar to that of the mean filter. However, the edges tend to remain sharper (fig. 4.9.f, g, h).

$$(K_\sigma)(i, j, k) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{-\frac{x'^2+y'^2+z'^2}{2\sigma^2}}$$

where  $x' = (i - i_c)v_x$   
 $y' = (j - j_c)v_y$   
 $z' = (k - k_c)v_z$   
 $(i_c, j_c, k_c)$  , center of the kernel  
 $(v_x, v_y, v_z)$  , size of voxel

(4.26)

#### 4.3.2.3 Gabor's, enhancing texture characteristics

Gabor introduced Gabor's filters in 1946 to analyze signals using both time and frequency [Gabor 1946]. Gabor's filters since showed a good adequacy for the extraction of texture features and for the segmentation of objects that differ by their textures [Clausi 2000, Bovik 1990]. These filters are thus relevant to distinguish healthy from non-healthy hepatic tissues, for which texture plays an important role. Jones and Palmer showed that Gabor's filters are *strongly related with the human visual system*, because they are *narrow band filters tuned to different frequencies* [Glatard 2004, Jones 1987]. Because these filters may be set up for many different textures, they are often used inside banks of filters, where Gabor's filters with various parameters are considered.

On spatial domain 2D Gabor's filters are obtained by convolution, with kernels defined as Gaussians modulated by a sinusoid. These kernels are defined for a Cartesian basis, whose origin is the center of the kernel, and abscissa obtained by following an orientation  $\theta$ . The kernel for spatial Gabor's filter  $K_{\theta,\sigma,\gamma,\lambda,\varphi}$  is composed of two terms, a Gaussian and a sinusoid that are customized independently (eq. 4.27). First, the Gaussian component is parameterized by its deviation  $\sigma$  and a spatial aspect ratio  $\gamma$  that defines the ellipticity of the circular Gaussian. Then, the sinusoid is customized with a spatial wavelength  $\lambda$  and a phase offset  $\varphi$ .

$$K_{\theta,\sigma,\gamma,\lambda,\varphi}(x, y) = \exp\left(-\frac{x^2 + \gamma^2 y^2}{2\sigma^2}\right) \cos\left(2\pi\frac{x}{\lambda} + \varphi\right) \quad (4.27)$$

Hence for a global basis, the Gabor's kernel is expressed using the orientation  $\theta$ , and with a change of both scale defined by the size of pixel  $(v_x, v_y)$  and basis by a translation of the center of the kernel  $(i_c, j_c)$ .

$$\begin{aligned}
K_{\theta,\sigma,\gamma,\lambda,\varphi}(i,j) &= \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi\frac{x'}{\lambda} + \varphi\right) \\
x' &= (i - i_c)v_x \cos\theta + (j - j_c)v_y \sin\theta \\
y' &= -(i - i_c)v_x \sin\theta + (j - j_c)v_y \cos\theta
\end{aligned} \tag{4.28}$$

The choice of the parameters for Gabor's filter is a critical step, because it is a determining factor for the adequacy between one filter and specific texture appearances. A Gabor kernel is composed of parallel stripes with different weights, inside an ellipsoidal envelope (fig. 4.10). The parameters of the kernel control the size, the orientation, and the position of these stripes.

The wavelength  $\lambda$  is the preferred wavelength of the filter, specified in pixels. This value is a scale factor for the stripes. By modifying the wavelength, the stripes keep same orientation and relative dimensions but the global sizes of the stripes are modified (fig. 4.10.a-c). This wavelength should be superior to 2 and is often inferior to the fifth of the image size to avoid problems on borders.

The orientation  $\theta$  specifies the angle of the parallel stripes of the Gabor's function (fig. 4.10.d-f). Modifying the orientation equals to a rotation of the kernel. The kernel for an orientation  $\theta_2$  is indeed obtained by applying a rotation with an angle  $(\theta_2 - \theta_1)$  from the kernel with the orientation  $\theta_1$ .

The phase offset  $\varphi$  represents the shift of the cosine factor. This shift determines the symmetry of the kernel by modifying the positions of the inhibitory and excitatory stripes. For a phase shift of  $\varphi = 0$  the kernel is symmetric, for a shift of  $\varphi = \frac{\pi}{2}$  it becomes antisymmetric (fig. 4.10.g-i).

The aspect ratio  $\gamma$  represents the ellipticity of the support. For a ratio of 1, the support is circular. A larger ratio then tends to give more but smaller stripes, whereas a smaller ratio defines longer stripes (fig. 4.10.j-l).

The bandwidth  $b$  is a parameter often introduced to replace the Gaussian deviation  $\sigma$  by analogy to the animal visual system [Kruizinga 1999, Glatard 2004, Clausi 2000]. This bandwidth is specified in octaves and gives the half response spatial frequency, and allows computing  $\sigma$ , which cannot be chosen directly.

$$\frac{\sigma}{\lambda} = \frac{1}{\pi} \sqrt{\frac{\ln 2 \cdot 2^b + 1}{2 \cdot 2^b - 1}} \tag{4.29}$$

#### 4.3.2.4 Relevant filters in the liver case

Only linear filters were presented in previous sections. Two non-linear filters with proven relevance in the liver case will now be presented, the median and Nagao's filter. First, the median filter aims to replace the intensity of each pixel by the median value inside a neighborhood. This approach better reduces the noise than linear methods for some types of noise, in particular salt and pepper or speckle noises. Moreover, this filter is better at preserving the edges than the previous smoothing filters. Then, Nagao's filters improve median filtering with the introduction of multiple masks in order to keep sharper edges.

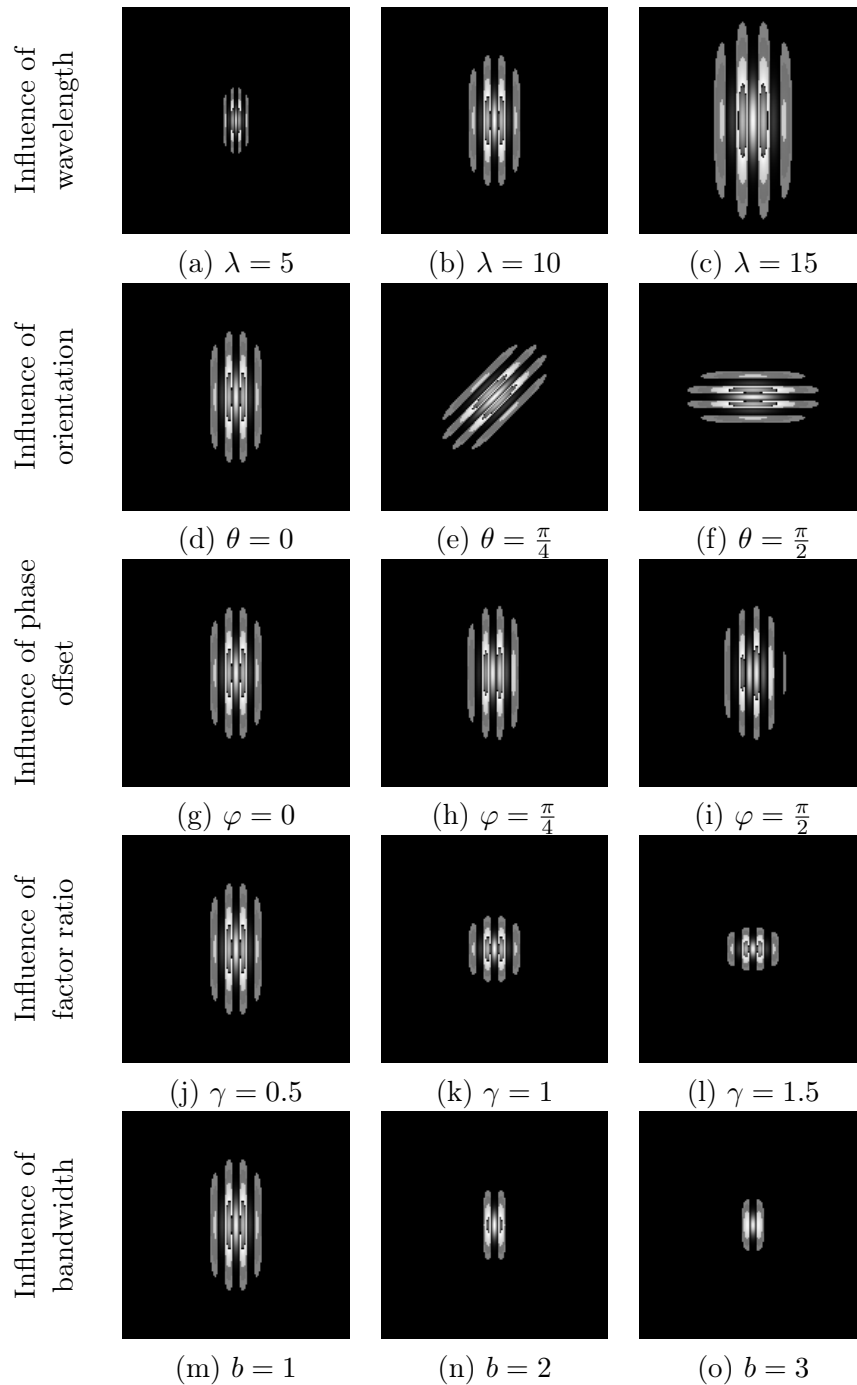


Figure 4.10: Impact of the parameters on a Gabor's kernel. The impact of the parameters on a Gabor's kernel is shown by modifying one parameter at a time from an initial set of parameters. For this initial setting, the orientation is  $\theta = 0$ , the wavelength  $\lambda = 10$ , the spatial aspect ratio  $\gamma = 0.5$ , the bandwidth  $b = 1$ , and the phase offset  $\varphi = 0$ . The intensities displayed in the images were modified to improve the visualization.

### Median filter

The median filter is a non linear filter with good ability to remove salt and pepper noise [Tukey 1977, Perreault 2007]. This image processing tool was shown to be relevant in the liver case due to the random noise induced by the contrast product inside the parenchyma [Sørli 2005, Wong 2008, Fernández-de Manuel 2009]. Given a neighborhood patch  $\mathcal{P}_n$ , median filtering replaces the intensity of each pixel by the median value inside a patch centered on this pixel.

The median approach allows ignoring the extreme percentiles of intensities inside the patch  $\mathcal{P}_n$ . Because only the median value is relevant, extreme values (high or low) may be incorrect without any influence on the final result, whereas they would have a big impact with linear filters. Thus, only the most significant pixels matter, where the significant pixels are the one inside the biggest cluster of intensities. As a side effect, median filtering tends to keep more visible edges. Instead of smoothing the edges by using pixels with large differences in intensity, median filtering will set pixels to one side of the edge only. However, this may be the wrong one.

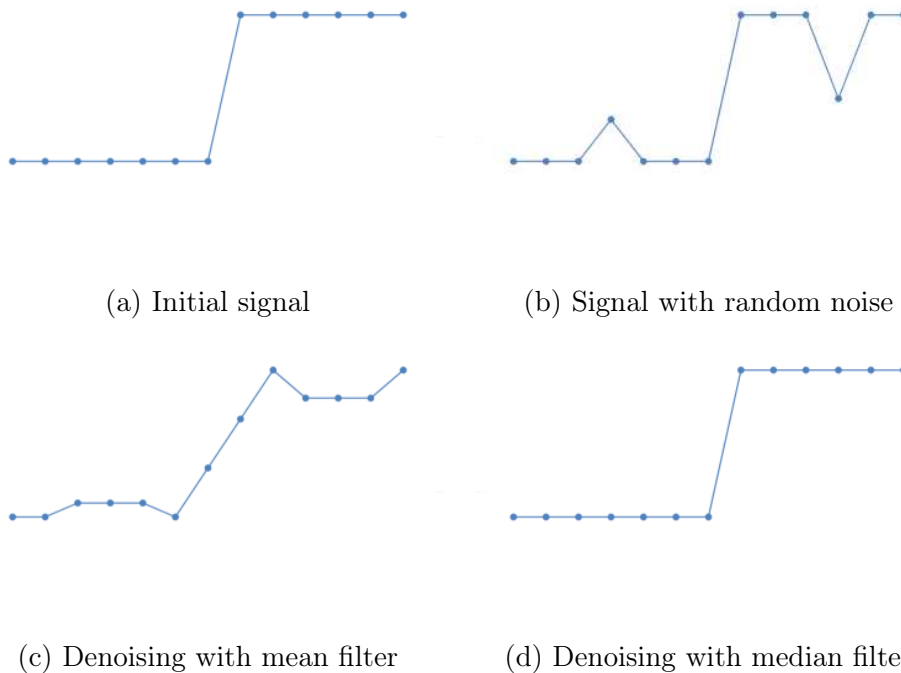


Figure 4.11: Comparison of median and mean filters on a 1D signal. A random noise is added to a Heaviside step function (a, b). Noise reduction is then compared for mean (c) and median (d) filters using a neighborhood of 3 points.

The difference between a linear and a median filter is shown for a simple example on a 1D signal. Some salt and pepper noise is added to the Heaviside step function (fig. 4.11.a, b). Denoising with median and mean filters gives very different results. While the mean filter transforms the signal around the noisy points

and smooths the boundary between the two segments of the signal, median filtering removes the artificial noise with no side effect (fig. 4.11.c, d).

In this study, the neighborhood patch  $\mathcal{P}_n$  is defined as a cube on a millimetric basis. Given a radius  $r$  expressed in millimeters the neighborhood patch is defined by computing the number of pixels on each axis for the current size of voxel  $(v_x, v_y, v_z)$ .

$$\mathcal{P}_n = \left(2 \left\lfloor \frac{r}{v_x} \right\rfloor + 1\right) \times \left(2 \left\lfloor \frac{r}{v_y} \right\rfloor + 1\right) \times \left(2 \left\lfloor \frac{r}{v_z} \right\rfloor + 1\right) \quad (4.30)$$

As mentioned before, median filters keep clear edges, but may modify the location of the edges. Moreover, this displacement increases with the size of the patch of the filter, which is a problem. The effects of median filtering are shown for an artificial binary image (fig. 4.12). First, boundaries of thin elements are modified (fig. 4.12.b). Then, corners are trimmed (fig. 4.12.c,d). In fact, when two regions are in contact the biggest one tends to erode the smaller one. To prevent this behavior and keep more stable edges Nagao's filters were introduced.



(a) Initial image    (b)  $5 \times 5$  patch    (c)  $11 \times 11$  patch    (d)  $21 \times 21$  patch

Figure 4.12: Influence of median filtering on a binary image ( $400 \times 400$ ) for diverse sizes of patch.

### Nagao's filter

Nagao introduced a non-linear smoothing filter that preserves the edges, called Nagao's or Kuwahara-Nagao's filter, which has been shown to be relevant in the liver case [Nagao 1979, Kuwahara 1976, Chemouny 1999]. Nagao's filter is founded on the assumption that intensities remain more homogeneous inside a single region than between different objects. A set of mask  $\mathcal{B}_{Nagao}$  is thus introduced to simulate possible shapes of objects and both median value and variance are computed within every mask. Based on the assumption of homogeneity inside a region, the mask with the smallest variance should be the one with the fewest regions, thus the median value for this mask should smooth the image without any impact on the edges. Moreover, this approach has low impact in centers of regions. Thanks to the homogeneity assumption, median intensity should indeed keep a similar value when using the entire patch or only some pixels inside the same patch.

A simple synthetic example shows the benefits of masks to preserve the edges while filtering (fig. 4.13). A pixel inside the gray region is filtered using neighbors in a local square patch (fig. 4.13.a). A median filter would set this central pixel to the

intensity of the white region, because there are more pixels in this white region than in the gray one. However, Nagao's filter allows keeping the pixel inside the gray region. The variance inside the three masks (fig. 4.13.b,c,d) is indeed minimal for the mask that is entirely inside the gray region (fig. 4.13.d). The first mask indeed contains multiple pixels from both regions (fig. 4.13.b) and the variance inside the last mask (fig. 4.13.d) will be smaller than for the second one (fig. 4.13.c) because of the central black pixel. Thus, the variance will be minimal for the last case, and the pixel will remain inside the gray region, which prevents the blurring of edges.

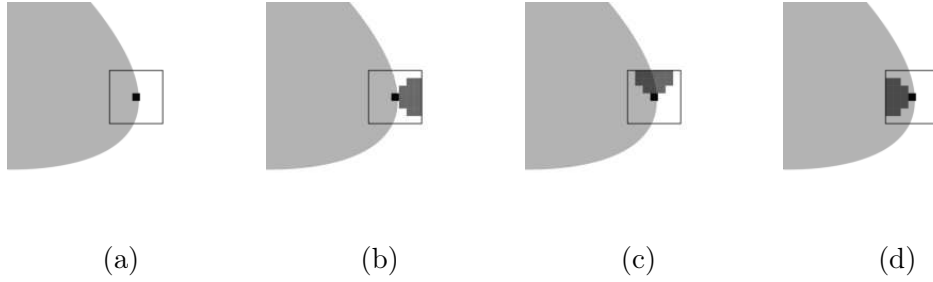


Figure 4.13: Contribution of Nagao's filter on a simple example. A pixel (black square), located on the edge between two regions is filtered using a local neighborhood (black square contours). Different masks that may be considered for Nagao's are shown in medium gray (b,c,d).

Given a set of masks  $\mathcal{B}_{Nagao}$ , called Nagao's masks, where each mask  $b(i, j, k)$  is defined relatively to any pixel  $\mathcal{V}(i, j, k)$ , an image  $\mathcal{V}$  is filtered by Nagao's such that each filtered pixel  $(f(\mathcal{V}))(i, j, k)$  becomes the median value inside the mask with smaller variance (compared to the median value) (eq. 4.31).

$$(f(\mathcal{V}))(i, j, k) = \text{median}(b^*(i, j, k))$$

$$\text{where } b^*(i, j, k) = \underset{b \in \mathcal{B}_{Nagao}}{\operatorname{argmin}} \sum_{\mathbf{x} \in b(i, j, k)} (\mathbf{x} - \text{median}(b))^2 \quad (4.31)$$

The choice of the set of mask for Nagao's filter is truly important, because the masks should be designed to remain inside a single region in order to preserve the edges. For example, masks chosen as lines with diverse orientations would not bring any improvement to the previous example (fig. 4.13). Thus Nagao's masks should be chosen to reflect the possible interfaces between the regions. These masks should also not be too many; otherwise filtering would become too slow.

For this study 27 dynamically computed Nagao's masks were chosen to reflect the possible interfaces between the lesions, and handle regions one above the other. The Nagao's filter is defined in a millimetric basis to better take in account the anisotropy of voxels and the variations between the images. Thus Nagao's masks are computed dynamically by scaling shapes, functions of the radius of the Nagao's filter and the dimensions of voxels. First, 3 masks centered on the patches are selected in order to treat pixels in center of a region, with a z shift to take relative



positions in account (fig. 4.14.a-c). Then, 12 masks are added to manage regions in contact by the diagonals of patches (angles of  $\{-\frac{3\pi}{4}, -\frac{\pi}{4}, \frac{\pi}{4}, \frac{3\pi}{4}\}$ ). These masks are created by rotation of an initial mask that exists with 3 different positions in the z axis (fig. 4.14.d-f). The final 12 masks are chosen to handle regions in contact by the side of patches (angles of  $\{-\frac{\pi}{2}, 0, \frac{\pi}{2}, \pi\}$ ), and are also created by rotation of an initial mask (fig. 4.14.g-i).

The comparison of these two non-linear filters with a linear one is shown for denoising with different sizes of patches (fig. 4.15). An artificial noise was added to the pepper image, as salt and pepper noise (fig. 4.15.b). While denoising is incorrect with the mean filter, results of filtering are better with the non-linear ones. Indeed, with the mean filter the image remains very different from the source image whatever the size of kernel. For smaller kernels the image remains very noisy (fig. 4.15.c) and with larger kernels the boundaries of the objects become blurry without even giving homogeneous regions (fig. 4.15.e). Median and Nagao's filter exhibit better denoising abilities. For smaller kernels the results are very close to the initial image (fig. 4.15.f,h). Then these results do not impair for larger patches. A difference is nevertheless visible between the two filters. The median filter gives smoother boundaries (fig. 4.15.h) than the Nagao's one (fig. 4.15.j), for which the boundaries are quite fluffy. However, the boundaries remain more marked with Nagao, for example on the left of the bottom left red pepper or on the top left of the long yellow pepper.

### 4.3.3 Defining texture features

#### 4.3.3.1 Defining the features

Features are values that characterize a sample and are defined for this study as a cross-product between a set of filters and a set of texture descriptors. This definition means that each feature describing a texture patch is obtained by first filtering the image and then computing a texture descriptor on this patch.

