



**HAL**  
open science

# Modélisation du mouvement des quadrupèdes à partir de la vidéo

Laurent Favreau

► **To cite this version:**

Laurent Favreau. Modélisation du mouvement des quadrupèdes à partir de la vidéo. Modélisation et simulation. Institut National Polytechnique de Grenoble - INPG, 2006. Français. NNT: . tel-00379321

**HAL Id: tel-00379321**

**<https://theses.hal.science/tel-00379321>**

Submitted on 28 Apr 2009

**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.

INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

*N° attribué par la bibliothèque*

--	--	--	--	--	--	--	--	--	--

THÈSE

pour obtenir le grade de

DOCTEUR DE L'INP GRENOBLE

Spécialité **Imagerie, Vision, Robotique**

Préparée au Laboratoire GRAVIR/IMAG-INRIA, UMR CNRS C5527,  
dans le cadre de l'École Doctorale "Mathématiques, Sciences et Technologies de  
l'Information, Informatique"

Présentée et soutenue publiquement

par

Laurent FAVREAU

le 30 Novembre 2006

---

Modélisation du mouvement des quadrupèdes à partir de la vidéo

---

*Directrice de thèse :* Marie-Paule CANI

*Co-directeur :* Lionel REVERET

**JURY**

M. Pascal PERRIER	Professeur - INP Grenoble	, Président
M. Ronan BOULIC	Adjoint Scientifique - EPFL	, Rapporteur
M. Jean-Paul LAUMOND	Directeur de Recherches - CNRS	, Rapporteur
Mme Sabine RENOUS	Directeur de Recherches - CNRS	, Examineur
Mme Marie-Paule CANI	Professeur - INP Grenoble	, Directrice de thèse
M. Lionel REVERET	Chargé de Recherches - INRIA	, Co-directeur de thèse



---

# Remerciements

---

Je souhaite tout d'abord remercier mes directeurs de thèse, Marie-Paule Cani et Lionel Reveret pour m'avoir permis d'effectuer cette thèse au sein de l'équipe EVASION, pour leurs conseils et leur soutien.

Je remercie également Ronan Boulic et Jean-Paul Laumond pour avoir accepté d'être les rapporteurs de ma thèse, et Pascal Perrier pour m'avoir fait l'honneur de présider mon jury. Je tiens aussi à remercier Sabine Renous pour avoir accepté de faire partie de mon jury.

Cette thèse m'a permis d'effectuer de nombreuses collaborations. Je remercie Christine Depraz qui a apporté son talent artistique à ce travail et grandement amélioré la qualité finale de mes résultats. Je remercie Zoran Popović et Steve Seitz pour m'avoir accueilli pendant mon séjour à Seattle, ainsi que l'équipe du Muséum National d'Histoire Naturelle, dont le travail m'a ouvert de nouveaux horizons.

Je remercie tous les membres des équipes Evasion et Artis, pour tous les moments passés tout au long de ma thèse, et en particulier tous ceux qui ont partagé mon bureau pendant un moment.

Enfin, je remercie ma famille et mes proches pour leur soutien et leur enthousiasme constant durant ces trois années.



---

# SOMMAIRE

---

<b>Introduction</b>	<b>7</b>
<b>I Etat de l'art</b>	<b>11</b>
1 Modélisation de quadrupèdes virtuels pour l'animation	13
2 Techniques de contrôle du mouvement	33
3 Utilisation de données réelles pour l'animation	49
<b>II Modélisation de squelettes de quadrupèdes</b>	<b>61</b>
4 Reconstruction d'un squelette de quadrupède à partir d'images	63
5 Construction d'un squelette générique de quadrupèdes à partir d'images	73
<b>III Animation de quadrupèdes à partir de vidéos</b>	<b>89</b>
6 Animation de mouvements cycliques à partir de vidéos	91
7 Analyse des séquences de posés à partir de vidéos.	117

<b>Conclusion</b>	<b>133</b>
<b>A Hiérarchie de repères</b>	<b>137</b>
<b>B Quaternions</b>	<b>141</b>

---

# Introduction

---



ANIMAUX VIRTUELS sont de plus en plus présents dans les films d'animation et les effets spéciaux, comme on a pu le voir dans des productions récentes comme *Le monde de Narnia*, *l'âge de glace 2* ou *Le seigneur des anneaux* (Figure 1).

Créer des animaux en 3D permet d'ouvrir des possibilités très riches : animer des animaux fantastiques ou disparus, avoir un contrôle total sur le mouvement d'un animal sauvage, animer des milliers d'animaux simultanément. Les quadrupèdes, par la richesse de leur mouvement et leurs liens privilégiés avec l'homme, sont particulièrement intéressants : félins, chevaux, éléphants sont des animaux très attractifs visuellement.

## Motivations et enjeux

La modélisation et l'animation de modèles 3D de quadrupèdes est un processus fastidieux qui inclus de longues analyses de documentation, et représente des semaines de travail pour la création et l'animation d'un quadrupède effectuant un mouvement de locomotion. A l'autre bout du spectre, des techniques existent pour capturer le mouvement de créatures réelles. Elles donnent de bons résultats, mais demandent une collaboration absolue de la part du sujet, qui doit se déplacer dans un espace confiné (quelques dizaines de mètres carrés), porter des marqueurs optiques ou magnétiques, et parfois même courir sur un tapis roulant. Si elles sont bien adaptées aux humains, elles sont donc difficiles à appliquer aux animaux sauvages comme des fauves.



FIG. 1 – Images extraites de *l'âge de glace* et du *monde de Narnia*. Que ce soit dans les films d'animation ou les effets spéciaux, les animaux virtuels sont de plus en plus présents.

En revanche, les vidéos d'animaux sauvages représentent un fond documentaire abondant et facilement accessible, très utilisé par les animateurs, et qui ne nécessite pas d'avoir un réel contact avec l'animal à animer. L'objectif de cette thèse est d'utiliser des séquences vidéo, et en particulier des documentaires animaliers, pour automatiser autant que possible la création de quadrupèdes 3D en mouvement. Nous nous proposons de développer des méthodes permettant d'extraire, à partir de vidéos, des informations pertinentes en vue du contrôle de mouvements articulés en 3D. Cette thèse se concentre sur le contrôle de mouvement, et ne vise pas à reconstruire en 3D la forme géométrique des animaux en mouvement.

Les documents vidéo, s'ils existent, restent un défi pour l'exploitation automatique du fait de grandes sources de variabilités : éclairage, prise de vue, occlusions, manque de résolution. Nous cherchons donc à savoir sous quelles conditions il est possible d'utiliser la vidéo pour créer des animations 3D de quadrupèdes. Fournir des outils permettant d'aider l'animateur à créer de telles animations constitue une contribution importante dans le domaine de la synthèse de scènes naturelles.

## Contributions de la thèse

Les premières contributions s'appliquent à la création de squelettes 3D de quadrupèdes. Nous validons un protocole de reconstruction 3D de squelette réels à partir d'images. Ensuite, nous montrons comment construire un modèle générique de squelette d'animation 3D qui permettra de créer rapidement des squelettes d'animation de quadrupèdes de morphologie différente à partir de quelques mesures prises sur des images ou des surfaces 3D.

Ensuite nous traitons de l'animation de quadrupèdes à partir de séquences vidéo. Dans

un premier temps nous montrons comment animer un mouvement cyclique de locomotion à partir d'une vidéo. En particulier, nous proposons un algorithme de sélection automatique des images-clés à partir de vidéo, et résolvons des ambiguïtés dues au passage de 2D vers 3D. Enfin, nous proposons une nouvelle méthode d'analyse de vidéos qui permet de détecter les allures présentes sur une vidéo ainsi que les transitions entre allures. Cette méthode valide une nouvelle méthode de formalisation des allures mise au point par nos partenaires du Muséum National d'Histoire Naturelle dans le cadre du projet ANR KAMELEON.

## Organisation du document

Ce mémoire est divisé en sept chapitres regroupés en trois parties.

La première partie s'attache tout d'abord, au chapitre 1, à présenter la morphologie des quadrupèdes réels, puis à détailler les méthodes classiques permettant de modéliser des quadrupèdes virtuels en 3D. Au chapitre 2, nous introduisons la locomotion des quadrupèdes et présentons les techniques existantes de contrôle du mouvement de créatures virtuelles. Le chapitre 3 recense l'ensemble des méthodes existantes qui visent à utiliser des données réelles dans un cadre d'animation.

La seconde partie présente les contributions à la modélisation du squelette des quadrupèdes. Au chapitre 4, nous montrons comment reconstruire un squelette 3D de quadrupède à partir de photos. Le chapitre 5 détaille comment nous avons pu construire un modèle générique de squelette de quadrupèdes qui puissent s'adapter très facilement à différents animaux.

Enfin, la troisième et dernière partie présente les recherches effectuées sur l'animation de quadrupèdes à partir de séquences vidéo. Le chapitre 6 présente notre méthode pour animer des mouvements cycliques d'animaux à partir de vidéo. Au chapitre 7, nous montrons comment extraire les séquences de posés des pieds sur une vidéo et, en utilisant un nouveau formalisme pour décrire les différentes allures, détecter les allures présentes sur la vidéo.

Nous dressons en conclusion le bilan de ces travaux, et donnons des pistes pour poursuivre plus avant ces recherches.



**Première partie**

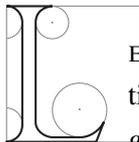
**Etat de l'art**



---

# Modélisation de quadrupèdes virtuels pour l'animation

---



LES ANIMAUX VIRTUELS sont de plus en plus présents dans les films d'animation et les effets spéciaux, avec des films comme *Le monde de Narnia*, *Le seigneur des anneaux* ou *King Kong*. Créer des animaux en 3D permet d'ouvrir de nouvelles possibilités : animer des animaux fantastiques ou disparus, avoir un contrôle total sur le mouvement d'un animal sauvage, animer des milliers d'animaux simultanément. Les quadrupèdes, par la richesse de leur mouvement et leurs liens privilégiés avec l'homme, sont particulièrement intéressants : félins, chevaux, éléphants sont des animaux très attractifs visuellement.

Mais créer un modèle 3D d'un quadrupède donné est une tâche difficile. Tout d'abord, l'anatomie des animaux réels est complexe, et nos connaissances intuitives sont beaucoup plus réduites pour les animaux que pour les humains. Cela rend difficile aussi bien la création du modèle 3D que la validation du résultat. Ensuite, comme le modèle est destiné à être animé, il doit pouvoir être manipulé avec des paramètres de contrôle bien choisis. Les paramètres de contrôle doivent être à la fois pertinents et peu nombreux pour que l'utilisateur puisse les

manipuler de manière intuitive, et porteurs d'assez d'information pour pouvoir décrire des mouvements de qualité.

Ce chapitre présente tout d'abord des notions essentielles sur la structure des quadrupèdes réels. Ensuite, nous dressons un état de l'art des méthodes existantes pour la modélisation de quadrupèdes 3D.

## 1 Quadrupèdes : du réel au virtuel

Par définition, les quadrupèdes sont les animaux qui utilisent quatre pattes pour se déplacer. Il est important de noter que tous les *tétrapodes*, c'est-à-dire tous les animaux qui possèdent quatre membres, ne sont pas des quadrupèdes : les humains et les kangourous, par exemple, ont quatre membres mais ne les utilisent pas tous pour la locomotion. Dans cette thèse, nous nous limiterons aux mammifères quadrupèdes, mais de nombreux reptiles (lézards, tortues, crocodiles) et amphibiens (grenouilles, crapauds) le sont aussi.

De nombreux ouvrages proposent de la documentation détaillée sur l'anatomie des quadrupèdes [Feh00, Muy57, EDB56, Cal75], sous forme de photos ou de dessins d'anatomie. Pour une introduction détaillée, nous vous renvoyons à [Ren94].

### 1.1 Structure anatomique des quadrupèdes



FIG. 1.1 – A gauche : le megazostrodon, mammifère primitif. A droite : le cheval est historiquement le quadrupède le plus étudié.

Historiquement, les premiers quadrupèdes terrestres ont été des reptiles de la famille des lézards, apparus il y a environ 360 millions d'années. Le Megazostrodon, apparu il y a 200 millions d'années, est considéré comme le premier mammifère (Figure 1.1). C'est un lointain ancêtre des animaux que nous allons étudier. Des millions d'années d'évolution ont donné

naissance à des animaux aussi sophistiqués que le cheval, mais la structure de base des quadrupèdes est restée la même (Figure 1.2). Le corps est en appui sur les quatre membres, sur lesquels repose la colonne vertébrale, sorte de "poutre" qui soutient tout le corps. Le tronc repose sur les membres postérieurs au niveau du bassin, et il est suspendu aux membres antérieurs par des muscles. La colonne dessine un arc à concavité ventrale tendu par le poids des viscères et des muscles. Les membres sont des chaînes articulées qui servent de propulseurs. Ils exercent une poussée sur le sol qui, par réaction, exerce une force propulsive sur le corps.

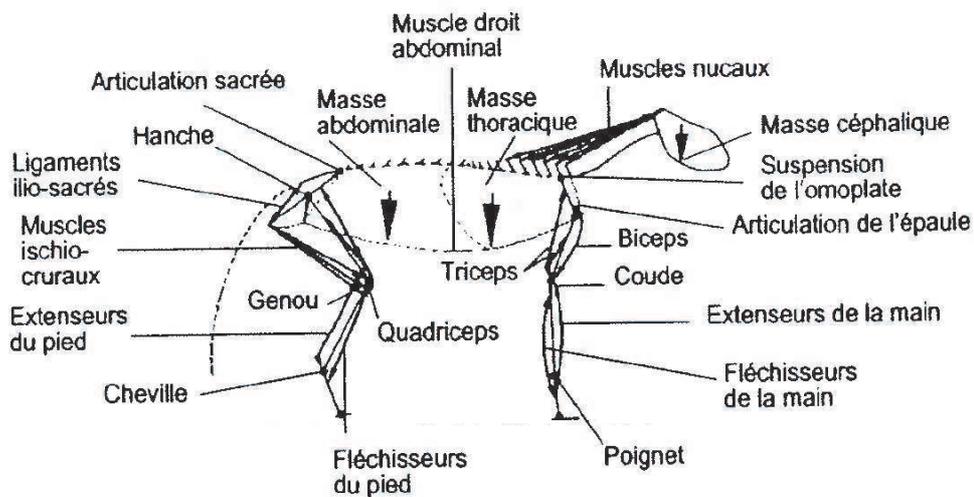


FIG. 1.2 – Schéma général de construction d'un mammifère (d'après [Ren94])

Les quadrupèdes, comme tous les animaux vertébrés, sont composés de trois systèmes distincts :

- Le squelette. Il est composé d'éléments rigides articulés entre eux. La rigidité des os leur permet de transmettre et multiplier les forces qui leur sont appliquées, mais aussi de résister aux forces extérieures.
- Les muscles sont attachés au squelette. En se contractant, ils appliquent des forces sur les os et permettent d'initier et de contrôler le mouvement.
- Le système nerveux stimule et contrôle la contraction des muscles. C'est l'organe de commande du mouvement.

En animation, le squelette est le système qui mobilise le plus d'attention car son mouvement permet de décrire le mouvement de tout le corps de façon compacte. Les membres des vertébrés terrestres sont des chaînes articulées dont les segments forment des angles alternés. Chez les quadrupèdes les plus primitifs comme les amphibiens, les membres s'articulent dans un plan transversal (type *transversal*), alors que les formes les plus évoluées (type *parasagittal*) ont des membres articulés dans le plan sagittal (Figure 1.3) . Ceci leur permet de mieux assurer la

fonction d'élément propulseur. Dans cette thèse, nous étudierons uniquement les quadrupèdes de type parasagittal. Pour une étude du mouvement d'un quadrupède transversal, voir [ICC05].

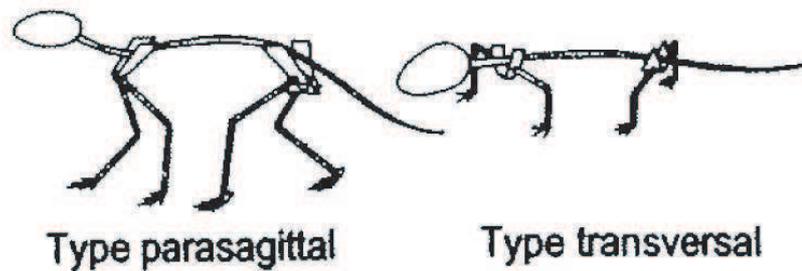


FIG. 1.3 – Schéma de deux types d'orientation des membres (d'après [Ren94]). Les membres sont de type transversal chez les quadrupèdes plus primitifs comme les amphibiens, et de type parasagittal chez les mammifères.

On peut séparer les mammifères en trois catégories (Figure 1.4) :

- les *plantigrades*, qui se déplacent en prenant appui sur la plante des pieds, comme les hommes, mais aussi les ours, les lapins, les rats,
- les *digitigrades*, qui appuient les doigts sur le sol, comme les chats et les chiens,
- les *ongulés*, qui marchent sur la pointe des pieds (les ongles), comme les chevaux et les moutons.

De manière générale, pour les quadrupèdes, les animaux les plus "hauts" (le centre de gravité est élevé par rapport au sol) sont ongulés ou digitigrades, et les animaux les plus "effondrés" sont plantigrades. Cette règle a bien sûr des exceptions, l'exemple de l'ours étant le plus connu.

## 1.2 Vers un quadrupède virtuel

Les quadrupèdes vertébrés, comme les humains, sont constitués d'une hiérarchie de segments articulés, le squelette, qui détermine leur structure, et de tissus qui entourent le squelette et leur donnent leur aspect extérieur. Les différents systèmes ne sont pas traités de la même manière lorsqu'on passe à un quadrupède virtuel :

- On fait abstraction du système nerveux, qui occupe pour la plupart un volume très réduit, pendant la phase de modélisation
- Le squelette, qui est la structure de contrôle du mouvement, est représenté. Comme le squelette réel est très complexe, on le représente de manière simplifiée et adaptée par un squelette d'animation.
- Les muscles (et les autres tissus) sont aussi représentés dans un but de réalisme, pour

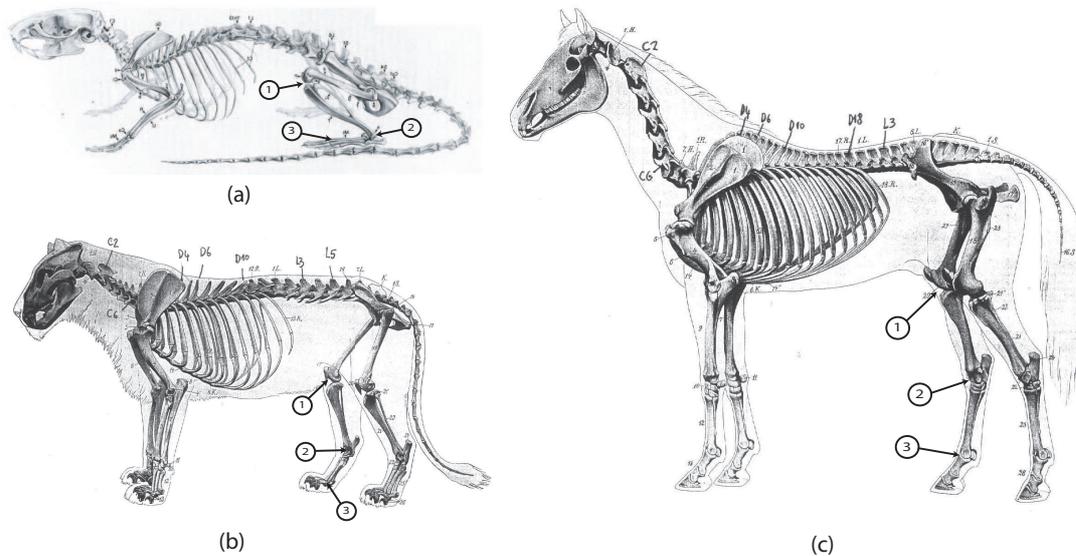


FIG. 1.4 – exemples de plantigrades (le rat) (a), de digitigrade (le tigre) (b) et d'ongulé (le cheval)(c) (d'après [EDB56], modifié). Notez les positions respectives des articulations du genou (1), du tarse (2), et de la première phalange (3). Ces positions peuvent donner l'illusion que les membres ont un segment de plus que des membres d'humains. Il n'en est rien, seule la longueur et la position des segments change.

représenter le volume du corps. Comme nous le verrons, cette représentation peut être plus ou moins complexe.

Il y a une grande différence lorsqu'on passe d'un quadrupède réel à un quadrupède virtuel. Dans le cas d'un quadrupède réel, ce sont les muscles qui, en se contractant, exercent des forces sur le squelette et créent le mouvement. Dans le cas des animaux virtuels, c'est généralement le squelette qui porte l'information de mouvement, et les muscles suivent ce mouvement en se déformant éventuellement. D'où la notion fondamentale de *squelette d'animation* que nous allons détailler dans la section suivante.

## 2 Modélisation du squelette d'animation

Un animal en 3D modélisé de façon classique est composé de deux parties : un *squelette d'animation*, qui est la structure permettant de contrôler le mouvement, et une *enveloppe* qui entoure le squelette et sert à donner à l'animal une apparence réaliste ( Figure 1.5).

Le squelette d'un animal réel est très complexe : les articulations ne sont pas de simples pivots, mais sont des emboîtements d'os, qui se déplacent les uns par rapport aux autres de manière complexe. Les forces s'appliquent en plusieurs points, les os glissant les uns sur les

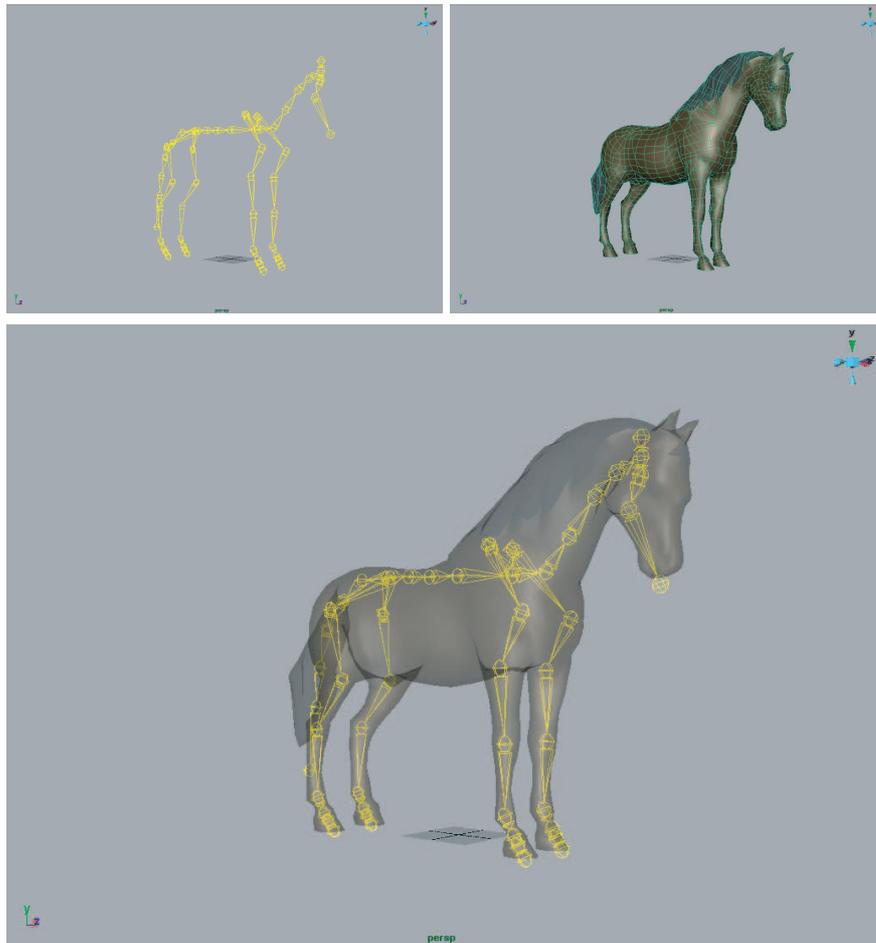


FIG. 1.5 – Structure de base d'un quadrupède 3D : Un squelette d'animation (en jaune) entouré d'une surface 3D (en gris).

autres, et l'axe de rotation se déplace au cours du mouvement. Physiologiquement, il n'existe donc pas d'axe unique et fixe de rotation pour une articulation donnée [Mau99]. Cependant, les translations étant la plupart du temps négligeables par rapport aux rotations, il est possible de modéliser une articulation par une articulation idéale à 1,2 ou 3 degrés de liberté (Figure 1.6).

## 2.1 Le squelette comme une hiérarchie de repères

En considérant les articulations comme des liaisons rotules, et en supposant que la longueur des segments du squelette est constante pour un animal donné, on peut considérer le squelette comme une hiérarchie (ou un arbre) de repères en 3D. Le repère père (souvent appelé *root*), définit la position et l'orientation globale de l'animal dans le monde 3D. Ce nœud père a plusieurs fils, et chacun de ces fils peuvent eux-même avoir des fils, etc. Ceci donne une struc-

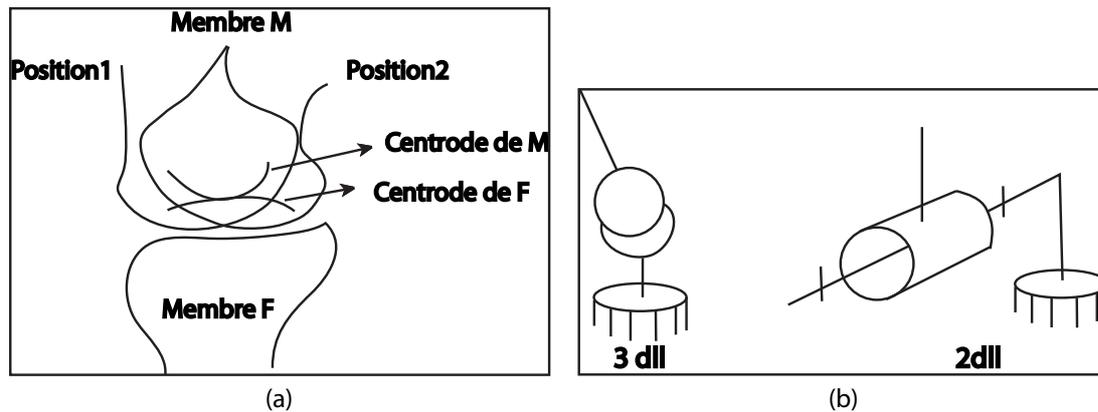


FIG. 1.6 – Articulations réelles et articulations idéales : (a)-lorsque deux membres d'une articulation réelle sont en mouvement relatif, les centrodes (c'est-à-dire les lieux des centres instantanés de rotation des deux objets) montrent le glissement qui intervient entre les deux membres (image extraite de [KG83]); (b)-en négligeant les translations, l'articulation peut être ramenée à une rotule (3 degrés de liberté) ou à un modèle plus contraint (images extraites de [Mau99].)

ture arborescente au modèle. La hiérarchie que nous utilisons dans ce travail est détaillée dans l'annexe A. Par convention et par commodité, le pelvis est généralement choisi comme nœud père.

Les positions/orientations des repères sont relatives les unes aux autres. Si on déplace ou tourne un repère, tous ses repères fils se déplaceront relativement à lui. Par exemple, si on bouge le repère placé au niveau de l'épaule gauche, tous les repères de l'avant-bras gauche suivront le mouvement (Figure 1.7).

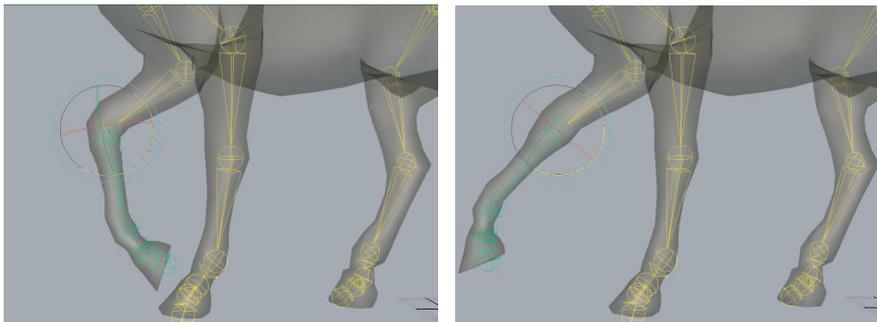


FIG. 1.7 – Les positions des repères sont relatives. En conséquence, changer l'orientation d'un repère déplacera tous ces repères fils.