Let us introduce a bank of filters  $\mathcal{F} = \{f_{\Theta}\}$  and a bank of descriptors  $\mathcal{D} = \{d_{\Theta'}\}$ , where  $\Theta$  (resp.  $\Theta'$ ) defines the type of the filter (resp. a texture descriptor) and its possible parameters. The feature  $\phi_{\Theta, \Theta'}(\mathbf{x})$  is defined for any voxel  $\mathbf{x} \in \mathcal{V}$  inside the image  $\mathcal{V}$  by computing the descriptor  $d_{\Theta'}$  on a texture  $T_{\mathbf{x}}$  centered on the voxel  $\mathbf{x}$  for the filtered image  $f_{\Theta}(\mathcal{V})$ .

$$\phi_{\Theta, \Theta'}(\mathbf{x}) = d_{\Theta'} \circ f_{\Theta}(T_{\mathbf{x}}) \quad (4.32)$$

The use of filters is consistent with the framework. First, work is done on noisy images, thus filtering is recommended. Then, 3D filters may have a normalization effect between the various sizes of slice. Finally, filters may be chosen to enhance some features. The retained filters are detailed in previous sections but will be briefly reviewed (sect. 4.3.2). The usual filters (Gaussian, Mean, etc) are used, with diverse radius. In addition, 3D Gabor's filters were chosen to consider the texture information. Gabor's filters are strongly related with the human visual system, and

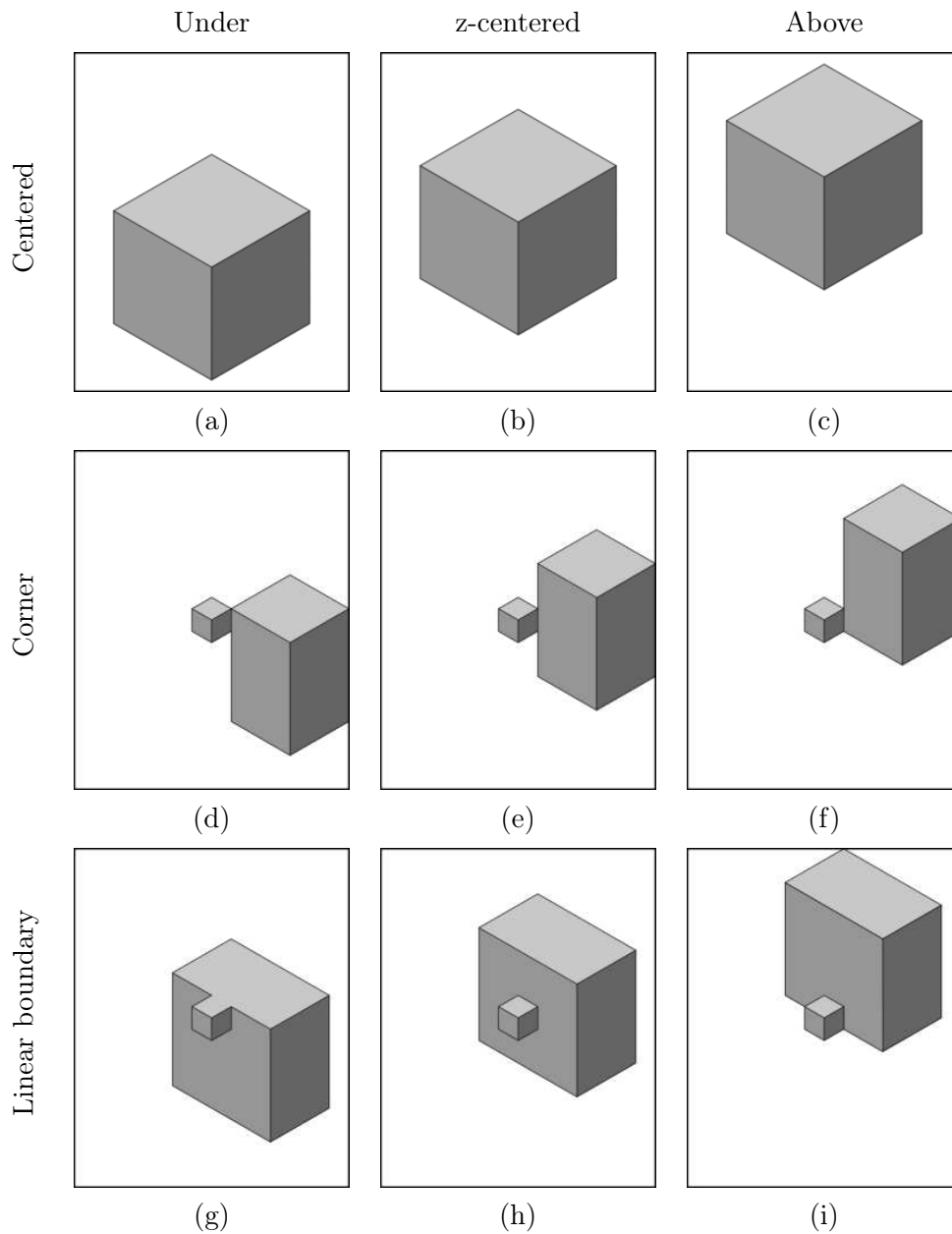


Figure 4.14: The 3 kinds of Nagao's masks retained in our study, each type with 3 possible locations on  $z$ . Centered masks (a,b,c) are boxes including the central voxel of the patch. Corner masks (d,e,f) and masks for linear boundaries (g,h,i) are respectively designed for regions in contact at the corners and on sides of patches.

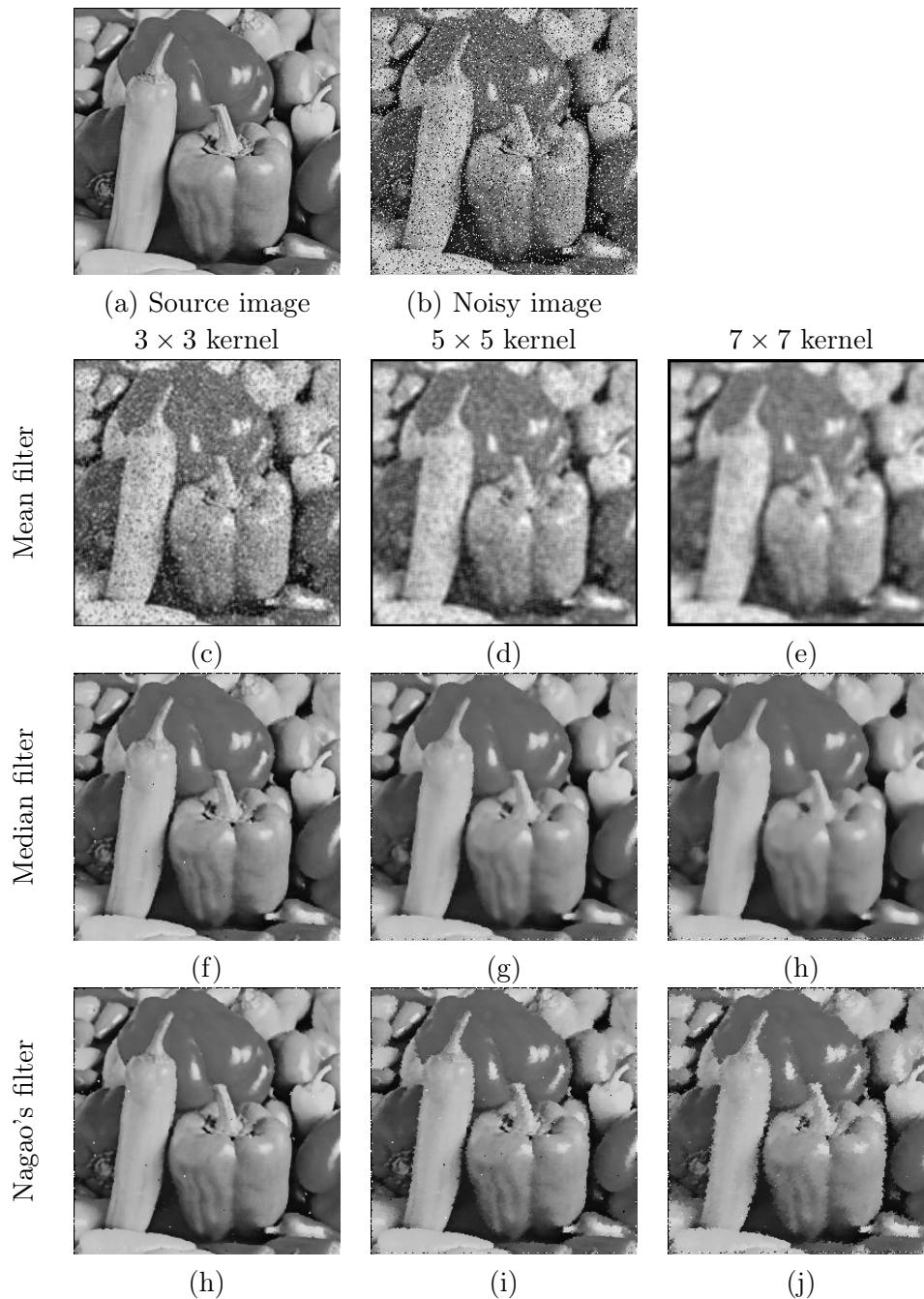


Figure 4.15: Impact of mean, median and Nagao's filters on a noisy image. The pepper image (a) was modified by adding salt and pepper noise (b). The noisy image is smoothed using a mean filter (c,d,e), a median filter (f,g,h) and Nagao's filter with Nagao's initial masks(i,j,k), all with diverse sizes of kernel [Nagao 1979].

are often used for problems with texture constraints. Finally, useful filters in the liver case were retained, namely median and Nagao's filters [Chemouny 2001].

Statistical and Haralick's descriptors are used as texture descriptors, but with some refinements. The idea is to introduce some kind of multi scale approach at the texture level, meaning to use diverse sizes of texture at the same time. Instead of manipulating a set of textures with diverse sizes, the texture descriptors were modified to account for diverse sizes at the same time. In terms of statistical descriptors, one can account for the above modification through a histogram computation within a radius from the center. For Haralick's descriptors the change is made in the co-occurrence matrices by adding a radius  $r$  to their definition that becomes  $P(d, \theta, r)$ , while the definition of descriptors remain unchanged. This new matrix gives the probability of pixels pairs at distance  $d$ , for direction  $\theta$  and for pixels at distance lesser than  $r$  from the texture center. This radius  $r$  may take any value between 1 and the texture radius. A radius of 0 makes no sense, because pixels pairs are required and the co-occurrence matrix should be computed within the texture patch. The other parameters, the distance  $d$  and the direction  $\theta$ , keep their previous values, namely  $d = 1$  and  $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ . It should be noted that the computation of a co-occurrence matrix  $P(d, \theta, r + 1)$  is eased when  $P(d, \theta, r)$  is known; one only has to add pairs on the edges.

For computation of these co-occurrence matrices, gray levels are reset to an admitted number of gray levels  $m = 8$  following a linear mapping. Keeping the raw intensities would indeed lead to big and scarce co-occurrence matrices, which would offer very little information gain. Moreover, these huge and scarce matrices would induce additional computational costs and higher memory consumption. Thus, intensities of each texture are linearly reset to an admitted number of gray levels  $m = 8$ , meaning that for each texture the range of intensity is divided in  $m$  parts, and each intensity is reset depending on the interval it is in. The choice of a number of admitted gray levels was done empirically, by increasing the number of admitted gray levels until no further improvement can be obtained by using more admitted gray levels. A more elaborate reset may be more informative, in particular with a reset based on the texture histogram, but this change would be time consuming. This lead may be nevertheless worthwhile to later explore, but the information gain should be balanced with the additional computation time.

#### 4.3.3.2 Choosing a texture size

The choice of a size of texture is a critical step. One has to find a balance between the computation time that is lower for smaller textures, and the quality of segmentation that should increase with texture size. For the problem of segmentation of healthy liver against lesions, the optimal size of texture was obtained for a size of  $13 \times 13$  pixels.

The definition of a size of texture in pixels and not in millimeters is justified by the type of the images for this study. While retained filters are expressed in millimeters for taking in account the anisotropy of voxels, textures may be defined

in a pixel basis because of the low variability of pixel dimensions on slices. Volumes used for diagnostic or follow-up of liver cancers are indeed in an axial basis, where voxel depth may vary a lot (1-5mm), but the dimensions of the pixel on each slice remain similar (average of 0.74mm, standard deviation of 0.07mm). Thus, introducing a variable size of texture would make the problem more complex without any predictable improvement.

The choice of the size of texture is empirically done by comparison of classification results for diverse sizes of texture. Given two sets of images, feature samples are extracted for a set size of texture, in order to create a training and a validation set. The training set is then used to train a classification function, whose quality is later evaluated as a balance between sensitivity and specificity on the validation set.

The quality was defined to insure a good sensitivity without loss in term of specificity. Thus, a comparison metric is defined as a weighted sum of these square two terms, allowing increasing the penalty for small values (eq. 4.33). The balance term is chosen to give more weight to sensitivity that matters more in a clinical context by setting  $\beta = \frac{2}{3}$ .

$$\beta [\text{sensitivity}]^2 + (1 - \beta) [\text{specificity}]^2 \quad (4.33)$$

The results shows that the best texture size is  $13 \times 13$  pixels (fig. 4.16). At first the quality of classification improves with the size of the texture until a size of  $13 \times 13$  is reached (fig. 4.16.a). Then, increasing the size of the texture over  $13 \times 13$  decreases the quality of classification, because the gain in sensitivity is more than negated by the loss of specificity (fig. 4.16.b). The retained size of texture is consistent with sizes chosen in similar problems. Pham retained  $9 \times 9$  patches for classification of healthy liver tissues [Pham 2007] and Smutek's experiments showed that classification of HCC vs. cysts is best achieved with textures samples from  $9 \times 9$  to  $13 \times 13$  pixels [Smutek 2006].

#### 4.3.4 Feature selection

##### 4.3.4.1 Worth of feature selection

Feature selection refers to the choice of a small number of features that better describes an object inside a bigger set of features. In this study, the selection relates to the choice of the filters inside the bank of filters  $\mathcal{F}$  as well as the features  $\phi_{\Theta, \Theta'}$  that are the most discriminative ones for the classification process.

The combination of a bank of filters and a bank of texture descriptors generates a huge number of candidate features (slightly more than 4000 at first). However, most of these features are useless or of very small relevance. Most features will indeed never be selected by the learning process, because only between 150 and 200 features are experimentally retained by the AdaBoost algorithm. The selection of features refers to the selection of a smaller subset of these candidate features where only remain the most significant ones for the problem. For the particular case of

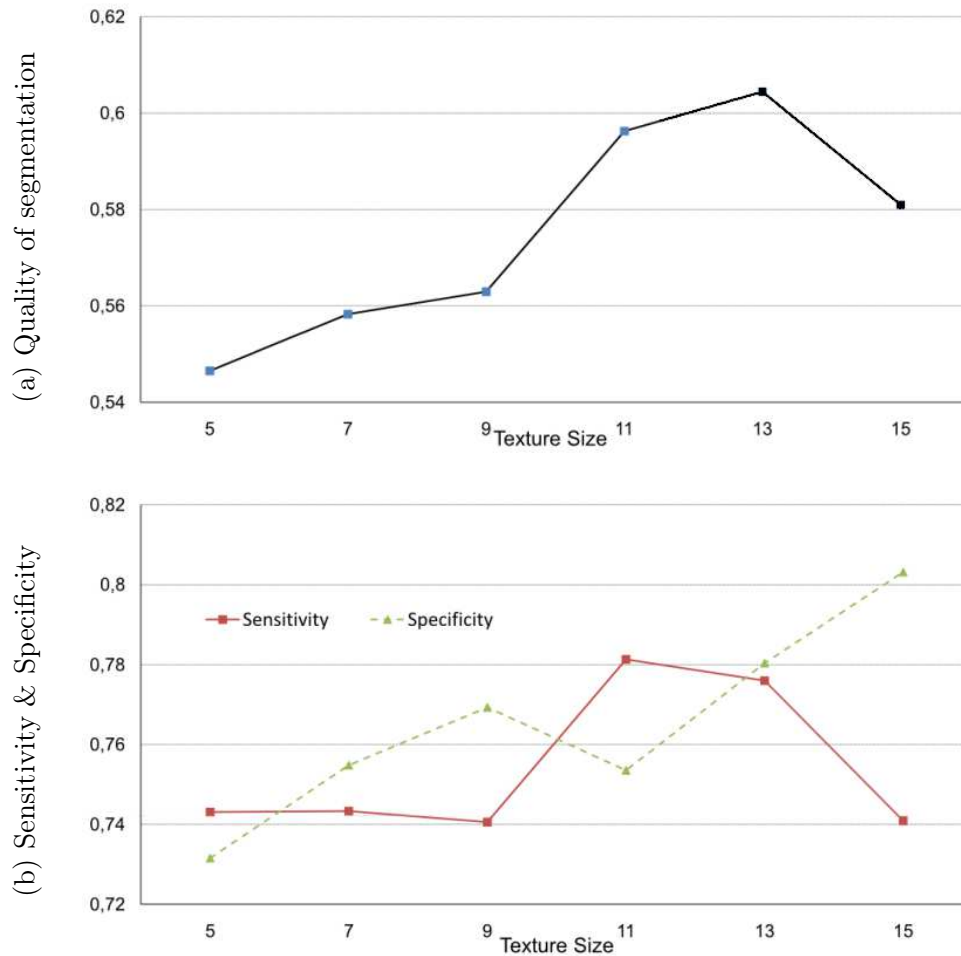


Figure 4.16: Quality of segmentation functions of the size of texture. First, the global quality is given for diverse sizes of texture (a). Then, the underlying metrics, namely sensitivity and specificity are given (b).

this study not only features are selected, but filters too. As features are composed of one filter and one texture descriptor, many features may require a same filter, meaning that the computational cost for one filter is shared between all the features that require this filter. Due to the speed constraint, removing features with a high computational cost is worthwhile. This pruning of features with high computational cost amounts to selecting relevant filters inside the filter bank  $\mathcal{F}$  and favoring features using the same intermediate objects (e.g. same co-occurrence matrix).

The selection of features offers several advantages; it will better fit with the clinical prospect, it may improve the robustness of the classification, and may allow a better understanding of the differences between the tissues. First, computing features takes some time, thus the less features the faster the classification, which will favor the clinical prospect of this study. Then, keeping only the most relevant features may improve the final classification. Removing less relevant features or

redundant ones will favor the use of more robust features inside the classification function, and the combination of more robust features should give a more robust classifier. Finally, the selection of features is similar to Principal Component Analysis in that features that better explain the difference between classes are selected. These features will be the ones where the difference between the classes is the more visible, which might be interesting from a research perspective.

#### 4.3.4.2 Method

##### Method

The selection of features is done to satisfy two constraints, keeping the most relevant features, while removing time consuming ones. This selection is done in an iterative process by training successive classifiers and selecting relevant features through a selection heuristic.

The selection is done using two different sets of samples described by their features, a training set and a validation set. First, training is done for a list of features, while tracing the best features at each step of the learning process. Then, the classification function is validated on the validation set. Finally, a smaller set of features is selected using a heuristic, according to the relevant features during the training process and the computational cost of these features. This iterative selection finishes when removing features implies a significant loss of quality for the learned classification function.

The AdaBoost algorithm iteratively selects the best features inside a set of features and for a weighted training set. For selection, the best feature is not the sole relevant feature, because many other features with similar discrimination ability may exist. Thus at step  $t$  of the selection, a classification function is trained using a set of features  $\phi^{(t)} = \{\phi_{\Theta, \Theta'}\}$ , while tracing the best features for this step of the learning process. This trace allows quantifying the relevance of each feature for this training step. Then, the global relevance of each feature for the whole classification is defined by combining the relevance of this feature at each step of the training process.

The validation aims to define the relevant *steps* of the learning process and to evaluate the quality of classification. As seen before, the quality of classification tends to be asymptotic after a while, thus features used in the later stages of the classification bring less improvement to the overall classification. Hence these later features are less significant for the selection process, because it is reasonable to assume that the small gain brought by these features may be obtained with other features. Consequently, the relevance of the features used in the last components of the training process should not be retained during the selection. The validation also aims at evaluating the loss of quality induced by the last set of features, which will trigger the end of selection when a significant loss of quality is detected. To summary, the quality of classification is drawn functions of the number of components inside the classification function. The quality of classification is then assessed. If this quality decreased too much, the selection process is aborted and the last correct set

of features is kept. Otherwise, the beginning of the asymptotic part in the graph is located and only the previous components are taken in account for the next selection step, as well as the subsequent trace from training.