En animation 3D, on définit donc une articulation comme un repère défini relativement à un repère père. Il existe plusieurs méthodes permettant de formaliser les transformations

d'une articulation à une autre. Dans cette thèse, nous utilisons la convention suivante : Un joint est défini par plusieurs transformations successives qui s'effectuent dans un ordre précis ( Figure 1.8) :

- Sa translation 3D par rapport au joint précédent. Cette valeur est fixe au cours du temps, sauf pour le nœud père dont la translation donne la position globale en 3D du quadrupède dans le monde. Rarement, quelques autres repères peuvent être animés en translation, par exemples les clavicules peuvent l'être pour tenir compte des mouvements de glissement des homoplates dans le cas de quadrupèdes réels.
- Son orientation. Elle définit l'orientation naturelle du joint à la pose de repos. Cette valeur est fixe.
- La rotation. Elle varie. Les changements apportés à cette valeur permet de faire bouger le squelette au cours du temps.

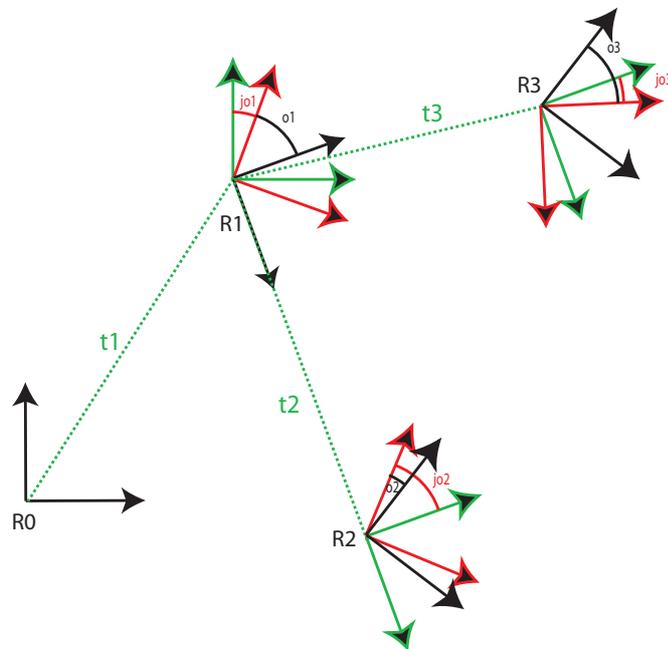


FIG. 1.8 – L'organisation en joints, présentée en 2D pour plus de clarté. Un joint est défini par plusieurs transformations dans un ordre précis : d'abord la translation (en vert) , puis une rotation fixe qui définit l'orientation naturelle du joints(en rouge), puis une rotation qui varie au cours du temps et définit le mouvement du joint (en noir)

Plusieurs conventions peuvent être prises pour le placement des axes de rotation. Dans cette thèse nous utiliserons les conventions suivantes :

- L'axe  $x$  est dirigé vers le joint fils si il n'y en a qu'un. C'est le cas pour tous les joints sauf

le pelvis (qui a pour fils les hanches, le sacrum et la colonne, on choisit alors la colonne) et le cou (qui a pour fils les épaules et le cou, on choisit alors le cou). Dans ce cas, l'axe  $x$  est placé dans le prolongement de la colonne vertébrale.

- L'axe  $z$  est placé pour rendre compte de la principale direction de rotation de l'articulation. Par exemple, pour le coude qui est une articulation à un degré de liberté, l'axe  $z$  est placé pour que la rotation d'axe  $z$  plie le coude.
- L'axe  $y$  est choisit pour que  $(x,y,z)$  forme un repère direct.

Au final, nous obtenons une hiérarchie de repère 3D telle que celle présentée dans l'annexe A.

Naturellement, un joint réel à  $n$  degrés de liberté ( $n \leq 3$ ) ne pourra explorer qu'un sous-ensemble des rotations possibles. Il y a en effet des bornes aux mouvements réalisables par chaque joint. Par exemple, le coude, qui est considéré comme une articulation à un degré de liberté, ne peut pas se "retourner". Ces limites sont d'abord dues au squelette, dont la morphologie induit des limites sur l'amplitude des mouvements possibles (Figure 1.9). Les forces de résistance des muscles, des tendons et des autres tissus s'opposent aussi au mouvement, ainsi que les collisions entre les zones du corps rapprochées par le mouvement. On peut donc définir des limites sur les rotations qu'un joint peut effectuer, afin d'éviter que le squelette puisse prendre des poses irréalistes.

On définit ainsi une *pose* comme étant la donnée d'une position de tout le squelette. Une pose est donc définie par la position en 3D du nœud père ainsi que des rotations de tous les joints. On définit aussi ce qu'on appelle la *pose de repos*, c'est-à-dire la pose qu'on considère comme la plus naturelle possible pour l'animal.

## 2.2 Représentations des rotations

Comme nous l'avons vu, positionner le squelette d'animation revient à définir sa position dans le repère monde et l'orientation de tous ses joint. Il existe plusieurs formalismes standard permettant de décrire les orientations d'un repère 3D que nous allons à présent les détailler. Il est possible de faire des conversions entre tous ces formalismes, chacun ayant ses avantages et ses inconvénients.

### Angles d'Euler

Les angles d'Euler permettent de décrire une rotation à partir de rotations successives autour des trois axes principaux  $x$ ,  $y$  et  $z$ . L'ordre le plus souvent utilisé consiste à tourner d'abord autour de l'axe  $x$  d'un angle  $\alpha$ , puis de tourner autour du nouvel axe  $y$  d'un angle  $\beta$ , puis autour du nouvel axe  $z$  d'un angle  $\gamma$  (Figure 1.10). On note alors l'ordre de rotation **XYZ**. On peut aussi choisir d'autres ordres de rotation, par exemple **YXZ**, **ZYX**, **XYX**. Dans le cas de **XYX**,

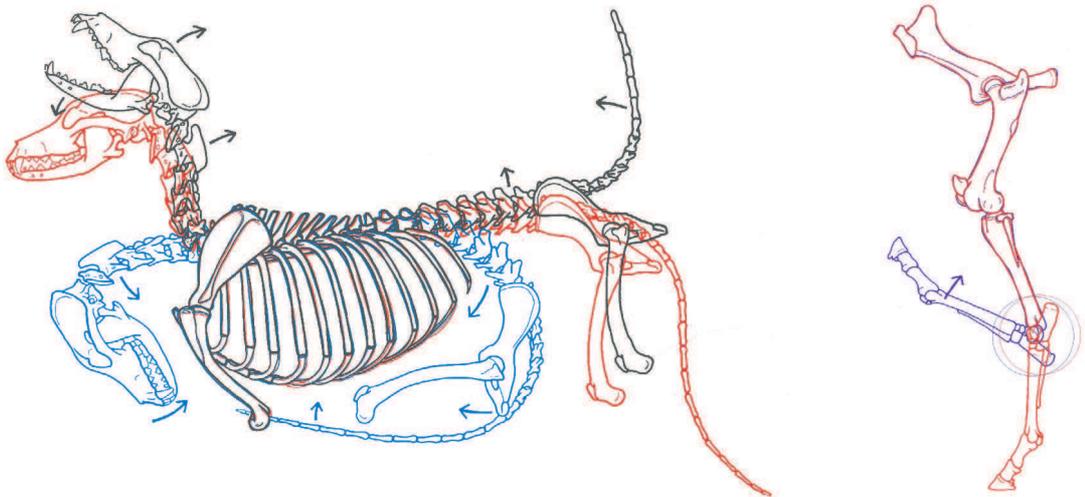


FIG. 1.9 – Les limites induites sur l'amplitudes de quelques articulations par la morphologie du squelette.

on tourne autour de l'axe  $x$ , puis autour de l'axe  $y$ , puis autour du nouvel axe  $x$  obtenu. Pour une présentation détaillée, voir [Dun02].

Notez que toute rotation en 3D peut être décrite par cet ensemble  $(\alpha, \beta, \gamma)$  de 3 paramètres. Les avantages des angles d'Euler sont leur côté intuitif et leur facilité d'utilisation. Ils sont très facile à manipuler dans les logiciels d'infographie, et faciles à visualiser : on peut en effet afficher le graphe des rotations selon  $x$ ,  $y$  et  $z$  au cours du temps. Cependant, cette notation est ambiguë : tous les angles sont définis à  $2\pi$  près (en radians), et plusieurs valeurs de  $(\alpha, \beta, \gamma)$  différentes peuvent représenter la même rotation. De plus, interpoler directement les angles d'Euler donne une vitesse de rotation qui n'est pas constante.

## Matrices

Les rotations peuvent aussi être représentées sous forme matricielle. Soient les rotations de base introduites au cours de la section précédente, les rotations d'angle  $\alpha$ ,  $\beta$  et  $\gamma$  autour des axes  $x$ ,  $y$ , et  $z$ . On peut les représenter respectivement par les matrices :

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix}, \begin{pmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{pmatrix}, \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Comme le produit de deux matrices de rotation est une matrice de rotation, la multiplication matricielle permet de composer les rotations. En multipliant ces trois matrices de base on

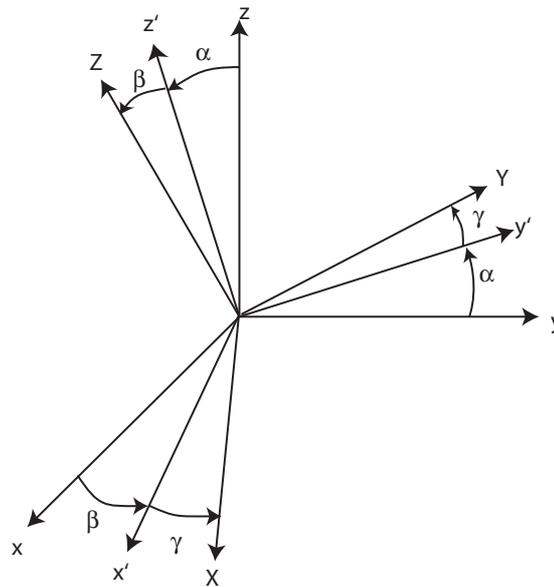


FIG. 1.10 – Les angles d'Euler : exemple de rotation selon la convention XYZ. Trois rotations successives sont effectuées : d'abord le repère  $(x, y, z)$  est tourné d'un angle  $\alpha$  autour de l'axe  $x$ . On obtient le repère  $(x', y', z')$ . Ensuite on tourne d'un angle  $\beta$  autour de l'axe  $y'$  et on obtient le repère  $(x'', y'', z'')$ . Enfin on tourne d'un angle  $\gamma$  autour de l'axe  $z''$  et on obtient le nouveau repère  $(X, Y, Z)$ .

peut obtenir une représentation matricielle de chaque rotation représentée par les angles d'Euler  $(\alpha, \beta, \gamma)$ . Les matrices permettent donc de formaliser la composition des rotations, avantage indéniable car nous travaillons avec une hiérarchie de repères. Le passage en notation matricielle permet de mettre en évidence un défaut des angles d'Euler, le *gimbal lock*. Dans le cas d'une rotation de paramètres  $(\alpha, \pi/2, \gamma)$ , la matrice de transformation devient :

$$\begin{aligned} & \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix} \\ &= \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & \sin\alpha & -\cos\alpha \\ 0 & \cos\alpha & \sin\alpha \\ 1 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & \sin\alpha\cos\gamma + \cos\alpha\sin\gamma & \sin\alpha\sin\gamma - \cos\alpha\cos\gamma \\ 0 & \cos\alpha\cos\gamma - \sin\alpha\sin\gamma & \sin\alpha\cos\gamma + \cos\alpha\sin\gamma \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & \sin(\alpha+\gamma) & -\cos(\alpha+\gamma) \\ 0 & \cos(\alpha+\gamma) & \sin(\alpha+\gamma) \\ 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

On obtient donc une transformation à un seul degré de liberté en  $\alpha + \gamma$  : on a perdu un degré de liberté. C'est dû au fait que la rotation d'axe  $y$  et d'angle  $\pi/2$  amène l'axe  $z$  sur l'ancien axe  $x$  : les deux rotations d'angle  $\alpha$  et  $\gamma$  s'effectuent autour du même axe.

Contrairement aux angles d'Euler, la notation par matrices est unique : chaque rotation 3D a une représentation matricielle unique. Cependant les informations sont redondantes : on utilise 9 paramètres pour stocker une transformation à 3 degrés de liberté. Et surtout, l'interpolation linéaire de deux matrices de rotations ne donne pas une matrice de rotation car les matrices obtenues alors ne sont pas forcément orthogonales. Le problème de l'interpolation reste donc à résoudre. C'est possible en introduisant la notation axe/angle.

### Axe-angle

Toute rotation 3D peut être représentée par une seule rotation d'angle  $\theta$  et d'axe  $u$ , où  $u$  est un vecteur 3D unitaire (Figure 1.11). Cette notation, contrairement aux précédentes, permet d'interpoler linéairement entre deux rotations. En effet, pour passer continûment d'une base à une autre, il suffit de calculer la rotation  $(u, \theta)$  qui passe d'une base à l'autre, puis d'appliquer les rotations  $(u, a)$  avec  $a \in [0, \theta]$ . Cette notation a deux inconvénients. D'abord, elle est ambiguë : la rotation  $(u, \theta)$  est la même que la rotation  $(-u, -\theta)$ . De plus on ne peut pas composer facilement les rotations définies de cette façon sans faire de conversion. Nous allons le faire en introduisant la notation par quaternions.

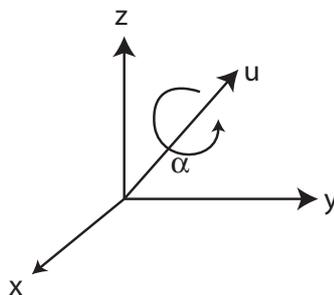


FIG. 1.11 – . Toute rotation 3D peut être représentée par une seule rotation d'angle  $\theta$  et d'axe  $u$ , où  $u$  est un vecteur 3D unitaire.

### Quaternions

Les rotations peuvent aussi être représentées par les quaternions unitaires, c'est-à-dire par les points situés sur la boule unité d'un espace de dimension 4. Un vecteur 3D  $(x, y, z)$  est représenté par le quaternion :

$$p = (0; x; y; z)$$

et la rotation d'axe  $theta$  et d'angle  $u$  est représentée par :

$$q_{\theta,u} = \left( \cos \frac{\theta}{2}, u_x \sin \frac{\theta}{2}, u_y \sin \frac{\theta}{2}, u_z \sin \frac{\theta}{2} \right)$$

La rotation de la rotation  $q$  au vecteur  $p$  est alors notée :

$$qp\bar{q}$$

Voir B pour un formulaire sur les quaternions, et [Wat99] pour une présentation détaillée. cette notation permet d'interpoler et de composer les rotations. Il reste la même ambiguïté que pour la notation axe/angle : on peut obtenir la même rotation en prenant l'angle et l'axe opposés, mais cela peut-être résolu en prenant des conventions simples. Cette notation est donc très pratique d'un point de vue mathématique, mais elle n'est pas facile à manipuler pour les artistes. On peut donc utiliser les quaternions en interne et les transformer en angles d'Euler lorsqu'une intervention utilisateur est nécessaire.

### 2.3 Les cartes exponentielles

Les *cartes exponentielles* consistent à représenter une rotation par un vecteur 3D dont l'orientation donne l'axe de rotation et la norme donne l'angle. On peut définir une fonction qui associe une rotation à chaque vecteur 3D [Gra98]. Cette notation permet d'éviter le *gimbal lock* et ne nécessite pas de renormalisation comme les quaternions unitaires. Cependant, l'interpolation n'est possible que dans un voisinage très réduit.

Nous venons de terminer de passer en revue les différentes manières de représenter l'orientation d'un repère 3D, c'est-à-dire de définir la pose de notre squelette d'animation 3D.

A présent que nous savons comment décrire le squelette d'animation, nous allons voir comment le recouvrir d'une peau pour lui donner une apparence réaliste.

## 3 Habillage du squelette

Le squelette d'animation est une structure utile pour contrôler et décrire un mouvement. Mais il n'a pas de volume en lui-même, et n'est habituellement visualisé qu'en représentant les joints par des chaînes de solides, ou même seulement des segments de droites (Figure 1.5). Cette méthode permet de visualiser simplement le mouvement, mais ne donne pas de résultat

satisfaisant pour le rendu, notamment à cause du manque de volume. Une solution simple consiste à créer des ellipsoïdes 3D qui sont attachées à chaque joint (Figure 1.12), chaque ellipsoïde étant attachée au repère d'un joint. C'est une technique simple pour donner du volume au modèle, mais elle donne un résultat peu réaliste visuellement.

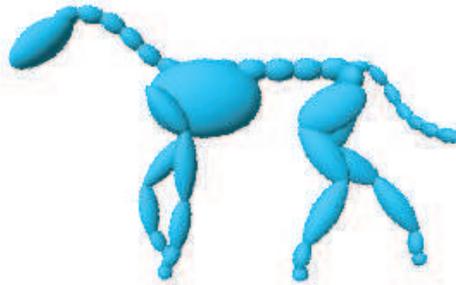


FIG. 1.12 – Habillage d'un squelette d'animation par une chaîne de solides.

Une technique classique mais plus avancée consiste à prendre une représentation visuellement réaliste de la surface de l'animal dans sa pose de repos, donnée sous la forme d'un maillage polygonal (Figure 1.5). Une image qui représente l'aspect de la surface de l'animal (sa peau) peut être plaquée sur ce maillage (Figure 1.13). Cette image est appelée *texture*. Pour cela, à chaque point 3D du maillage est attribué des *coordonnées de texture*, c'est-à-dire une coordonnée correspondante sur l'image. Des techniques d'interpolation permettent alors de déduire une couleur pour chaque position sur le maillage. Bien entendu, comme l'image correspond à la surface de l'animal mise à plat, elle subit des déformations, une partie importante du problème consiste donc à bien choisir ces coordonnées. Pour un état de l'art complet sur le plaquage de texture et la synthèse à partir d'exemples, voir [Lef05].

Il nous faut des outils permettant de spécifier comment déformer ce maillage lorsque l'animal se déplace. Nous allons présenter d'abord les méthodes par interpolation de formes-clés, qui ne nécessitent pas de squelette d'animation, puis les méthodes de déformation par *skinning*, qui sont plus utilisées pour les personnages articulés et correspondent mieux à notre approche, pour en venir enfin aux méthodes anatomiques, reposant sur une modélisation fine des déformations musculaires.

### 3.1 Interpolation de formes-clés

L'interpolation de formes-clés peut s'appliquer à un maillage sans squelette interne. Cette technique est donc très utilisée pour l'animation faciale, mais aussi pour l'animation de muscles

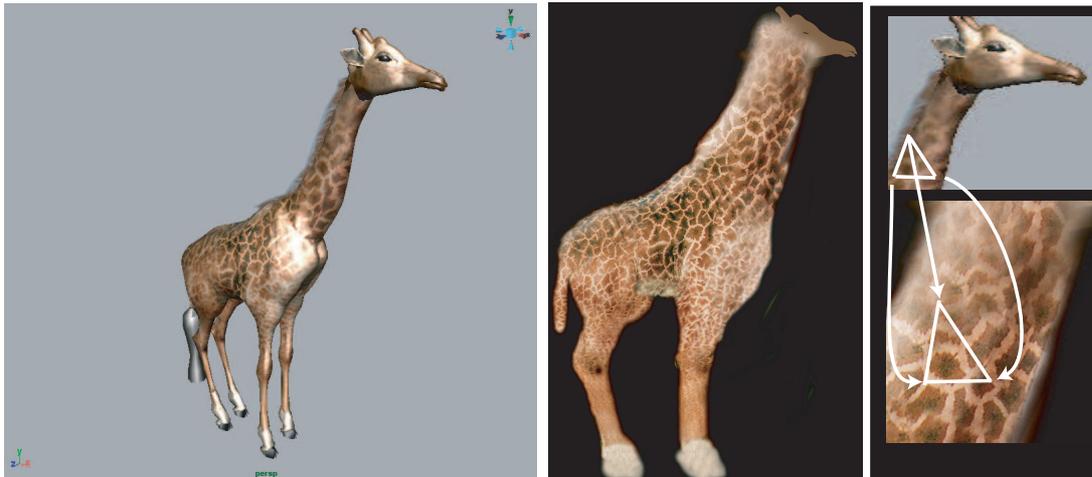


FIG. 1.13 – La texture est une image plaquée sur le maillage 3D. A chaque sommet 3D du maillage est associé une position dans l'image, ou coordonnée de texture.

[LrN98]. Elle consiste à mélanger plusieurs formes de base, appelées *formes-clés*, qui correspondent par exemple à différentes expressions : la tristesse, la joie, la colère pour les visages, ou bien des positions articulaires extrêmes pour les personnages articulés. La méthode la plus classique consiste à obtenir la position des sommets du maillage lors d'un mouvement quelconque par combinaison linéaire de la position des sommets dans les formes-clés. Avec  $n$  formes clés  $S_i$  ( $1 \leq i \leq n$ ) auxquelles on donne des poids respectifs  $w_i$ , la forme finale  $S_f$  est égale à :

$$S_f = \sum_{i=1}^n w_i S_i$$

Des méthodes plus récentes [ZOTX03, SZGP05, DSJ06] procurent des techniques d'interpolation plus sophistiquées que la combinaison linéaire, qui permettent d'effectuer des déformations plus adaptées aux exemples (Figure 1.14). Les problèmes liés à cette approche sont divers. Tout d'abord, les différentes formes clés ne sont pas indépendantes : Si certaines formes clés présentent des caractéristiques qui s'ajoutent ou s'annulent lors de l'interpolation, le résultat obtenu sera inattendu. Il est possible de corriger ces problèmes d'interpolation en ajoutant plus de formes-clés, mais cela représente un travail fastidieux, surtout que le risque de conflit entre formes augmente avec le nombre de formes-clés. De plus, changer de personnage demande de redessiner toutes les formes-clés. Enfin, les paramètres de contrôle utilisés par ces méthodes ne sont pas adaptés à l'animation de mouvements de locomotion.

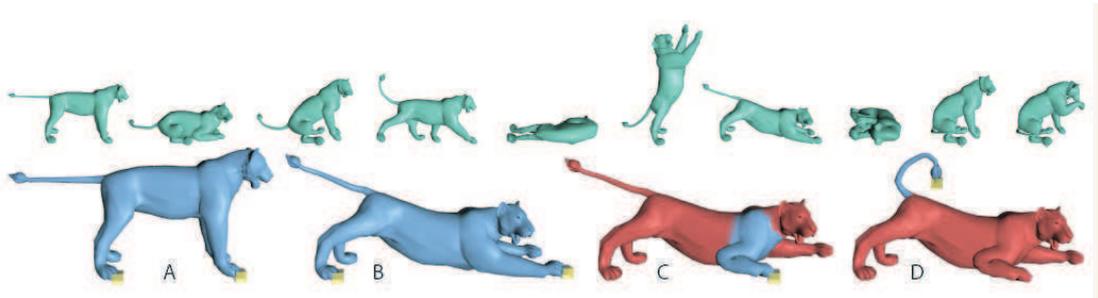


FIG. 1.14 – En haut : des poses-clés choisies pour un maillage de lion. En bas : un champ de déformation appris à partir des exemples permet d’interpoler entre les formes-clés pour déformer de façon réaliste le maillage en n’en déplaçant que certains sommets (d’après [SZGP05]).

### 3.2 Techniques de skinning

Le *skinning* [LCF00, Blo02] est une technique classique qui permet d’animer les sommets de la surface d’un objet en réponse au mouvement d’un squelette interne. Elle est donc bien adaptée aux créatures articulées 3D comme nos quadrupèdes, qui sont constitués d’un squelette d’animation et d’une peau. Dans le *skinning*, chaque sommet est influencé par le mouvement d’un ou plusieurs joints.

Dans le cas le plus simple, chaque sommet est attaché à un joint, et il subit la même transformation rigide que le joint auquel il est attaché. On parle alors de *skinning rigide*. Le maillage est découpé en morceaux rigides, qui se déplacent comme un armure qui entourerait le sujet. Cela conduit à des déformations non réalistes comme des étirements ou bien des auto-intersections du maillage, voire des trous autour des articulations (voir Figure 1.15).

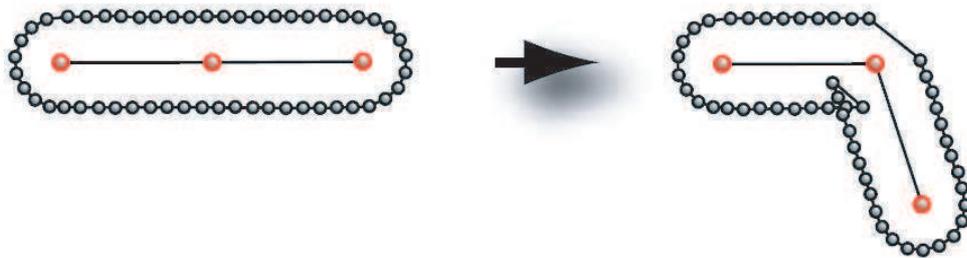


FIG. 1.15 – La technique du *skinning rigide* : chaque point du maillage se déplace avec le joint auquel il est attaché. Le maillage est étiré d’un côté de l’articulation et les morceaux s’intersectent de l’autre côté. Si le maillage a été découpé en membres, des trous peuvent même apparaître (image extraite de [Blo02]).

Une solution pour éviter les problèmes autour des articulations est d’utiliser un *smooth*

*skinning*, que nous désignerons simplement par *skinning* dans tout le reste de ce document. Le *skinning* autorise un sommet du maillage à être influencé par plusieurs repères. Dans ce cas, la position d'un sommet  $i$  est définie comme étant une somme pondérée des positions qu'il aurait si il était influencé par un seul des repères dont il dépend (Figure 1.16). Pour un sommet  $P_i$  influencé par  $n$  repères, la position finale  $P'_i$  après *skinning* est donnée par :

$$P'_i = \sum_{j=1}^n (w_{ij} T_j) v_i$$

avec :

- $v_i$  la position du sommet  $P_i$  avant transformation (position neutre)
- $T_j$  la matrice de transformation associée au  $j$ -ième repère.
- $w_{ij}$  le poids associé  $j$ -ième repère pour le sommet  $P_i$ .

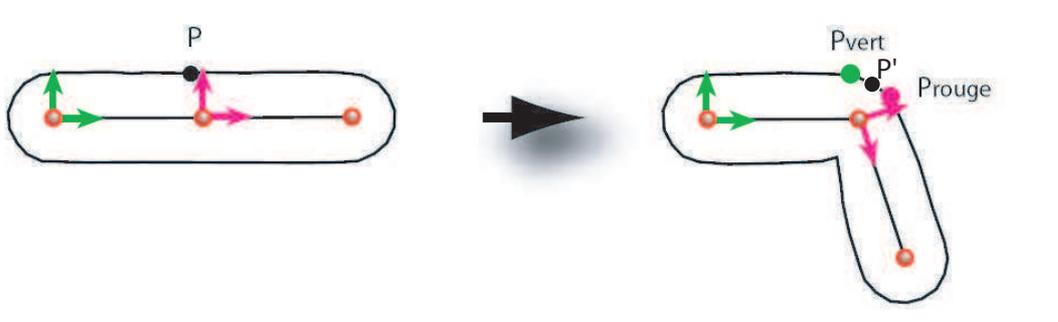


FIG. 1.16 – Illustration de la technique du *skinning* avec deux repères. Le point obtenu  $P'$  est une combinaison linéaire des positions qu'il aurait si il était influencé seulement par le premier repère ( $P_{vert}$ ) ou par le second ( $P_{rouge}$ ).  $P'$  se trouve obligatoirement sur la droite reliant  $P_{vert}$  à  $P_{rouge}$ .