Finally, the selection of a new set of features is achieved using several heuristics along with several rules on the global relevance of each feature. This global relevance, or relevance for the entire classifier, is defined as the cumulative relevance of each feature used through the learning process and before the asymptotic part.

### Selection heuristic

The selection is done in two steps. First, some heuristic metrics are used to preselect relevant features. Then, rules are applied to remove time consuming features, while taking in account the gain they bring.

The selection starts with a preselection of the best features for a number of metrics. More than a selection, this step aims at removing the least informative features. This removal is aimed at decreasing around a third of the initial features in order to avoid removing too many features at the same time. Three metrics were used to evaluate the relevance of each feature at each learning step. These metrics were then summed to define the relevance of each feature for the entire classifier. Next, features were sorted according to each global metric and the best features for each global metric were preselected. Three metrics were retained to evaluate the value of each feature. The two first metrics relate to an absolute gain, namely whether or not the feature is useful. Then, the last heuristic value aims at better describing the relative contribution of one feature by quantifying the information loss compared to the optimal feature at one step of the learning process. These metrics are defined as follows.

- The number of times a feature is among the best ones, which characterizes how often a feature brings a gain without taking in account the amount of gain.
- The number of times a feature is the best feature for one step.
- The relevance of a feature compared to the best one, which takes in account the gain brought by a feature. Let us consider the successive errors  $\epsilon^t$  of one feature through the  $T$  steps of a learning process, along with the error  $\epsilon_{best}^t$  for the optimal feature at each step of the training process. The relevance of this feature is then defined as a sum of inverse exponentials of the gain brought by the feature compared to the optimal one, with an additional factor  $\alpha = 5$  aimed at defining the relevance of the gain (eq. 4.34).

$$\sum_{t=1}^T \exp^{-\alpha \frac{\epsilon^t - \epsilon_{best}^t}{\epsilon_{best}^t}} \quad (4.34)$$

The selection is then finalized by applying a number of rules to remove time-consuming features from the preselected set. Preselected features are indeed relevant



		Initial	Final
Features (count)	Filters	13	8
	First order descriptor	57	16
	Second order descriptor	101	4
	Total	158	20
Computation time (s)		129	33
Quality of segmentation	Sensitivity	0.818	0.826
	Specificity	0.911	0.887

Table 4.1: Selection of features, gain and cost. Given an initial set where useless features were removed beforehand, the best features are selected. This selection offers a gain of speed by decreasing both the number of features and the number of filters, to the cost of a small loss in specificity. Given values were obtained for the segmentation of liver colorectal metastases with a core 2Duo 2.8GHz CPU.

ones, but may have very different computational costs. The final selection is thus done to penalize time-consuming features by favoring the share of computational costs. This choice is achieved by penalizing filters that are not required by many features, or by favoring second order descriptors on same co-occurrence matrices. This selection is done by applying some rules, while insuring that no significant information is lost through the process.

- A feature is removed when it relies on a filter that is not used for any other feature. However, when this feature is very useful (often chosen as best descriptor), a comparison between learning with and without this feature should be done beforehand, because its removal might induce a significant loss of quality.
- The features for a same filter and with second order descriptors defined on a same co-occurrence matrix are favored over features computed over different matrices. In order to further share the calculi required to compute a co-occurrence matrix, processes common to many co-occurrence matrices are favored. Thus features composed of a same filter and of any second order descriptors are favored when they are computed on co-occurrence matrices defined for a same radius  $r$  first and for a same direction  $\theta$  otherwise.

### Selection gain

The selection of features allows faster classification, with small loss of quality. A comparison of classification without and after feature selection is given for colorectal metastases (fig. 4.1). The selection of features allows decreasing the number of both features and filters, which induces a huge speed boost of almost 45%. This substantial speed gain is obtained at the cost of a small worsening of quality: 2% decrease of specificity, but 1% improvement for sensitivity.

## 4.4 Segmentation of tumoral tissues

### 4.4.1 Creation of a classification function

A classification function to distinguish healthy tissues from tumoral ones inside liver is learned. This classification function is trained using a machine learning technique previously introduced, namely AdaBoost. This technique aims at classifying each voxel using selected features that characterize the surrounding texture of this voxel. Features were previously introduced, thus only three tasks remain before obtaining an adequate classification function. First, the weak learners for the AdaBoost process should be defined. Then, a well chosen training set should be created. Finally, the classification function should be learned and validated.

#### 4.4.1.1 Generation of a good training set

The creation of a training set is a crucial step. In order to achieve good generalization error, tumoral and healthy textures should be correctly sampled, meaning that the samples should depict well the possible appearances of each class, while sticking close to the real distribution of these appearances. First, using samples that depict the wide range of possible appearances for each class is important to get robust classifiers. Even if AdaBoost classifiers have the ability to generalize to new samples, training on a biased set will not provide good results. Thus the training set should be a good sampling of the possible appearances of the tissues. This sampling should in particular follow the real life distribution of the appearances. Training is indeed done by minimizing the error on the training set. If some appearances are overrepresented on the training set, the learning process will tend to better classify these appearances, which could be at the expense of other patterns. Thus an incorrect distribution of samples may lead to better classification ability for appearances with rare occurrence, while having lower classification ability for more common appearances.

The creation of a good training set as previously defined is insured by the selection of a representative set of images first, and then by a good sampling of texture patches inside this set.

A good sampling of possible textures is first dependent on a set of images that represents well the possible clinical cases and environments. The differences of appearances between images have two main causes, technical ones and anatomical ones. First, images differ between various scanners, levels of enhancement vary in function of the injection protocols and the size of slice has a big impact on the appearances of tissues. Then, the appearance of lesions varies between two types of tumors and even within a single type. Thus, images coming from various hospitals, with diverse injection protocols, various sizes of slice and diverse scanners were retained. These images were chosen to cover the possible types of tumors, while insuring that there were more examples of more common tumors.

Sampling texture patches on a set of images while well depicting the possible appearances is done under the assumption that a good spatial distribution of the

samples implies a good distribution of the appearances. This assumption seems valid because sampling texture patches regularly located inside a region will provide samples from diverse part of this region, which should provide a good sampling of the appearances. This assumption will not be valid for regions with a regular organization of appearances (like a chessboard). However, no regular organization seems to exist inside the liver envelope. Thus, the creation of a training set consists in taking an equal number of patches from tumoral and healthy tissues while sampling patches with a regular distribution within each region, tumoral or healthy.

Introducing a bias while sampling texture patches may sometimes have a positive contribution. This bias will allow giving more weight to some appearances, which might be relevant to correct some classification errors. In particular the boundaries of the lesions are difficult to classify and are not given an important weight by a regular distribution of samples. More weight may be given to these locations by adding a regular distribution of samples inside lesions boundaries to the regular distributions inside lesions and healthy liver.

#### 4.4.1.2 Definition of weak learners

Weak learners are functionals defining simple ways to discriminate the samples and are defined in this study as a simple comparison of one feature with a threshold. No complex weak learners are required, because the texture features already contain the information. The role of the weak learners is only to discriminate using one feature at a time. Thus, the weak learners are defined as the comparison of the texture feature  $\phi = \phi_{\Theta, \Theta'}$  of any voxel  $\mathbf{x}$  to a threshold value  $\gamma$ , where the sign of comparison is given by  $\delta \in \{-1; 1\}$ .

$$h_{\phi, \delta, \gamma} : x \rightarrow \begin{cases} 1 & \text{if } \delta \cdot \phi(\mathbf{x}) \leq \delta \cdot \gamma \\ -1 & \text{otherwise} \end{cases} \quad (4.35)$$

#### 4.4.1.3 Learning the classification function

The classification function is learned using the AdaBoost approach. The only difficulty is the selection of the best weak learner at each step of the algorithm. This selection is equivalent to the research of optimal parameters for the weak learners  $(\delta, \phi, \gamma)$ ; while defining a sign of comparison  $\delta$  is easy, finding the best feature  $\phi$  and the threshold  $\gamma$  is more complex (eq. 4.35).

The research of the best features is done with a heuristic value that measures the relevance of each feature. Two heuristics were tried to select the best feature during the AdaBoost process. First, the relevance of each feature  $\phi$  was defined as the classification error for the optimal weak learner  $h_{\phi, \dots}$  on the weighted training set. Then, the relevance of each feature was defined by the separability between the classes, using a weighted Fisher's discriminant to measure this separability. This last heuristic was tried expecting more robust classification functions. However, the classification results were worse. Moreover, no significant improvement of the

robustness was obtained. Thus, the best feature was chosen as the one for which the optimal weak learner has the smallest error on the weighted training set.

The optimal threshold  $\gamma$  is computed with a brute force approach. Let us consider the training set  $\chi = \{(x_1, y_1), \dots, (x_m, y_m)\}$  previously defined (sect. 4.2.4). At step  $t$  of the AdaBoost algorithm, the weights are distributed according to the distribution  $D_t = \{D_t(1), \dots, D_t(n)\}$ . The research of the optimal threshold is then done by taking in account only one chosen feature  $\phi$  at a time. First, the  $n$  values of this feature  $\{\phi_1, \dots, \phi_n\}$  inside the training set are sorted, which defines a mapping  $\Psi$  from the set of features to the sorted set. Then,  $n-1$  candidate thresholds are defined as average thresholds between two successive feature values inside the sorted set:  $\gamma_i = \frac{\Psi(\phi_i) + \Psi(\phi_{i+1})}{2}$ . The optimal threshold is finally defined as the candidate threshold  $\gamma_i$  with minimal weighted classification error (eq. 4.36).

$$\gamma = \underset{\gamma_i}{\operatorname{argmin}} \sum_{j=1}^n \Psi(D_t(j)) \cdot [h_{\phi, \delta, \gamma_i}(\Psi(\phi_j)) \neq \Psi(y_j)] \quad (4.36)$$

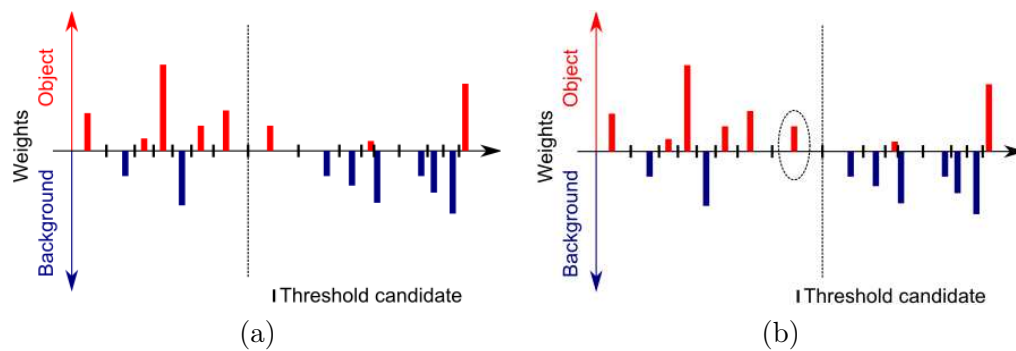


Figure 4.17: Computation of the parameters for a weak learners. The values of the features are shown on a horizontal axis as bars, whose size shows the weight associated to each sample. The candidate thresholds are displayed as green strokes.

The advantage of sorting is twofold. First, it allows defining a finite set of threshold candidates. Then, the ordering allows computing the weighted classification error in a fast and simple manner, following a recursive scheme. Indeed, knowing the weighted classification error for a threshold  $\gamma_i$ , the error for the next threshold  $\gamma_{i+1}$  is directly obtained by updating the error while considering only the weighted difference of error between the two thresholds  $\gamma_i$  and  $\gamma_{i+1}$  for the feature  $\phi_{i+1}$ . This update is illustrated for the case where comparison is done for  $\delta = 1$  (fig. 4.17). Given the weighted classification error for a threshold candidate shown by a vertical line (fig. 4.17.a), only one feature value (encircled) has to be taken in account to compute the error for the next candidate threshold (fig. 4.17.b). Indeed, the feature values inferior to the first threshold remain inferior to the new one and so do superior feature values, apart from the encircled one. Thus, only the modification of the classification for this feature value has to be taken in account for updating the weighted classification error.

$$\sum_{j=1}^n \beta_{i+1,j} = (\beta_{i,i+1} - \beta_{i+1,i+1}) + \sum_{j=1}^n \beta_{i,j} \quad (4.37)$$

where  $\beta_{p,q} = \Psi(D_t(q)) \cdot [h_{\phi,\delta,\gamma_p}(\Psi(\phi_q)) \neq \Psi(y_q)]$

#### 4.4.1.4 Validation

The validation is an intermediate step between learning of a classification function and use of this function on new data sets. This step aims at checking that the learned classifier is relevant for the problem and sometimes to do some final tuning before freezing the classifier for later use.

This step has several uses that were previously presented (sect. 4.2.4.3). First, the validation allows verifying that the classification function is relevant for the considered problem. Then, the validation may be used to tune up the classification function. This tuning allows reducing the problems due to overfitting, it also improves the classification speed and allows detecting non optimal uses of the learning process.

### 4.4.2 Segmentation based on pixel classification

The segmentation aims at detecting and defining the boundaries between the tumoral regions and the healthy ones. From a computational perspective, segmentation amounts to defining the class of each pixel inside an image by labeling the pixels. For this study, segmentation is done inside a liver mask, where intensities are normalized, in order to decrease the intensity variability between the images.

Let us assume that a liver envelope is known. Such envelope may come from automatic processes as detailed in previous sections (sect. 3.3) (sect. 3.5.4) (sect. 3.6.3) or may be manually done in order to rule out any influence of the initial envelope on the segmentation results.

Before any labeling of pixels, intensities are normalized within the liver envelope. This normalization allows working on a common basis for any image by giving a similar intensity to similar anatomical structures. This normalization is done with non-linear histogram matching, aiming for the transformation of the histogram inside the liver envelope into a reference histogram (sect. 2.6).

#### 4.4.2.1 Straight segmentation

Applying a classifier to each pixel of the image is the easiest way to segment lesions inside this envelope. Using a classification function trained to distinguish healthy from tumoral tissues, pixels may be labeled in function of the binary result of the classifier, thereby giving the class of each pixel. The creation of such classifier was previously detailed and does not involve any additional difficulty, because no additional problem is introduced by the work on normalized intensities instead of

raw ones. The only additional constraint is the definition of a texture patch that characterizes each pixel.

Any pixel inside the mask may be characterized by a texture patch defined with the surrounding pixels, with the exception of pixels on the boundaries of the image. Because of the spatial relations, neighbor pixels are always known and may be used to define a square patch around a central pixel. Such texture patch will characterize the central pixel inside this patch. However, no texture patch can be defined for pixels on the boundaries of the image because some pixels required for the texture patch may be outside the image. These pixels may nevertheless be excluded without any drawback, because the liver remains far from these boundaries for usual exams.

#### 4.4.2.2 MRF for better segmentation

In this section the segmentation is done while taking in account both the probability of belonging to a class at pixel level and the likely classes of neighbor pixels. With the previous approach the value of the AdaBoost score (eq. 4.4) did not matter, meaning that two pixels with very close scores could be assigned to different classes. Indeed, the reliability of the classification was not taken in account with this rough estimate.

To compensate for this previous lack, the segmentation problem is expressed as a balance between the probability of belonging to a class and the class of the neighbor pixels expressed as a problem of MRF minimization. First, this new formulation requires the definition of the class probability at pixel level. These probabilities of belonging to a class are introduced and computed from the AdaBoost score using Friedman formula (eq. 4.5) [Friedman 2000]. As in previous case, a classification function is used to define the class of each pixel using its surrounding texture. However, the information on the reliability of classification is retained instead of using binary information. Then, the neighbor pixels are introduced to favor the local homogeneity of classes through a MRF formulation. Indeed, MRF are well suited for this type of problem and provide a fast solving method. Thus this problem is stated as a pairwise MRF.

The proposed approach is presented for a 2-class problem, but extends to any number of classes. However, doing so will require other methods to compute class probabilities, because standard AdaBoost deals only with binary problems.

#### Introducing the MRF problem

Let us consider a discrete set of labeled tissues  $\mathcal{L} = \{u_L, u_O\}$ , where  $u_L$  stands for the liver tissues, and  $u_O$  for everything else. Let us consider now a set of nodes  $\Omega'$  given by the set of voxels inside the liver envelope and an associated neighborhood system  $\mathcal{N}_n$  that describes the spatial relations between the pixels. The segmentation of the image is given for the labeling  $\mathcal{C}^*$  that minimizes a MRF energy, where any labeling  $\mathcal{C}$  gives the label  $u_{\mathbf{x}}$  of a node  $\mathbf{x}$  inside  $\Omega'$ ,  $\mathcal{C} = \{u_{\mathbf{x}} : \mathbf{x} \in \Omega', u_{\mathbf{x}} \in \mathcal{L}\}$ .

$$\begin{aligned}
\mathcal{C}^* &= \underset{\mathcal{C}}{\operatorname{argmin}} E_{seg}(\mathcal{C}) \\
E_{seg}(\mathcal{C}) &= \sum_{\mathbf{x} \in \Omega'} V_{\mathbf{x}}(u_{\mathbf{x}}) + \beta \sum_{\mathbf{x}, \mathbf{y} \in \mathcal{N}_n} V_{\mathbf{x}, \mathbf{y}}(u_{\mathbf{x}}, u_{\mathbf{y}})
\end{aligned} \tag{4.38}$$

The segmentation energy (eq. 4.38) is defined as a sum of two terms, a data term and a regularization term, balanced by a factor  $\beta$ . The first term aims to give the best fitting label to each voxel of the set of nodes  $\Omega'$ , while the second component aims to penalize neighbor voxels with different labels.

### Defining the data term

The data term aims at maximizing the adequacy between each voxel and a class. The global adequacy is thus expressed as the sum of the individual adequacies for each voxel of the image, where the adequacy  $V_{\mathbf{x}}(u_{\mathbf{x}})$  measures the probability of belonging to a class  $c_{\mathbf{x}}$  for a voxel  $\mathbf{x}$ , with a negative log likelihood to be able to proceed to minimization.

$$V_{\mathbf{x}}(u_{\mathbf{x}}) = -\log P(\mathbf{x} | c_{\mathbf{x}}) \tag{4.39}$$

### Defining a regularization term

The regularization term aims at making up for classification errors by taking advantage of the neighbor pixels. This spatial regularization is achieved with the introduction of a penalty  $V_{\mathbf{x}, \mathbf{y}}(u_{\mathbf{x}}, u_{\mathbf{y}})$  for the change of labels between two neighbor voxels  $\mathbf{x}$  and  $\mathbf{y}$  ( $\mathbf{x}, \mathbf{y} \in \mathcal{N}_n$ ).

Several regularization terms have been tried, beginning with the exact definition of the penalty, meaning a penalty in case of discontinuity and nothing otherwise, which is the opposite of the definition of Kronecker's delta.

$$\begin{aligned}
V_{\mathbf{x}, \mathbf{y}}(u_{\mathbf{x}}, u_{\mathbf{y}}) &= \bar{\delta}_{u_{\mathbf{x}}, u_{\mathbf{y}}} \\
\bar{\delta}_{i, j} &= \begin{cases} 1 & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases}
\end{aligned} \tag{4.40}$$

This regularization term may be also chosen to take more information in account, in particular the Euclidean distance between pixels  $\|\mathbf{x} - \mathbf{y}\|_e$  and the intensities of voxels  $V(\mathbf{x})$ .

$$V_{\mathbf{x}, \mathbf{y}}(u_{\mathbf{x}}, u_{\mathbf{y}}) = \frac{1}{\|\mathbf{x} - \mathbf{y}\|_e} \exp^{-\frac{(V(\mathbf{x}) - V(\mathbf{y}))^2}{2\sigma^2}} \bar{\delta}_{u_{\mathbf{x}}, u_{\mathbf{y}}} \tag{4.41}$$

where  $\sigma$  characterizes the image noise. This formulation term is often used for MRF segmentation using intensity distributions as models. The use of intensity

may indeed bring over some information for homogeneous lesions. However this regularization term may be detrimental to heterogeneous tumors.

The first proposition of regularization term (eq. 4.40) might improve with the addition of a distance constraint that could be a relevant way to better take in account the anisotropy of voxels. In particular for thick slices, it would avoid giving same importance to neighbor voxels on a same slice and on different slices.

## 4.5 Protocol

### 4.5.1 Defining the classification function

Several guidelines for learning a good classification function will be detailed with an emphasis on the critical points. These instructions will be first given as generic instructions for any kind of problem, and then applied to the classification of healthy vs. non healthy tissues within the liver. Then, the choices of parameters or datasets, either generic or specific to this study will be justified or reviewed.

The process of creation of a classification function is shared by any problems, but the settings are specific to each problem. Before beginning to create a classification function, two sets of images should be chosen for training and validation. Then, normalization parameters should be defined, and the images inside the data sets should be normalized. After normalization, relevant features for the problem should be selected. Finally, a classification function is learned and later validated.

A schematic view of the creation process is shown in (fig. 4.18). The user begins by defining a training and a validation set, as well as a bank of features and some parameters for learning. First, a reference histogram is computed and used to normalize the sets of images. This reference histogram is computed using only the training set in order to avoid the introduction of a bias. Then, features are selected inside the provided bank of features while using samples from the normalized training set. Learning is done next, using the selected features and samples from the normalized training set. The learned classifier is finally validated on the normalized validation set. The result of validation is truly important. If validation succeeds, the classifier is ready to use. However, when validation does not provide the expected results, the whole process should be done again after examining the reasons of the failure. An exception exists when the quality of classification still improves with the last components of the classifier. In this case the classifier might be improved by returning to the learning step while using a bigger number of rounds for training.

#### 4.5.1.1 Choice of image sets

Before any learning, a set of images for training and one for validation are chosen. These sets should contain various examples of tumoral and healthy regions and should be chosen to reflect both the possible cases encountered in real life and the probability distribution of these cases. The intersection between these two sets should also be empty to prevent any bias.



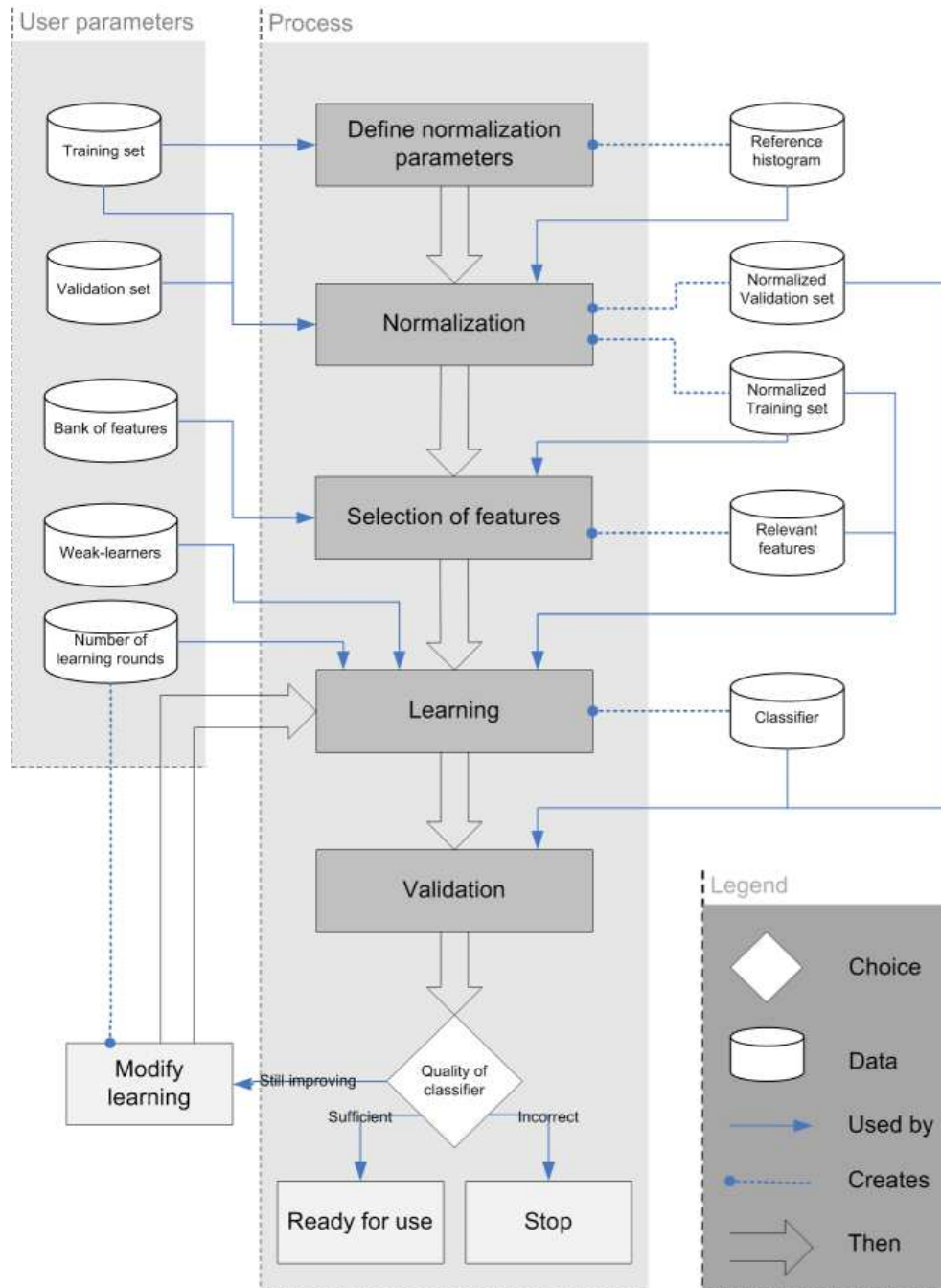


Figure 4.18: Creation process of a correct classification function.

A realistic representation of images encountered in clinical context is required to hope for learning a robust classifier. Indeed, the chosen images should depict as much as possible the diversity of technical and anatomical conditions, while following the distribution of this diversity in real life (sect. 4.4.1.1).

Different images should be used for training and for validation. Using a same image twice would indeed introduce a bias, because this image would not allow evaluating the generalization of any learned classifier. A same image does not imply same texture samples, but even with different samples a bias remain. By using a same image twice, many problems to address disappear; the anatomical variability between patients, and the technical variations (enhancement, reconstruction and scanner...). Thus, training and validating with samples extracted from a same image does not allow any conclusion regarding the robustness and the classification ability of any classifier learned with such a bias. Biases introduced by the use of a same image are still relevant for the test phase, where no image should be used for test and be contained inside one of these two previous sets.

#### 4.5.1.2 Normalization

Normalization is introduced to reduce the intensity variability between the images as histogram matching between the histogram inside an envelope and a reference one. This reference histogram is first defined as a mean histogram and then used as reference for the normalization of all images inside training and validation sets.

The reference histogram is defined as a mean histogram inside all liver envelopes from the training set. Images from the validation set are excluded to prevent any bias. Using histograms extracted from the validation set would indeed exclude the errors of the normalization step from the evaluation of classification, which would induce a bias by not favoring features that are more robust to normalization errors. A mean histogram is a better approximation than the use of a random image as reference histogram. However, a more robust reference histogram might be obtained by applying matching for histogram population. Matching of histogram population, meaning research of a histogram for which matching is optimal for all histograms inside the training set, could indeed allow better normalization by defining a reference that is more suitable for the retained matching method.

The normalization of training and test sets does not introduce any particular difficulties. Histogram matching was previously detailed (sect. 2.6) and normalization consists only in applying the mapping function to the images inside the sets.

#### 4.5.1.3 Feature Selection

The normalized training set is used to select the best features inside a bank of features, i.e. to build a subset of relevant features for a problem from a bank of features. Three main steps form this stage, the choice of a bank of features, the definition of a set of samples from the training set and the selection process.

A bank of features relevant for the problem should be first defined by the user. For this study, the features were chosen as texture descriptors applied to a filtered

image (sect. 4.3.3). Then, the images from the normalized training set were sampled while imposing a regular spacing of patches taken from tumoral and healthy tissues within the liver (sect. 4.4.1.1). Finally, features were selected following the approach described in a previous section (sect. 4.3.4.2).

#### 4.5.1.4 Learning

Learning consists in combining weak learners provided by the user to create a classifier that has good classification ability on the normalized training set. Learning is done with the AdaBoost algorithm (sect. 4.2.4) and is composed of a number of rounds (chosen by the user) where the best weak learner (sect. 4.4.1.2) for the weighted training set is chosen and parameterized (sect. 4.4.1.3). For this study, the number of training rounds was initially set to 250 and later set to 400 to prevent any suspicion on the end of the learning process.

#### 4.5.1.5 Validation

The validation is the last stage of the creation of a classification function, where is decided on whether the classification function is suitable or should be improved or even rejected. The classification previously learned is first applied to the normalized validation set. Then, the classifier is validated by plotting the quality of the classification, functions of the number of components inside the classifier (sect. 4.2.4.3).

The results of segmentation may trigger off three possible actions. If the classification showed sufficient results, the classifier is suitable for segmentation. Otherwise, learning again should be considered when the quality is not sufficient and is still improving. A more accurate classifier may indeed be learned when the quality of segmentation is still improving with the last components of the classifier. Thus, the learning process is done again with a higher number of learning rounds. For the other cases of failure, an explanation should be researched on the adequacy of the bank of features first, and then on the suitability between the proposed method and the problem. The bank of features should be checked first. This verification implies evaluating the relevance of chosen features and if necessary adding new ones as well as assessing the texture descriptors, in particular the size of texture patches and the number of admitted gray levels. Then, the relevance of training and validation set should be verified. Finally, the normalization process may be questioned by evaluating the contribution of normalization. If none of these trails allows improving the classification function, maybe the proposed approach is not suitable for the considered problem.

### 4.5.2 Implementation issues

The methods previously presented were developed using the C++ language, with Visual Studio 2005 as Integrated Development Environment and applied as a segmentation module for Myrian<sup>®</sup>, software for review of medical images and aided diagnosis. Binary files were generated by the compiler of Visual C++ 2005 with

support of parallelization done with OpenMP. The segmentation module was build either as 32 or 64 bits executable; this latter case allows solving MRF problems for huge images without memory issues during the optimization of the MRF.

A special attention was given to critical parts of the algorithm. Indeed, the clinical prospect of the study implies a speed constraint, thus time consuming parts are implemented using optimized methods. Coding tricks are not considered here; focus will be on choices of algorithmic methods or implementations to simplify the complexity of the tasks. Only the classification and MRF solving are treated here, because other tasks are non significant from a length perspective.

#### 4.5.2.1 Classification

The classification requires the computation of filters, of texture descriptors and the combination of the features to compute the function response. First, filtering and computation of texture descriptors will be presented in later sections. Then, no particular optimization is done for the computation of the response of the classification function. The calculus of the response of a weak learner is too simple to be improved. However, the classification may be done to favor the computation of features. Instead of computing each feature following the order inside the classification function, the components of the classifier may be sorted to take advantage of common computations. For example, all components using features on a same filter should be consecutive, and so should features computed on a same co-occurrence matrix. . .

#### 4.5.2.2 Filtering

The optimization of filtering involves mainly parallelization and code optimization. . . in particular for filters based on convolution. However, the choice of a method to get the median value for Nagao's and median filters was an important algorithmic choice. Radix sort was retained to sort values in order to get the median one.

Radix sort relies on the digits of a number instead of the whole number to sort a list of numbers. Such approach provides a complexity gain for the sorting process. Indeed, the complexity of radix sort is asymptotically optimal for a table of  $n$  numbers, where each number is coded by  $k$  keys of same size (the Bytes defining an integer in C++), with a time complexity in  $O(nk)$  and a space complexity of  $O(\log n)$ . Moreover, radix sort is the most efficient sorting method for alphanumeric keys provided that the keys are not too long according to Knuth [Knuth 1973]. The parallelization provides sorting algorithms with lower complexity. However, such approaches do not offer any advantages in this particular situation. Due to the limited size of the arrays to sort, the complexity gain brought by these parallel approaches is indeed negated by the cost of launching the parallel threads.

### 4.5.2.3 Fast computation of first order features

The optimization of the calculi of first order features is first done by grouping computations, and then by using integral images to compute means. Every first order metric requires the mean value inside a texture patch, thus metrics are grouped by size of texture patch in order to compute each mean only once. Then, integral images are introduced to compute mean values inside a texture patch. Integral images allow computing each mean value with only 4 additions and 1 division, which is a huge gain.

#### Introducing Integral Images

Viola and Jones introduced integral images in order to compute sums on rectangular 2D regions in a fast manner [Viola 2004]. Integral images are defined as the sum of intensities above and on the left inclusive, which is shown by a gray rectangle inside an example (fig. 4.19.a). Then, the computation of the sum inside any rectangle is done with only 4 additions and one division. For example (fig. 4.19.b) the sum of the intensities inside A is computed by subtracting the values of the integral image in 2 and 4 to those in 1 and 3. Noting  $ii$  the value of the integral image, the values of the integral image are indeed defined by summing the intensities over the regions A, B, C, D.

$$\begin{aligned}
 ii(1) &= A + B + C + D \\
 ii(2) &= B + C \\
 ii(3) &= C \\
 ii(4) &= C + D
 \end{aligned} \tag{4.42}$$

hence

$$A = ii(1) + ii(3) - (ii(2) + ii(4)) \tag{4.43}$$

Given an image  $I : (x, y) \rightarrow i(x, y)$ , Viola and Jones introduced the integral image  $II : (x, y) \rightarrow ii(x, y)$  as the sum of intensities above and on the left inclusive of each pixel (eq. 4.44). The authors also introduced a recursive scheme to initialize the integral image that constructs the integral image by first summing rows and then propagating these sums on columns (eq. 4.45).

$$ii(x, y) = \sum_{\substack{x' \leq x \\ y' \leq y}} i(x', y') \tag{4.44}$$

$$\begin{aligned}
 s(x, y) &= s(x - 1, y) + i(x, y) \\
 ii(x, y) &= ii(x, y - 1) + s(x, y)
 \end{aligned} \tag{4.45}$$

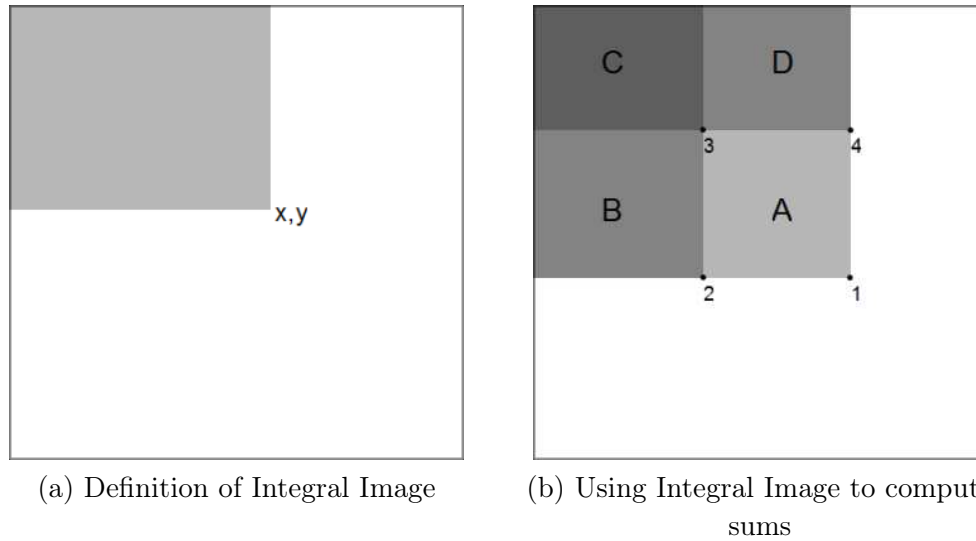


Figure 4.19: Definition and use of an integral image. The integral image at point  $(x,y)$  is defined as the sum of pixel intensities on the upper left part of this point, and shown as a gray region in image (a). Then, sums on squares may be computed using an integral image by using only the values of the integral image on the corner of this square (b).

#### Gain for first order features

The introduction of integral images allows computing mean values with a substantive gain of time, but to the cost of additional memory requirements. First, the computation of means with an integral image is faster than without it, even when the creation of the integral image is added to the evaluation of means. This is verified experimentally and will be proven in next paragraph through the computational complexity at each pixel of the image. Then, an additional memory requirement comes from the integral image that should be kept in memory. However, only one integral image for one slice at a time is required, because textures are taken on plane only, which is not too expensive.

Let us consider an image  $\mathcal{V}$  where some means have to be computed for each pixel of the image and for diverse sizes of patch  $\{n_1, \dots, n_p\}$ . Only additions will be retained to compare the complexity of both approaches, because the number of multiplications remains the same, i.e. 1 per texture patch and per pixel. Computing the means with a naive method requires  $n_i^2$  additions per texture patch and per pixel. Hence  $\sum_{i=1}^p n_i^2$  additions are required to compute all the required means. With the other method the integral image has first to be computed, requiring 2 additions per pixels with the recursive construction (eq. 4.45), and then 4 additional additions are required to compute the mean for each texture. Hence a total of  $2 + 4p$  additions per pixel is required to compute mean values with an integral image. To conclude, integral images allow simplifying the complexity from polynomial compared to the size of texture to constant time.

#### 4.5.2.4 Fast computation of second order features

As for first order metrics, features are first grouped to share common computations; features are computed in order to avoid multiple normalizations of intensities inside the texture patches and to prevent computing the same co-occurrence matrix twice. Then, some metrics are computed at same time to take advantage of common computations. Finally, probabilities inside the co-occurrence matrices are replaced by discrete counters, which lead to a speed gain because of the differences of treatments inside the CPUs.

Miyamoto and Merryman studied fast computation of Haralick's texture features [Miyamoto 2006]. Their contribution is twofold. First, they proposed a method to construct co-occurrence matrix faster, but their approach does not apply for this study because it was build for use with multiple distances  $d$  for the matrices. Then, they also introduced optimizations for the computations of second order features that are relevant for this study. Indeed, Miyamoto and Merryman described ways to concatenate the Haralick's features in order to avoid multiple accesses to the same values inside the co-occurrence matrices. They finally took in account the symmetry of the co-occurrence matrices to decrease the number of accesses to values inside these matrices.

Instead of working with co-occurrence matrices defining the probability of each pixel pair, work is done directly on the number of occurrences for each pixel pair. This modification has little influence on the complexity of the calculi of the metrics. However, computing on float instead of integer has a significant impact on the computation time, because of the difference of treatments by a CPU. In particular, doing additions on integer is twice as fast as the same operation on float.

The Haralick's features previously retained (sect. 4.3.1.4) may all be expressed for a discrete co-occurrence matrix  $Q$ . First, the entropy is simplified by the replacement of  $P_{i,j} \log P_{i,j}$  by accesses inside a discrete table. Then, many features may be computed entirely using integers before a normalization step, for example Entropy, Sum Average, Energy, Contrast and Maximum Probability. The normalization term requires some more time, but for patches with a reasonable size the gain brought by the integers more than offsets the cost of normalization. Sometimes, like for the Sum Average the normalization term even disappears inside an already existing term. Finally, the other features are still computed as floats, with no additional calculi, apart from one additional multiplication for the Inverse Different Moment or the Cluster Tendency. This additional multiplication should even be offset by the gain while computing the auxiliary means.