Les  $w_{ij}$  expriment la dépendance de  $P_i$  au repère  $j$ . Ils définissent la manière dont le maillage va se déformer sous l'influence du squelette, et jouent un rôle essentiel au niveau de l'apparence visuelle du résultat. Classiquement, les poids correspondants à chaque joint sont "peints" à la main par l'infographiste sur la surface du modèle. Cette opération est longue et demande énormément de savoir-faire. On peut cependant calculer les poids automatiquement en fonction de la distance du sommet donné aux différents joints. Ceci permet d'obtenir des poids raisonnables qui varient sans discontinuité le long du maillage, tant que les parties du maillage considérées sont simples (comme un bras, une jambe). Cette technique ne fonctionne plus du tout au niveau du torse et des épaules par exemple. Généralement, on utilise donc les poids calculés automatiquement comme une première approximation qu'on ajuste ensuite à la main. Notez que, comme on déforme le maillage placé en pose de repos, le choix de cette pose est très important.

Un des problèmes de cette technique est que dans le cas où un point dépend de deux repères, sa position finale se situe obligatoirement sur le segment reliant les deux points représentant sa position dans chacun des repères. Dans le cas du coude par exemple, cela ne correspond pas à une déformation naturelle (Figure 1.17(a)). D'autre part, l'édition des poids à la main est un travail fastidieux qui demande beaucoup d'essais-erreurs pour adapter les poids de skinning à tous les mouvements possible. L'amélioration automatique des poids [Blo02] ou l'utilisation du skinning en combinaison avec d'autres techniques [LCF00, KJP02] permettent de corriger certains de ces problèmes.

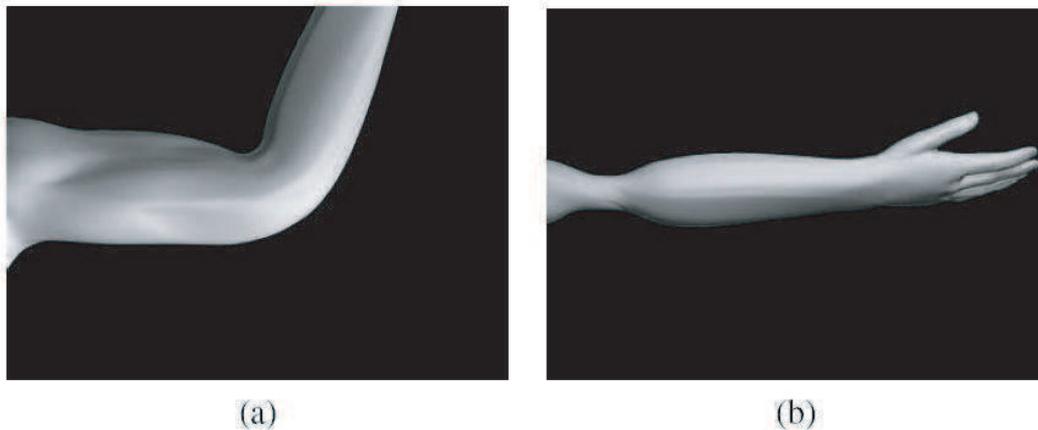


FIG. 1.17 – Problèmes liés à la technique du skinning : (a)-aplatissement au niveau du coude ; (b)-amincissement au niveau de l'articulation quand le bras est vrillé.

Certaines méthodes utilisent des modèles dynamiques pour simuler la déformation de la peau. Ces modèles dynamiques peuvent utiliser des systèmes masses-ressorts [LTW95], des éléments finis [CGC<sup>+</sup>02] ou de l'analyse modale [JP02]. Cependant ces techniques dynamiques laissent moins de contrôle à l'artiste et sont donc moins populaires. Malgré ses défauts, la technique du skinning reste la plus utilisée lorsqu'il s'agit d'animer des créatures articulés.

### 3.3 Techniques de modélisation anatomique

Contrairement au skinning qui lie directement la pose du squelette d'animation avec les déformations de la peau, les méthodes anatomiques décrivent explicitement la forme des muscles et des tissus entre le squelette et la peau. Le modèle proposé par Wilhelms et al. [JG97] s'intéresse aux animaux et comprend les os, les muscles, les tissus non musculaires et enfin la peau (Fig. 1.18). Les muscles sont des cylindres déformés qui changent de forme avec le mouvement du squelette. Plus récemment, [SWG02] permet, à partir d'un modèle anatomique

canonique de l'animal souhaité, et de la donnée d'une série de mesures, de déformer le modèle canonique pour d'adapter à un animal précis.

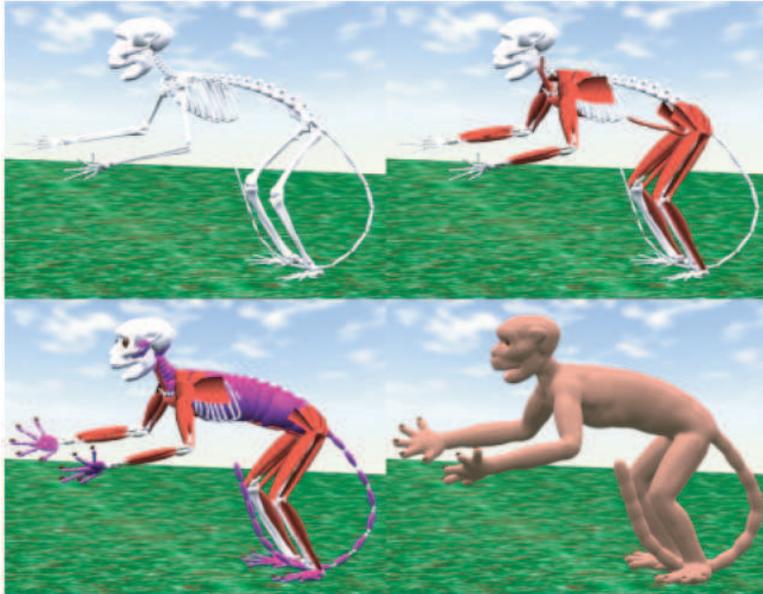


FIG. 1.18 – Le modèle anatomique de Wilhelms et al. comprend le squelette (en blanc), les muscles (en rouge), des tissus non musculaires (en violet) et enfin la peau (en bas à droite).

Pratscher et al. [PCLS05] adaptent un modèle générique de squelette muni de muscles à l'intérieur d'un maillage 3D muni d'un squelette d'animation, ce qui leur permet d'obtenir des déformations réalistes (Figure 1.19). Des techniques de simulation des muscles existent [CZ92, TBHF03, TSB<sup>+</sup>05], mais leur complexité calculatoire et le manque de contrôle laissés à l'utilisateur font qu'elles sont peu utilisées.

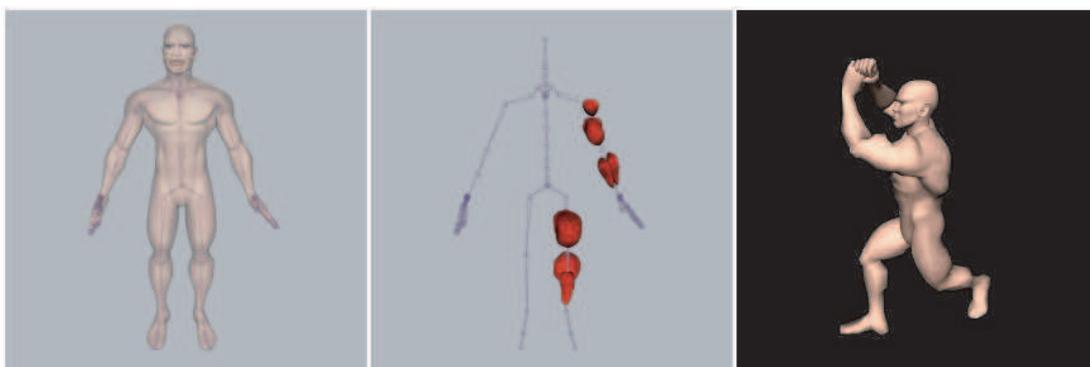


FIG. 1.19 – Pratscher et al. [PCLS05] adaptent un modèle générique de muscles à l'intérieur d'un maillage 3D muni d'un squelette d'animation.

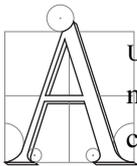
## **4 Bilan**

Ce chapitre a d'abord présenté la structure anatomique des quadrupèdes réels, en introduisant les différentes structures qui leur permettent de se déplacer. Ensuite nous avons dressé un état de l'art des techniques existantes qui permettent de modéliser ces quadrupèdes. Nous avons vu qu'un quadrupède virtuel est constitué d'un squelette d'animation proche du squelette réel mais adapté, auquel on adjoint une surface qui se déforme en fonction des mouvements du squelette d'animation. Dans le chapitre suivant nous allons détailler les méthodes permettant de donner vie à ces quadrupèdes 3D en les dotant de mouvement.

---

# Techniques de contrôle du mouvement

---



DU CHAPITRE PRÉCÉDENT, nous avons détaillé les méthodes classiques qui permettent de construire un modèle 3D de quadrupède : Ce dernier est modélisé comme un squelette d'animation auquel est liée une surface 3D qui représente la peau. Pour doter ce modèle 3D de mouvement, il suffit de faire varier la pose du squelette d'animation au cours du temps. La surface de l'animal, qui est attaché au squelette d'animation, se déformera alors pour créer du mouvement 3D volumique. Dans ce chapitre, nous présentons d'abord une étude théorique du mouvement des quadrupèdes réels, et détaillons les différentes allures qui font la richesse de leur mouvement. Ensuite, nous dressons un état de l'art des méthodes existantes qui permettent d'animer des créatures 3D.

## 1 Caractéristiques du mouvement des quadrupèdes réels

Pouvoir se déplacer est indispensable aux créatures terrestres, pour la recherche de nourriture, la chasse, la fuite devant un prédateur et la reproduction. Un quadrupède est toujours

en interaction avec son environnement. Comme il subit l'effet de la gravité, il doit utiliser ses pattes comme des points d'appui pour soutenir son corps, ce qui lui permet de se maintenir dans un état d'équilibre. Il exerce une poussée sur le sol, qui par réaction produit une force de propulsion.

Lorsqu'il se déplace, l'animal doit lever successivement ses pattes. La manière de coordonner les déplacements de ces appuis dans l'espace et dans le temps définit l'*allure* qu'il utilise. Les allures les plus connues incluent la marche, le trot et le galop. Nous les présenterons plus loin dans cette section. Tous les quadrupèdes n'utilisent pas toutes les allures possibles : les allures utilisées varient selon les espèces, et sont étroitement corrélées à la morphologie de l'animal (proportions du corps, répartition de masse, etc.). Les allures sélectionnées pour une vitesse de déplacement donnée sont celles qui procurent un meilleur compromis entre *stabilité* et *manoeuvrabilité*.

### **Stabilité**

En équilibre statique, la somme des forces de réaction du sol sur les quatre membres s'oppose au poids du corps. La stabilité du corps dépend de la capacité de résistance de l'axe vertébral aux forces qui s'exercent sur lui. Pour avoir un équilibre statique stable, le quadrupède doit avoir un centre de gravité situé près du sol (donc des membres courts), et une masse importante. Lorsque le quadrupède se déplace, son équilibre dépend de la durée des appuis des membres, des forces appliquées au sol et de la vitesse de progression. A vitesse lente, le corps sera souvent soutenu par trois membres (appui tripode), et l'animal cherche à chaque instant à être en équilibre statique. Parmi tous les ordres d'entrée en action des membres qui sont possibles, le quadrupède choisit celui qui assure l'appui tripode le plus large possible sur le sol. A vitesse lente, les deux séquences d'appuis possibles sont (Figure 2.1) :

- La *séquence latérale*, pour laquelle le posé d'un antérieur précède celui du postérieur diagonal, et l'appui d'un postérieur est suivi de l'appui de l'antérieur du même côté.
- La *séquence diagonale*, pour laquelle le posé d'un antérieur précède celui du postérieur du même côté, et l'appui d'un postérieur est suivi de l'appui de l'antérieur diagonal.

La contrainte de stabilité conduit la plupart des quadrupèdes à préférer la séquence latérale, car l'ordre de déplacement des pattes conduit à maintenir le centre de gravité du corps au-dessus du triangle d'appui formé par les trois membres de soutien (Figure 2.1). A vitesse plus grande, le corps tend à n'être plus soutenu que par deux membres et le support n'est plus un triangle, mais une ligne (appui latéral ou appui diagonal). Les temps de contact avec le sol sont plus courts. L'équilibre n'est plus statique, mais dynamique.

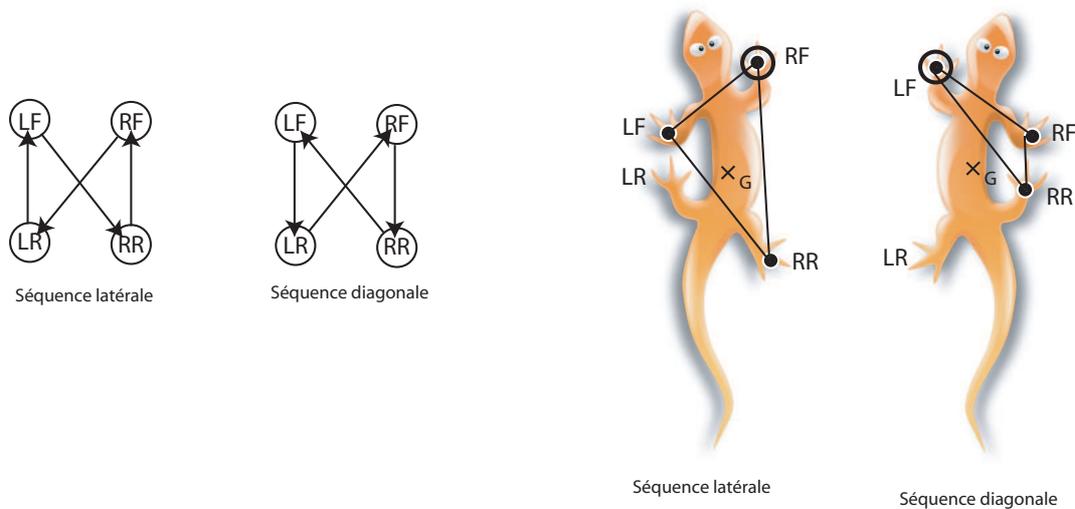


FIG. 2.1 – A gauche : ordre des posés pour la séquence latérale et la séquence diagonale. A droite : lorsque la patte la plus en avant (celle marquée d'un double cercle) vient d'être posée, la séquence latérale donne un meilleur équilibre statique que la séquence diagonale. Les pieds sont notés RF pour right front (avant droit), LF pour left front (avant gauche), RR pour right rear (arrière droit), et LR pour left rear (arrière gauche).

### Manœuvrabilité

La stabilité n'est pas la seule qualité recherchée pour la locomotion. A cause des conditions externes, comme la présence d'obstacles, ou une situation de fuite ou de poursuite, l'animal doit pouvoir contrôler rapidement sa direction et sa vitesse. C'est la contrainte de *manœuvrabilité*. Pour obtenir une bonne manœuvrabilité, l'animal doit avoir un centre de gravité élevé, des membres longs et une masse faible. Pour atteindre des vitesses très élevées il devra avoir une grande souplesse, surtout au niveau de la colonne vertébrale. Ces conditions sont antagonistes avec la contrainte de stabilité, l'animal doit donc constamment rechercher un compromis.

A présent que nous sommes familiers avec les notions de base concernant la locomotion des quadrupèdes, nous allons présenter les principales allures qu'ils utilisent.

### 1.1 Les allures

Au paragraphe précédent, nous avons défini une allure comme étant une manière de coordonner les déplacements des appuis dans l'espace et dans le temps. Renous [Ren94] définit une allure comme étant *un déplacement séquentiel membres dans l'espace et le temps*. C'est donc la séquence de déplacement des membres qui définit une allure. Les caractéristiques communes des allures des quadrupèdes sont les suivantes :

- Chaque membre passe par deux phases répétitives : le *posé*, où le membre est en contact

avec le sol, et le *levé*, où l'extrémité du membre est suspendue en l'air.

- La succession d'un posé et d'un levé définit un *cycle de mouvement*. La durée d'un cycle est la même pour les quatre membres. En revanche, la durée du posé et du levé à l'intérieur d'un cycle peut varier.
- Au cours d'un *cycle de révolution*, les mouvements des quatre membres se succèdent dans l'ordre de leur entrée en action.
- Le déplacement est diagonal en général, c'est-à-dire qu'au cours d'un demi-cycle de révolution, deux membres diagonaux se succèdent.

Une allure se définit par les durées des phases de posé et de lever et par les décalages qui séparent les levés (ou les posés) des différents membres. C'est une information complexe et difficile à analyser. Pour rendre l'analyse plus facile, on peut par exemple représenter les allures de manière visuelle, en représentant la succession dans le temps des différents types d'appuis (Figure 2.2).

Les allures utilisées par les quadrupèdes se divisent en deux groupes : les allures *symétriques* et les allures *asymétriques*.

### Les allures symétriques

Une allure est *symétrique* quand le décalage de temps entre les posés des membres antérieurs est le même qu'entre les posés des membres postérieurs (Figure 2.3). Le décalage temporel entre pattes d'une même paire et le décalage temporel entre la paire avant et la paire arrière permet alors de définir la démarche. Selon la vitesse, il y a alternance de 3 et 4 membres en contact (vitesse lente), 3 et 2 (vitesse moyenne), ou moins (vitesse rapide).

**La marche** De toutes les allures, la marche est la plus lente et celle qui vient la première à l'esprit lorsqu'on évoque la locomotion. Elle est souvent appelée *pas*, dans le cas du cheval par exemple. Par défaut le terme marche désigne la marche latérale, c'est-à-dire une allure à quatre temps qui utilise la séquence latérale, selon la succession de posés détaillée sur la Figure 2.4. Il existe une marche diagonale, mais elle ne serait utilisée que par les primates quadrupèdes et quelques autres espèces. A une vitesse normale, l'animal est toujours en appui sur 2 ou 3 jambes, ce qui fait de la marche une allure très stable. Les appuis latéraux et diagonaux sont utilisés. Si l'animal marche très lentement, il alterne entre des phases à 3 et 4 appuis.

**L'amble** L'amble est une variante de la marche qui permet de se déplacer plus rapidement. C'est une marche accélérée, avec la même suite de posés, mais les appuis sont sur 2 ou 1 pattes (Figure 2.5). Les appuis latéraux et diagonaux sont utilisés.

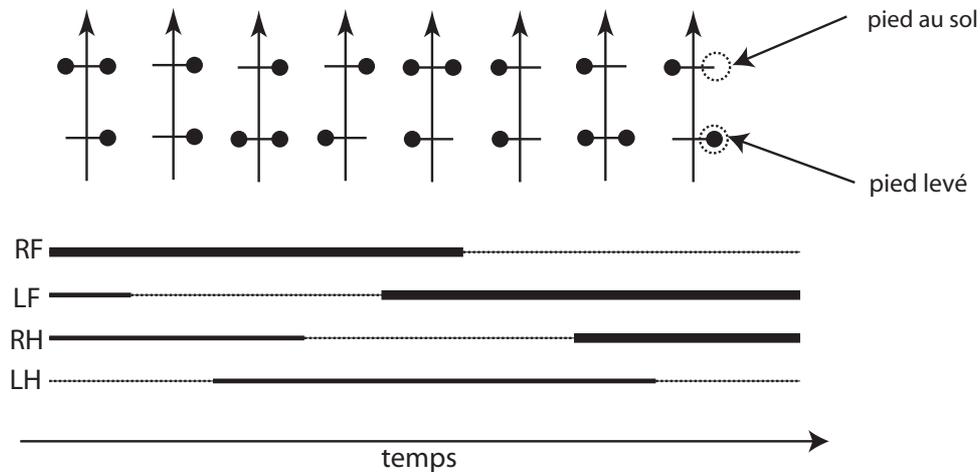


FIG. 2.2 – Exemples de représentations d’une allure de manière visuelle. Une représentation doit inclure la coordination entre les membres, donc les successions leur levés et posés. *En haut* : On représente l’allure comme une succession de phases d’appui ou de suspension. A chaque phase, une combinaison différente de membres est en contact avec le sol. La flèche représente la tête de l’animal. Pour chaque phase, les disques noirs désignent les pattes qui sont au sol, et les phases se succèdent dans le temps de gauche à droite. *En bas* : Le temps se déroule de gauche à droite. Pour chaque membre, les segments en traits pleins désignent les moments où le membre est au sol et les segments en pointillés désignent les moments où les membres sont levés. A noter que cette notation est plus riche car elle permet d’exprimer la longueur des phases. Les segments en gras désignent les posés qui font partie d’une même *séquence*, comme nous le définirons au chapitre 7.

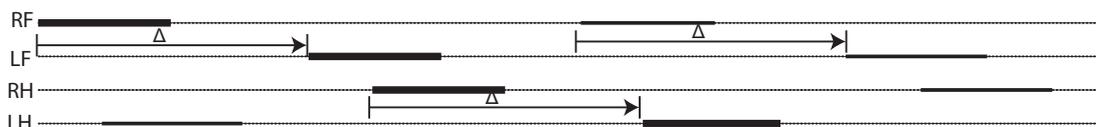


FIG. 2.3 – Allure symétrique : le décalage  $\Delta$  entre les posés des pattes avant est le même qu’entre les posés des pattes arrières.

**Le trot** Le trot est une allure plus rapide. C’est une allure à deux temps, où les pattes diagonales sont soulevées en même temps (Figure 2.6). Au cours d’un cycle de trot on observe deux phases de suspension, c’est-à-dire deux phases où aucune patte n’est posée au sol.

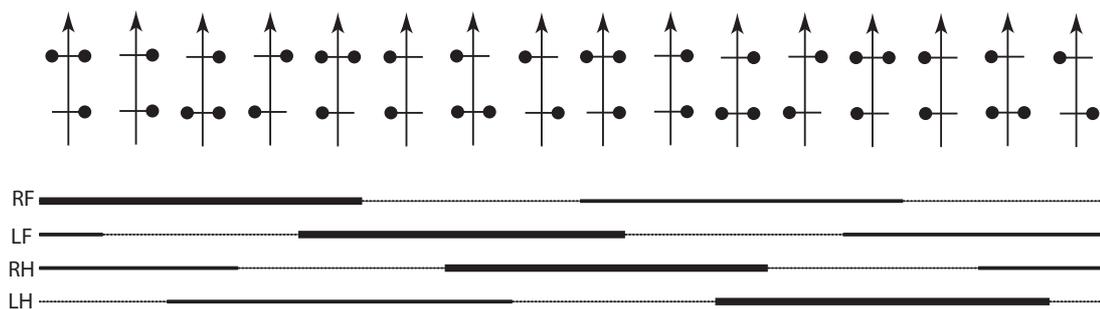


FIG. 2.4 – La marche.

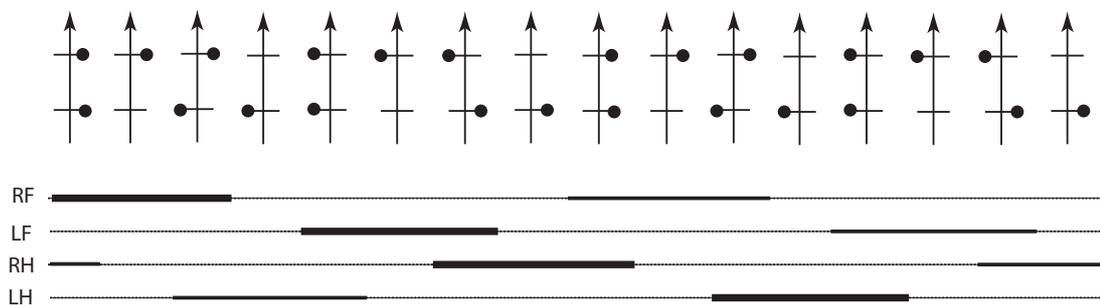


FIG. 2.5 – L'amble.

**Le rack** Comme le trot, le rack est une allure à deux temps qui inclut deux phases de suspension, mais ce sont les pattes latérales qui se déplacent en même temps (Figure 2.7). Cette allure est utilisée naturellement par le chameau et la girafe, et parfois par certains animaux comme le chien lorsqu'ils sont fatigués, car elle leur permet de se reposer. Les chevaux sont rarement dressés pour le rack car c'est une démarche désagréable pour le cavalier. Si on observe de profil la silhouette d'un animal qui utilise le rack ou le trot, il est très difficile de distinguer ces allures.

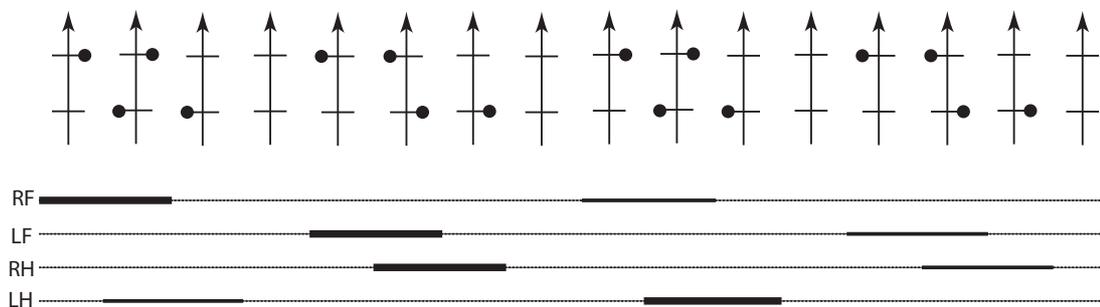


FIG. 2.6 – Le trot.

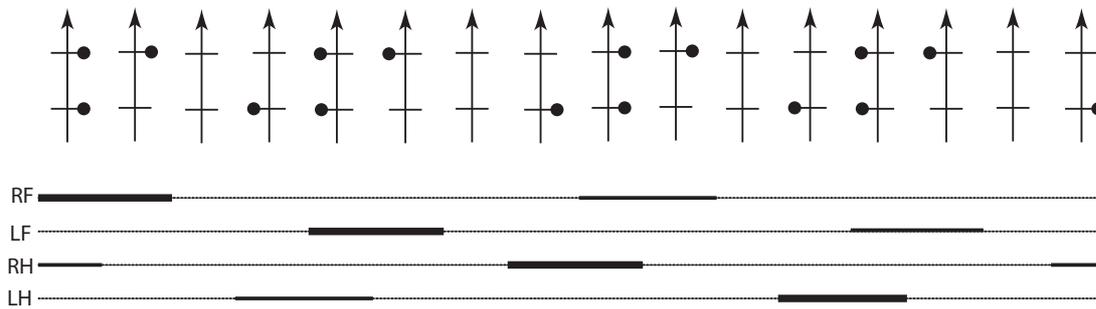


FIG. 2.7 – Le rack.

### Les allures asymétriques

On parle d'allure asymétrique quand le décalage temporel entre les pattes avant n'est pas le même que de décalage temporel entre les pattes arrière (Figure 2.8). Ces allures sont en général plus rapides que les allures symétriques.

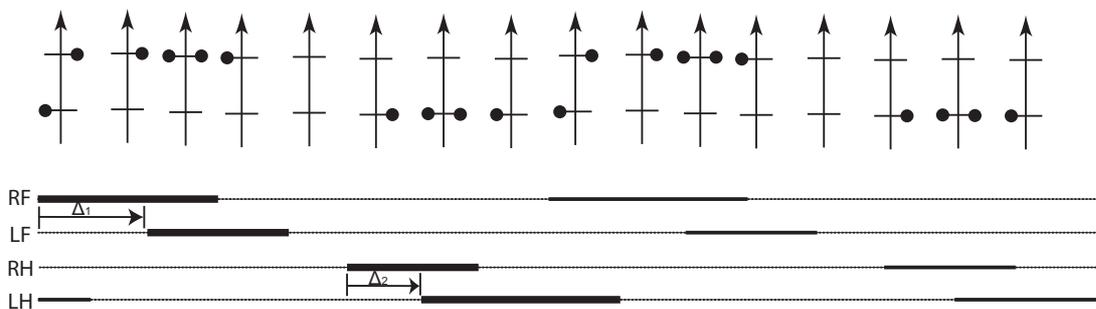


FIG. 2.8 – Allure asymétrique : le décalage entre les posés des pattes avant n'est pas le même qu'entre les posés des pattes arrière.

**Le canter** Le canter est une allure asymétrique utilisée par les chevaux, plus rapide que le trot mais plus lente que le galop. C'est une allure non symétrique à trois temps. Elle est caractérisée par les posés successifs d'une patte arrière, puis de la diagonale ne contenant pas cette patte arrière, puis de la patte avant restante (Figure 2.9).