Entropy :

$$- \sum_{i=1}^m \sum_{j=1}^m P_{i,j} \log P_{i,j} = \sum_{i=1}^m \sum_{j=1}^m f(Q_{i,j}) \quad (4.46)$$

$$\text{where } f : q \rightarrow \frac{q}{m^2} \log \frac{q}{m^2}$$

Sum Average :

$$\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m iP_{i,j} + jP_{i,j} = \frac{1}{2m^2} \sum_{i=1}^m \sum_{j=1}^m iQ_{i,j} + jQ_{i,j} \quad (4.47)$$

Energy :

$$\sum_{i=1}^m \sum_{j=1}^m P_{i,j}^2 = \frac{1}{m^4} \sum_{i=1}^m \sum_{j=1}^m Q_{i,j}^2 \quad (4.48)$$

Contrast :

$$\sum_{i=1}^m \sum_{j=1}^m (i-j)^2 P_{i,j} = \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m (i-j)^2 Q_{i,j} \quad (4.49)$$

Maximum Probability :

$$\max P_{i,j} = \frac{1}{m^2} \max Q_{i,j} \quad (4.50)$$

Variance :

$$\begin{aligned} \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (i - \mu_r)^2 P_{i,j} + (j - \mu_c)^2 P_{i,j} \\ = \frac{1}{2m^2} \sum_{i=1}^m \sum_{j=1}^m (i - \mu_r)^2 Q_{i,j} + (j - \mu_c)^2 Q_{i,j} \end{aligned} \quad (4.51)$$

Correlation :

$$\sum_{i=1}^m \sum_{j=1}^m \frac{(i - \mu_r)(j - \mu_c)P_{i,j}}{\sqrt{\sigma_r^2 \sigma_c^2}} = \sum_{i=1}^m \sum_{j=1}^m \frac{(i - \mu_r)(j - \mu_c)Q_{i,j}}{m^2 \sqrt{\sigma_r^2 \sigma_c^2}} \quad (4.52)$$

Inverse Difference Moment :

$$\sum_{i=1}^m \sum_{j=1}^m \frac{P_{i,j}}{1 + (i-j)^2} = \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m \frac{Q_{i,j}}{1 + (i-j)^2} \quad (4.53)$$

Cluster Tendency :

$$\sum_{i=1}^m \sum_{j=1}^m (i - \mu_r + j - \mu_c)P_{i,j} = \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m (i - \mu_r + j - \mu_c)Q_{i,j} \quad (4.54)$$

where  $\mu_r$ ,  $\mu_c$ ,  $\sigma_r^2$  and  $\sigma_c^2$  are means and variance for rows and columns.

$$\mu_r = \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m iQ_{i,j} \quad \sigma_r^2 = \sum_{i=1}^m \sum_{j=1}^m (i - \mu_r)^2 P_{i,j} \quad (4.55)$$

$$\mu_c = \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m jQ_{i,j} \quad \sigma_c^2 = \sum_{i=1}^m \sum_{j=1}^m (j - \mu_c)^2 P_{i,j} \quad (4.56)$$



#### 4.5.2.5 Solving the MRF

Solving MRFs is a challenging problem, for which the best optimization method was developed by Komodakis [Komodakis 2008, Komodakis 2007a]. This state of the art optimization technique provides a good solution for MRF problems with a significant speed gain compared to other methods; a magnitude of ten is indeed claimed by the author [Komodakis 2008]. Moreover, this optimization technique is proved to give as good results as the  $\alpha$ -expansion algorithm when the pairwise potentials are metric, which is the case in this study. Komodakis made this optimization method available on the Internet under the name of the FastPD MRF Optimization library, which was retained as solving library for the MRF segmentation [Komodakis 2009b]. For more details, please refer to the previous chapter (sect. 3.5.2.1).

### 4.5.3 Test

The protocol of creation of a classification function was previously detailed given a set bank of features and parameters for the learning process (sect. 4.5.1). Only remains the choice of a training and a test set for a specific problem. Two problems are handled, with different data set. First, the segmentation of any kind of tumor in liver is evaluated for straight segmentation with a classification function. Then, the segmentation of liver colorectal metastases is achieved using a MRF approach.

#### 4.5.3.1 Straight segmentation of tumors

The quality of segmentation is first evaluated for straight segmentation from a classification function (sect. 4.4.2.1). First, a classification function was trained using 15 manually annotated volumes. Then, the segmentation was evaluated on 798 slices.

A training set of 15 volumes that were manually segmented by physicians was chosen. This training set was created while insuring that the volumes were representative of the diversity of the CT images in clinical routine. Indeed, several metastases and the three main types of primary liver cancers were present inside the training set, namely HCC, Adenoma and Cholangiocarcinomas for a total of 56 lesions. Moreover, the CT images were from diverse phases, and various slice thicknesses. . . Diagnoses of the lesions were confirmed by anatomical pathology and reference segmentations carried out or validated by M.D.

Straight segmentation was then applied to a set of 798 slices, containing metastases from diverse primary sites. This straight segmentation was followed by several basic morphologic operations (opening/closure, removal of smallest components) to remove some errors in an automatic process.

The images used during tests were coming from a high number of centers: the volumes were coming from no less than 14 centers. Moreover, many different machines were used during the acquisitions: 10 different models belonging to the 4 main manufacturers (GE, Siemens, Philips and Toshiba). Slice thicknesses in this data set were varying between 1.25 and 5 mm. All segmentations were done or validated by M.D.. The segmentations were done by Dr L. Pupulim, Dr C. Bonnel,

Dr M. Zappa or S. Hemm-Ode. Any segmentation done was finally validated by a M.D.

#### 4.5.3.2 Segmentation of tumors, MRF based

Then, the segmentation of liver colorectal metastases is evaluated using segmentation with MRF (sect. 4.4.2.2). First, a classification function was learned using 20 volumes. Then, 9 new volumes were segmented. And finally, the results were evaluated visually and quantitatively. Both datasets were created in order to have a good representation of the possible cases, namely various slice thicknesses, and volumes with diverse sizes, localizations and numbers of lesions. The images inside both sets were parts of a medical study that retained 5 lesions at most. Thus, no more than five segmentations of lesions were available for any image. This limit might have introduced a slight bias in the test. Indeed, the training set does not contain exams with more than 5 lesions, whereas the test set contains many multi-metastatic cases that are more complex to solve.

For this test, segmentations were done by A. René or Pr. B. Gallix; all segmentations were validated by B. Gallix or a member of his team. All CT images were acquired in portal phase. The pixel spacing for the volumes was varying between 0.6 and 0.8mm, with an average of 0.72mm. The slice thickness was varying between 1 and 5 mm. The CT images were taken from 3 different centers, and with machines from 3 manufacturers (GE, Philips, Toshiba).

First, a classification function was learned on 20 CT volumes. Compared to the previous test, the process was slightly modified. Indeed, three classes of tissues were considered during feature selection and training, lesions, healthy liver, and a rim around each lesion. However, only two classes were kept for selection and training. Indeed, the healthy tissues and the rim of tissues around the lesions were merged during these tasks. This modification was done to prevent incorrect classification around the lesions, which was quite common in some preliminary tests. This change indeed allows giving more weight to the neighborhood of the lesions during the training steps. Thus, the classification on this part improved.

Then, 9 new volumes were segmented using the MRF method. The pairwise cost during segmentation was chosen as constant when neighbor labels are different (eq. 4.40) and this cost was given a weight of 0.1. The segmentations were achieved inside manual ROIs of the liver, and using the 64 bits version of Myrian<sup>®</sup>.

Finally, the quality of segmentation was evaluated visually and quantitatively. First, the results were visually compared to the reference segmentations. Then, several lesions were selected for quantitative evaluation. Indeed, the available measures were limited by the absence of complete references in many cases due to the maximum of five reference segmentations by CT image. Moreover, the multi-metastatic livers often contain confluent lesions, i.e. lesions that are in contact, or sometimes very close. Thus, only disjoint lesions were retained for the evaluation. Morphological openings were sometimes applied in order to disjoin the lesions, but many could not be used due to the existence of confluent lesions. Finally, the quality

of segmentation was evaluated for 18 lesions that could be distinguished from the others. The four metrics previously introduced were retained to quantify the segmentations (sect. 3.5.5.1), as well as some medical measures of the lesions, namely RECIST and WHO criteria along with the volume of the lesions. These medical metrics were evaluated using build in tools inside Myrian<sup>®</sup>.

## 4.6 Discussion, Results

### 4.6.1 Straight segmentation of tumors

Straight segmentation followed by morphology provides good results of segmentation with few changes between tumor types. The method offers good detection rates for larger lesions, but the detection is poor for the smallest lesions. First, the results will be qualitatively discussed. Then, they will be analyzed by quantifying the influence of both the type of tumor and the size of lesion.

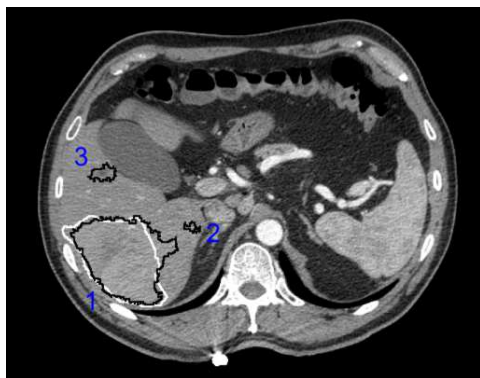
#### 4.6.1.1 Visual results

The results will be first presented using some examples and later discussed for the general case. Straight segmentation from a classification function was applied to a set of images with diverse types of tumors. Three slices were chosen to show the quality of segmentation, by comparison between the automatic segmentation outlined in black and the ground truth shown as a white outline (fig. 4.20).

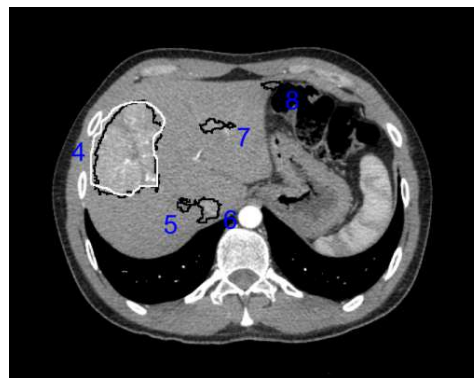
#### Some results

The segmentation of a metastasis in portal phase is first considered (fig. 4.20.a). For this image, the segmentation is similar to the reference one, but with two false positives. The automatic segmentation gives a region (1) that remains close to the ground truth, most of the time with a small defect, apart for the right part of the lesion, where there is an important excess with no anatomical explanation. Then there are two false positives. The smallest one (2) cannot be explained by any other reason than a false positive of the classifier, because there are no visible structures close to this region. However, the medium region (3) contains a hypodense region. This hypodense region exists on several consecutive slices and might be a missed lesion from the reference segmentation. Thus this second *false positive* might actually be a correct detection of one lesion with some segmentation excesses.

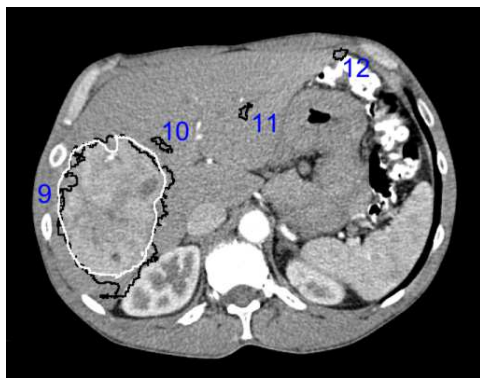
The second example displays the result of segmentation for a Hepatocellular Carcinoma (HCC) in arterial phase (fig. 4.20.b). Five regions including two connected ones are detected as tumoral by the segmentation process, while there is only one lesion segmented by the physician. From left to right, the first region (4) is the one segmented by the physician. Automatic segmentation is close to the ground truth, with a small excess on left part and on the upper right side too. Then, a lesion (5,6) on the bottom of the liver is a false positive but is explained by existing anatomical



(a) Metastasis, portal phase, slice thickness 1.3mm



(b) HCC, arterial phase, slice thickness 1.3mm



(c) Adenoma, arterial phase, slice thickness 2mm

Figure 4.20: Straight segmentation of tumors for a texture size of  $13 \times 13$  without removing the vessels. Ground truth is outlined in white and the result of the segmentation in black. CT images are courtesy from V. Vilgrain's Department (Beaujon, Paris).

structures. This region seems composed of two parts; a left circular part (5) contains a hypodense region that is a biliary duct, and the right part (6) contains an enhanced part of the hepatic network. The next wrongly detected region (7) also matches with a segmentation of vessels, but in this case there are two networks; the portal network is slightly enhanced while the arterial network is displayed with some small hyperdense spots. Finally, a false positive, probably due to the partial volume effect, is located on the upper right part of the liver (8). Because the liver is thin on the end of the liver, the intensity in this part is lower than it should be. Scanner indeed averages the blocking of X-Rays by the tissues inside cubes, thus tissues outside liver, in particular inside lungs, impact the intensities on this region by creating a hypodense region. This is called the partial volume effect and it may explain the incorrect detection as tumoral.

The third example concerns an Adenoma in arterial phase, where the actual

lesion is detected along with three false positive (fig. 4.20.c). The real lesion is well segmented (9) with a small default on the left and some excess on the right and bottom of the region. In particular, the excessive segmentation on bottom is bigger and reaches the boundary of the liver. No anatomical structure can explain this particular behavior. However, it might be caused by the proximity of two structures that make classification difficult: the boundary of the lesion and the boundary of the liver. Indeed, the classification around the boundary of lesions is complex because the texture change is seldom marked between healthy and tumoral tissues. Moreover, specific appearances are sometimes found on the boundary. The difficulty of classification around the liver boundary has a different source. This cause is twofold. First, the intensities on the boundary may be incorrect because of the effect of partial volume. Secondly, the texture patches on liver boundary contain patterns that are outside the liver, and thus were not learned. Consequently, the classification will be strongly affected by the structures outside liver. When these structures have a similar appearance to healthy tissues there is no error. On the opposite when appearance is similar to the tumoral one, the classification will give incorrect results. This last case happens for this lesion, because the segmentation excess is found near the kidney that has an appearance close to some tumors. Three false positive regions were detected in addition to the correct one, probably because of the underlying anatomical structures. The lesion (10) is probably due to a thin planar and hypodense structure inside liver. The lesion (11) seems caused by a biliary duct, and the (12) by a partial volume effect.

### Observation and corrections of segmentation errors

In a more general analysis, defaults seem to occur for small lesions and on boundary of lesions, and excesses seem to occur on liver or lesion boundaries as well as around anatomical structures. First, the defaults/excess on boundary of lesions will be analyzed. Then, excesses around other anatomical structures will be studied. Finally, the link between the size of lesions and the accuracy of segmentation will be studied in a subsequent section (sect. 4.6.1.3).

The defaults or excesses of segmentation on boundaries of lesions come from a same source, the fickleness of appearance on this boundary. Lesion boundaries indeed contain distinct structures (capsule, calcifications...) but may also gradually pass from tumoral to healthy. This behavior may be corrected with another classification function where more weight is given to texture samples on such boundaries. However such an approach does not seem to be the most efficient one, because it amounts to training a classifier using mostly unusual cases, which is the opposite of the aim of machine learning techniques. A subtler approach would be to use the local information to correct these defaults/excesses in order to take advantage of well classified voxels. This was done with the introduction of MRFs (sect. 4.4.2.2).

The excesses around other anatomical structures have two main sources that may be combined. First, the anatomical structures may have a similar appearance to that of lesions, for example for vessels or biliary ducts. Then, the appearance may be too

different from the usual healthy appearance, as for problems with partial volumes or on liver boundaries. These excesses may be corrected either with pre- or post process. Indeed, these structures might be segmented and removed as preprocess, or false positive that cannot be tumors could be removed in a post-process.

The segmentation of anatomical structures that are often mistaken for tumors allows avoiding any problem around these structures by removing them. However, this approach would add another constraint to the proposed method by propagating the errors of the initial segmentation to the segmentation of tumors. Besides, this additional segmentation will require some time. This approach might be nevertheless considered for structures that contribute to the diagnostic and therapeutic choice. The value of this approach was evaluated for the case of liver vessels. An initial segmentation of the vessels induce a slight decrease of the number of false positive but this improvement remains small (fig. 4.21). This step is indeed useless for the classification of the neighborhood of vessels, because tumors are often close to vessels and even growing new ones.

The removal of the regions that are false positive subsequently to the segmentation of tumors is another possible approach. Many anatomical structures wrongly detected as tumors have indeed a very specific shape; for example false positive on liver boundaries are thin, often look like sheets of paper, false positive on vessel or biliary duct have a tubular shape. . . These previous shapes cannot in general happen for tumors. However, the tumors inside the vessels may have these shapes (epitheloid hemangioendotheliomas), but they are rare and chances are they will not be detected in the first place. The removal of some false positive may thus be done by removing regions whose shapes are not common for tumors. This characterization by the shape was done in several papers [Soler 2001, Choudhary 2008].

#### 4.6.1.2 Quantifying the results

The overall results are quite good, and with slight differences between the tumor types (fig. 4.21). The best results are obtained for less common tumor types but it may not be truly significant because the test set contains few slices of these types. Then, better results of segmentation are obtained for metastases than for HCC. This is no surprise, because HCC are visually more difficult to distinguish. However, the difference of quality between these two types of tumors remains small and in both cases sensitivity and specificity remain high (84-87% for sensitivity and 79-81% for specificity).

Segmenting vessels in a first step to prevent a common case of false positive allows improving the specificity of the segmentation, with no impact on sensitivity. Indeed, the vessels are often mistaken for tumors, thus removing these vessels from the problem was an obvious move for improvement. This removal was done using approaches described by Selle and Soler, but with manual corrections when the approach did not provide correct segmentations [Selle 2002, Soler 2001, Soler 1998, Pescia 2006]. The removal of the vessels inside the liver improved slightly the specificity of the classification, but this improvement remained small (fig. 4.21). The small level of

	Whole Set	HCC	Metastases	Other
Sensitivity	86%	84%	87%	93%
Specificity	80%	79%	81%	84%
Specificity (with segmented vessels)	82%	80%	84%	84%

Figure 4.21: Segmentation results obtained with the same classifier for diverse tumors. The segmentation of the vessel networks as a first step does not modify the sensitivity, whereas specificity improves.

improvement may be explained because the preliminary segmentation cannot improve the classification close to the vessels, in particular when lesions and vessels are in contact. Unfortunately, this detrimental case often happens as tumors are often close to vessels and even grows new ones. Thus, the contribution of vessel segmentation is smaller than first expected and introduces an additional risk. The errors of this initial segmentation will indeed impact the final results.

#### 4.6.1.3 Impact of lesion size

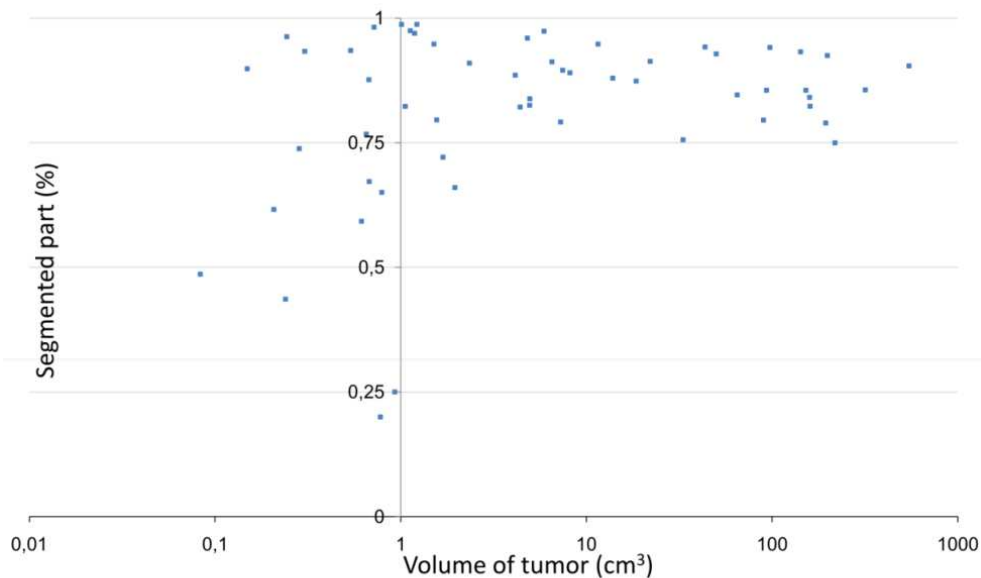


Figure 4.22: Percentage in volume of tumors accurately segmented, functions of the volume of the lesions. The volumes of the texture patches span between 0.61 and 3.05 cm<sup>3</sup>, with an average volume of 1.54 cm<sup>3</sup>.

The volume of the tumor impacts the quality of the segmentation, but two worthwhile facts should be noted. First, when tumors are bigger than the size of the texture, the detection rate is quickly optimal. Then, even tumors smaller than

the patches are partially segmented, but the accuracy of the segmentation is lower and more random (fig. 4.22).

A poor detection of the lesions of small size was expected as a consequence of the classification method. The classification indeed relies on the texture information contained inside patches of set size. Therefore, the classification cannot give good results when the lesions are smaller than the patches, because the patches do not contain only tumoral tissues.

The limitation of the approach due to the size of patches did not impact the detection of lesions of medium/small volume. Poor detection due to the presence of few tumoral voxels inside patches might have spread from small lesions to medium lesions, because these last lesions contain relatively many patches on boundaries. However, the approach proved able to deal well with these medium lesions.

#### 4.6.2 Segmentation of tumors using MRFs

The quality of liver tumor segmentation using MRFs will now be evaluated for liver colorectal metastases. As first observation, one may note that the speed constraint of the approach was respected, as segmentation was achieved in 2-3min for inframillimetric images on a core 2Duo 2.8GHz CPU. First, qualitative results will be presented using several examples. Then, the quality of segmentation will be quantified.

##### 4.6.2.1 Visual observations

The segmentation seems globally correct, *prima facie*. However, the segmentations show some common defects. First, the boundary of many lesions is not well located. Then, the region between two lesions is often wrongly detected as tumoral tissue. Next, some long and thin false positives are obtained on some boundaries of the liver. Finally, huge lesions are only partially detected.

*Prima facie*, the lesions are well detected. However, the boundaries of the lesions are sometimes offset compared to the ground truth. Indeed, the lesions are often well defined, for example in (fig. 4.23.c-f). However, with a closer look, a gap is sometimes found between automatic and reference segmentations. For example for the previous volumes, the test ROI is bigger than the reference in the first case (fig. 4.23.c,d), and so are the two lesions on the right for the second one (fig. 4.23.e,f). The difference between the lesions is nevertheless not always like that; the test lesions are sometimes oversegmented compared to the references.

The region between two lesions is also often wrongly set as tumoral. This behavior has two main causes, but may be negligible in clinical routine. First, the lesions induce changes of appearance around them. Thus, the neighborhood of the lesions is difficult to classify, which was the reason behind the learning process with three classes of tissues. Consequently, the neighborhood between two lesions is doubly abnormal, and hence is more difficult to classify. Moreover, this configuration was rarely found inside the training set because of the limit of 5 lesions per exam. Then,



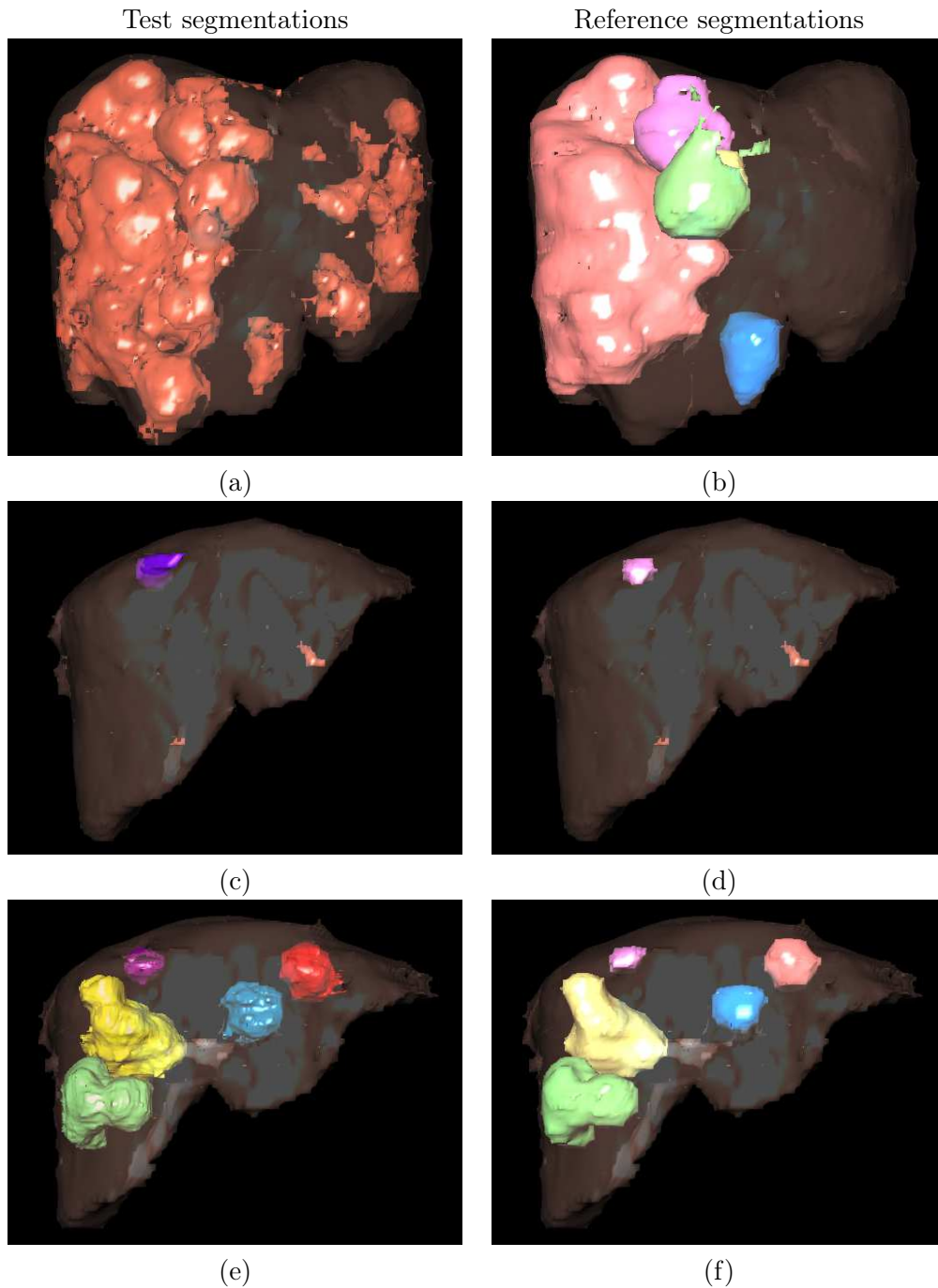


Figure 4.23: Segmentation of colorectal metastases using MRFs. Results of segmentation and references are given for three volumes. For the first volume, all ROIs obtained by our approach are displayed (a), while only 5 are shown as reference (b). On the second volume is only shown a single non confluent lesion (c,d). For the last volume, the 5 lesions chosen in the medical protocol are shown (f), along with the same lesions for the automatic approach (e). The ROIs of the lesions are displayed with surface rendering using Myrian<sup>®</sup>.

this behavior is amplified by the regularization term of the MRF. Indeed, the voxels between two lesions are difficult to classify, and thus their likelihoods for each class are almost equals. Therefore, the regularization term becomes predominant for the choice of a label, which is impacted by the proximity with two lesions. Finally, the aforementioned problem may be negligible for clinical routine. Indeed, many cases of confluent lesions are found in multi metastatic livers, for which a precise evaluation lesion by lesion has small value. However, the problem should not be ignored because this particular case is valuable for the evaluation of treatments. Moreover, this configuration may still happen with few lesions.

Long and thin false positive are sometimes found at the surface of the liver envelope. These false positive are probably due to the special appearance that sometimes happens on liver boundary, and which is not well learned. However, the correction of such problems should not be done with a change of training set, but would be better handled as a post process. Indeed, a change of training set induces a risk for the entire classification process. Moreover, this problem was handled by several authors in the liver case; Moltz proposes a method to remove the vessels in contact with lesions [Moltz 2008], and Choudhary characterizes the lesions with shape constraints [Choudhary 2008]. The first approach seems more valuable as it would allow dealing with lesions on the boundary of the liver that sometimes wrongly include the tissues between the lesion and the liver boundary.

Finally, huge lesions are not well segmented. Indeed, these lesions often contain many regions with necrosis that are well segmented, while the tumoral tissues between the necrotic zones are detected as healthy. For example, in the first volume (fig. 4.23.a,b) the automatic segmentation for the lesion on the left is more globular and slightly smaller than the reference one. This behavior comes from the intensity normalization that does not handle cases with large differences compared to the usual case. For example, the relative volumes of healthy and tumoral tissues are opposite to the usual case for this previous volume.

#### 4.6.2.2 Evaluation for several lesions

18 lesions were retained to evaluate the quality of segmentation. The selection of these lesions was not driven by a personal choice, but imposed by the non confluent lesions in the volumes. First, the segmentation will be evaluated using metrics from Computer Vision. Then, several metrics used in the medical domain will be compared for reference and automatic segmentation.

Quantitatively, the proposed approach compares with state of the art methods. Indeed, an overlap error of 38-40% is obtained depending on the inclusion of the smaller lesions ( $\geq 3\text{cm}^3$ ). With such values, the proposed method is the third best one when comparing with the automatic methods inside MICCAI segmentation challenge. However, a direct comparison of the results is of little value because the type of the lesions and the exams are very different.

The quality of segmentation was also evaluated for several medical measures, the volume of the lesions, RECIST, and the WHO criteria. First, the difference

	Sensitivity	DSC	Jaccard
All lesions	$0.76 \pm 0.03$	$0.74 \pm 0.01$	$0.60 \pm 0.02$
Lesions $\geq 3\text{cm}^3$	$0.75 \pm 0.02$	$0.76 \pm 0.01$	$0.62 \pm 0.02$

Figure 4.24: Segmentation of colorectal metastases using MRFs, results. The quality of segmentation is evaluated using measures from Computer Vision with the exception of specificity that is non significant due to the relative volumes of lesions and images. These measures are given for all lesions, or for all lesions  $\geq 3\text{cm}^3$ .

of the evaluation of volumes is too big to be useful in clinical routine. Indeed, an average variation of 30% is obtained between test and reference segmentations. Then, the current measures for the evaluation of tumors still vary, but less than for the evaluation of the volume. Indeed, the variation of the longer dimensions is around 20%. Moreover, a significant improvement is obtained when excluding the smaller lesions for the evaluation of RECIST or the first dimension of the WHO criteria. Indeed, this variation decreases from 19% to 14%.

	Volume	RECIST	WHO second axis
All lesions	$0.30 \pm 0.08$	$0.19 \pm 0.03$	$0.20 \pm 0.03$
Lesions $\geq 3\text{cm}^3$	$0.22 \pm 0.03$	$0.14 \pm 0.01$	$0.21 \pm 0.02$

Figure 4.25: Segmentation of colorectal metastases using MRFs, results. The quality of segmentation is evaluated using several medical metrics for the evaluation of solid tumors. These measures are given for all lesions, or for all lesions  $\geq 3\text{cm}^3$ .

## 4.7 Conclusion, Summary

A method for the segmentation of the liver tumors within a liver envelope was proposed. First, a complete technique was defined for this segmentation. Indeed, features were introduced that are later selected and combined for the classification of tissues, before being applied for segmentation. Then, the proposed technique offers quickly relatively good results. Finally, the proposed approach is not limited to the liver tumor, but applies to other problems

A complete framework for the segmentation of the liver tumors within a liver envelope was introduced. First, texture features were defined as a combination between texture descriptors and filters. Then, a heuristic method was proposed to select the most relevant features from a bank of features, while taking in account both the contribution and the computational cost of each feature. Next, a fast and accurate machine learning technique was chosen to combine these features towards an optimal separation of tumoral and healthy tissues in a feature space. Finally, two methods were proposed for the segmentation. The first one simply applies the classification function to an image in order to obtain a binary classification of this image. And the second retains the reliability of the classification and introduces

spatial constraints inside a MRF expression of the problem in order to obtain optimal solutions in a short length.

While the results of segmentation are difficult to compare with other methods due to the variability of the image sets, the proposed method offers comparable results to other methods. Moreover, the segmentation of the lesions does not take more than a few minutes for inframillimetric images. However, the segmentations have several defaults. First, the huge lesions are only partially segmented. However, these cases have small medical value, and this problem might be addressed when using a more reliable normalization of intensities. Then, the segmentation is often in excess between lesions, or on the boundary of the liver. These problems might be solved by imposing shape constraints on the lesions, or with post-treatments aiming at removing obvious false positives and dividing lesions from other structures.

The proposed method relies on texture features and filters to classify the tissues through machine learning techniques in a noisy medium. The whole process presented in this chapter is limited to the liver only by the choice of the filters and the chosen settings. Thus, this method could apply to other problems in medical imaging, where the structures of interest differ from the surrounding tissues by texture changes. Moreover, the amplitude of the noise the method has to deal with in the liver case should make the method robust for other organs.



# Conclusion, Future Work

---

To summarize, a complex problem of medical imaging has been dealt with in this study, the segmentation of the liver tumors. First, the contributions of this work will be presented. And then, future improvement of the proposed techniques will be detailed.

## 5.1 Contributions

The main contribution of this thesis is a framework to segment complex anatomical structures that differ from surrounding tissues by texture changes. Then, a new definition of statistical atlas was proposed, and its registration was handled. Finally, an image processing tool was developed, which computes a smooth ROI that wraps around another ROI.

The main contribution of this study is a fast method to segment complex structures that differ by texture change in a noisy medium. This method divides in several steps. First, the proposed method includes the definition of texture features as the combination of texture descriptors and filters. Then, the selection of relevant features from a bank of features is dealt with in a heuristic approach while considering the relevance, and the computational cost of the features. Next, a classification function is computed using a state of the art machine learning technique. The segmentation is later achieved directly with the classification function, or while considering both the reliability of the classification and the expected classes of the neighbor voxels. Finally, the whole segmentation algorithm is optimized in order to have small computational complexity. Therefore the proposed approach is fast. Moreover, the method was applied with some success to the segmentation of the liver tumors.

Then, a statistical atlas was proposed along with its registration on image volumes. First, a generalization of the statistical atlas proposed by Glocker was proposed. This new definition uses optimal Gaussian Mixture Models to model intensity distributions in the space. Moreover, the proposed definition is not limited to GMM, but also applies to other distribution models. Furthermore, the optimality condition even applies to multiple models at same time. Thus, the proposed atlas may use any intensity models equally. Then, a fast registration technique for this atlas definition was proposed in a coarse-to-fine manner. This registration was achieved by minimizing a MRF energy

Finally, an image processing tool was proposed, which computes a smooth envelope that wraps around a binary ROI. While only partially implemented, this

approach offers promising results, and may potentially apply to many other problems as post process. Indeed, this low level tool allows detecting and filling cavities that break the smoothness of a ROI surface. As structures with a smooth contour are common in medical imaging, this approach could apply to many other organs, such as heart, spleen, vessels. . . The proposed approach use a representative mesh for a binary ROI, and artificially introduces holes on this mesh where cavities are found inside the binary ROI. Then, these holes are filled by continuity, following principles from techniques of digital reconstruction.

## 5.2 Future work

As future work, the proposed techniques will be improved and validated for other problems. First, the validation on other problems will not be further developed, because these tests depend on the availability of databases for other problems. The proposed segmentation method will nevertheless be applied to the segmentation of lesions inside the brain. Then, several possible improvements have been considered for the wrapping tool, and the segmentation framework. Because the statistical atlas was abandoned for the liver, no future development of the related methods is planned yet.

First, no radical modification of the hole filling process is proposed. However, some slight modifications are planned for the detection of hole boundaries. Moreover, the smooth fill will be finished and evaluated. First, some modifications are planned in order to better detect the hole boundaries. These modifications include the improvement of the initial contour through better sampling of points, and the use of better heuristics during the detection of hole boundaries with maybe the introduction of machine learning techniques. Then, the holes will be filled by continuity with the direction of the contour on both sides of a hole, as this was proposed before but not yet evaluated.

Then, most of the future work focuses on the proposed segmentation framework. First, the speed of the method will be increased by simplifying some filters. Then, several steps of the proposed method will be modified in order to correct some lacks of the current approach. Finally, several additions might contribute to the value of the method.

During the segmentation, most of the running time comes from the filters and the computation of texture descriptors, while the other tasks are significantly less time consuming. In particular, median filters and Nagao's filters respectively account for 15% and 40% of the total computation time for millimetric images. Thus, a recent algorithm by Perreault *et al.* could induce a significant speed gain. Indeed, Perreault proposed an algorithm that computes the median filter in constant time [Perreault 2007]. This algorithm could induce significant gains for the median filters due to the sizes of the neighborhoods that are used to compute the features. Moreover, the proposed approach might also adapt to Nagao's filters, which could offer even higher speed gains.

Several weaknesses have been found in the current approach. First, the normalization of the intensities is not optimal, and then the robustness of the feature selection might be improved. First the normalization of the intensities will be improved. Two possible methods are envisaged, either with a special process to handle the different enhancement phases as was done by Linguraru and Nugroho [Linguraru 2010, Nugroho 2008], or by using multiple intensity models. Indeed, several intensity models have been proposed for the intensity distribution within the liver, either with 1, 2, or 3 modes that correspond to hypodense lesions, healthy parenchyma, hyperdense lesions and vessels [Soler 1998, Qi 2008, Nugroho 2008]. It might be possible to evaluate the number of modes of the intensity distribution for any volume, before using an adequate intensity model for the normalization. Moreover, one may imagine selecting an adequate classification function depending on the intensity profile. Then, the robustness of the feature selection might be improved by using k-fold cross validation. Indeed, the measures of the relevance of every feature would become more reliable when many training and test rounds are done.

Several evolutions of the methods have been retained in order to improve the segmentation framework. First, new texture features might be introduced. Then, the regularization of the MRF segmentation might be modified in order to impose a shape constraint. First, 3D texture features might improve the detection, and allow working without the axial constraint for the images. Indeed, Tesar *et al.* extended the Haralick's descriptors to 3D, and proved that this extension significantly improves the classification of abdominal tissues and in particular the liver tissues [Tesar 2008]. Moreover, this modification could allow getting rid of the constraint of axial images, because the features could be chosen as invariant by rotation. Then, a shape constraint might be imposed as a regularization term during the segmentation. Imposing such constraint might prevent merging close lesions. Indeed, close lesions are often merged through a small region. With a shape constraint, the lesions might be distinguished, which would simplify the use of their segmentations. Moreover, a shape constraint might also have some impact on the liver boundary, because the false positives at these locations do not have a usual tumor shape.





# Bibliography

- [Adams 2005] Leon A. Adams, Paul Angulo and Keith D. Lindor. *Nonalcoholic fatty liver disease*. Canadian Medical Association Journal, vol. 172(7), pages 899–905, 2005. 20
- [Andronache 2006] Adrian Stefan Andronache. *Multi-Modal Non-Rigid Registration of Volumetric Medical Images*. PhD thesis, Swiss Federal Institute of Technology Zurich, 2006. 78
- [Ashraf 2006] Mohamed Ashraf, Evangelia I. Zacharaki, Dinggang Shen and Christos Davatzikos. *Deformable registration of brain tumor images via a statistical model of tumor-induced deformation*. Medical Image Analysis, vol. 10, pages 752–763, 2006. 125
- [Bakir 2004] Gökhan.H. Bakir, Jason Weston and Bernhard Schölkopf. *Learning to Find Pre-Images*. Advances in Neural Information Processing Systems, vol. 16, pages 449–456, 2004. 138
- [Barron 1998] A. Barron, J. Rissanen and B. Yu. *The minimum description length principle in coding and modeling*. IEEE Transactions on Information Theory, vol. 44, no. 6, pages 2743–2760, 1998. 86
- [Ben-Dan 2008] Itay Ben-Dan and Elior Shenhav. *Liver Tumor segmentation in CT images using probabilistic methods*. In MICCAI Workshop, 2008. 41, 43
- [Binder-Foucard 2007] Florence Binder-Foucard, Michel Dofföel and Michel Velten. *Epidemiology of hepatocellular carcinoma in Bas-Rhin: analysis of all incident cases from 1990 to 1999*. Gastroentérologie Clinique et Biologique, vol. 31(10), pages 838–843, 2007. 26
- [Birnbaum 2007] Bernard A. Birnbaum, Nicole Hindman, Julie Lee and James S. Babb. *Multi-Detector Row CT Attenuation Measurements: Assessment of Intra- and Interscanner Variability with an Anthropomorphic Body CT Phantom*. Radiology, vol. 242(1), pages 109–119, 2007. 46
- [Blum 2002] A. Blum. *Scanographie volumique multicoupe : Principes, applications et perspectives*. Masson, 2002. xiii, 24
- [Blum 2005] H.E. Blum. *Hepatocellular carcinoma: Therapy and prevention*. World J. of Gastro., vol. 11(47), pages 7391–7400, 2005. 21, 28, 131
- [Boser 1992] B.E. Boser, I.M. Guyon and V.N. Vapnik. *A Training Algorithm for Optimal Margin Classifiers*. In COLT '92: Proceedings of the fifth annual workshop on Computational learning theory, pages 144–152, 1992. 136