**Le gallop transverse** Le galop est une allure très rapide qui se présente en deux variantes. La première est le galop transverse. C'est une allure qui utilise la séquence latérale, le posé d'une patte avant étant suivi du posé de la patte arrière diagonale (Figure 2.10). Le galop transverse est utilisé par le cheval et la majorité des quadrupèdes. En général un cycle de galop transverse comporte une phase de suspension, mais dans quelques cas il peut en présenter deux.

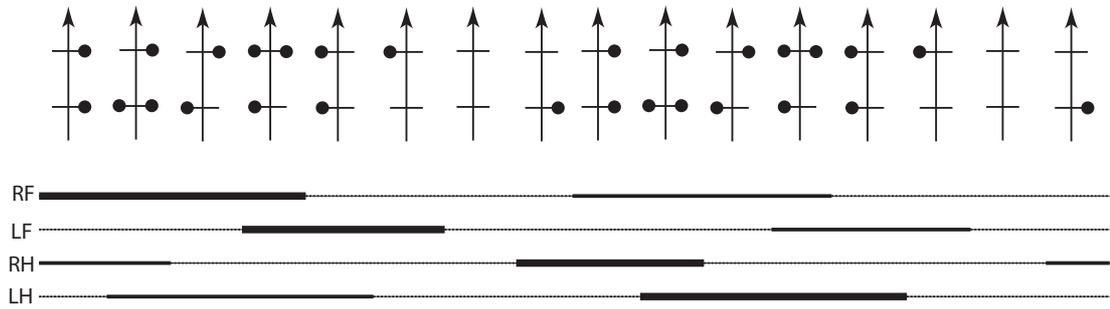


FIG. 2.9 – Le canter.

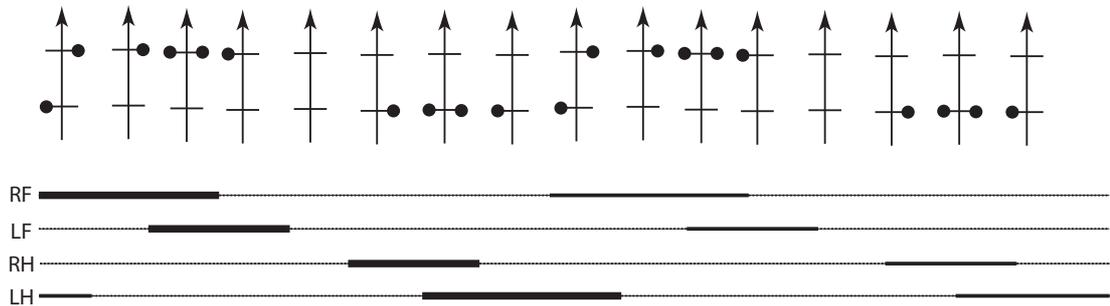


FIG. 2.10 – Le galop transverse.

**Le galop rotatoire** Le galop rotatoire est une deuxième variante du galop dans laquelle les posés se font dans un ordre circulaire (Figure 2.11). Le galop rotatoire permet d'atteindre des vitesses plus élevées, probablement parce qu'il implique un mouvement de la colonne vertébrale plus propice à la vitesse. En général le galop rotatoire présente deux phases de suspension. C'est une allure utilisée par le chien et le cerf par exemple.

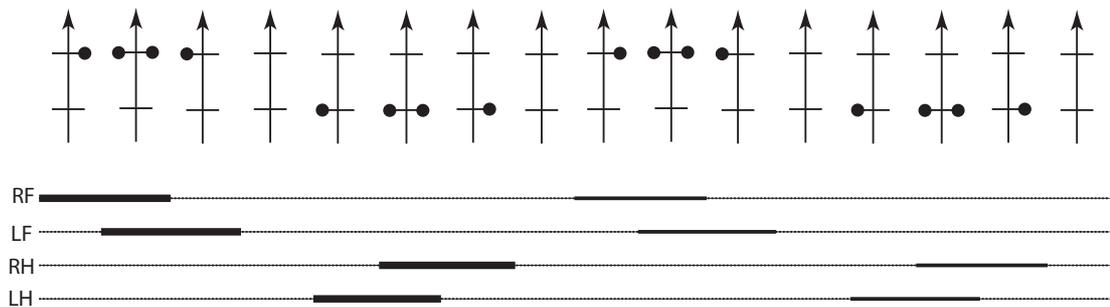


FIG. 2.11 – Le galop rotatoire.

**Le demi-bond et le bond** Le demi-bond et le bond sont des cas extrêmes de galop. Dans le demi-bond, les deux pattes arrière prennent appui en même temps pour donner l'impulsion (Figure 2.12). Dans le bond, les pattes avant prennent aussi appui en même temps (Fi-

gure 2.13). Ces allures permettent d'atteindre des vitesses extrêmes.

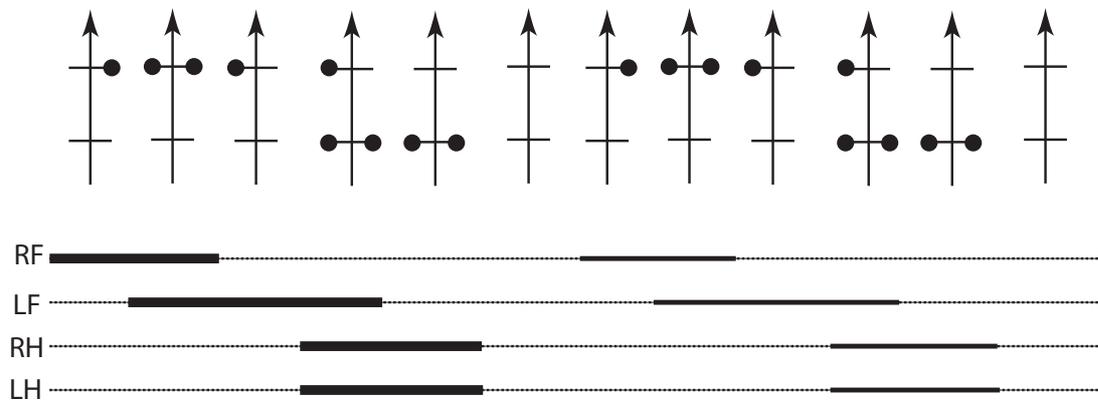


FIG. 2.12 – Le demi-bond.

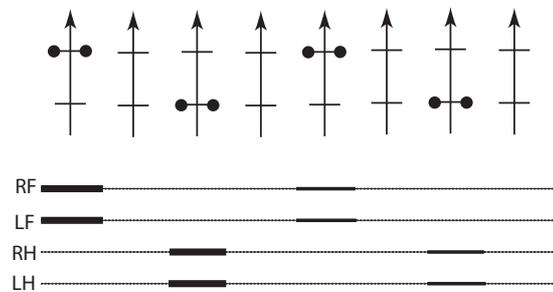


FIG. 2.13 – Le bond.

Nous avons passé en revue les principales allures utilisées par les quadrupèdes. Il existe d'autres allures plus rares ou non naturelles, le cheval par exemple étant capable d'apprendre un grand nombre d'allures non standard par dressage. Cependant nous n'aborderons que ces allures communes. A présent que nous sommes plus familiers avec le mouvement des quadrupèdes, nous allons aborder les différentes techniques qui permettent de contrôler le mouvement de quadrupèdes virtuels.

## 2 Animation Cinématique

L'animation cinématique d'un objet 3D rigide consiste à donner, pour chaque instant, sa position et son orientation en 3D. Le temps est discrétisé en instants, ou pas de temps, et on définit une fonction qui donne la position et l'orientation de l'objet pour chaque pas de temps discret. Dans le cas de créatures articulées comme les quadrupèdes, on contrôle la pose du squelette d'animation à chaque pas de temps. Pour contrôler la pose d'une articulation, deux

méthodes peuvent être utilisées : la cinématique directe et la cinématique inverse.

## 2.1 Cinématique directe

La cinématique directe consiste à positionner une chaîne articulée en donnant explicitement une valeur de rotation 3D pour chaque élément de la chaîne (Figure 2.14).

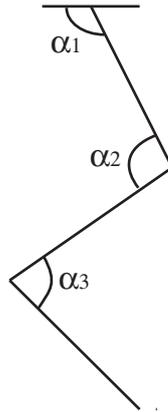


FIG. 2.14 – Cinématique directe : on définit le positionnement des éléments d'une chaîne articulée en réglant les orientations de chaque articulation. Amener l'extrémité de la chaîne à un endroit précis est difficile.

Lorsqu'on définit un mouvement de cette façon, on évite de donner explicitement une pose pour chaque pas de temps. On précise la pose correspondant aux étapes importantes du mouvement, les *poses-clés*, et les poses intermédiaires sont calculées par interpolation. Par exemple, si on veut qu'une créature plie le coude entre l'instant  $t_1$  et l'instant  $t_2$ , on définit la valeur de rotation du coude à  $t_1$  et à  $t_2$ , et on interpole les valeurs de rotation pour tout l'intervalle de temps  $[t_1, t_2]$ . Cette technique permet de gagner du temps en se concentrant sur les positions importantes du mouvement au lieu de préciser tous les détails d'un mouvement continu. De plus l'édition des poses-clés permet de modifier un mouvement beaucoup plus rapidement que si on avait à redéfinir la pose du quadrupède à chaque instant. On peut notamment de changer très rapidement le timing d'un geste en modifiant seulement la position temporelle d'une ou plusieurs clés.

L'animation par interpolation de poses clés est énormément utilisée en pratique car elle donne un contrôle total sur le mouvement. En général, l'interpolation est effectuée à l'aide de courbes splines (splines dans l'espace des quaternions pour les rotations) pour donner un mouvement lisse. Les outils d'animation permettent aussi de régler les tangentes des courbes aux positions clés, ce qui est très utile pour régler la dynamique du mouvement et obtenir un résultat expressif. Il est aussi possible d'utiliser du mouvement 3D issu de capture de mouve-

ment (voir Section 1) et de l'ajuster aux positions-clés données par l'utilisateur [WP95]. Ceci permet de synthétiser un mouvement relativement complexe avec peu de clés.

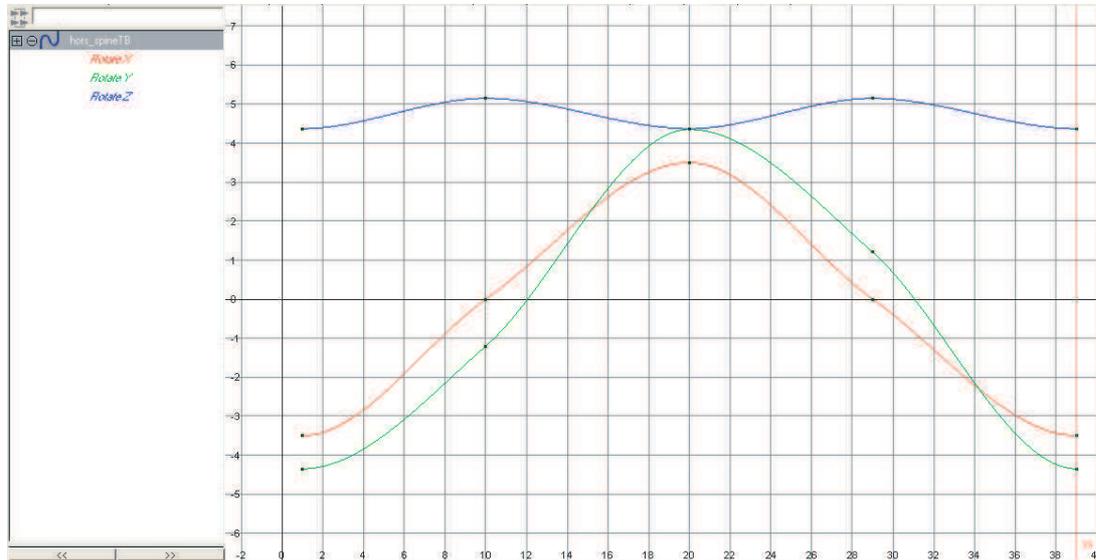


FIG. 2.15 – Animation par position-clés. Les valeurs des trois rotations du joint sont définies à quelques positions dans le temps (les carrés sur les courbes sont les clés). Ensuite toutes les valeurs intermédiaires sont interpolées.

Le principal défaut la cinématique directe est qu'il est très difficile d'amener les extrémités des chaînes articulées à une position précise dans l'espace. Comme on n'agit que sur des valeurs de rotations, on ne peut pas agir facilement sur les translations. Par exemple, il est très difficile d'éviter qu'un pied ne pénètre dans le sol entre deux clés. Dans ce cas, on utilise la cinématique inverse.

## 2.2 Cinématique inverse

La cinématique inverse est une technique issue de la robotique. Elle consiste à calculer le positionnement des éléments d'une chaîne articulée à partir des positions des deux extrémités de la chaîne [GM85] (Figure 2.16). Le mouvement de toute la chaîne articulée est alors contrôlé en déplaçant ses extrémités.

Calculer de façon analytique les rotations des éléments de la chaîne en fonction des positions des extrémités est possible dans certains cas particuliers, par exemple pour les mains et les pieds de personnages [BKK<sup>+</sup>85, TGB00]. Des solveurs analytiques spécialisés ont été utilisés pour diverses applications précises, comme la marche [SM01]. Mais dans le cas général, les solveurs analytiques ne sont pas adaptés et on utilise des méthodes numériques itératives.

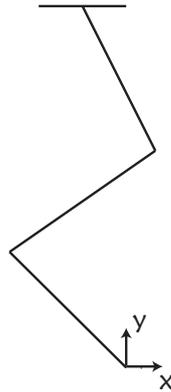


FIG. 2.16 – Cinématique inverse : Le mouvement de la chaîne articulée est contrôlé par les positions de ses extrémités. Les valeurs de rotations de toutes les articulations intermédiaires sont calculées en fonctions de ces positions.

Les méthodes classiques utilisent l'inverse de la matrice jacobienne [MK85], ou bien sa transposée [WE84, SS88]. D'autres méthodes résolvent ce problème comme une minimisation d'énergie [BMW87, ZB94].

Souvent on veut définir plusieurs contraintes de cinématique inverse en même temps, par exemple le pied d'un personnage au sol tout en le faisant saisir un objet avec la main. Dans ce cas, des conflits peuvent apparaître. Une solution consiste à pondérer les différentes tâches [BMW87, ZB94] ou bien d'utiliser un système de priorités entre contraintes [MK85, Bae01]. En définissant des contraintes différentes, ainsi que des priorités entre les contraintes, il est possible de générer des mouvements plus réalistes tout en laissant le soin à l'utilisateur de spécifier ces différentes contraintes et leurs priorités dans le cas de conflits [YN03b, CB04].

Kulpa et al. [KMA05, Kul05] définissent une représentation normalisée du squelette. Cette représentation permet de représenter la pose d'un squelette de façon indépendante de la longueur des membres (Figure 2.17). Ceci permet de positionner une extrémité et de définir des contraintes indépendamment de la morphologie.

Nous venons de voir comment animer un quadrupède 3D en contrôlant la cinématique de son squelette d'animation. Pour cela, nous définissons des positions-clés qui représentent les étapes importantes du mouvement, et les poses situées entre les poses-clés sont calculées par interpolation. Selon les besoins, ces poses-clés peuvent être définies à l'aide de cinématique directe et/ou de cinématique inverse. Par exemple, on pourra contrôler le mouvement d'une patte en utilisant la cinématique inverse pendant les phases de posé pour pouvoir facilement fixer le pied au sol, et la cinématique directe pendant les phases de levé pour contrôler le mouvement articulation par articulation. Dans la section suivante, nous allons aborder une autre technique d'animation : l'animation par modèle physique.

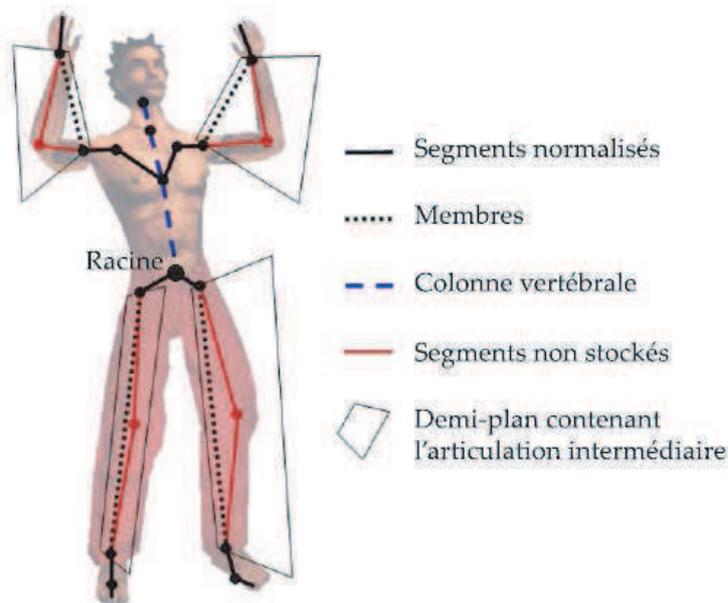


FIG. 2.17 – . Le squelette normalisé introduit par Kulpa et al. [KMA05]

### 3 Animation Physique

Si l'animation cinématique permet d'avoir un contrôle total sur le mouvement observé, l'utilisateur est le seul garant de la qualité du mouvement final. Les lois fondamentales de la dynamique, notamment, n'ont a priori aucune raison d'être respectée par le mouvement obtenu. Les techniques d'animation par modèles physiques introduisent les notions de masse et de force dans l'animation. La créature à animer n'est plus seulement un squelette d'animation entouré d'un maillage, mais est doté d'une distribution de masse. Les articulations, elles, sont capables d'exercer des forces qui vont mettre le corps en mouvement. Les méthodes d'animation physique peuvent être séparées en deux grandes classes : les méthodes d'animation dynamique et les méthodes d'animation par optimisation.

#### 3.1 Animation Dynamique

L'animation dynamique, introduite par [IC87, AG85], permet de calculer un mouvement qui dépend des lois de la mécanique et qui prend en compte les causes qui ont généré ce mouvement. La position, la vitesse et l'accélération des différents repères du squelette sont directement liés aux forces qui leur sont appliquées. Ces forces sont extérieures, comme la gravité ou l'effet du vent et intérieures, comme les forces appliquées par les muscles sur les articulations. En appliquant les lois de la mécanique des corps rigides, on peut prédire un

mouvement physiquement réaliste. Mais cela se paie par un manque de contrôle sur le mouvement généré et des équations souvent difficiles et longues à résoudre. Cependant, Hodgins et al. [Hod96, HWBO95] ont parvenus à générer des mouvements complexes de course et de saut en prenant en compte des données réalistes de masse et d'inertie, tirées de la littérature biomécanique.

### 3.2 Animation par optimisation

Les méthodes d'animation par optimisation ne considèrent plus le mouvement comme une succession d'états, mais comme un ensemble. Elles consistent à chercher, parmi l'espace de tous les mouvements possibles, celui qui satisfera le mieux des *contraintes* définies par l'utilisateur tout en minimisant une *fonction de coût*.

Les contraintes s'expriment formellement par des fonctions qui doivent garder un signe ou une valeur constants tout au long du mouvement. Par exemple, une fonction devra rester positive pour exprimer que les pieds doivent toujours rester au-dessus du sol.

La *fonction de coût* évalue la qualité d'un mouvement donné, elle dépend de l'application choisie. La plupart des méthodes utilisent des fonctions de coût qui préservent un réalisme physique. On cherchera donc à minimiser le déplacement de masse au cours du mouvement, la vitesse des articulations, et à préserver l'équilibre de la créature [WK88, LP02]. Gleicher et al. [Gle97] utilisent une fonction de coût qui favorise les mouvements proches d'un mouvement de référence. Safonova et al. [SHP04] utilisent une combinaison de ces deux fonctions, pour assurer à la fois réalisme physique et proximité d'un espace de mouvements de référence.

Alors, un *solveur*, souvent itératif, tente de trouver le mouvement qui minimise la fonction de coût tout en respectant les contraintes. Cette technique a été introduite par Witkin et Kass [WK88], qui animent une lampe qui effectue un saut. Le mouvement obtenu est satisfaisant visuellement car la lampe se baisse avant d'effectuer le saut et se plie à l'atterrissage pour amortir l'impact. Cependant le modèle utilisé n'avait que quelques degrés de liberté et résoudre ce problème pour des créatures articulées complexes pose des problèmes de complexité de calcul.

Pour des créatures complexes comme les humains ou les quadrupèdes, des petits segments d'animation peuvent être calculés : Rose et al. calculent des transitions entre des mouvements qui ont des poses proches au moment de la transition [RGBC96]. Pour des mouvements complets, on cherche à diminuer le nombre de degrés de liberté. Plusieurs articles ont montré que des mouvements de plongeon [LC94, CS95, ASBP00] ou de lancer du poids [HML00] peuvent être calculés efficacement pour un personnage simplifié. Popović et Witkin [PW99] appliquent l'optimisation à un personnage ne comportant que les degrés de liberté spécifiques au mouvement à animer. Ensuite le mouvement calculé est réajusté au personnage complet. Safonova

et al. [SHP04] utilisent une analyse en composantes principale pour projeter les 60 degrés de liberté d'un squelette d'humain sur espace spécifique au mouvement de plus petite dimension. Fang et al. [FP03] décrivent une manière de calculer plus rapidement les dérivées.

Liu et Cohen [LC95] parviennent à résoudre le problème en autorisant le solveur à modifier légèrement la position temporelle des contraintes. De même, en considérant un mouvement comme un ensemble de positions-clés très proches qui sont des contraintes douces, il est possible d'utiliser l'optimisation pour filtrer un mouvement et le rendre physiquement réaliste [DN99, YN03a, PR01].

Liu et Popović (Figure 2.18) utilisent cette technique en introduisant des contraintes physiques simplifiée. L'animation d'entrée est séparée en phases de contact avec l'environnement et en phases de vol. Ensuite des connaissances issues de la biomécanique sur l'évolution des moments (moment cinétique et quantité de mouvement) dans chacune de ces phases permet de contraindre le mouvement sans calculer explicitement des forces et des couples. Les résultats sont très convaincants pour des mouvements tels que les sauts ou la gymnastique (Figure 2.18), qui peuvent facilement être séparés en phases de contact et phases de vol. Cependant, les mouvements comme la marche, où au moins un pied (et plus pour les animaux) est en contact avec le sol, ne sont pas formalisables dans ce cadre. De plus, nous croyons que dans le cadre de la locomotion surtout à des vitesses moyennes ou faible, le contrôle moteur et musculaire joue un rôle bien plus grand que les facteurs dynamiques comme l'inertie.

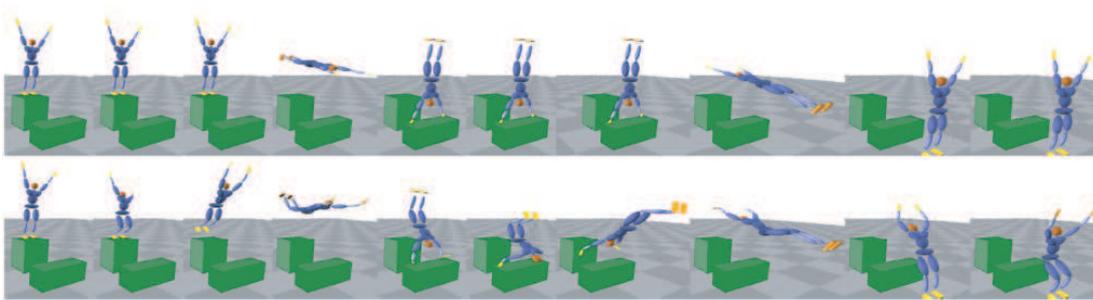


FIG. 2.18 – Illustration de la méthode de Liu & Popović [LP02]. En haut : l'animation simplifiée donnée par l'utilisateur. En bas : le résultat.

## 4 Bilan

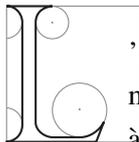
Ce chapitre a traité des façons de doter les quadrupèdes 3D de mouvement. Dans un premier temps nous avons présenté une brève étude de la locomotion des quadrupèdes réels et notamment des différentes allures qu'ils utilisent. Ensuite nous avons présenté les méthodes existantes pouvant permettre de contrôler le mouvement de quadrupèdes. L'animation cinématique, qui

consiste à manipuler directement les poses du squelette autour du temps, est la plus utilisée car elle donne un grand contrôle à l'artiste. L'animation par modèles physiques introduit la notion de force et de masses et permet de créer des mouvements assurant un certain réalisme physique, souvent au détriment de la qualité du contrôle. Dans tous les cas, la création d'un mouvement complet reste une tâche fastidieuse. Dans le chapitre suivant, nous allons voir comment il est possible d'utiliser des données réelles, notamment extraites de vidéo, pour faciliter la création d'animations 3D de quadrupèdes.

---

# Utilisation de données réelles pour l'animation

---



L'OBJECTIF DE CETTE THÈSE est de créer des animaux virtuels dont le mouvement est visuellement semblable à celui des animaux réels. Même si nous avons à présent à notre disposition tous les outils permettant de créer un quadrupède virtuel en mouvement, la création de mouvement réaliste reste une tâche difficile. Nos connaissances théoriques sur la morphologie et le mouvement des quadrupèdes sont assez limitées, et créer un mouvement réaliste implique de s'inspirer de la réalité. Dans ce chapitre, nous présentons un état de l'art des méthodes qui visent à capturer du mouvement réel, et permettent donc d'animer une créature 3D.

Le plus souvent, l'analyse de mouvement commence par la retranscription du mouvement sous forme d'une succession d'images. Mais la plupart des mouvements de locomotion sont trop rapides pour être décomposés à l'oeil nu. Les premières représentations d'animaux en mouvement datent de l'époque préhistorique, et les chevaux notamment sont souvent représentés en peinture, souvent avec des poses non réalistes. L'étude théorique du mouvement des animaux commence au XVII<sup>ème</sup> siècle : dans son livre *De muto animalium*, Giovanni Al-

fonso Borelli, qui est considéré comme le père de la biomécanique, tente d'expliquer les mouvements du corps des animaux à partir de principes de mécanique. Mais la plupart des données recueillies sur la locomotion animales ne sont que des suppositions, faute de techniques d'analyse objectives et performantes. Un tournant majeur arrivera au XIX<sup>ème</sup> siècle avec l'invention de la photographie et les travaux de Muybridge puis de Marey. A cette époque, une polémique porte sur la locomotion du cheval. Le physiologiste français Étienne-Jules Marey affirme qu'un cheval au galop passe par une phase où toutes ses pattes sont décollées du sol, ce qui est vivement mis en doute par la communauté. Un jour, le photographe Eadweard Muybridge rencontre le riche Leland Stanford, passionné par les chevaux de course, éleveur et entraîneur, qui promet un prix à celui qui résoudra ce problème. En 1878, Muybridge commande 24 appareils photographiques qu'il dispose le long d'une piste équestre. Les appareils sont déclenchés par des fils tendus sur la piste. A l'aide de ce dispositif, Muybridge obtient *le galop de Daisy*, le fameux cliché qui confirme la théorie de Marey. Par la suite il réalisera plus de 100000 photographies (Figure 3.1). Ces planches sont encore considérées aujourd'hui comme des références et sont encore largement utilisées pour l'analyse de mouvement. Après sa rencontre avec Muybridge, Marey mettra au point le *fusil photographique* qui lui permet de photographier "sur nature" un être en mouvement sur douze poses, puis la *chronophotographie* : Avec un seul objectif, la même plaque est exposée plusieurs fois, ce qui permet de voir toutes les phases d'un mouvement superposées sur une même image (Figure 3.2).



FIG. 3.1 – Une des planches réalisées par Muybridge : Plusieurs appareils photo sont déclenchés par des fils tendus sur le chemin de l'animal. La phase de suspension est clairement visible sur les premières images.