- [Bovik 1990] A.C. Bovik, M. Clark and W.S. Geisler. *Multichannel texture analysis using localized spatial filters*. IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12(1), pages 55–73, 1990. 156
- [Boykov 2001a] Y. Boykov, O. Veksler and R. Zabih. *Fast Approximate Energy Minimization via Graph Cuts*. IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, no. 11, pages 1222–1239, November 2001. 98, 101
- [Boykov 2001b] Yuri Boykov and Marie-Pierre Jolly. *Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images*. In Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on, volume 1, pages 105–112, 2001. xvi, 97, 98, 100, 101
- [Boykov 2006] Yuri Boykov and Gareth Funka-Lea. *Graph Cuts and Efficient N-D Image Segmentation*. IJCV, vol. 70(2), pages 109–131, 2006. 97
- [Bruix 2005] J. Bruix and M. Sherman. *Management of hepatocellular carcinoma*. Hepatology, vol. 42(5), pages 1208–1236, 2005. 23, 26, 27, 28, 32, 33, 34, 131, 132
- [Casciola 2005] G. Casciola, D. Lazzaro, L.B. Montefusco and S. Morigi. *Fast surface reconstruction and hole filling using positive definite radial basis functions*. Numerical Algorithms, vol. 39, pages 289–305, 2005. 67
- [Chemouny 1999] Stéphane Chemouny, Henri Joyeux, Bruno Masson, Frédéric Borne, Marc Jaeger and Olivier Monga. *Advanced 3D image processing techniques for liver and hepatic Tumors location and volumetry*. SPIE Conference on Image Processing, vol. 3661, pages 761–771, 1999. 60, 160
- [Chemouny 2001] Stéphane Chemouny. *Filtrage et segmentation d’images tridimensionnelles : Application à la détection et à la caractérisation des structures anatomiques et pathologiques du foie*. PhD thesis, CIRAD, 2001. 60, 165
- [Chen 1998] C. H. Chen and L. F. Pau. The handbook of pattern recognition and computer vision. World Scientific Publishing, 1998. 148
- [Choi 2000] Yongchoel Choi and Seungyong Lee. *Injectivity Conditions of 2D and 3D Uniform Cubic B-spline Functions*. Graphical Models, vol. 62, no. 6, pages 411–427, 2000. 106
- [Choi 2006] J. Choi. *Imaging of Hepatic Metastases*. Cancer Control, vol. 13(1), pages 6–12, 2006. 10, 29, 31, 33, 58, 131
- [Choudhary 2008] Anirudh Choudhary, Nicola Moretto, Francesca Pizzorni Ferrarese and Giulia A. Zamboni. *An Entropy Based Multi-Thresholding Method for Semi-Automatic Segmentation of Liver Tumors*. In MICCAI Workshop, 2008. 41, 43, 94, 191, 195

- [Ciecholewski 2007] Marcin Ciecholewski and Marek R. Ogiela. *Automatic Segmentation of Single and Multiple Neoplastic Hepatic Lesions in CT Images*. In IWINAC '07: Proceedings of the 2nd international work-conference on Nature Inspired Problem-Solving Methods in Knowledge Engineering, pages 63–71, 2007. 41, 44, 48
- [Clark 2005] Hollins P. Clark, W. Forrest Carson, Peter V. Kavanagh, Coty P. H. Ho, Perry Shen and Ronald J. Zagoria. *Staging and Current Treatment of Hepatocellular Carcinoma*. RadioGraphics, vol. 25, pages S3–S23, 2005. 23, 24, 26, 28, 33, 34, 131
- [Clausi 2000] D.A. Clausi and M.E. Jernigan. *Designing Gabor filters for optimal texture separability*. Pattern Recognition, vol. 33, pages 1835–1849, 2000. 156, 157
- [Couinaud 1957] C. Couinaud. *Le foie : études anatomiques et chirurgicales*. Masson Edition, 1957. 18
- [Couinaud 1999] Claude Couinaud. *Tell me more about liver anatomy*. Paris , C. Couinaud, 1999. 32
- [Daniels 2008] J. Daniels, T. Ochotta, L.K. Ha and C.T. Silva. *Spline-based feature curves from point-sampled geometry*. Visual Comput, vol. 24, pages 449–462, 2008. 66
- [Dasgupta 2008] Sanjoy Dasgupta. *The hardness of k-means clustering*. Technical report, University of California, San Diego, Department of Computer Science and Engineering, 2008. 83
- [Demetriou 2004] A.A. Demetriou, R.S. Jr Brown, R.W. Busuttil, J. Fair, B.M. McGuire, P. Rosenthal, J.S. II Am Esch, J. Lerut, S.L. Nyberg, M. Salizzoni, E.A. Fagan, B. de Hemptinne, C.E. Broelsch, M. Muraca, J.M. Salmeron, J.M. Rabkin, H.J. Metselaar, D. Pratt, M. De La Mata, L.P. McChesney, G.T. Everson, P.T. Lavin, A.C. Stevens, Z. Pitkin and B.A. Solomon. *Prospective, Randomized, Multicenter, Controlled Trial of a Bioartificial Liver in Treating Acute Liver Failure*. Annals of Surgery, vol. 239(5), pages 660–670, 2004. 131
- [Dempster 1977] A.P. Dempster, N.M. Laird and D.B. Rubin. *Maximum Likelihood from Incomplete Data via the EM Algorithm*. Journal of the Royal Statistical Society: Series B (Methodological), vol. 39, no. 1, pages 1–38, 1977. 84
- [Eisenhauer 2009] E.A. Eisenhauer, P. Therasse, J. Bogaerts, L.H. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, M. Mooney, L. Rubinstein, L. Shankar, L. Dodd, R. Kaplan, D. Lacombe and J. Verweij. *New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1)*. European Journal of Cancer, vol. 45, pages 228–247, 2009. 36, 39, 132, 133

- [El Serag 1999] H.B. El Serag and A.C. Mason. *Rising incidence of hepatocellular carcinoma in the United States*. N Engl J Med, vol. 340, pages 745–750, 1999. 26, 131
- [Ernst 2005] O. Ernst. *Imagerie du foie, des voies biliaires et du pancréas*. Masson, Imagerie Médicale Diagnostic, 2005. 20, 21, 23, 26, 29, 30
- [Fernández-de Manuel 2009] L. Fernández-de Manuel, J.L. Rubio, M.J. Ledesma-Carbayo, J. Pascau, J.M. Tellado, E. Ramón, M. Desco and A. Santos. *3D liver segmentation in preoperative CT images using a levelsets active surface method*. In Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE, pages 3625–3628, 2009. 41, 159
- [Freeman 1974] Herbert Freeman. *Computer processing of line-drawing images*. Computing Surveys, vol. 6(1), pages 57–97, 1974. 69
- [Freund 1997] Y. Freund and R.E. Schapire. *A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting*. Journal of Computer and System Sciences, vol. 55(1), pages 119–139, 1997. 134, 137, 139, 143
- [Friedman 2000] J. Friedman, T. Hastie and R. Tibshirani. *Additive logistic regression: a statistical view of boosting*. Annals of Statistics, vol. 28, no. 2, pages 337–407, 2000. 143, 175
- [Fritsch 2003] Helga Fritsch and Wolfgang Kühnel. *Atlas de poche d’anatomie : 2. les viscères*. Flammarion Médecine Science, 2003. 20, 130
- [Gabor 1946] D. Gabor. *Theory of Communication*. Journal of the Institute of Electrical Engineers, vol. 93(26), pages 429–457, 1946. 156
- [Glatard 2004] T. Glatard, J. Montagnat and I. Magnin. *Texture based medical image indexing and retrieval: application to cardiac imaging*. In Proceedings of ACM Multimedia 2004, workshop on Multimedia Information Retrieval (MIR), pages 135–142, 2004. 156, 157
- [GLOBOCAN 2002] GLOBOCAN. *International Agency for Research on Cancer (IARC-WHO)*. CANCERmondial Website, 2002. <http://www-dep.iarc.fr/>. xiii, 25
- [Glocker 2007a] B. Glocker, N. Komodakis, N. Paragios, C. Glaser, G. Tziritas and N. Navab. *Primal/Dual Linear Programming and Statistical Atlases for Cartilage Segmentation*. In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2007), pages 536–543, 2007. 75, 76, 80, 87, 96
- [Glocker 2007b] B. Glocker, N. Komodakis, N. Paragios, G. Tziritas and N. Navab. *Inter and Intra-Modal Deformable Registration: Continuous Deformations*

- Meet Efficient Optimal Linear Programming*. In Information Processing in Medical Imaging, 2007. 80, 101
- [Glocker 2008] B. Glocker, N. Komodakis, G. Tziritas, N. Navab and N. Paragios. *Dense Image Registration through MRFs and Efficient Linear Programming*. Medical Image Analysis, vol. 12, no. 6, pages 731–741, 2008. 80, 96, 101, 119
- [Glocker 2009] Ben Glocker. *Drop Registration library*. Website, 2009. <http://www.mrf-registration.net/index.html>. xv, xvi, 77, 88, 94, 110
- [Gomaa 2008] Asmaa Ibrahim Gomaa, Shahid A Khan, Mireille B Toledano, Imam Waked and Simon D Taylor-Robinson. *Hepatocellular carcinoma: Epidemiology, risk factors and pathogenesis*. World Journal of Gastroenterology, vol. 14(27), pages 4300–4308, 2008. 21, 27
- [Häme 2008] Yrjö Häme. *Liver Tumor Segmentation Using Implicit Surface Evolution*. In MICCAI Workshop, 2008. 41, 43
- [Haralick 1973] R. Haralick, K. Shanmugam and I. Dinstein. *Textural features for image classification*. IEEE Transactions Systems on Man and Cybernetics, vol. 3(6), pages 610–621, 1973. 150
- [Heimann 2006] Tobias Heimann, Ivo Wolf and Hans-Peter Meinzer. *Active shape models for a fully automated 3D segmentation of the liver - an evaluation on clinical data*. In Medical Image Computing and Computer-Assisted Intervention - MICCAI, pages 41–48, 2006. 61, 62
- [Heimann 2007] T. Heimann, S. Münzing, H.-P. Meinzer and I. Wolf. *A Shape-Guided Deformable Model with Evolutionary Algorithm Initialization for 3D Soft Tissue Segmentation*. In Information Processing in Medical Imaging, pages 1–12, 2007. 61, 62
- [Heimann 2009] T. Heimann, B. van Ginneken, M.A. Styner, Y. Arzhaeva, V. Aurich, C. Bauer, A. Beck, C. Becker, R. Beichel, G. Bekes, F. Bello, G. Binnig, H. Bischof, A. Bornik, P. Cashman, Ying Chi, A. Cordova, B.M. Dawant, M. Fidrich, J.D. Furst, D. Furukawa, L. Grenacher, J. Hornegger, D. Kainmuller, R.I. Kitney, H. Kobatake, H. Lamecker, T. Lange, Jeongjin Lee, B. Lennon, Rui Li, Senhu Li, H.-P. Meinzer, G. Nemeth, D.S. Raicu, A.-M. Rau, E.M. van Rikxoort, M. Rousson, L. Rusko, K.A. Saddi, G. Schmidt, D. Seghers, A. Shimizu, P. Slagmolen, E. Sorantin, G. Soza, R. Susomboon, J.M. Waite, A. Wimmer and I. Wolf. *Comparison and Evaluation of Methods for Liver Segmentation From CT Datasets*. IEEE Transactions on Medical Imaging, vol. 28(8), no. 8, pages 1251–1265, 2009. 61
- [Hélénon 2003] O. Hélénon and P. Biclet. *Imagerie des tumeurs du foie*. Elsevier, 2003. 10, 20, 25, 26, 27, 28, 29, 30, 31, 32, 33, 58, 131

- [HepaLife Technologies 2009] Inc. HepaLife Technologies. *HepaLife Plans for Phase III Clinical Trial for HepaMate Bioartificial Liver System*. Press Release, 2009. [http://www.hepalife.com/press\\_releases/20090217.html.php](http://www.hepalife.com/press_releases/20090217.html.php). 131
- [Herdilou 1999] Sandra Herdilou. *Caractérisation tissulaire par l'analyse de texture*. PhD thesis, Groupe de Recherche sur l'Analyse Multimodale de la Fonction Cérébrale, 1999. 147
- [Honda 1993] Hiroshi Honda, K. Ochiai, E. Adachi, K. Yasumori, T. Hayashi, A. Kawashima, T. Fukuya, M. Gibo, T. Matsumata, M. Tsuneyoshi and K. Masuda. *Hepatocellular carcinoma: correlation of CT, angiographic, and histopathologic findings*. *Radiology*, vol. 189(3), pages 857–862, 1993. xiii, 27
- [Ising 1925] Ernst Ising. *Beitrag zur Theorie des Ferromagnetismus*. *Zeitschrift für Physik*, vol. 31, pages 253–258, 1925. 96
- [Jolly 2008] L. Jolly M.-P. Grady. *3D GENERAL LESION SEGMENTATION IN CT*. In *Proceedings of the 2008 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 796–799, 2008. 42, 43
- [Jones 1987] J. P. Jones and L. A. Palmer. *An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex*. *Journal of Neurophysiology*, vol. 58, pages 1233–1258, December 1987. 156
- [Joyeux 2003] H. Joyeux, J. Berticelli, B. Chemouny S.and Masson and P. Borianne. *Mesure semi-automatique des différents lobes hépatiques. Application à la recherche d'une corrélation entre volumes des lobes du foie. Une étude sur 50 angioscanners de foies sains*. *Annales de chirurgie*, vol. 128, pages 251–255, 2003. 54, 58
- [Julesz 1981] Bela Julesz. *Textons, the elements of texture perception, and their interactions*. *Nature*, vol. 290, pages 91–97, 1981. 148
- [Kainmüller 2007] Dagmar Kainmüller, Thomas Lange and Hans Lamecker. *Shape Constrained Automatic Segmentation of the Liver based on a Heuristic Intensity Model*. In *MICCAI Workshop. 3D Segmentation in the Clinic: A Grand Challenge*, pages 109–116, 2007. 62
- [Kindermann 1980] Ross Kindermann and J. Laurie Snell. *Markov random fields and their applications*. American Mathematical Society, 1980. 96, 97
- [Knuth 1973] Donald E. Knuth. *The art of computer programming, volume 3: Sorting and Searching*. Addison-Wesley, 2 édition, 1973. 181
- [Kolmogorov 2004] V. Kolmogorov and R. Zabih. *What Energy Functions Can Be Minimized via Graph Cuts?* *PAMI*, vol. 26, no. 2, pages 147–159, february 2004. 97, 98

- [Komodakis 2007a] N. Komodakis and G. Tziritas. *Approximate Labeling via Graph-Cuts Based on Linear Programming*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 8, pages 1436–1453, 2007. 98, 186
- [Komodakis 2007b] Nikos Komodakis, Georgios Tziritas and Nikos Paragios. *Fast, Approximately Optimal Solutions for Single and Dynamic MRFs*. In IEEE Computer Vision and Pattern Recognition, pages 1–8, 2007. 101, 119
- [Komodakis 2008] Nikos Komodakis, Georgios Tziritas and Nikos Paragios. *Performance vs Computational Efficiency for Optimizing Single and Dynamic MRFs: Setting the State of the Art with Primal Dual Strategies*. Computer Vision and Image Understanding, vol. 112, pages 14–29, 2008. 96, 97, 98, 120, 186
- [Komodakis 2009a] N. Komodakis and N. Paragios. *Beyond Pairwise Energies: Efficient Optimization for Higher-order MRFs*. In IEEE Computer Vision and Pattern Recognition, pages 2985–2992, 2009. 97
- [Komodakis 2009b] Nikos Komodakis. *FastPD MRF Optimization Library*. Website, 2009. <http://www.csd.uoc.gr/~komod/FastPD/index.html>. xv, 77, 120, 186
- [Kruizinga 1999] P. Kruizinga and N. Petkov. *Nonlinear Operator for Oriented Texture*. IEEE transactions on image processing, vol. 8(10), pages 1395–1407, 1999. 157
- [Kubota 1997] K. Kubota, M. Makuuchi, K. Kusaka, T. Kobayashi, K. Miki, K. Hasegawa, Y. Harihara and T. Takayama. *Measurement of liver volume and hepatic functional reserve as a guide to decision-making in resectional surgery for hepatic tumors*. Hepatology., vol. 26(5), pages 1176–1181, 1997. 58
- [Kubota 2008] Toshiro Kubota. *Efficient Automated Detection and Segmentation of Medium and Large Liver Tumors: CAD Approach*. In MICCAI Workshop, 2008. 43
- [Kumar 2007] Amitesh Kumar, Alan M. Shih, Yasushi Ito, Douglas H. Ross and Bharat K. Soni. *A Hole-Filling Algorithm Using Non-Uniform Rational B-Splines*. In Proceedings of 16th International Meshing Roundtable, pages 169–182, 2007. 66
- [Kuszyk 1996] B.S. Kuszyk, D.A. Bluemke, B.A. Urban, M.A. Choti, R.H. Hruban, J.V. Sitzmann and E.K. Fishman. *Portal-phase contrast-enhanced helical CT for the detection of malignant hepatic tumors: sensitivity based on comparison with intraoperative and pathologic findings*. Am. J. Roentgenol., vol. 166, pages 91–95, 1996. 30



- [Kuwahara 1976] M. Kuwahara, K. Hachimura, S. Eiho and M. Kinoshita. Digital processing of biomedical images, chapter Processing of RI-angiocardigraphic images, pages 187–203. Plenum Press., 1976. 160
- [Kwok 2004] J.T.-Y. Kwok and I.W.-H. Tsang. *The pre-image problem in kernel methods*. Neural Networks, vol. 15(6), pages 1517–1525, 2004. 138
- [Kyrgyzov 2007] I.O. Kyrgyzov, O.O. Kyrgyzov, H. Maitre and M. Campedel. *Kernel mdl to determine the number of clusters*. In Proceedings of the 5th international conference on Machine Learning and Data Mining in Pattern Recognition, pages 203–217, 2007. 86
- [Lamecker 2002] Hans Lamecker, Thomas Lange and Martin Seebass. *A Statistical Shape Model for the Liver*. In Miccai, pages 421–427, 2002. 62, 75, 114
- [Lamecker 2004] H. Lamecker, T. Lange and M. Seebae. *Segmentation of the Liver using a 3D Statistical Shape Model*. Technical report, Zuse Institute Berlin, 2004. 62
- [Levi 1982] C. Levi, JE. Gray and RR. McCullough EC.and Hattery. *The unreliability of CT numbers as absolute values*. American Journal of Roentgenology, vol. 139, pages 443–447, 1982. 46
- [Li 2006] Y. Li, S. Hara and K. Shimura. *A Machine Learning Approach for Locating Boundaries of Liver Tumors in CT Images*. Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, pages 400–403, 2006. 41, 43
- [Lin 1998] XZ Lin, YN Sun, YH Liu, BS Sheu, BN Cheng, CY Chen, HM Tsai and CL Shen. *Liver volume in patients with or without chronic liver diseases*. Hepatogastroenterology., vol. 45(22), pages 1069–1074, 1998. 54, 58
- [Ling 2008] H. Ling, S.K. Zhou, Y. Zheng, B. Georgescu, M. Suehling and D. Comaniciu. *Hierarchical, Learning-based Automatic Liver Segmentation*. In CVPR 2008, pages 1–8, 2008. 54, 62, 125
- [Linguraru 2010] M.G. Linguraru, J. K. Sandberg, Z. Li, F. Shah and R.M. Summers. *Automated segmentation and quantification of liver and spleen from CT images using normalized probabilistic atlases and enhancement estimation*. Medical Physics, vol. 37(2), pages 771–783, 2010. 54, 61, 122, 201
- [Lupberger 2007] Joachim Lupberger and Eberhard Hildt. *Hepatitis B virus-induced oncogenesis*. World J. of Gastro., vol. 13(1), pages 74–81, 2007. 21, 27
- [MacQueen 1967] J. MacQueen. *Some methods for classification and analysis of multivariate observations*. In Proc. Fifth Berkeley Symp. on Math. Statist. and Prob., pages 281–297, 1967. 82, 137