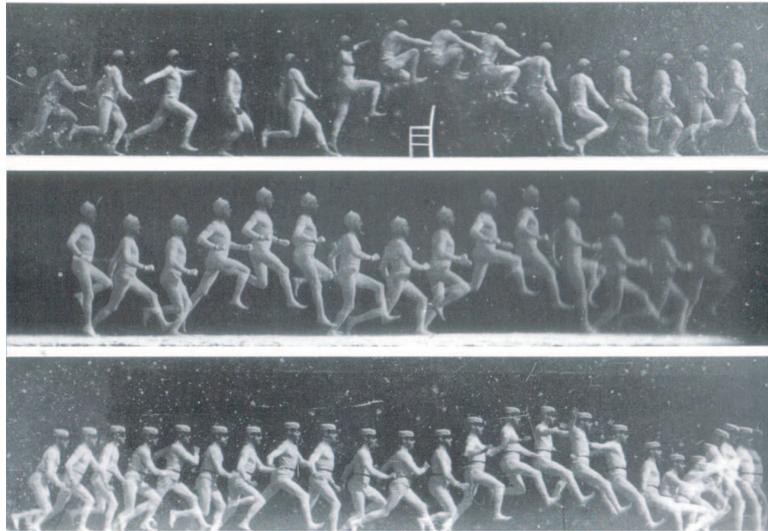


FIG. 3.2 – Quelques planches réalisées par Marey : Avec un sujet clair sur fond foncé, la même plaque est exposée plusieurs fois.

Depuis cette époque, il est possible de visualiser avec précision les détails d'un mouvement. Mais recréer un mouvement complet à partir d'une séquence d'images reste un processus long. Le rotoscope (Figure 3.3), inventé en 1914, est utilisé en animation 2D. Les images d'un film sont projetées une par une sur un écran, et l'artiste dessine un personnage sur chacune d'elles pour créer un dessin animé. En animation, il existe plusieurs méthodes qui visent à utiliser des données réelles en entrée. La première est la capture de mouvement à l'aide de marqueurs.

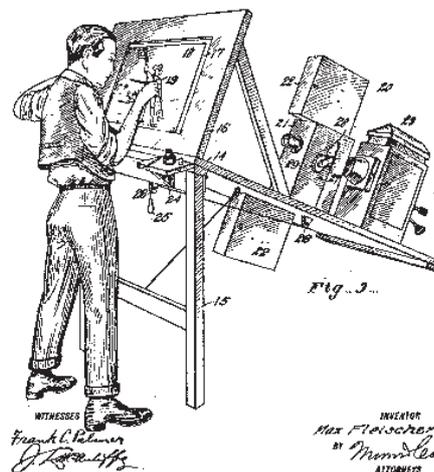


FIG. 3.3 – Le Rotoscope (Image extraite du brevet de 1914).

## 1 Capture de mouvement à l'aide de marqueurs

La capture de mouvement à l'aide de marqueurs (ou *motion capture*) a été d'abord développée comme outil d'analyse en biomécanique, avant de se populariser dans les domaines de l'animation 3D et de la médecine du sport. Le mouvement est capturé sur un acteur qui porte des marqueurs à des positions précises, la plupart du temps aux articulations (Figure 3.4). Les marqueurs peuvent être, entre autres, visuels, magnétiques ou inertiels. La position de ces marqueurs est suivie au cours du temps, et on déduit les positions, orientations, vitesses et accélérations de chaque joint au cours du temps. On ne capture pas l'apparence de l'acteur, mais les positions et orientations des joints, qu'on applique alors à un modèle 3D pour reproduire le mouvement. Pour une présentation détaillée de cette technique et ses applications, voir [MM99].



FIG. 3.4 – Un acteur muni de marqueurs.

En animation cette technique permet d'obtenir le mouvement d'un personnage entier sans avoir besoin de positionner chaque joint à la main pour chaque image. Les mouvements capturés sont plus naturels, ce qui peut être un avantage si on cherche le réalisme, mais aussi un inconvénient si on veut créer des mouvements "cartoon", où le corps se compresserait par exemple. En biomécanique ou en médecine du sport, ces données de mouvement peuvent four-

nir des informations pour diagnostiquer des problèmes ou pour fournir un moyen d'améliorer les performances.

La position 3D de chaque marqueur est triangulée grâce aux images données par plusieurs caméras calibrées. Le nombre de caméras dépend de la taille de la zone de capture et du nombre de marqueurs à suivre. Les informations d'orientation des joints sont déduites de ces positions : par exemple les positions de l'épaule, du coude et du poignet permettent de déduire l'angle formé par le coude. Plusieurs types de marqueurs très utilisés sont les marqueurs optiques, magnétiques et mécaniques.

### **Systemes optiques**

Les systèmes optiques passifs utilisent des marqueurs réfléchissants, souvent des sphères ou des demi-sphères en plastique ou en mousse. Comme ces marqueurs ont tous la même apparence, il existe un fort risque de confusion entre les marqueurs. Il existe des techniques pour lever ces confusions. Les données peuvent aussi être nettoyées à l'aide de contraintes cinématiques. Les systèmes par marqueurs passifs ne nécessitent pas de porter des fils ou de l'équipement électronique, d'où une grande liberté de mouvement. Ces systèmes peuvent capturer un grand nombre de marqueurs jusqu'à 2000 fps avec une bonne précision 3D.

Il existe aussi des systèmes par marqueurs actifs qui permet de lever le doute sur l'identité des marqueurs. En général la précision temporelle dépend du nombre de marqueurs. Ces systèmes sont très populaires en biomécanique.

### **Systemes magnétiques**

Les systèmes magnétiques calculent la position et l'orientation des marqueurs à partir du flux magnétique de trois bobines orthogonales situés sur les marqueurs et sur le récepteur. Comme les capteurs donnent des positions et des orientations, on peut obtenir les mêmes résultats que précédemment avec moins de capteurs (la position du coude, par exemple, peut être déduit de la position et de l'orientation de l'épaule et du poignet). Cependant il y a de gros problèmes d'interférences avec les objets métalliques situés dans la zone de capture où les pièces électroniques.

### **Systemes mécaniques**

Les systèmes mécaniques qui capturent directement les angles aux joints, sont souvent appelés *exo-squelettes* à cause de la façon dont les capteurs sont attachés sur le corps. L'acteur est muni de plusieurs structures qui ont la morphologie d'un squelette. Ces structures se déplacent en même temps que lui, et capturent le mouvement. Ces systèmes sont en temps réel, peu coûteux, libres d'occlusions et sans fils et ont un volume de capture infini. Ce sont des structures de pièces de métal ou de plastique liées entre elles.

Le mouvement capturé est souvent bruité et nécessite souvent un post-traitement manuel. Un autre problème est que le mouvement capturé est spécifique au squelette de l'acteur capturé et au mouvement précis effectué. De nombreuses techniques [Gle98, MBBT00] permettent de traiter ce problème en adaptant automatiquement les paramètres de l'animation capturée à la morphologie du personnage que l'artiste souhaite animer. Il existe également des techniques pour synthétiser des mouvements réalistes à partir de bases de données de mouvements capturés comme la technique des motion graphs [KGP02] ou les techniques de pondération [KG04].

Par rapport aux techniques traditionnelles, la capture de mouvement par marqueurs a plusieurs avantages. D'abord, dans de bonnes conditions, elle permet d'obtenir une grande quantité de mouvements en un temps très court. Ensuite, on peut capturer instantanément du mouvement secondaire et des nuances que des animateurs traditionnels ne reproduiraient pas forcément. Enfin, cette technique peut capturer avec précision des mouvements difficiles à modéliser, comme des mouvements très rapides, ou des mouvements très dynamiques comme la gymnastique ou les arts martiaux. Le premier inconvénient de cette méthode est que les données sont très difficiles à éditer à la main une fois capturées. Il est donc plus facile de refaire toute la capture plutôt que d'effectuer une modification manuelle. Ensuite, l'équipement utilisé peut coûter cher (caméras, logiciels, projecteurs), peut être obsolète très vite, et nécessite plusieurs personnes pour le faire fonctionner. Enfin, et c'est le point qui nous intéresse le plus, il est facile de capturer de cette façon le mouvement d'humains ou bien d'animaux dociles (Figure 3.5), mais l'appliquer à des animaux sauvages serait beaucoup plus difficile, car il faudrait les équiper de marqueurs et les faire se déplacer dans des endroits confinés. La course d'un tigre fondant sur sa proie, par exemple, ne peut pas être capturée de cette façon. Et même si on réussit à équiper un animal sauvage d'un tel système, il n'y a aucune garantie que le mouvement obtenu sera réaliste.

## 2 Extraction de mouvement 3D à partir de vidéos

Une autre source de mouvement envisageable est la vidéo. Extraire du mouvement 3D à partir de vidéos est intéressant car les vidéos constituent une bonne base documentaire sur le mouvement des animaux sauvages : les données sont nombreuses et accessibles, représentent un grand nombre d'animaux et de mouvements différents, y compris des mouvements qui seraient impossible à capturer à l'aide de marqueurs.

Le vidéos présentent un défi au traitement automatique car elles ont une faible résolution spatiale et surtout une très faible résolution temporelle. Les mouvements les plus rapides, comme la course, ne peuvent être capturés qu'imparfaitement lorsqu'ils sont échantillonnés à 25 ou 30 images par seconde. De plus, comme nous voulons utiliser des documentaires ani-



FIG. 3.5 – Un animal muni de marqueurs.

maliers, les vidéos sont tournées en plein air, avec de grandes variations d'éclairage. Ceci met à l'épreuve la robustesse des méthodes classiques. La plupart des travaux dédiés spécifiquement aux animaux se contentent de suivre la position rigide en 2D [ZHDM05, Eva03, dGITT03, Per01, TK02] ou en 3D [ZC06] d'animaux à partir de vidéo monoculaire, et ne donnent pas des résultats d'animation. Cependant les méthodes permettant d'extraire des mouvements articulés humains à partir de vidéo sont très nombreuses, et sont en théorie applicables telles quelles aux quadrupèdes.

## 2.1 Extraction de formes 3D à partir de vidéos

Certaines techniques visent à reconstruire une surface 3D à partir de vidéos monoculaires. Des points d'intérêts sont suivis sur l'image [LK81]. Les points d'intérêts correspondent à des petites zones sur la vidéo qui ont un aspect particulier qui les rend faciles à suivre au cours du temps, comme des coins ou des motifs en croix. Ensuite, la géométrie 3D correspondante à ces points 2D est reconstruite.

Dans le cas où l'objet à suivre est rigide, Tomasi et Kanade [TT92] montrent que la matrice contenant les positions 2D de tous les points au cours du temps est de rang 3, et peut être factorisée en 2 matrices. La première matrice donne la position de l'objet dans l'espace au cours du temps, et la deuxième matrice sa forme 3D. Costeira et al. étendent cette méthode pour prendre en compte plusieurs objets rigides [CK98]. Bregler et al. [BHB99, TYAB01] factorisent la matrice de suivi en 3 matrices : une matrice de formes-clés, une matrice de poids qui définit la configuration de l'objet comme une somme pondérée des formes-clés, et une matrice de

positions. (Figure 3.6). Cette méthode a été appliquée à des animaux, mais la précision des résultats est trop faible pour permettre une application en animation.

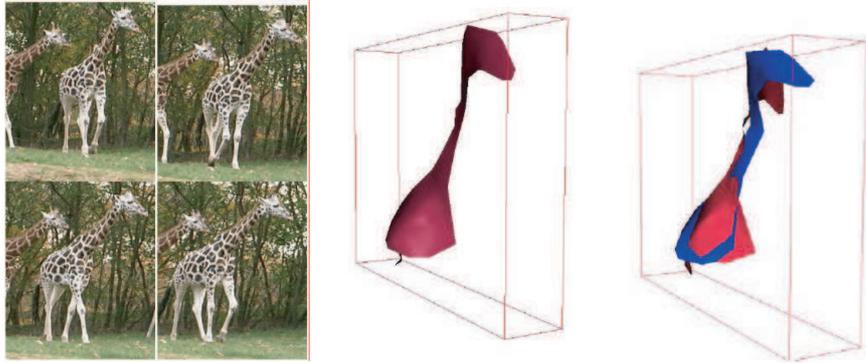


FIG. 3.6 – D'après[BHB99]. A partir d'un suivi de points, la forme 3D de la girafe est extraite(au centre), ainsi que les modes principaux de variation de la forme (à droite)

Ramanan et al. [RF03, RFK06] reconstruisent des modèles articulés simples d'animaux à partir de vidéo (Figure 3.7). Cependant, les modèles sont seulement constitués de quelques segments et ne peuvent donc pas être utilisés pour de l'animation. Pour obtenir des résultats plus précis, il faut introduire des connaissances a priori sur la morphologie du sujet à capturer.



FIG. 3.7 – Ramanan et al. [RF03] reconstruisent des modèles articulés simples d'animaux à partir de vidéo.

## 2.2 Suivi de mouvement articulé 3D à partir de vidéos

Un domaine de recherche très actif en vision par ordinateur est l'extraction de mouvement 3D articulé à partir d'une vidéo, en utilisant un modèle 3D a priori du sujet à suivre. Là encore, la plupart des travaux effectués se sont concentrés sur les humains [OSBH00, SBF00, BM98,

FB02, ST03]. Le traitement des animaux est similaire dans son cadre général : on dispose d'un modèle 3D *a priori* de la créature à suivre, et d'une vidéo monoculaire. A chaque pas de temps, on cherche à trouver la pose du modèle 3D qui va maximiser la corrélation entre le rendu du modèle 3D et l'image (Figure 3.8).

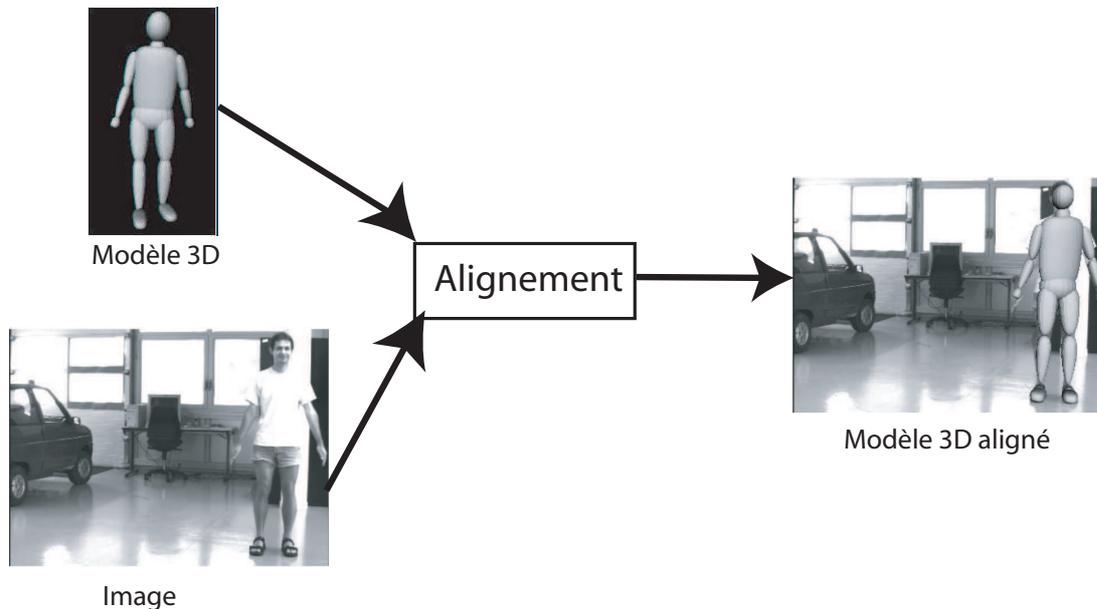


FIG. 3.8 – Suivi 3D à l'aide d'un modèle 3D a priori : à chaque instant, on cherche la pose du modèle 3D qui maximise la corrélation avec l'image (D'après [ST03], modifié).

L'un des problèmes de cette approche est que la reconstruction d'informations en trois dimensions à partir d'une image 2D est un problème largement sous contraint notamment à cause de nombreuses ambiguïtés et occlusions : des poses différentes d'un même modèle 3D peuvent donner des images très proche. Contrairement aux méthodes multivues qui sont relativement robustes, le suivi monoculaire est un problème mal conditionné [ST03], donc très sensible au bruit.

Ensuite, la recherche de la meilleure pose est faite à l'aide d'une descente de gradient ou d'une de ses variantes, alors que la fonction à minimiser présente de nombreux minima locaux (Figure 3.9). Trouver le minimum global est donc très difficile, surtout compte tenu du grand nombre de degrés de libertés du modèle (au moins une trentaine d'articulations). De nombreux algorithmes ont été utilisés comme le recuit simulé [KGV83], les algorithmes génétiques, la recherche tabou[GL93], des techniques d'échantillonnage[ST03], le repérage de chemins entre les minima locaux proches [ST05], l'utilisation de cartes exponentielles [BMP], et des techniques issues de la reconnaissance de la parole [GG03].

Pour rendre la recherche plus efficace, il est possible d'introduire plus d'informations a

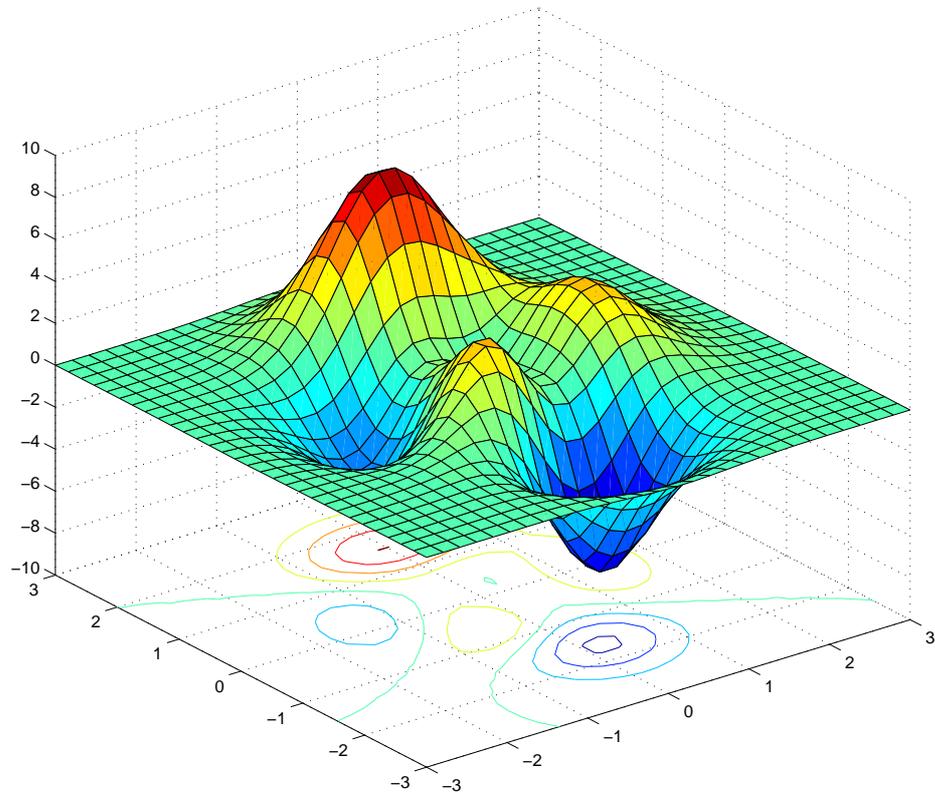


FIG. 3.9 – Représentation schématique d'un espace de recherche. Le présence de minima locaux rend la recherche du minimum global difficile.

priori sur le mouvement à suivre. Black et al. [OSBH00] utilisent des données de capture de mouvement. Ces données sont d'abord découpées en cycles, puis une analyse en composantes principales (ACP) permet de construire un modèle de faible dimension pour contraindre les espaces de recherche.

Il est aussi possible d'introduire des connaissances a priori sur les variations de l'image. Le *flot optique* est un outil souvent utilisé pour obtenir des informations dynamiques au niveau image. Il donne le vecteur vitesse correspondant à chaque pixel de l'écran [HS81](Figure 3.10). Black et al.[FB02] utilisent en entrée des données de capture de mouvement. Ces données sont utilisées pour animer un modèle 3D, puis l'image de ce modèle est projetée sur plusieurs plans tout autour de l'objet. A partir de ces images, ils construisent des modèles de flot optique de faible dimension à l'aide d'une ACP. Ce modèle donne une information dynamique *a priori*

sur le mouvement au niveau image.

Toutes ces méthodes souffrent du manque d'information contenu dans les vidéos. Même en introduisant beaucoup de connaissances, les vidéos ne contiennent pas assez d'information pour extraire les orientations individuelles de chaque articulation à partir d'une série d'images.

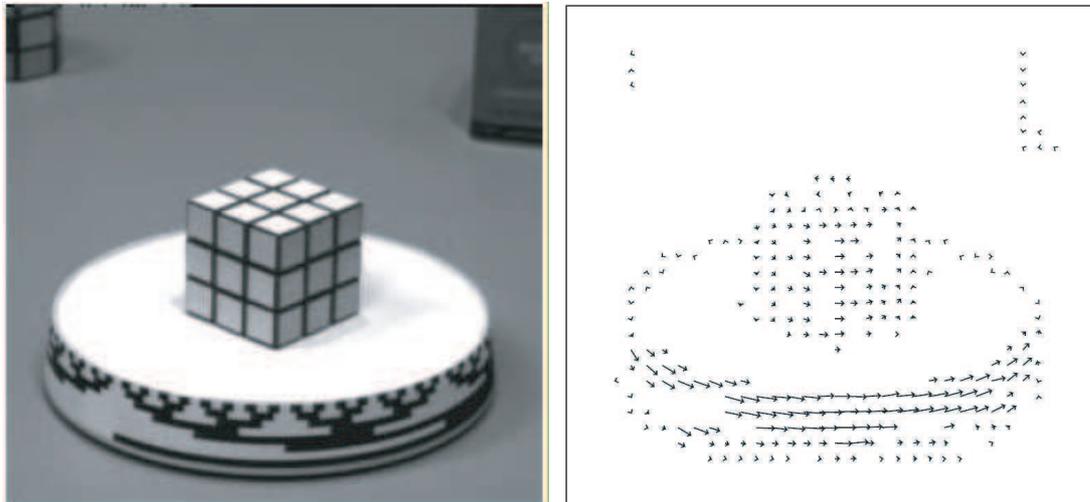


FIG. 3.10 – Le flot optique donne le mouvement local en 2D pour chaque position sur l'image.

### 3 Animation à partir de vidéo

Certaines méthodes tentent donc d'inférer du mouvement 3D à partir de vidéo sans extraire explicitement les rotations de chaque articulation. Elles visent à extraire un petit nombre de paramètres pertinents à partir de la vidéo. Ces paramètres, qui sont choisis pour décrire le mouvement de façon compacte, sont alors utilisés pour animer un modèle 3D.

Bregler et al.[BLCD02] créent des animations 3D à partir de dessins animés. L'utilisateur choisit sur un dessin animé les *images-clés*, c'est-à-dire les images qui correspondent à des étapes importantes du mouvement. Ensuite, il donne les formes 3D qui correspondent aux images-clés. Du mouvement 3D est alors recrée par interpolation de formes clés (Fig.3.11), à partir de la vidéo. Malheureusement, comme l'interpolation linéaire de formes-clés donne des formes incorrectes dès qu'on s'éloigne un peu des formes-clés, il est nécessaire de générer un grand nombre de clés de façon semi-automatique puis de les corriger à la main.

Dans le cas de Bregler et al., chaque image de la vidéo est interprétée comme une combinaison linéaire d'images-clés définies par l'utilisateur, à laquelle on ajoute une déformation affine. Reveret et Essa [RE01] animent des visages et extrayant à partir de vidéo un petit nombre de paramètres issus du domaine de la phonétique qui décrivent le mouvement du visage. Ensuite

ces paramètres peuvent être utilisés pour animer un modèle 2D ou 3D. Diener et al. [DRE06] extraient une structure hiérarchique de mouvement à partir d'une vidéo d'arbre sous le vent et animent un modèle 3D d'arbre.

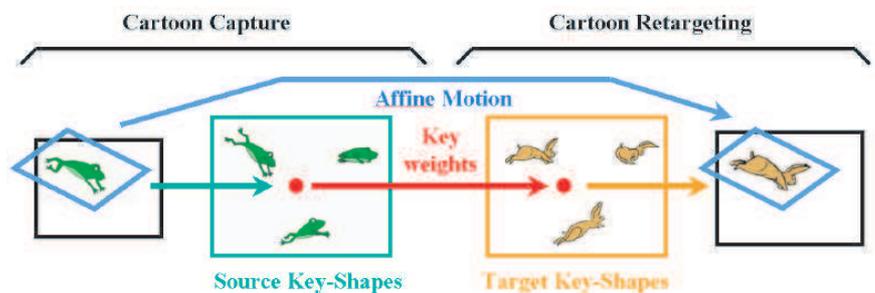


FIG. 3.11 – D'après Bregler et al.[BLCD02]. Chaque image de la vidéo est décrite comme une combinaison d'images-clés. L'utilisateur donne une pose-clé 3D qui correspond à chacune des images-clés, et les poses-clés sont alors interpolées pour créer de l'animation 3D à partir de vidéo.

## 4 Bilan

Dans ce chapitre, nous avons présenté les méthodes qui visent à utiliser des données réelles pour faciliter la création d'animations 3D. La capture de mouvement à l'aide de marqueurs est une technique bien maîtrisée mais qui n'est pas applicable aux quadrupèdes sauvages. Au contraire, les documentaires animaliers constituent une grande base d'informations sur le mouvement des quadrupèdes. Les techniques de capture de forme ou de mouvement articulé à partir de vidéo abondent, mais les vidéos naturelles sont très difficiles à utiliser pour obtenir de l'animation joint par joint. C'est pourquoi certaines méthodes récentes d'animation à partir de vidéo permettent d'animer un modèle 3D directement à partir d'une séquence vidéo, en extrayant de la vidéo des paramètres pertinents pour recréer de l'animation 3D sans extraire explicitement des valeurs d'orientation des articulations.

## **Deuxième partie**

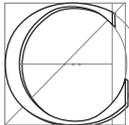
# **Modélisation de squelettes de quadrapèdes**



---

# Reconstruction d'un squelette de quadrupède à partir d'images

---



COMME NOUS L'AVONS VU au chapitre précédent, obtenir de l'animation 3D exploitable à partir de vidéos 2D nécessite d'introduire de la connaissance a priori sur le mouvement que l'on cherche à capturer. Dans notre formalisme où le quadrupède est modélisé comme un squelette d'animation entouré d'une peau, il est très important d'avoir un squelette d'animation de qualité, avec des os de longueur correcte et surtout une bonne orientation des axes de rotation pour que le modèle s'articule selon des degrés de liberté optimaux.

Ce chapitre présente une méthode pour reconstruire un squelette d'animation 3D à partir d'une série de photographies d'un squelette de quadrupède réel. Ce projet a eu lieu lors d'une collaboration avec le Muséum National d'Histoire Naturelle dans le cadre du projet ANR Kameleon, qui a pour but l'étude des structures squelettiques des vertébrés. Obtenir des squelettes 3D précis de quadrupèdes est une question importante dans plusieurs domaines de recherche : en anatomie pour la description du vivant, en recherche clinique pour le diagnostique articulaire, et en infographie 3D pour l'animation de créatures virtuelles. Dans le cadre de l'anima-

tion de quadrupèdes à partir de vidéo, disposer de ces squelettes nous donne une information fiable sur la structure du quadrupède à animer. En particulier, cette information est indispensable avant d'aborder dans la troisième partie de cette thèse l'analyse de vidéo naturelles, donc très bruitées.

Dans la suite de ce chapitre, nous présentons d'abord le modèle de caméra sténopé, très utilisé en informatique graphique. Ensuite, nous présentons la méthode de reconstruction 3D à partir d'images que nous avons utilisée. Enfin, nous présentons nos premiers résultats : un squelette 3D de rat est reconstruit à partir d'images.

## 1 Le modèle de caméra sténopé

Le modèle de caméra sténopé ou en tête d'épingle est un modèle simplifié de caméra qui est largement utilisé en informatique graphique. Ce modèle est largement inspiré par les premiers appareils photographiques, qui étaient composés d'une boîte noire percée d'un minuscule trou. La lumière rentre dans la boîte par ce trou et va imprimer sur une face une image inversée de la scène observée, image qu'on peut alors capturer à l'aide d'un matériau photosensible. Dans le modèle de caméra sténopé, le principe est le même : tous les rayons lumineux convergent vers un même point, le foyer. Cependant, l'écran est un écran virtuel situé entre le foyer et l'objet observé, à une distance du foyer égale à la distance focale (Figure 4.1). L'image obtenue n'est donc pas renversée comme avec une vraie caméra.

Formellement, un point en 3D de coordonnées cartésiennes  $(x, y, z)$  est projeté sur l'écran  $E$  pour donner un point en 2D de coordonnées homogènes  $(u, v, w)$ , par la formule :

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

Ou encore, en utilisant les coordonnées cartésiennes  $(U, V)$  du point 2D :

$$\begin{cases} U = \frac{u}{w} = \frac{f_x x + c_x z}{z} = f_x \frac{x}{z} + c_x \\ V = \frac{v}{w} = \frac{f_y y + c_y z}{z} = f_y \frac{y}{z} + c_y \end{cases}$$

Les scalaires  $f_x$  et  $f_y$  sont les distances focales de la caméra en  $x$  et en  $y$ , et  $(c_x, c_y)$  est la position en 2D sur l'écran du centre de l'image, en général le milieu de l'écran. Ces paramètres  $(f_x, f_y, c_x, c_y)$  sont appelés les paramètres *intrinsèques* de la caméra. En ajoutant une matrice de transformation rigide à la transformation, on obtient :

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} (R|t) \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

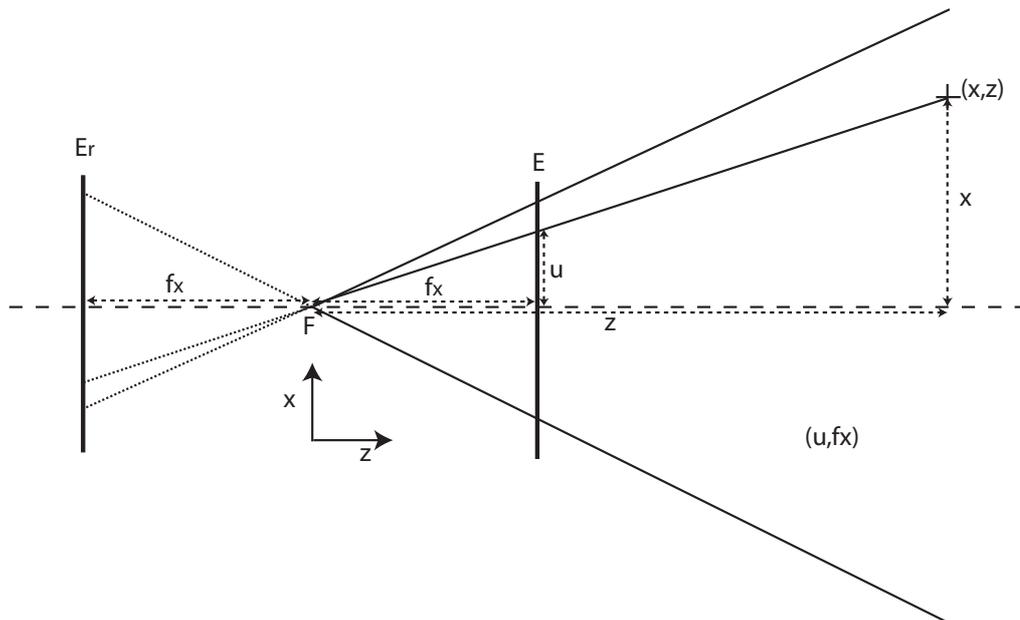


FIG. 4.1 – Le modèle de caméra sténopé, présenté ici en 2D. Tous les rayons convergent vers le foyer  $F$ . Contrairement aux appareils sténopé réels, les points ne sont pas projetés sur un écran  $E_r$  situé derrière le foyer, mais sur un écran virtuel  $E$  situé entre le foyer et la scène, à une distance du foyer égale à la distance focale  $f_x$  est la distance focale de la caméra. Le point de coordonnées  $(x, z)$  est projeté sur l'écran  $E$  en un point 1D de coordonnées homogènes  $(u, w)$ .

La matrice de transformation rigide  $(R|t)$  est appelée matrice de paramètres *extrinsèques*. Elle sert à décrire le mouvement de la caméra autour d'une scène statique ou bien, de façon symétrique, à décrire le mouvement d'un objet dans un repère fixé à la caméra.

Calculer les paramètres intrinsèques et extrinsèques de la caméra est ce qu'on appelle le processus de *calibration*. Pour calibrer une caméra, on utilise un objet dont la structure et les dimensions en 3D sont connues. La plupart du temps, on utilise une mire de calibration, c'est-à-dire un damier dont les dimensions sont connues (Figure 4.2). En prenant des photographies de la mire de calibration dans plusieurs positions différentes, on suppose que la caméra se déplace de façon rigide autour d'une scène 3D connue. La position des coins du damier peut être facilement détectée de façon automatique, ce qui donne les positions 2D projetées correspondant aux positions 3D connues des points pour chaque position de la caméra. Ces contraintes permettent de calculer les paramètres intrinsèques et extrinsèques de la caméra [Zha99].

Ce modèle de caméra est abondamment utilisé en vision par ordinateur. Maintenant que nous sommes familiers avec ce modèle, nous allons étudier la méthode de reconstruction 3D que nous avons utilisée.

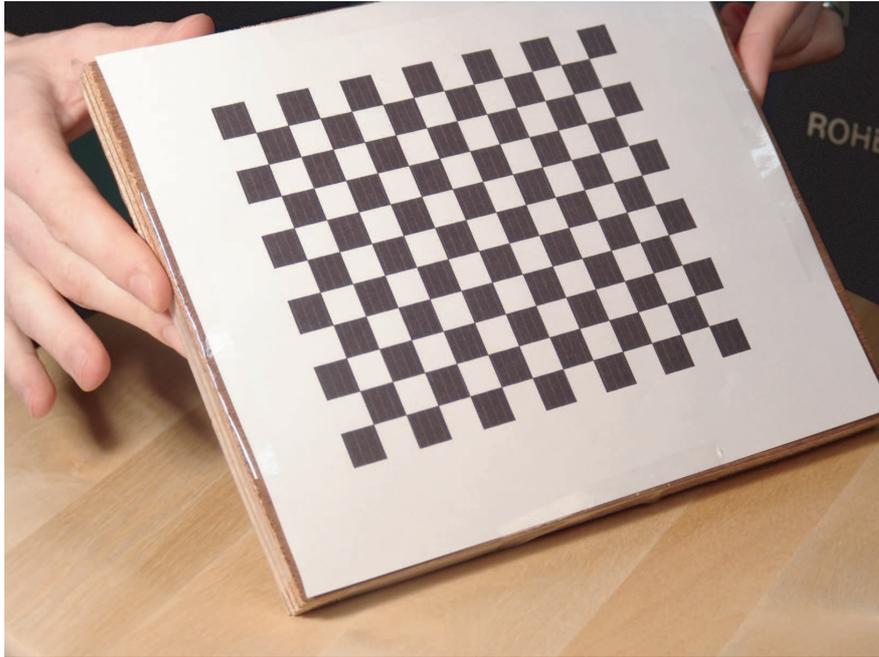


FIG. 4.2 – Une mire de calibration.

## 2 Reconstruction d'un squelette 3D de rat à partir d'images

Notre objectif est de reconstruire un squelette de rat en 3D à partir de plusieurs images d'un squelette réel. Ces images sont des photographies prises depuis plusieurs endroits autour d'un squelette de rat fourni par nos partenaires du Muséum National d'Histoire Naturelle. Un exemple de photographies obtenues est visible sur la Figure 4.3. Les paramètres intrinsèques de la caméra sont calibrés une fois pour toute : la focale de la caméra est réglée avant la phase de calibration et n'est pas changée ensuite. La reconstruction est faite en deux étapes. Tout d'abord, les paramètres extrinsèques de la caméra correspondant à chaque image sont retrouvés. Cela peut être fait en désignant sur chaque image la position 2D d'au moins 3 points dont les positions 3D sont connues. Pour faciliter ce processus, le support sur lequel le rat est posé est recouvert d'un damier. Ce damier correspond au plan 3D d'équation  $y = 0$ , et nous cliquons sur au moins 3 coins du damier pour retrouver les paramètres extrinsèques de la caméra.

La deuxième étape consiste à reconstruire la position 3D de chaque articulation. Pour cela, nous désignons la position 2D de chaque articulation sur au moins deux images. Chaque contrainte ainsi définie force le point 3D à être sur une droite : la droite reliant le foyer de la caméra à la position du point sur l'écran de la caméra (Figure 4.4). A partir de la donnée de 2 projections, on peut estimer la position du point 3D comme étant la position 3D la plus proche

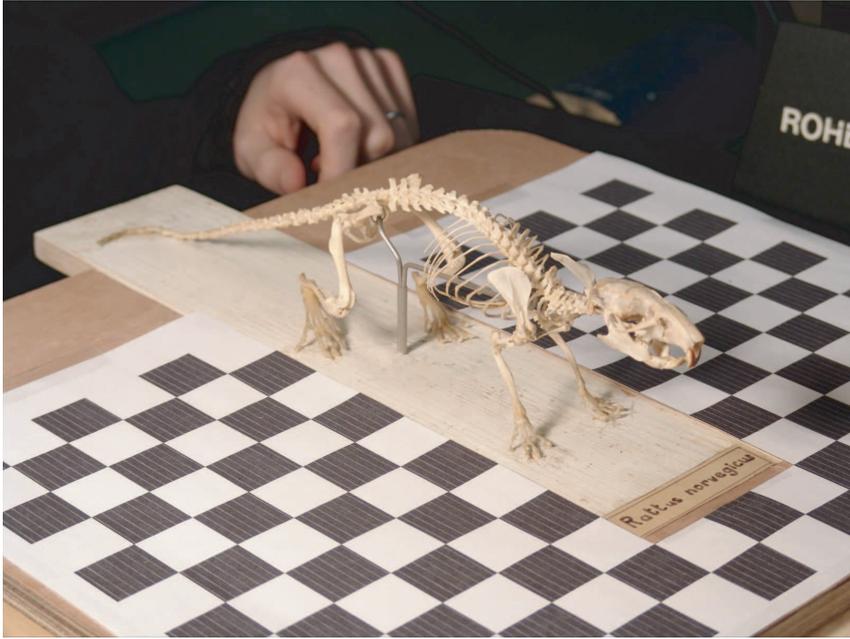


FIG. 4.3 – Une photographie du squelette réel de rat.

des droites. La méthode que nous avons utilisée, ainsi que d'autres techniques de triangulation, sont présentées dans [HP97].

Formellement, la position du point 3D est estimée de la façon suivante : Supposons que le point 3D ait pour coordonnées homogènes  $X = (x, y, z, \omega)$ . La  $i$ -ème caméra, associée à la matrice de projection  $A^i$ , projette ce point en un point 2D de coordonnées homogènes  $(u_i, v_i, w_i)$ .

On a donc :

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} A_1^i \\ A_2^i \\ A_3^i \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ \omega \end{pmatrix} = \begin{pmatrix} A_1^i X \\ A_2^i X \\ A_3^i X \end{pmatrix}$$

où  $A_j^i$  est la  $j$ -ème ligne de la matrice  $A^i$ . Les coordonnées cartésiennes  $(U_i, V_i)$  du point projeté vérifient :

$$\begin{cases} U_i = \frac{u_i}{w_i} = \frac{A_1^i X}{A_3^i X} \\ V_i = \frac{v_i}{w_i} = \frac{A_2^i X}{A_3^i X} \end{cases}$$

soit :

$$\begin{cases} A_1^i X - U_i A_3^i X = 0 \\ A_2^i X - V_i A_3^i X = 0 \end{cases}$$

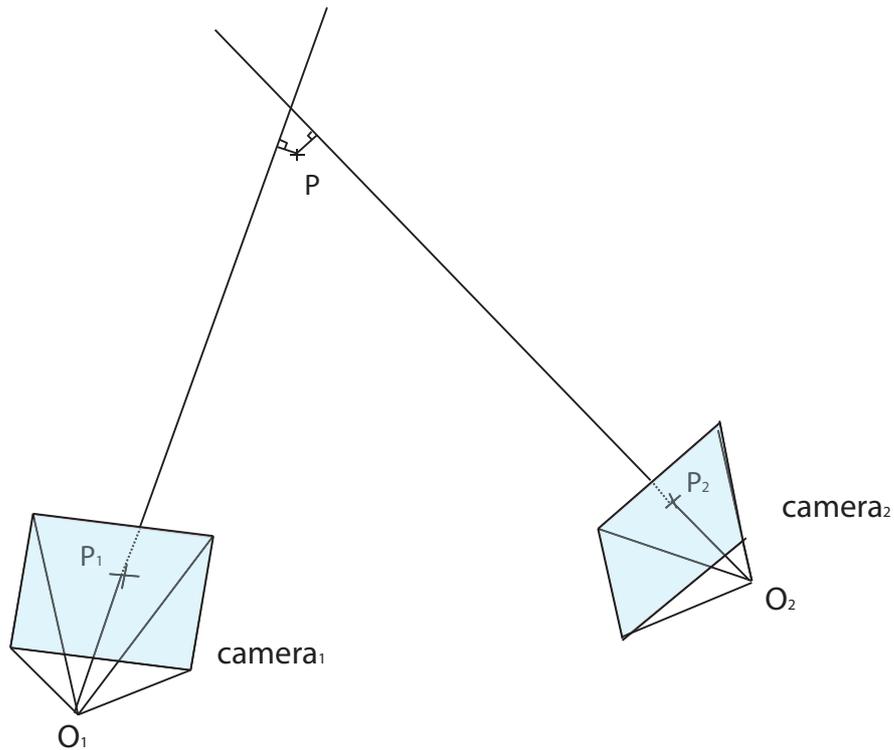


FIG. 4.4 – La donnée de la projection du point 3D par une caméra contraint le point 3D à être sur une droite : la droite reliant le foyer de la caméra à la projection du point 3D sur l'écran. A partir de la donnée de 2 projections, on peut estimer la position du point 3D comme étant la position 3D la plus proche des droites.

Au final la position 3D  $X$  du point que l'on cherche à reconstruire vérifie :

$$\begin{pmatrix} A_1^i - U_i A_3^i \\ A_2^i - V_i A_3^i \end{pmatrix} X = 0$$

Cette contrainte impose que  $X$  soit situé sur la droite reliant le foyer de la caméra à la projection du point sur l'écran. A présent empilons les contraintes données par toutes les caméras pour lesquelles la position de la projection 2D du point est connue, et posons la matrice de

contraintes :

$$A = \begin{pmatrix} A_1^1 - uA_3^1 \\ A_2^1 - vA_3^1 \\ A_1^2 - uA_3^2 \\ A_2^2 - vA_3^2 \\ \vdots \\ A_1^n - uA_3^n \\ A_2^n - vA_3^n \end{pmatrix}$$

La solution idéale serait  $X$  tel que  $AX = 0$ . Malheureusement, ce ne sera jamais le cas à cause de l'imprécision sur les positions des projections. On cherche donc le point 3D  $X$  qui minimisera la norme de la contrainte  $AX$  :

$$X = \arg \min_x \|Ax\|$$

En prenant la norme euclidienne, cette équation peut être résolue directement à l'aide de la décomposition en valeurs singulières (SVD) de  $A$ . En effet, cette décomposition s'écrit :

$$A = U \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \lambda_3 & \\ & & & \lambda_4 \end{pmatrix} V^T$$

Où  $\lambda_1, \dots, \lambda_4$  sont les valeurs singulières de la matrice classées par ordre décroissant. Si on note  $V_i$  la  $i$ -ème colonne de  $V$  ( $1 \leq i \leq 4$ ), on a alors :

$$AV_4 = U_{n4} \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \lambda_3 & \\ & & & \lambda_4 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \\ V_3^T \\ V_4^T \end{pmatrix} V_4 = U \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \lambda_3 & \\ & & & \lambda_4 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = U \begin{pmatrix} 0 \\ 0 \\ 0 \\ \lambda_4 \end{pmatrix}$$

On en déduit que  $\|AV_4\| = |\lambda_4|$ . De même,  $\|AV_i\| = |\lambda_i|$  ( $1 \leq i \leq 4$ ). Comme  $\lambda_4$  est la plus petite valeur singulière de  $A$ ,  $V_4$  est la solution cherchée :

$$X = V_4$$

Nous venons de montrer comment reconstruire la position du point 3D à partir de ses projections dans les plans images d'au moins 2 caméras calibrées. Dans la section suivante, nous présentons nos résultats.

### 3 Résultats

Pour notre expérience, nous avons pris 36 photos tout autour du rat. La caméra est à une position fixe. Une fois la caméra calibrée, le squelette de rat est placé sur un plateau que l'on peut faire pivoter sur lui-même (Figure 4.5). A partir de ces 36 vues, nous avons reconstruit avec succès le squelette 3D correspondant à l'animal réel. La Figure 4.6 présente notre squelette 3D reconstruit, vu depuis une caméra virtuelle. La caméra virtuelle est réglée avec les paramètres intrinsèques et extrinsèques calculés à partir des images, et la photographie correspondante est superposée à l'image. Comme on peut le voir, la qualité visuelle de la reconstruction est très bonne. Numériquement, la précision de cette méthode est de quelques millimètres.

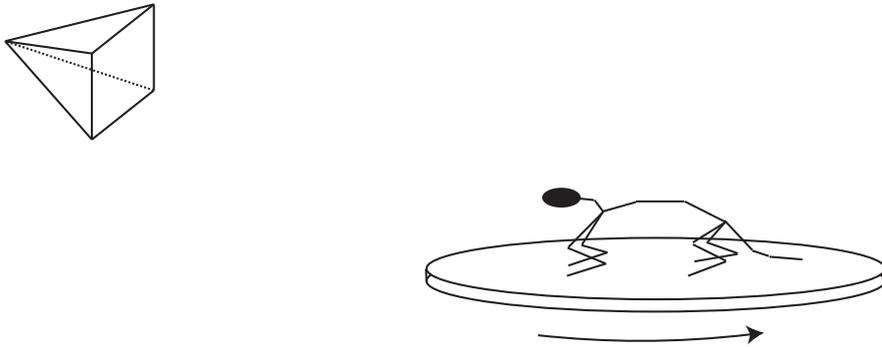


FIG. 4.5 – La caméra est immobile, et le rat est placé sur un plateau qui peut pivoter sur lui-même.

### 4 Bilan

Dans ce chapitre, nous avons reconstruit précisément un squelette 3D de rat à partir de photographies. La construction de squelettes d'animation précis est une étape indispensable pour fournir des données fiables préalablement à la création d'animations 3D. Nos résultats ont montré que la méthode est robuste. Cependant, elle demande la mise en place d'un protocole relativement lourd qui doit être recommencé en totalité chaque fois qu'on s'intéresse à un nouvel animal. Il est en particulier nécessaire de désigner à la main la position de la projection 2D de nombreux points. De plus, il faut disposer du squelette de l'animal que l'on souhaite animer. Dans le chapitre suivant, nous allons construire un modèle générique de squelette de quadrupèdes, dans le but de faciliter et d'automatiser au maximum la création de squelettes d'animation.

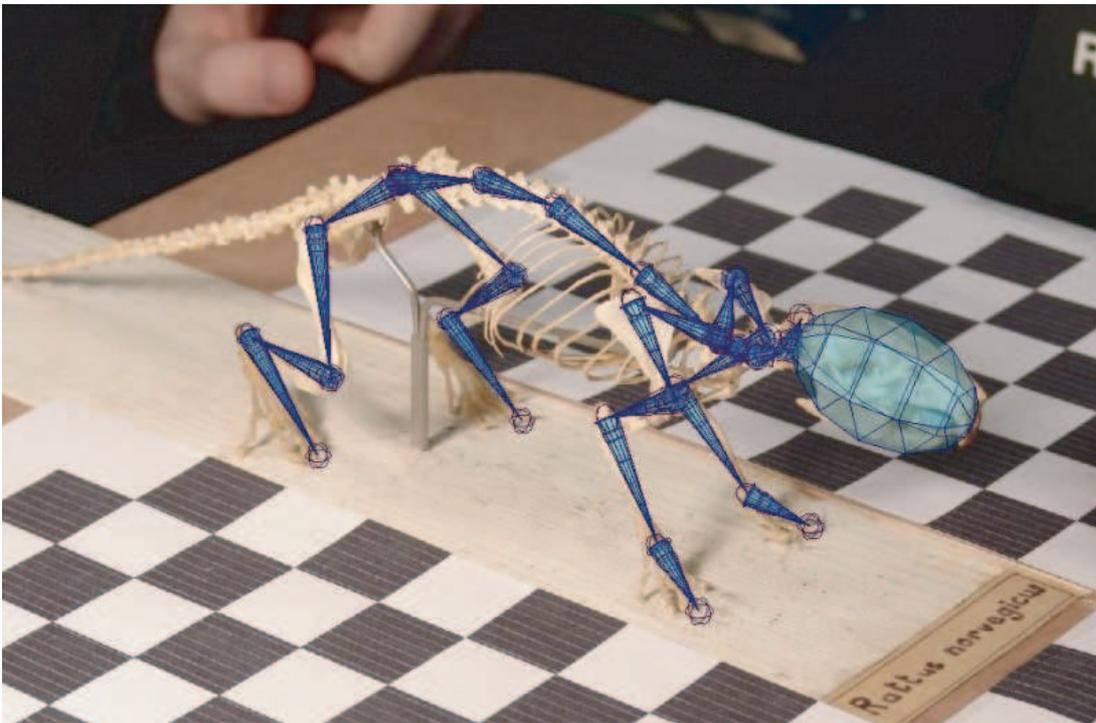


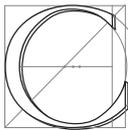
FIG. 4.6 – Le modèle 3D de rat reconstruit, vu depuis une caméra virtuelle correspondant aux paramètres de caméra extraits des images. Une fois le modèle superposé aux images d'origine, on peut apprécier visuellement la qualité de la reconstruction.



---

# Construction d'un squelette générique de quadrupèdes à partir d'images

---



CRÉER LE MODÈLE 3D d'un personnage commence toujours par la construction du squelette et le placement des articulations. Cette première étape implique la définition et le réglage de nombreux degrés de libertés, c'est-à-dire la position et l'orientation en 3D de chaque joint du squelette. C'est une tâche complexe qui est généralement réalisée par un animateur expérimenté. Dans le cas des animaux virtuels le problème est encore plus difficile, car nos connaissances sont plus réduites et les données anatomiques sont plus difficiles à obtenir.

Dans ce chapitre, nous montrons qu'une analyse statistique peut être appliquée sur un petit nombre de squelettes construits par un animateur expérimenté pour générer un modèle déformable de squelettes de quadrupèdes facile à adapter à un grand nombre d'animaux.

Les modèles déformables sont récemment devenus très populaires en informatique graphique. Ils permettent de créer, par analyse statistique, une paramétrisation pratique d'un en-

semble de données. L'espace des données peut alors être parcouru de manière intuitive. Dans [BV99], un modèle déformable de formes et de textures de visages 3D est appris à partir de milliers de modèles de visages obtenus avec un scanner 3D. Il permet d'explorer l'espace des visages en contrôlant des paramètres intuitifs tels que l'âge, le sexe, l'humeur, etc. De façon similaire un modèle déformable de formes de corps humains a été construit à partir de données issues de scanner 3D [ACP03]. Les modèles déformables maintiennent les résultats dans un espace de résultats plausible caractérisé par les exemples d'apprentissage. Ils donnent donc de meilleurs résultats que des techniques de déformation telles que les Déformations de Formes Libres (FFD).

Ce chapitre étudie pour la première fois la génération d'un modèle déformable dans le cas des squelettes d'animation de quadrupèdes. Ceci suppose d'utiliser une interpolation continue entre des données qui peuvent représenter à la fois des angles et des longueurs de joints. En particulier, la paramétrisation des rotations 3D peut présenter des singularités (comme le *gimbal lock* pour les angles d'Euler) et des ambiguïtés ( $2\pi$ -périodicité), ce qui la rend difficile à utiliser dans un modèle statistique. De plus, le modèle déformable doit prendre en compte des valeurs définies en différentes unités (par exemple des distances et des angles). Enfin, pour être utilisable, un modèle déformable doit offrir des paramètres de contrôle pertinents, c'est-à-dire assez généraux pour permettre suffisamment de variabilité dans les résultats et assez spécifiques pour permettre une utilisation intuitive. Ce chapitre discute ces problèmes et montre comment construire un modèle déformable de squelettes de quadrupèdes.

Les travaux présentés dans ce chapitre ont été publiés à la conférence *ACM SIGGRAPH / Eurographics Symposium on Computer Animation* [RFDC05].

## 1 Contexte

Wilhems et Van Gelder ont présenté l'une des premières méthodes de modélisation d'animaux 3D [JG97]. Leur approche est anatomique et repose donc sur une modélisation précise de la géométrie des os, des muscles et de la peau. Notre objectif est différent : nous souhaitons permettre la construction automatique de squelettes d'animation (définis comme une hiérarchie de joints) prêts à être animés à l'aide de méthodes d'habillage standard comme le skinning. Nous utilisons pour cela une analyse statistique sur un ensemble de squelettes d'animation plutôt que de la modélisation anatomique.

Wade and Parent utilisent un calcul de squelette topologique pour créer automatiquement des squelettes géométriques à l'intérieur de modèles polygonaux d'animaux [WP00]. Le squelette géométrique est nettoyé automatiquement afin d'être utilisable en animation. Dans notre cas, les squelettes de référence sont créés manuellement par un animateur expérimenté. Les articulations sont donc placées en fonction des caractéristiques anatomiques du quadrupède

correspondant, et la plupart du temps ne sont pas localisées sur le squelette topologique de la surface de l'animal (voir figure 5.1 par exemple). Notre approche, même si elle est moins automatique, vise à construire un squelette proche de celui d'un animateur expérimenté construirait.

Un modèle de croissance des animaux a été proposé par Walter et al. [WF97, WFM01]. Cette approche utilise le squelette, la forme du corps et des changements de texture pour représenter la croissance de l'animal. Nous étudions ici une autre source de variation, les changements dans la morphologie du squelette selon plusieurs espèces de quadrupèdes.

Sumner et Popović étudient le problème du transfert de l'animation d'un animal vers un autre animal [SJ04]. Les animaux sont modélisés par des maillages polygonaux, et transfert de mouvement est réalisé directement sur des maillages sans utiliser de squelette d'animation. Notre choix est plutôt de se concentrer sur la structure de squelette sous-jacente comme le contrôleur essentiel pour l'animation de personnages. Ceci nous permet d'insérer notre modèle déformable dans le cadre de travail standard de l'animation 3D.

L'approche théorique utilisée par Grochow et al. [GMHP04] est très proche de la notre. Ce travail combine un grand ensemble de mouvements humains capturés dans un modèle probabiliste sophistiqué qui permet de positionner un squelette 3D à partir de paramètres géométriques tels que des positions de cinématique inverse. Dans notre cas, même si nous voulons un modèle statistique contrôlé par des paramètres géométriques, nous travaillons sur des variations de morphologie entre squelettes d'animaux en pose de repos plutôt que sur du mouvement. En conséquence, nous n'allons pas seulement capturer des variations d'angles, mais aussi les longueurs des joints. Utiliser des données de différentes unités a soulevé des problèmes spécifiques que nous détaillons dans ce chapitre.

## 2 Apprentissage des données

Pour être utiles, les modèles déformables ont besoin de données précises pour la phase d'apprentissage. Les scanners 3D et les systèmes de capture de mouvement avec marqueurs peuvent fournir de telles données pour les humains. Obtenir des données sur les squelettes d'animaux est plus difficile. Comme nous voulons construire des modèles prêts à être animés, nous avons décidé d'apprendre notre modèle déformable à partir de quelques squelettes de référence construits par un animateur expérimenté (15 années d'expérience en production).

### 2.1 Les modèles de squelettes de référence

Nous avons considéré neuf quadrupèdes, couvrant un large spectre de morphologies : le cheval, la chèvre, l'ours, le lion, le rat, l'éléphant, la vache, le chien et le cochon. Les squelettes de références ont d'abord été construits à partir de références anatomiques [EDB56, Cal75].

Ces références donnent des dessins précis de l'anatomie interne et externe des animaux. Cependant, ces dessins ne donnent des informations que pour quelques vues (Côté, Face, Dessus). Pour pouvoir construire des squelettes 3D de manière plus précise, nous avons donc utilisé des modèles 3D de la surface des animaux. Afin d'assurer un alignement correct entre les données pendant la phase d'apprentissage, tous les squelettes d'animation ont la même topologie : même nombre d'articulations et même hiérarchie de joints. Nous avons utilisé la convention classique en animation qui est de choisir le pelvis comme racine de la hiérarchie. Chaque squelette consiste en 58 articulations avec 6 degrés de liberté chacune, 3 pour la position et 3 pour la rotation. La Figure 5.1 montre quelques étapes de création du squelette dans le cas de la vache.