- [Maes 1996] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal and P. Suetens. *Multi-modality image registration by maximization of mutual information*. In Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis, pages 14–22, 1996. 79
- [Mahajan 2009] Meena Mahajan, Prajakta Nimbhorkar and Kasturi Varadarajan. *The Planar k-means Problem is NP-hard*. In WALCOM '09: Proceedings of the 3rd International Workshop on Algorithms and Computation, pages 274–285, 2009. 83
- [Massoptier 2008] Laurent Massoptier and S. Casciaro. *A new fully automatic and robust algorithm for fast segmentation of liver tissue and tumors from CT scans*. European Radiology, vol. 18(8), pages 1658–1665, 2008. 41, 43, 122
- [Masutani 2006] Y. Masutani, K. Uozumi, Masaaki Akahane and Kuni Ohtomo. *Liver CT image processing: A short introduction of the technical elements*. European Journal of Radiology, vol. 58, pages 246–251, 2006. 54
- [Matsui 1991] O. Matsui, M. Kadoya, T. Kameyama, J. Yoshikawa, T. Takashima, Y. Nakanuma, M. Unoura, K. Kobayashi, R. Izumi and M. Ida. *Benign and malignant nodules in cirrhotic livers: distinction based on blood supply*. Radiology, vol. 178, pages 493–497, 1991. 23
- [McGuckin 2008] Colin P. McGuckin and Nicolas Forraz. Advances in tissue engineering, chapter Cord Blood Stem Cells - Potentials and Realities, pages 123–142. Imperial College Press, 2008. 131
- [Meir 2003] R. Meir and G. Rätsch. *An introduction to Boosting and Leveraging*. Advanced Lectures on Machine Learning, pages 118–183, 2003. 137, 141
- [Merle 2005] P. Merle. *Épidémiologie, histoire naturelle et pathogénèse du carcinome hépatocellulaire*. Cancer Radiothérapie, vol. 9, pages 452–457, 2005. 21, 23, 24, 26
- [Mika 1999] S. Mika, B. Schölkopf, A. J. Smola, K-R. Müller, M. Scholz and G. Rätsch. *Kernel PCA and De-Noising in Feature Spaces*. In Advances in Neural Information Processing Systems, volume 11, 1999. 138
- [Miller 1981] A. B. Miller, B. Hoogstraten, M. Staquet and A. Winkler. *Reporting results of cancer treatment*. Cancer, vol. 47, no. 1, pages 207–214, 1981. 36, 39, 132
- [Miyamoto 2006] Eizan Miyamoto and Thomas Jr Merryman. *Fast Calculation of Haralick Texture Features*. Technical report, Carnegie Mellon University, September 2006. 184

- [Moltz 2008] Jan Hendrik Moltz, Lars Bornemann, Volker Dicken and Heinz-Otto Peitgen. *Segmentation of Liver Metastases in CT Scans by Adaptive Thresholding and Morphological Processing*. In MICCAI Workshop, 2008. 41, 43, 195
- [Müller 2001] Mika Müller, Tsuda Rätsch and Bernhard Schölkopf. *An Introduction to Kernel-Based Learning Algorithms*. IEEE Transactions on Neural Networks, vol. 12(2), pages 181–201, 2001. 136
- [Mumford 1989] D. Mumford and J. Shah. *Optimal approximation by piecewise smooth functions and associated variational problems*. Comm. Pure Appl. Math., vol. 42, pages 577–685, 1989. 41
- [Nagao 1979] Makoto Nagao and Takashi Matsuyama. *Edge preserving smoothing*. Computer Graphics and Image Processing, vol. 9, pages 394–407, 1979. xix, 160, 164
- [Nino-Murcia 2000] M. Nino-Murcia, E.W. Olcott, R.B. Jeffrey, R.L. Lamm, C.F. Beaulieu and K.A. Jain. *Focal Liver Lesions: Pattern-based Classification Scheme for Enhancement at Arterial Phase CT*. Radiology, vol. 215, pages 746–751, 2000. xiii, 25, 30
- [NML 1989] National Library of Medicine NML. *The Visible Human Project*<sup>®</sup>, 1989. [http://www.nlm.nih.gov/research/visible/visible\\_human.html](http://www.nlm.nih.gov/research/visible/visible_human.html). xiii, 8
- [Nugroho 2008] Hanung Adi Nugroho, Dani Ihtatho and Hermawan Nugroho. *Contrast Enhancement for Liver Tumor Identification*. In MICCAI Workshop, 2008. 41, 43, 201
- [Obed 2007] A. Obed, A. Beham, K. Püllmann, H. Becker, H.J Schlitt and T. Lorf. *Patients without hepatocellular carcinoma progression after transarterial chemoembolization benefit from liver transplantation*. World J. of Gastro., vol. 13(5), pages 761–767, 2007. 131
- [Okada 1993] S Okada, N Okazaki, H Nose, K Aoki, N Kawano, J Yamamoto, K Shimada, T Takayama, T Kosuge and S. Yamasaki. *Follow-up examination schedule of postoperative HCC patients based on tumor volume doubling time*. Hepatogastroenterology., vol. 40(4), pages 311–315, 1993. 132
- [Okada 2007] T. Okada, R. Shimada, Y. Sato, M. Hori, K. Yokota, M. Nakamoto, Y.W. Chen, H. Nakamura and S. Tamura. *Automated Segmentation of the Liver from 3D CT Images Using Probabilistic Atlas and Multi-level Statistical Shape Model*. In MICCAI 2007, volume 4791, pages 86–93, 2007. 54, 61, 62, 75, 114
- [Park 2003] Hyunjin Park, Peyton H. Bland and Charles R. Meyer. *Construction of an Abdominal Probabilistic Atlas and its Application in Segmentation*. IEEE

- Transactions On Medical Imaging, vol. 22, no. 4, pages 483–492, April 2003. 61, 75
- [Park 2005] Seung-Jin Park, Kyung-Sik Seo and Jong-An Park. *Automatic Hepatic Tumor Segmentation Using Statistical Optimal Threshold*. In ICCS 2005, pages 934–940, 2005. 41, 44
- [Pearson 1901] K. Pearson. *On lines and planes of closest fit to systems of points in space*. The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science, vol. 2, pages 559–572, 1901. 137
- [Pérez 2008] Emiliano Pérez, Santiago Salamanca, Pilar Merchán, Antonio Adán, Carlos Cerrada and Inocente Cambero. *A Robust Method for Filling Holes in 3D Meshes Based on Image Restoration*. ACIVS, pages 742–751, 2008. 67
- [Perreault 2007] S. Perreault and P. Hebert. *Median Filtering in Constant Time*. IEEE Transactions on Image Processing, vol. 16, no. 9, pages 2389–2394, 2007. 159, 200
- [Pescia 2006] Daniel Pescia. Contribution au développement d’un module de détection/segmentation du réseau vasculaire hépatique dans des images 3d tomodensitométriques (ct-scan). Master’s thesis, Ecole Centrale Paris, Institut de Formation Supérieure BioMédicale, Institut d’Optique, 2006. 191
- [Pham 2007] M. Pham, R. Susomboon, T. Disney, D. Raicu and J. Furst. *A comparison of texture models for automatic liver segmentation*. Progress in biomedical optics and imaging, vol. 8(3), 2007. 60, 97, 150, 166
- [Pickren 1982] J.W. Pickren, Y. Tsukada and W.W. Lane. *Liver metastasis: Analysis of autopsy data*. Weiss L, Gilbert HA, eds., pages 2–18, 1982. 131
- [Preston 1974] Christopher J. Preston. Gibbs states on countable sets. Cambridge University Press, 1974. 96
- [Qi 2008] Yingyi Qi, Wei Xiong, Wee Keng Leow, Qi Tian, Jiayin Zhou, Jiang Liu, Thazin Han, Sudhakar K Venkatesh and Shih-chang Wang. *Semi-automatic Segmentation of Liver Tumors from CT Scans Using Bayesian Rule-based 3D Region Growing*. In MICCAI Workshop, 2008. 41, 43, 201
- [rapid i 2008] rapid i. *Rapid Miner*. Software, 2008. <http://www.rapidminer.com/>. 145
- [Rueckert 1998] D. Rueckert, C. Hayes, C. Studholme, P. Summers, M. Leach and D. J. Hawkes. *Non-rigid Registration of Breast MR Images Using Mutual Information*. In Medical Image Computing and Computer-Assisted Intervention – MICCAI 98, volume 1496, pages 1144–1153, 1998. 79

- [Ruskó 2007] L. Ruskó, G. Bekes, G. Németh and M. Fidrich. *Fully automatic liver segmentation for contrast-enhanced CT images*. In MICCAI Workshop. 3D Segmentation in the Clinic: A Grand Challenge, pages 143–150, 2007. 60
- [Schapire 1999] R.E. Schapire. *A brief Introduction to Boosting*. In IJCAI '99: Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, pages 1401–1406, 1999. 137
- [Schiano 2000] TD. Schiano, C. Bodian, ME. Schwartz, N. Glajchen and AD. Min. *Accuracy and significance of computed tomographic scan assessment of hepatic volume in patients undergoing liver transplantation*. Transplantation, vol. 69, pages 545–550, 2000. 54, 58
- [Schölkopf 1998a] B. Schölkopf, S. Mika, A. Smola, G. Rätsch and K-R. Müller. *Kernel PCA Pattern Reconstruction via Approximate Pre-Images*. In Proceedings of the 8th International Conference on Artificial Neural Networks, 1998. 138
- [Schölkopf 1998b] B. Schölkopf, A. Smola and K-R. Müller. *Nonlinear component analysis as a kernel eigenvalue problem*. Neural Computation, vol. 10, pages 1299–1319, 1998. 136, 137, 138
- [Schölkopf 1999] B. Schölkopf, A. Smola and K-R. Müller. *Kernel Principal Component Analysis*. In Proceedings of the 8th International Conference on Artificial Neural Networks, pages 327–352, 1999. 138
- [Schütte 2009] Kerstin Schütte, Jan Bornschein and Peter Malfertheiner. *Hepatocellular Carcinoma - Epidemiological Trends and Risk Factors*. Digestive Diseases, vol. 27, pages 80–92, 2009. 20, 21, 27
- [Selle 2002] Dirk Selle, Bernhard Preim, Andrea Schenk and Heinz-otto Peitgen. *Analysis of Vasculature for Liver Surgical Planning*. IEEE Transactions on Medical Image Analysis Imaging, vol. 21, pages 1344–1357, 2002. 191
- [Seo 2005] Kyung-Sik Seo and Tae-Woong Chung. *Automatic Boundary Tumor Segmentation of a Liver*. In ICCSA 2005, pages 836–842, 2005. 40, 44
- [Shimizu 2005] A. Shimizu, T. Kawamura and H. Kobatake. *Proposal of computer-aided detection system for three dimensional CT images of liver cancer*. International Congress Series, vol. 1281, pages 1157–1162, 2005. 44
- [Shimizu 2007] Akinobu Shimizu, Rena Ohno, Takaya Ikegami, Hidefumi Kobatake, Shigeru Nawano and Daniel Smutek. *Segmentation of multiple organs in non-contrast 3D abdominal CT images*. Int J CARS, vol. 2, pages 135–142, 2007. 61, 75, 94
- [Shimizu 2008] Akinobu Shimizu, Takuya Narihira, Daisuke Furukawa, Hidefumi Kobatake, Shigeru Nawano and Kenji Shinozaki. *Ensemble segmentation*

- using AdaBoost with application to liver lesion extraction from a CT volume.* In MICCAI Workshop, 2008. 41, 42, 43
- [Simon 2005] C.J. Simon, D.E. Dupuy and W.W. Mayo-Smith. *Microwave Ablation: Principles and Applications.* RadioGraphics, vol. 25, pages S69–S83, 2005. 33
- [Sitruk 2004] V. Sitruk, O. Seror and Trinchet J-C. *Traitement percutané du carcinome hépatocellulaire.* EMC-Hépatologie, vol. 1(4), pages 123–130, 2004. 33
- [Smeets 2008] Dirk Smeets, Bert Stijnen, Dirk Loeckx, Bart De Dobbelaer and Paul Suetens. *Segmentation of Liver Metastases Using a Level Set Method with Spiral-Scanning Technique and Supervised Fuzzy Pixel Classification.* In MICCAI Workshop, 2008. 42, 43
- [Smutek 2006] D. Smutek, A. Shimizu, H. Kobatake, S. Nawano and L. Tesar. *Texture Analysis of Hepatocellular Carcinoma and Liver Cysts in CT Images.* In Proceedings of the 24th IASTED international conference on Signal processing, pattern recognition, and applications, pages 56–59, 2006. 42, 44, 166
- [Soler 1998] Luc Soler, Grégoire Malandain and Hervé Delingette. *Automatic Segmentation: Application to 3D Angioscanners of the Liver.* Technical report, INRIA, 1998. 191, 201
- [Soler 2001] Luc Soler, H. Delingette, G. Malandain, J. Montagnat, N. Ayache, C. Koehl, O. Dourthe, B. Malassagne, M. Smith, D. Mutter and J. Marescaux. *Fully automatic anatomical, pathological, and functional segmentation from CT scans for hepatic surgery.* Computer Aided Surgery, vol. 6(3), pages 131–142, 2001. 40, 43, 94, 191
- [Spitzer 1971] Frank Spitzer. Random fields and interacting particle systems. Mathematical Association of America, 1971. 96
- [Srinivasan 2008] G. N. Srinivasan and G. Shobha. *Statistical Texture Analysis.* Proceedings Of World Academy Of Science, Engineering And Technology, vol. 36, pages 1264–1269, December 2008. 148
- [Sørli 2005] Rune Petter Sørli. Automatic segmentation of liver tumors from mri images. Master’s thesis, University Of Oslo, 2005. 159
- [Strong 2006] Russell W. Strong. *Living-donor liver transplantation: an overview.* J Hepatobiliary Pancreat Surg, vol. 13, pages 370–377, 2006. 58
- [Szeliski 2006] Richard Szeliski, Ramin Zabih, Daniel Scharstein, Olga Veksler, Vladimir Kolmogorov, Aseem Agarwala, Marshall Tappen and Carsten Rother. *A Comparative Study of Energy Minimization Methods for Markov*

- Random Fields*. Lecture Notes in Computer Science, vol. 3952, no. 6, pages 16–29, 2006. 98
- [Szeliski 2008] Richard Szeliski, Ramin Zabih, Daniel Scharstein, Olga Veksler, Vladimir Kolmogorov, Aseem Agarwala, Marshall Tappen and Carsten Rother. *A Comparative Study of Energy Minimization Methods for Markov Random Fields with Smoothness-Based Priors*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 6, pages 1068–1080, 2008. 98, 101
- [Taieb 2008] Y. Taieb, O. Eliassaf, M. Freiman, L. Joskowicz and J. Sosna. *An iterative Bayesian approach for liver analysis: tumors validation study*. In MICCAI Workshop, 2008. 41, 43
- [Tang 1989] Z-Y Tang, Y-Q Yu, X-D Zhou, Z-C Ma, R Yang, J-Z Lu, Z-Y Lin and B-H Yang. *Surgery of Small Hepatocellular Carcinoma: Analysis of 144 Cases*. Cancer, vol. 64(2), pages 536–541, 1989. 131
- [Tang 2001] Zhao-You Tang. *Hepatocellular Carcinoma-Cause, Treatment and Metastasis*. World J. of Gastro., vol. 7(4), pages 445–454, 2001. 23, 26, 34, 130, 131
- [Taylor-Robinson 1997] Simon D. Taylor-Robinson, GR Foster, S Arora, S Hargreaves and HC Thomas. *Increase in primary liver Cancer in the UK, 1979-1994*. Lancet, vol. 350, pages 1142–1143, 1997. 131
- [Tesar 2008] Ludvik Tesar, Akinobu Shimizu, Daniel Smutek, Hidefumi Kobatake and Shigeru Nawano. *Medical image analysis of 3D CT images based on extension of Haralick texture features*. Computerized Medical Imaging and Graphics, vol. 32, pages 513–520, 2008. 114, 201
- [Therasse 2000] Patrick Therasse, Susan G. Arbuck, Elizabeth A. Eisenhauer, Jantien Wanders, Richard S. Kaplan, Larry Rubinstein, Jaap Verweij, Martine Van Glabbeke, Allan T. van Oosterom, Michaele C. Christian and Steve G. Gwyther. *New Guidelines to Evaluate the Response to Treatment in Solid Tumors*. Journal of the National Cancer Institute, vol. 92(3), pages 205–216, 2000. 36, 132, 133
- [Tukey 1977] John W. Tukey. *Exploratory data analysis*. Addison-Wesley, 1977. 159
- [Turetsky 2003] R. Turetsky and D. Ellis. *Ground-Truth Transcriptions of Real Music from Force-Aligned MIDI Syntheses*. In 4th International Symposium on Music Information Retrieval ISMIR-03, pages 135–14, 2003. 45, 48, 49
- [Van Hoe 1997] L. Van Hoe, A.L. Baert, S. Gryspeerdt, G. Vandenbosh, F. Nevens, W. Van Steenbergen and G. Marchal. *Dual-phase helical CT of the liver:*

- value of an early-phase acquisition in the differential diagnosis of noncystic focal lesions.* Am. J. Roentgenol., vol. 168, pages 1185–1192, 1997. 26
- [Vilgrain 2000] V. Vilgrain, L. Boulos, M-P. Vullierme, A. Denys, B. Terris and Y. Menu. *Imaging of Atypical Hemangiomas of the Liver with Pathologic Correlation.* RadioGraphics, vol. 20, pages 379–397, 2000. 20, 25
- [Vilgrain 2002] Valérie Vilgrain and Yves Menu. *Imagerie du foie, des voies biliaires, du pancréas et de la rate.* Flammarion Médecine-Sciences, 2002. 20, 130
- [Viola 1995] P. Viola and W. M. III Wells. *Alignment by maximization of mutual information.* In Proceedings of the Fifth International Conference on Computer Vision, pages 16–23, 1995. 79
- [Viola 2004] P. Viola and M-J. Jones. *Robust Real-Time Face Detection.* International Journal of Computer Vision, vol. 57(2), pages 137–354, 2004. 182
- [Vogl 2006] TJ Vogl, A Scheller, U Jakob, S Zangos, M Ahmed and M Nabil. *Transarterial chemoembolization in the treatment of hepatoblastoma in children.* Eur Radiol., vol. 16(6), pages 1393–1396, 2006. 34
- [Wang 2007] Jianzhe Wang and Tianzi Jiang. *Nonrigid registration of brain MRI using NURBS.* Pattern Recognition Letters, vol. 28, no. 2, pages 214–223, 2007. 79
- [WHO 2009] World Health Organization WHO. *WHO Statistical Information System (WHOSIS).* Website, February 2009. <http://www.who.int/whosis/en/>. 20, 23
- [Wong 2008] Damon Wong, Jiang Liu, Yin Fengshou, Qi Tian, Wei Xiong, Jiayin Zhou, Yingyi Qi, Thazin Han, Sudhakar K Venkatesh and Shih-chang Wang. *A semi-automated method for liver tumor segmentation based on 2D region growing with knowledge-based constraints.* In MICCAI Workshop, 2008. 41, 43, 159
- [Wu 2008] Xiao J. Wu, Michael Y. Wang and B. Han. *An Automatic Hole-Filling Algorithm for Polygon Meshes.* Computer-Aided Design and Applications, vol. 5, no. 6, pages 889–899, 2008. 67
- [Xiang 2008] Deng Xiang and Du Guangwei. *3D Liver Tumor Segmentation Challenge 2008.* MICCAI Workshop, 2008. <http://lts08.bigr.nl/>. xiv, 40, 43
- [Yalcin 2004] S. Yalcin. *Diagnosis and management of cholangiocarcinomas: a comprehensive review.* Hepatogastroenterology, vol. 51(55), pages 43–50, 2004. 28



- [Yin 2004] Zhongwei Yin. *Reverse engineering of a NURBS surface from digitized points subject to boundary conditions*. *Computers & Graphics*, vol. 28, pages 207–212, 2004. 66
- [Yuki 1990] K. Yuki, S. Hirohashi, M. Sakamoto, T. Kanai and Y. Shimosato. *Growth and Spread of Hepatocellular Carcinoma: A Review of 240 Consecutive Autopsy Cases*. *Cancer*, vol. 66(10), pages 2174–2179, 1990. 27
- [Zhao 2007] Wei Zhao, Shuming Gao and Hongwei Lin. *A robust hole-filling algorithm for triangular mesh*. *The Visual Computer*, vol. 23, no. 12, pages 987–997, 2007. xiv, 67, 69
- [Zhou 2005] X. Zhou, T. Kitagawa, K. Okuo, T. Hara, H. Fujita, R. Yokoyama, M. Kanematsu and H. Hoshi. *Construction of a probabilistic atlas for automated liver segmentation in non-contrast torso CT images*. *International Congress Series*, vol. 1281, pages 1169–1174, 2005. 54, 61, 75, 94, 110
- [Zhou 2006] T. Zhou X. and Kitagawa, T. Hara, H. Fujita, X. Zhang, R. Yokoyama, H. Kondo, M. Kanematsu and H. Hoshi. *Constructing a Probabilistic Model for Automated Liver Region Segmentation Using Noncontrast X-Ray Torso CT images*. In *MICCAI*, pages 856–863, 2006. 61
- [Zhou 2008] Jiayin Zhou, Wei Xiong, Qi Tian, Yingyi Qi, Jiang Liu, Wee Keng Leow, Thazin Han, Sudhakar K Venkatesh and Shih-chang Wang. *Semi-automatic Segmentation of 3D Liver Tumors from CT Scans Using Voxel Classification and Propagational Learning*. In *MICCAI Workshop*, 2008. 41, 43