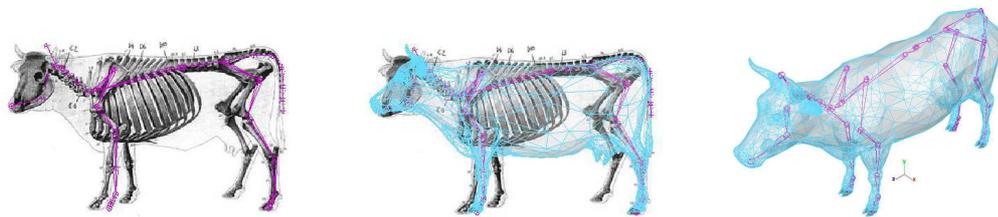


FIG. 5.1 – Construction des données d'apprentissage - l'exemple de la vache

## 2.2 Paramétrisation des données

Chaque squelette d'animation créé par l'animateur est paramétré comme un vecteur d'observation unique contenant la position et l'orientation de chaque articulation. Nous avons considéré deux alternatives pour la paramétrisation des données :

- utiliser les angles d'Euler ou les quaternions pour représenter les rotations,
- utiliser des coordonnées locales ou globales (dans le repère monde), pour représenter les rotations et les translations.

Dans les sections suivantes, nous montrons et discutons l'impact de ces choix. Pour plus de clarté, nous noterons ER pour les rotations en angles d'Euler, QR pour les rotations en quaternions, LC pour les coordonnées locales et GC pour les coordonnées globales.

Chaque squelette est représenté par un vecteur de  $58 * (3 + 3) = 348$  valeurs pour ER et de  $58 * (3 + 4) = 406$  pour QR. LC et GC sont obtenues par des manipulations standard de matrices de transformation, suivies d'une conversion de matrices en angles d'Euler ou en quaternions. Cela nous donne quatre conditions à tester : ER×LC, QR×LC, ER×GC, et QR×GC.

Lorsque nous rassemblons les données pour chaque animal, les paramètres de rotation doivent être traités avec un soin particulier pour éviter les discontinuités. L'analyse statistique va interpoler linéairement les données d'apprentissage, ce qui donnera de mauvais résultats si

	ER×LC	ER×GC	QR×LC	QR×GC
tx	1.1	0.6	1.1	0.6
ty	0.5	3.2	0.5	3.2
tz	0.6	1.6	0.6	1.6
rx	0.1	2.1	0.1	0.2
ry	0.2	1.2	0.1	0.2
rz	0.6	2.0	0.2	0.2
rs	n.a.	n.a.	0.2	0.2

TAB. 5.1 – Les écarts-types les plus grands rencontrés dans nos exemples, pour tous les paramètres de rotations et de translation et pour les 4 paramétrisations testées.

deux exemples ayant des matrices de rotation similaires sont représentés avec des paramètres différents, comme des angles d'Euler définis à  $2\pi$  près où des quaternions de signe opposé. De plus, les angles d'Euler perdent un degré de liberté quand ils sont près du *gimbal lock*, ce qui soulève des problèmes d'ambiguïtés. Pour chaque représentation, ER ou QR, nous cherchons et corrigeons ces problèmes sur les 9 exemples de nos données d'apprentissage.

Finalement, les données sont rassemblées dans une matrice  $\mathbf{X}$ . Les données sont arrangées en lignes, chaque ligne représentant un squelette.  $\mathbf{X}$  est donc une matrice  $9 \times 348$  for ER et  $9 \times 406$  pour QR.

### 2.3 Normalisation

Un problème important à régler avant de commencer l'analyse statistique est de spécifier clairement quelle variance est à considérer. En effet, le système métrique de l'animation 3D est sans unités. Un modèle peut avoir n'importe quelle unité de distance, et est donc défini à une constante près. Pour éliminer les effets de mise à l'échelle sur les données, tous les squelettes sont normalisés : les articulations du bassin sont positionnées au même endroit, et une mise à l'échelle globale est appliquée pour que la colonne vertébrale ait la même longueur (10) pour tous les animaux. Ceci permet d'explorer la variabilité des données indépendamment de la taille de l'animal.

D'autre part, nous utilisons des données définies en différentes unités comme des rotations et des translations. Les valeurs de rotation sont exprimées en radians pour ER et sont contraintes dans l'intervalle de  $[-1, +1]$  pour QR. Le Tableau 2.3 montre, pour chaque paramètre de translation et de rotation, le plus grand écart type calculé parmi toutes les articulations de tous les animaux (dans un souci de clarté nous ne présentons pas les résultats articulation par articulation). Ceci donne une idée de l'échelle des données d'apprentissage.

Une pratique courante en analyse statistique consiste à diviser chaque paramètre par son écart-type calculé sur les exemples d'apprentissage. Les valeurs pourraient aussi être normalisées par des valeurs différentes pour donner plus de poids soit aux rotations, soit aux translations. Nous avons testé plusieurs valeurs de normalisation et nous n'avons pas trouvé de différences significatives sur les résultats.

### 3 Etude de la dimensionnalité

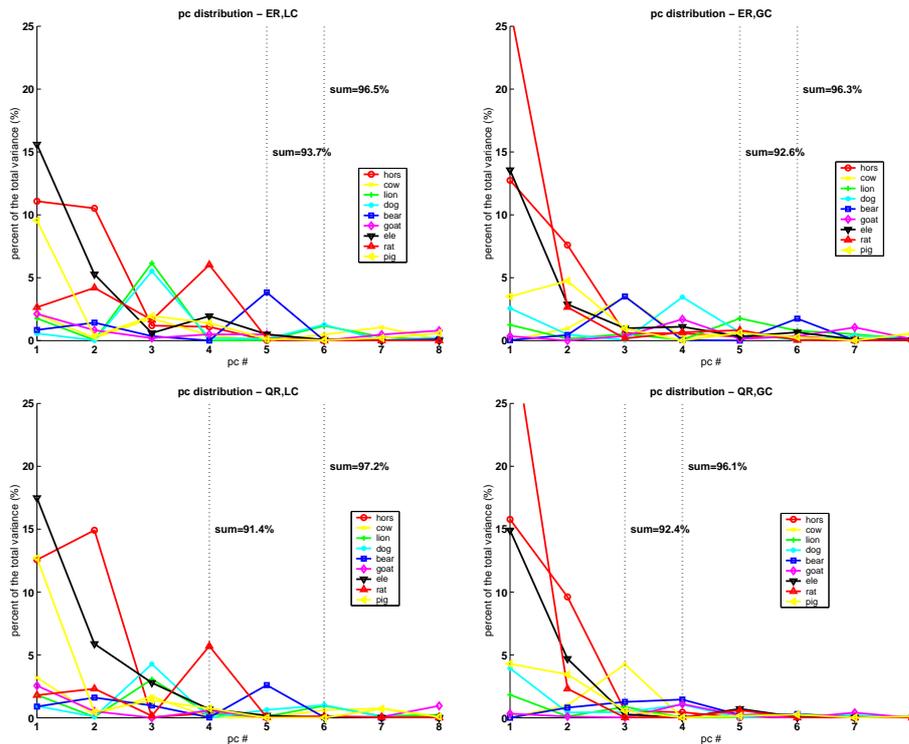


FIG. 5.2 – Distribution des composantes principales pour les quatre paramétrisations

Même si chaque squelette est paramétrisé par 348 (ou 406) valeurs scalaires, nos 9 exemples d'apprentissage donnent au modèle un maximum de 8 degrés de liberté linéaires, plus une pose moyenne. Dans cette section, nous voulons étudier la redondance entre les exemples. La question est de savoir si cette redondance peut être factorisée en moins de 8 paramètres de contrôles (l'interpolation de squelettes, de manière similaire à l'interpolation de formes, serait une paramétrisation canonique de ce modèle déformable). Nous avons étudié ce modèle en appliquant d'abord une ACP standard aux données d'apprentissage pour chacune des quatre conditions  $\{ER, QR\} \times \{LC, GC\}$ . La Figure 5.2 montre les distributions des composantes principales selon les exemples dans le cas d'une paramétrisation à 4 paramètres. Nous allons détailler

comment cette distribution a été estimée.

D'habitude, les résultats d'ACP sont étudiés en termes de la quantité de variance accumulée en ajoutant successivement des composantes principales. La Figure 5.2 suit cette idée mais ajoute des détails pour analyser au cas par cas les exemples d'apprentissages. Ces détails sont calculés de la manière suivante : l'ACP consiste à calculer les  $p$  valeurs propres et les  $p$  vecteurs propres d'un ensemble de données  $\mathbf{X}$  à  $n$  éléments. Les valeurs propres sont stockées dans une matrice diagonale  $\mathbf{D}$ . Les vecteurs propres sont stockés comme les colonnes d'une matrice  $\mathbf{E}$ .  $\mathbf{X}$  est centré sur le vecteur squelette moyen. Dans notre cas, nous avons un maximum de  $p = 8$  vecteurs propres associés à une variance non nulle.

$$\frac{1}{n}\mathbf{X}^T\mathbf{X}\mathbf{E} = \mathbf{E}\mathbf{D} \quad (5.1)$$

$$\mathbf{E}^T\mathbf{E} = \mathbf{I} \quad (5.2)$$

$\mathbf{D}$  est une matrice diagonale  $p \times p$  où chaque élément donne la variance totale des données expliquée par chacune des  $p$  composantes principales. La variance totale des données est donnée par  $\text{trace}(\mathbf{D})$ . Chaque colonne de la matrice  $\mathbf{P} = \mathbf{X}\mathbf{E}$  donne la projection de tous les individus sur chaque vecteur propre. La matrice  $\mathbf{D}$  peut être interprétée comme :

$$\mathbf{D} = \frac{1}{n}\mathbf{P}^T\mathbf{P} \quad (5.3)$$

Si nous développons  $\mathbf{D}$  selon les vecteurs lignes  $\mathbf{p}_i$  de  $\mathbf{P}$ , nous obtenons :

$$\mathbf{D} = \frac{1}{n} \sum_{i=1}^n \mathbf{p}_i^T \mathbf{p}_i \quad (5.4)$$

Finalement, comme  $\mathbf{D}$  est diagonale, nous avons pour chaque élément  $D_j$  :

$$D_j = \frac{1}{n} \sum_{i=1}^n P_{ij}^2 \quad (5.5)$$

La variance de chaque composante peut donc être décomposée selon la contribution de chacun des  $n$  individus. Sur la Figure 5.2, nous affichons le carré de la projection  $P_{ij}^2$  de chaque individu sur de la  $j$ -th composante, normalisé par la variance totale  $\text{trace}(\mathbf{D})$ . De plus, nous indiquons par une ligne verticale à quelle composante la variance cumulative de *tous les individus* portée par les composantes successive atteint 90% et 95% de la variance totale.

La figure Figure 5.2 nous apprend que certaines paramétrisations conduisent à des cas de surapprentissage. En effet, surtout dans les cas LC, la variance d'une certaine composante peut être liée en grande partie à un seul individu. Cela signifie que cette composante ne factorise pas d'informations parmi les données. Les meilleurs résultats sont obtenus pour la condition

GC×QR (coordonnées globales et représentation par quaternions). Dans cette paramétrisation, on constate peu de surapprentissage. De plus, 90% de la variance totale est capturée en utilisant les trois premières composantes. Ce résultat suggère que, dans les conditions GC×QR, toutes les données peuvent être réduites efficacement en un modèle linéaire contrôlé par 3 paramètres seulement.

Un nouveau squelette d'animation est généré simplement par combinaison linéaire des vecteurs propres, à laquelle on ajoute le squelette moyen  $\bar{\mathbf{x}}$  des exemples d'apprentissage :

$$\mathbf{x}(\mathbf{p}) = \bar{\mathbf{x}} + \mathbf{p}\mathbf{E}^T \quad (5.6)$$

Sur la Figure 5.3, nous montrons les résultats de la variation du modèle déformable selon les trois premiers modes linéaires à -2,-1,+1 et +2 fois l'écart-type du paramètre. La colonne centrale correspond à la forme moyenne.

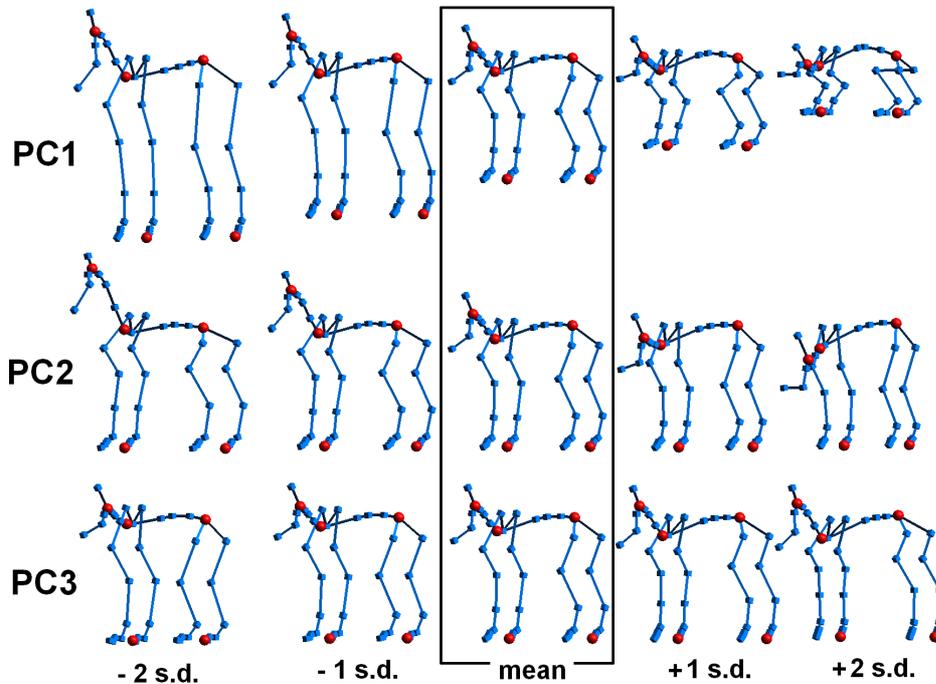


FIG. 5.3 – Les trois premiers modes de variation de l'ACP sur les squelettes. Étonnamment, ces modes sont faciles à identifier intuitivement comme la hauteur de l'animal au repos, l'inclinaison de la colonne vertébrale, et un paramètre ongulé/plantigrande.

Ces paramètres peuvent être implicitement interprétés comme étant :

- la hauteur de l'animal
- l'inclinaison de la colonne vertébrale

- une variation entre ongulé et plantigrade, qu'on pourrait caractériser par l'inclinaison des pieds par rapport au sol.

Cependant, ils ne sont pas faciles à aligner sur un modèle 3D : leurs valeurs seraient difficiles à régler à la main pour un maillage particulier.

Nous concluons cette section en rappelant que 3 paramètres semblent être suffisants. Dans la prochaine section, nous étudions 3 autres paramètres qui sont plus intuitifs : des mesures géométriques prises directement sur la surface de l'animal.

## 4 Paramétrisation géométrique

Dans [ACP03], Allen et al. mentionnent que même si l'ACP fournit des paramètres compacts, ils ne sont pas toujours intuitifs à utiliser. Au contraire, leur espace de formes 3D de corps humains est contrôlé avec des paramètres tels que le poids et la hauteur pour rendre le modèle déformable plus intuitif à utiliser. Nous utilisons la même approche pour notre modèle déformable de squelettes de quadrupèdes, ce qui rend l'alignement du squelette sur n'importe quel maillage plus simple à réaliser. Nous proposons de choisir comme paramètres de contrôle des paramètres géométriques intuitifs qui peuvent être mesurés sur le squelette anatomique de l'animal vu de côté (Figure 5.4) ou sur un maillage 3D. Le squelette généré sera alors aligné avec le maillage si le modèle déformable est stable, c'est-à-dire si les paramètres de contrôle géométriques génèrent un squelette qui ait les mêmes mesures géométriques.

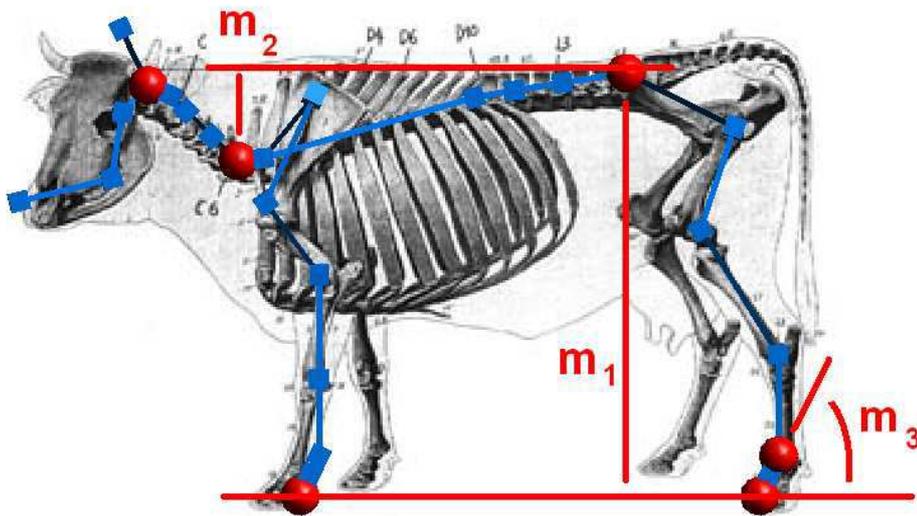


FIG. 5.4 – Les trois mesures géométriques qui contrôlent notre modèle déformable.

A partir de l'observation des trois premiers mode d'ACP et de l'intuition donnée dans

la section précédente, trois paramètres géométriques ont été testés ( à comparer avec les 9 paramètres utilisés dans [WF97, WFM01] pour la croissance des animaux ). Ces paramètres, visibles sur la Figure 5.4, sont :

- la hauteur de l'animal  $\mathbf{m}_1$ , mesurée sur un maillage comme la distance verticale entre une position estimée du pelvis et le sol ;
- l'inclinaison de la colonne vertébrale  $\mathbf{m}_2$  mesurée comme étant la différence entre la hauteur de l'animal (comme définie ci-dessus) et la hauteur de la base du cou (Rappelons que la longueur de la colonne vertébrale est gardée constante et est utilisée comme un facteur d'échelle) ;
- un paramètre *ongulé/plantigrade*, mesuré par l'angle  $\mathbf{m}_3$  entre le sol et la ligne qui joint le pied arrière à la cheville arrière.

Ces paramètres sont extraits des squelettes de référence en calculant les matrices de transformation globales sur le pelvis, la première vertèbre cervicale et la cheville arrière. Comme premier modèle, nous calculons une transformation linéaire de ces trois paramètres de mesure vers les rotations et translations du squelette. Des transformations plus complexes pourraient être appliquées, comme les Fonctions de Base Radiales (FBR) [LCF00] où encore le schéma sophistiqué utilisé par [GMHP04]. Une prédiction linéaire s'est révélée suffisante dans notre cas. La prédiction est faite depuis un vecteur de mesures géométriques  $\mathbf{m} = [m_1, m_2, m_3]$  vers un vecteur squelette  $\mathbf{x}$ . Ces deux vecteurs sont des vecteurs lignes :

$$\mathbf{x}(\mathbf{m}) = \bar{\mathbf{x}} + \mathbf{m}\mathbf{V} \quad (5.7)$$

Le modèle linéaire  $\mathbf{V}$  est estimé par un ajustement aux moindres carrés entre les squelettes de référence  $\mathbf{X}$  et la matrice  $\mathbf{M}$  composée des vecteurs lignes d'observations empilés. On en déduit :

$$\mathbf{V} = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T\mathbf{X} \quad (5.8)$$

La Figure 5.5 montre les variations de ce modèle déformable selon ses trois paramètres de contrôle. Notez que, contrairement au troisième mode d'ACP, le troisième paramètre corrèle l'inclinaison du pied par rapport au sol avec la longueur du cou. Cette corrélation existe effectivement sur nos exemples d'animaux ; nous n'avons pas encore cherché à voir si c'est valide sur tous les quadrupèdes, ce qui pourrait être vrai pour des raisons anatomiques. La longueur du cou est corrélée avec l'inclinaison de la colonne vertébrale, comme le montrent le second mode d'ACP sur la Figure 5.3 ainsi que le second paramètre géométrique sur la Figure 5.5.

D'autres paramètres pourraient aussi être utilisés comme paramètres de contrôle du modèle déformable. Nous avons fait des expériences en remplaçant le paramètre *ongulé/plantigrade*

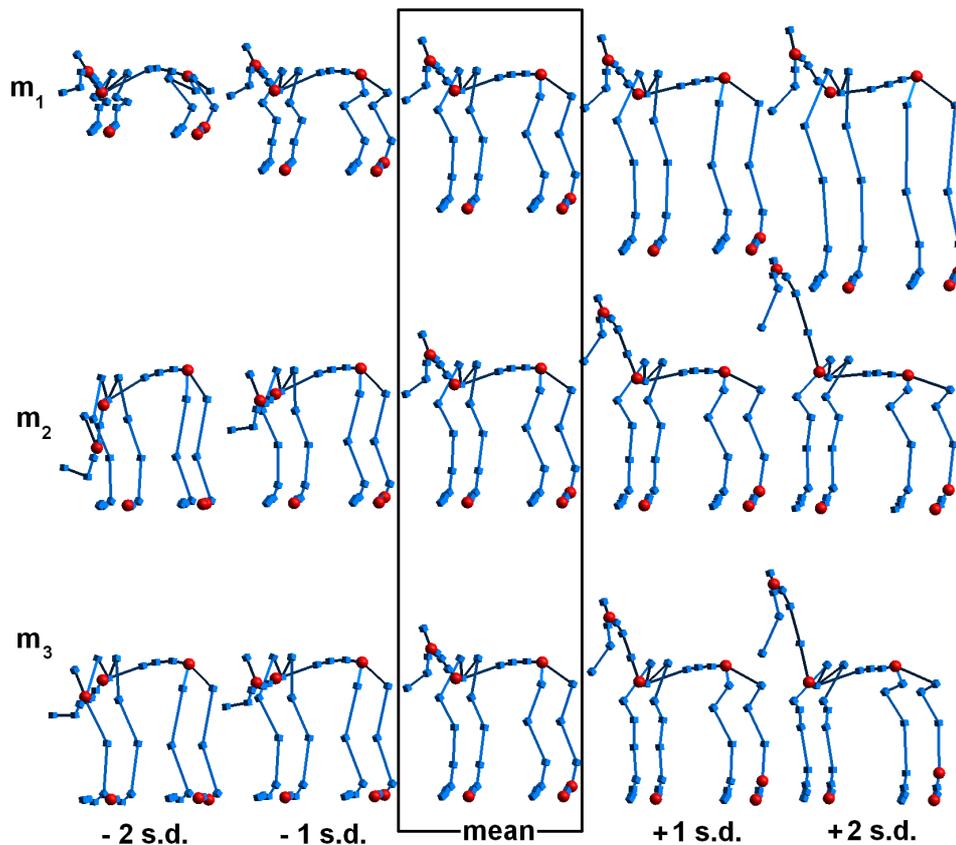


FIG. 5.5 – Les trois premiers modes de variation du modèle déformable, contrôlés par la hauteur de l’animal, l’inclinaison de la colonne vertébrale, et le paramètre ongulé/plantigrade (angle du pied).

par la longueur du cou. Ceci a pour résultat de garder la longueur du cou constante lorsqu’on change l’inclinaison de la colonne vertébrale, comme on peut le voir sur la Figure 5.6.

## 5 Résultats et discussion

### 5.1 Reconstruction des données

Dans cette section, nous discutons l’impact des 4 conditions  $ER$ ,  $QR \times LC$ ,  $GC$  sur les propriétés des modèles déformables obtenus par PCA et par nos modèles de prédiction. Nous évaluons les résultats en regardant si les exemples d’apprentissage sont correctement reconstruits par le modèle déformable à partir de leur paramètres d’entrée associés. Dans le cas de l’ACP, les paramètres d’entrée sont les coefficients de projection des exemples sur les vecteurs

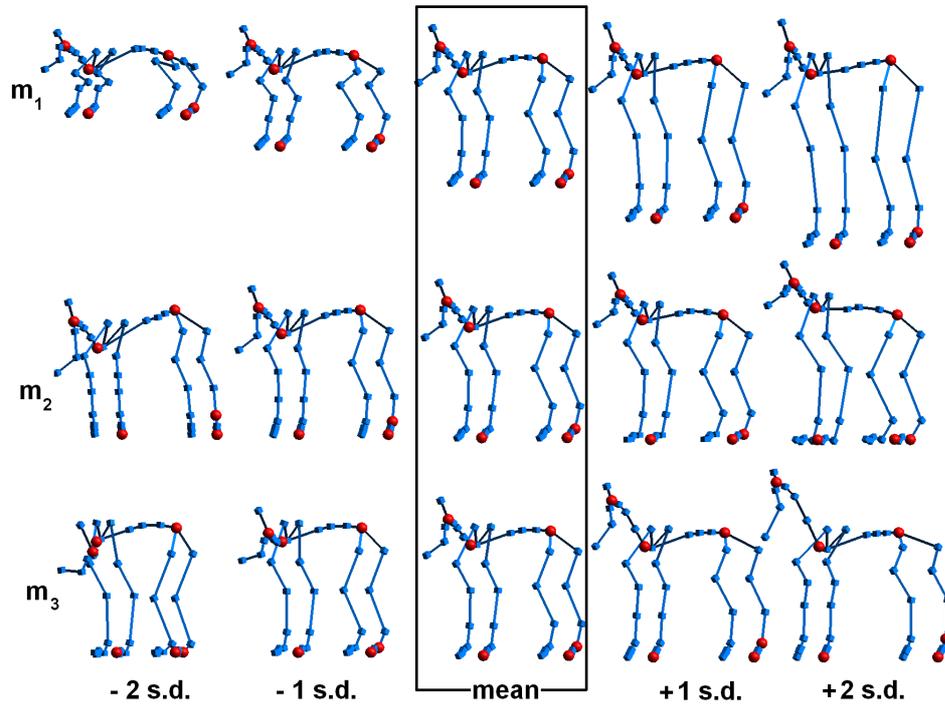


FIG. 5.6 – Les trois premiers modes de variation du modèle déformable, contrôle par la hauteur de l'animal, l'inclinaison de la colonne vertébrale, et la longueur du cou.

propres. Dans le cas du modèle déformable, ce sont les mesures effectuées sur les squelettes. Comme la taille des données est petite et que le nombre de paramètres (trois) est inférieur au nombre d'exemples d'apprentissage (neuf), nous nous attendons à une prédiction non exacte dans les deux cas. Nous séparons l'évaluation en deux parties :

1. l'erreur sur les translations
2. l'erreur sur les rotations

Nous prenons donc tous les squelettes d'apprentissage et les squelettes prédits en utilisant les 4 conditions. Les matrices globales de transformation  $4 \times 4$  sont calculées et la partie translation (vecteur  $3 \times 1$  issu de la quatrième colonne) et rotation (la première sous-matrice  $3 \times 3$ ) sont séparées. Pour chaque individu, l'erreur est estimée comme le maximum sur toutes les articulations de la norme de la différence entre le squelette de référence et le squelette prédit, calculée sur les vecteurs de translation pour les translation et sur les matrices de rotation pour les rotations (dans ce cas nous utilisons la plus grande valeur singulière).

La Figure 5.7 (resp. Figure 5.8) rassemble les résultats pour les 9 exemples sous les 4 conditions en utilisant le modèle d'ACP contrôlé par les trois premières composantes princi-

pales (resp. le modèle déformable). Elles montrent que la reconstruction des données est aussi bonne avec notre modèle basé sur des mesures anatomiques qu'en utilisant directement l'ACP.

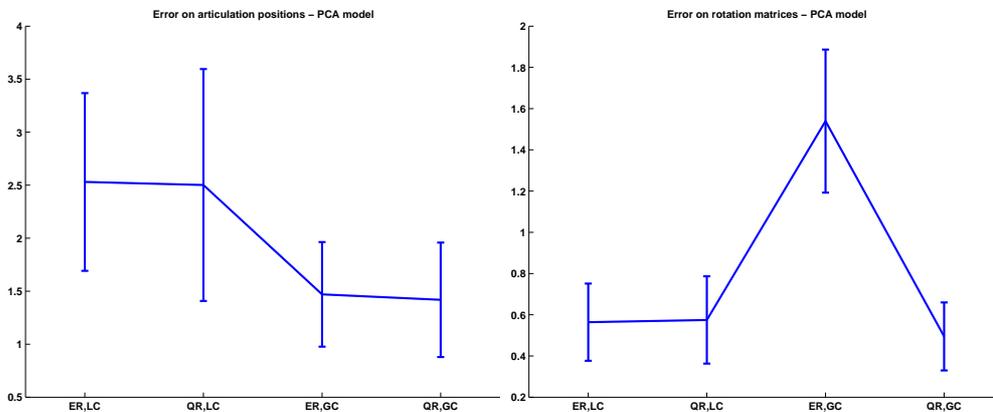


FIG. 5.7 – La moyenne et l'écart-type de l'erreur de prédiction pour les 4 conditions : translation (à gauche) et rotation (à droite) en utilisant le modèle contrôlé par ACP.

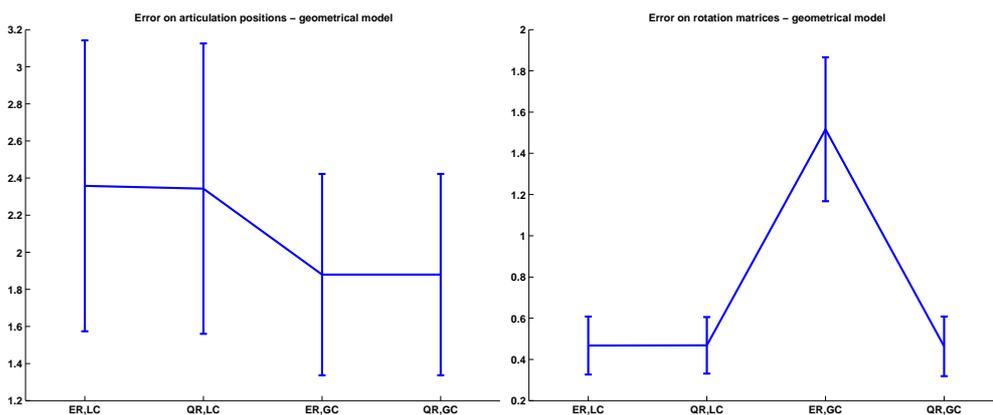


FIG. 5.8 – La moyenne et l'écart-type de l'erreur de prédiction pour les 4 conditions : translation (à gauche) et rotation (à droite) en utilisant notre modèle déformable.

Pour chaque modèle, ces résultats montrent que l'erreur sur les translations est plus petite en utilisant un repère global (condition GC), quelque soit la représentation des rotations. Ceci vient du fait que dans la condition LC, les erreurs de prédiction peuvent se cumuler à cause de la structure hiérarchique des squelettes. C'est particulièrement visible sur les joints extrémaux (le bout des pattes). La condition GC induit de plus grandes variations sur les données de rotation. Dans ce cas, les quaternions offrent une plus grande stabilité que les angles d'Euler. Ceci explique les meilleurs résultats de  $GC \times QR$  comparé à  $GC \times ER$ . La Figure 5.9 illustre ces deux représentations des données dans notre plus mauvais cas de prédiction, l'éléphant. Pour conclure, ces observations suggèrent que les paramètres de transformation globale et la

représentation par quaternions sont les meilleurs choix pour construire un modèle déformable de squelettes de quadrupèdes.

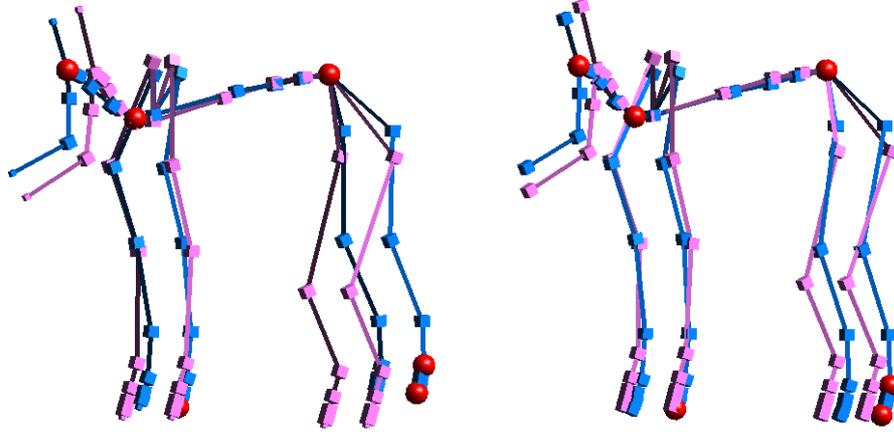


FIG. 5.9 – Comparaison entre les conditions  $QR \times LC$  (à gauche) et  $QR \times GC$  (à droite) dans le cas de l'éléphant. La référence est en violet et le squelette prédit est en bleu.

## 5.2 Animation d'un nouveau quadrupède

Pour valider le procédé en entier - à partir de la génération du squelette jusqu'à l'animation finale - nous avons appliqué notre modèle générique de squelette à un animal qui n'était pas présent dans notre base d'apprentissage : le chat. Nous avons généré le squelette en utilisant le modèle contrôlé par mesures géométriques à partir de mesures faites sur une vue de côté du maillage de l'animal. La peau du chat a été ensuite attachée au squelette en utilisant une technique de skinning (Figure 5.10). Nous avons également pu utiliser ce modèle pour faire du transfert de mouvement. La même animation de marche est appliquée sur le squelette pendant que les paramètres du modèle déformable varient continûment au cours de la séquence<sup>1</sup>.

## 6 Bilan

Dans ce chapitre, nous avons montré que les analyses statistiques peuvent être utilisées pour la génération automatique de squelettes d'animation. Nous avons traité le problème dans le cas spécifique des quadrupèdes, en utilisant des squelettes construits par un animateur expérimenté comme données d'apprentissage ainsi que pour la validation du modèle. Nous avons montré

<sup>1</sup>La vidéo correspondant à ce chapitre peut être consultée à l'adresse web suivante : <http://www-evasion.imag.fr/Publications/2005/RFDC05/>

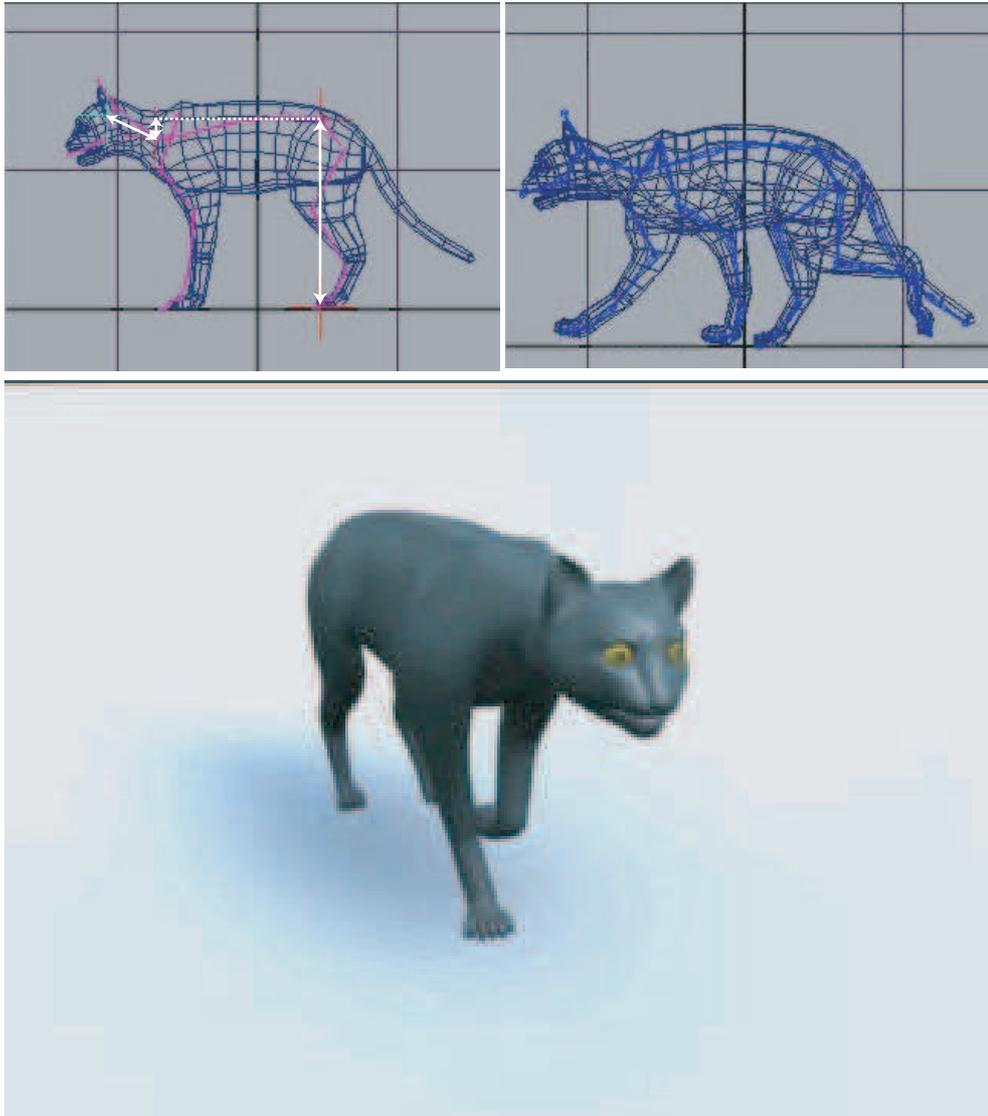


FIG. 5.10 – Animation d'un nouveau quadrupède : ajustement du squelette, habillage, rendu final.

qu'un espace à 3 dimensions est suffisant pour représenter cet ensemble de squelettes d'animation. Le modèle déformable de squelette obtenu peut être facilement ajusté sur un nouveau quadrupède en prenant trois mesures simples sur la vue de côté de l'animal. Nous avons aussi montré que les squelettes obtenus donnent de bons résultats d'animation quand on leur attache une peau grâce à une technique de skinning.

En pratique, un animateur expérimenté peut utiliser ce squelette généré automatiquement un point de départ et obtenir un modèle très réaliste en réglant ensuite finement chaque joint.



## **Troisième partie**

# **Animation de quadrupèdes à partir de vidéos**



---

# Animation de mouvements cycliques à partir de vidéos

---



LES MÉTHODES CLASSIQUES de capture de mouvement à l'aide de marqueurs, si elles donnent de bons résultats, ne peuvent être appliquées qu'à des sujets coopératifs : Le sujet doit porter des marqueurs et se déplacer dans un endroit confiné, parfois même sur un tapis roulant. Le nombre de mouvements d'animaux adaptés à ces méthodes est donc très limité : si il est possible de les appliquer à un chien ou à un cheval, la capture de mouvements d'animaux sauvages est impossible. Or ce type de mouvement est très intéressant pour les films d'animation 3D ou les effets spéciaux où, au-delà des animaux dangereux, des créatures fantastiques doivent être animées alors qu'aucune source de mouvement n'est disponible.

Dans ce chapitre, nous proposons une méthode d'extraction de mouvements 3D cycliques de quadrupèdes à partir de séquences vidéo monoculaires, notamment de séquences extraites de documentaires animaliers. Cette méthode a la particularité de permettre le transfert du mouvement : un mouvement capturé sur un animal donné pourra être rejoué sur tout modèle 3D de quadrupède auquel on souhaite appliquer un mouvement similaire.

Les méthodes classiques de capture de mouvement sans marqueurs sont encore difficiles à utiliser dans le cadre de l’animation. Nous proposons une technique alternative qui combine des techniques robustes existantes dans un nouveau pipe-line : nous combinons l’Analyse en Composantes Principales (ACP) sur les images et l’animation par interpolation d’exemples pour générer de l’animation 3D de mouvement cycliques d’animaux à partir de séquences vidéo. L’ACP sur les images est bien adaptée pour les mouvements cycliques car elle isole les modes principaux de variation des données en factorisant les variations corrélées entre elles. Nos expériences montrent que certaines contraintes et traitements additionnels peuvent être utilisés pour aider à n’encoder que les variations liées au mouvement. Notre objectif est d’isoler et de caractériser, en utilisant l’ACP sur les images, un ensemble minimal de paramètres permettant de décrire un mouvement cyclique et de générer l’animation 3D correspondante par interpolation d’exemples, guidée par la vidéo.

Notre méthode permet d’économiser beaucoup de temps du point de vue de l’animateur. Traditionnellement, pour l’animation des quadrupèdes, un artiste travaille effectivement avec une vidéo, mais il passe beaucoup de temps à analyser visuellement cette dernière et doit faire une série d’essais erreurs pour déterminer de quelles images s’inspirer pour créer les poses-clés de l’animal 3D. Notre méthode basée sur l’ACP permet de déterminer directement les images-clés<sup>1</sup> importantes pour lesquelles une pose 3D doit être fournie. La méthode d’interpolation permet de générer automatiquement de longues séquences de mouvement 3D qui suivent le rythme de la vidéo originale.

Les deux principales contributions de ce chapitre sont : un critère automatique de sélection des images de la vidéo auxquelles on devra associer un exemple de pose 3D, et un algorithme de résolution des ambiguïtés dans la prédiction de l’enchaînement des poses 3D à partir de vidéo 2D.

Ce travail a fait l’objet de publication à la conférence *ACM SIGGRAPH / Eurographics Symposium on Computer Animation* [FRDC04], et dans une version étendue au journal *Graphical Models* [FRDC06].

## 1 Contexte

Le travail de Wilhelms et al.[WG03] est l’un des premiers à s’intéresser à l’animation d’animaux à partir de vidéo. Des contours déformables (*snakes*) sont alignés sur une séquence vidéo d’un cheval au galop pour extraire le mouvement en 2D des jambes. Ensuite ce mouvement 2D est transformé en mouvement 3D en alignant les contours d’un modèle 3D sur les contours 2D. Les méthodes de contours actifs sont connues pour être très sensibles au bruit et dépendent de

---

<sup>1</sup>Dans ce chapitre, les terme *image clé* désigne une image d’une vidéo qui représente une pose importante pour le mouvement

paramètres difficiles à régler. Wilhelms et al. [WG03] mentionnent ce problème et permettent à l'utilisateur de réinitialiser les contours actifs. Ceci rend la méthode difficile à utiliser dans un cadre général, surtout lorsque de nombreuses occlusions entre les jambes se produisent. Plus généralement, Gleicher et al. [GF02] montrent que les méthodes actuelles de traitement automatique des vidéos ne peuvent pas donner d'informations 3D robustes telles que des rotations de joints stables dans le temps. Ils concluent qu'utiliser ce genre d'approche pour le contrôle direct d'animation 3D joint par joint n'est pas faisable.

Les méthodes à base d'exemples ont été récemment reconnues comme une bonne alternative aux méthodes classiques de modélisation et d'animation. L'idée de base est d'interpoler entre un ensemble donné d'exemples, ce qui donne une fonction d'un espace de paramètres d'entrée vers l'espace des exemples. Les exemples peuvent être des poses 3D ou des séquences de mouvement 3D entre autres. Rose et al. [RBC98] paramétrisent la synthèse de nouveaux mouvements à partir de séquences de mouvement capturé (*motion capture*) étiquetées avec des paramètres abstraits qui caractérisent le style du mouvement. Lewis et al. [LCF00] interpolent des surfaces d'un bras qui plie à partir de valeurs d'angle en utilisant des Fonctions de Base Radiales (FBR). Ils montrent que l'interpolation d'exemple permet d'éviter les défauts des méthodes classiques de skinning. Sloan et al. [SIC01] étendent cette formulation en combinant FBR et régression linéaire. Pyun et al. [PKC<sup>+</sup>03] montrent qu'un cadre de travail similaire peut être utilisé pour animer des nouveaux visages à partir de captures de mouvement facial. Dans ce cas, les paramètres abstraits sont remplacés par les données d'animation 3D. Des paramètres visuel extraits d'images 2D peuvent aussi être utilisés comme paramètres d'entrée pour contrôler l'interpolation : Bregler et al. [BLCD02] capturent des informations à partir de dessins animés pour contrôler l'animation par interpolation de formes-clés de personnages 3D.

## 2 Vue d'ensemble de la méthode

Notre méthode se range dans la catégorie des méthodes à base d'exemples. Nous voulons voir dans quelles conditions des vidéos réelles sont utilisables comme paramètres d'entrée pour contrôler de l'animation 3D. Les vidéos réelles représentent un défi particulier : elles manquent de contraste et de résolution. La détection automatique d'indices, tels que des points d'intérêt, est donc difficilement applicable et nécessiterait beaucoup d'interventions de l'utilisateur. Notre idée est de convertir les images originales en images binaires, sur lesquelles une Analyse en Composantes Principales (ACP) est appliquée. Les coordonnées des images dans l'espace des composantes principales représentent un bon ensemble de paramètres pour contrôler l'animation 3D.

Quand les paramètres d'entrée sont dérivés d'un grand ensemble de données, toutes les

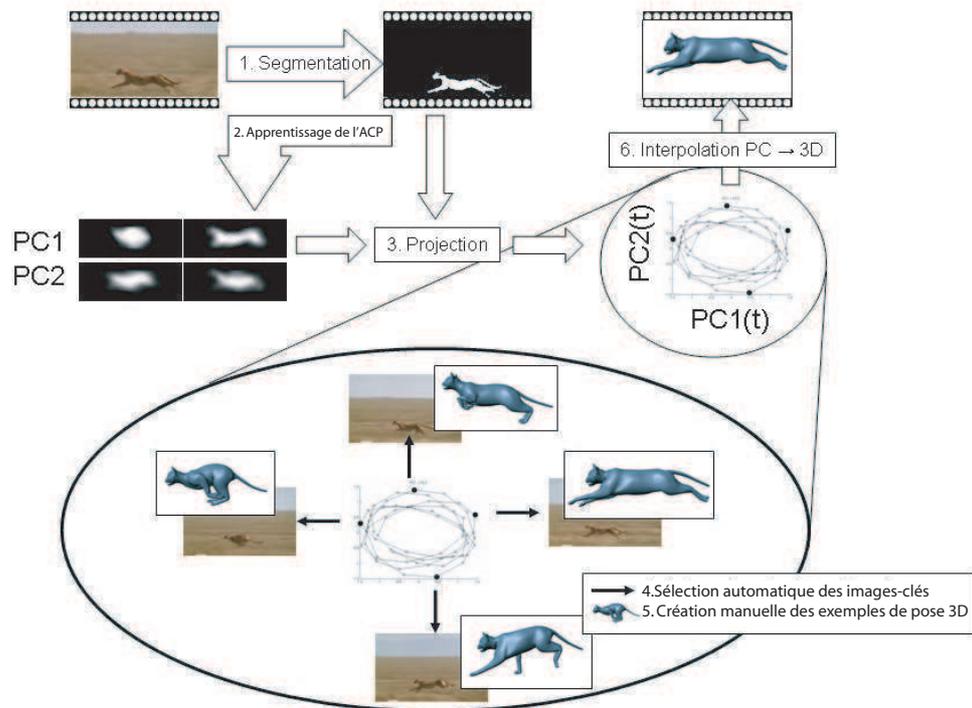


FIG. 6.1 – Vue d'ensemble de la méthode

méthodes à base d'exemples nécessitent que l'utilisateur choisisse les exemples à la main. Nous proposons un critère nouveau et automatique pour sélectionner ces exemples. Les Fonctions de Base Radiales (FBR) sont utilisées pour interpoler entre ces exemples au cours du temps, à partir de la séquence de paramètres extraits de la vidéo.

### 3 Prédiction d'animation 3D à partir d'ACP sur les images.

Notre approche combine analyse statistique d'images binaires par ACP et animation par interpolation d'exemples. Idéalement, les images binaires peuvent être générées par segmentation automatique. Lorsque la segmentation automatique n'est pas applicable, nous proposons un outil de dessin pour étiqueter la vidéo, et des traits blancs dessinés sur un fond noir créent l'image binaire. L'ACP est ensuite appliquée sur les images binaires, en considérant chaque image comme un vecteur d'observation unique. Les coefficients de projection sur les composantes principales sont analysés pour extraire les meilleurs exemples de pose 3D à interpoler. Ces coefficients de projection servent de paramètres d'entrée pour l'interpolation et contrôlent

l'évolution temporelle de l'animation (Figure 6.1).

### 3.1 Réduction de la variabilité dans des images binaires

Utiliser l'ACP directement sur les images encoderait toutes les variations en apparence des images : les variations dues au mouvement, mais aussi les variations d'éclairage et le mouvement de la caméra. Nous segmentons donc la vidéo en images binaires avant d'appliquer l'ACP pour filtrer les variations indésirables de l'image et séparer l'animal du décor.

Plusieurs méthodes peuvent être utilisées pour la segmentation d'images. Nous avons utilisé une segmentation d'images par rapport aux couleurs à base de mixture de gaussiennes, qui est facile à implémenter et donne de bons résultats sur nos séquences de test. Nous demandons à l'utilisateur d'indiquer grossièrement la position de l'animal et du décor en choisissant deux rectangles sur la première image de chaque séquence : un pour l'animal et un pour le décor. Les résultats sont bons sur nos séquences de test (Figure 6.1 et le haut de la Figure 6.2), mais nous pourrions bien sur utiliser des méthodes plus élaborées et obtenir des résultats encore plus précis [SM00, WBC<sup>+</sup>05].

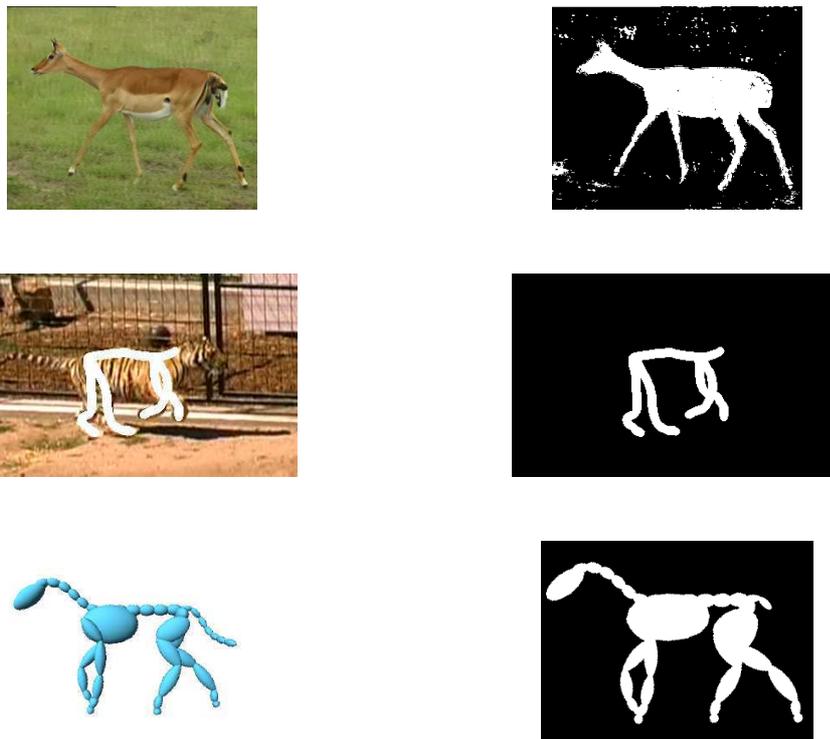


FIG. 6.2 – Les résultats de segmentation obtenus pour nos trois sources de données : vidéo, croquis et images de synthèse.

Parfois, la segmentation automatique n'est pas applicable à la vidéo : animal de la même couleur que le décor, images très texturées ou très bruitées. Dans ce cas nous proposons à l'utilisateur une interface de dessin pour tracer des traits sur la vidéo à l'emplacement de l'animal. Les croquis n'ont pas besoin d'être aussi précis que dans Davis et al. [DAC<sup>+</sup>03], où le dessin doit être assez précis pour que les joints puissent être identifiés automatiquement. Dans notre cas, les grandes variations d'éclairage et la fréquence des occlusions empêchent de compter sur un étiquetage précis des joints. Au contraire, nous nous basons sur un étiquetage rapide, avec l'outil de dessin, des caractéristiques les plus importantes de la morphologie de l'animal telles que la colonne, les pattes et la tête (Figure 6.2). Il n'est pas nécessaire de dessiner chaque joint individuellement si ils n'apparaissent pas sur l'image, l'important étant que l'utilisateur utilise le même protocole d'étiquetage sur toutes les images. L'idée est d'appliquer l'ACP sur les images de la même manière qu'elles soient issues d'une segmentation automatique, dessinées à la main ou, ce qui est indispensable pour la validation, générées à partir de données 3D.

Une fois que l'animal est isolé du décor, une région d'intérêt est automatiquement extraite autour de la silhouette par des opérations standard de morphologie mathématique et d'extraction de composantes connexes. Ce processus est appliqué à toutes les images de la séquence vidéo. Pour que la région d'intérêt soit toujours autour de la bonne composante connexe, nous utilisons comme première estimation de la position de l'animal la position du centre de masse de la silhouette au pas de temps précédent.

Cette étape nous permet de nous débarrasser de la variance due au mouvement de la caméra, qui ne correspond pas au mouvement en 3D de l'animal. Cependant, cela élimine aussi la translation de l'animal, qui fait partie intégrante du mouvement. Ceci ne pose pas de problème car nous ne voulons pas extraire la position 3D de l'animal comme un paramètre indépendant. Notre but est au contraire de capturer le timing général du mouvement et de prédire du mouvement 3D par interpolation d'exemple de poses 3D. En conséquence, si une translation verticale est spécifiée dans les exemples de poses et que cette translation est corrélée avec le reste du mouvement visible sur la séquence d'images, elle apparaîtra dans l'animal finale. Par exemple, une position en extension du corps sera corrélée avec une phase de suspension dans la démarche, ce qui devrait être suffisant pour retrouver la translation verticale de l'animal. La méthode pourrait être étendue pour capturer la translation horizontale à l'aide de techniques de flot optique. En pratique, nous avons simplement ajouté une translation adéquate correspondant à chaque exemple de pose.

À la fin de ce pré-traitement, nous obtenons une séquence d'images binaires. Le Tableau 3.1 donne, pour sept de nos séquences de test, le nombre d'images, la taille en pixels des images originales, et la taille en pixels de la fenêtre suivie autour de la silhouette.

Séquence	Nombre d'images	Image originale	Image binaire
Cheval, marche	100	320x240	320x240
Cheval, canter	100	320x240	320x240
Cheval, gallop	100	320x240	320x240
Guépard, gallop	137	192x144	90x34
Tigre, course	60	720x480	448x216
Antilope, marche	122	352x288	232x173
Girafe	73	352x288	195x253

TAB. 6.1 – Nombre d'images et taille des images pour quelques séquences de test.

### 3.2 Description du mouvement par les composantes principales

L'Analyse en Composantes Principales (ACP) est une méthode standard de réduction de dimension, que nous avons déjà utilisée au chapitre 5. Elle consiste à trouver un ensemble d'axes principaux, qui génèrent linéairement la variance de données multidimensionnelles. Une des premières utilisations de l'ACP sur les images a été introduite par Turk et Pentland dans le cadre de la reconnaissance de visages (eigen-faces) [TP91]. Dans ce cas, chaque image est considérée comme une observation indépendante dans laquelle les couleurs de tous les pixels sont empilées dans un seul vecteur. Les images propres (*Eigen-images*) ont été beaucoup utilisées pour classifier et reconnaître des formes sur les images. Une de nos contributions est de montrer que l'ACP sur les images peut encoder des variations dues au mouvement et peut être utilisée non seulement pour classifier des formes, mais aussi pour prédire des variations continues dans un mouvement. Nous tirerons avantage de cette propriété dans notre schéma d'interpolation.

L'ACP consiste à calculer les vecteurs propres et les valeurs propres de la matrice de covariance des données d'entrée. Dans notre cas, chaque image de la séquence est vue comme un vecteur ligne  $\mathbf{i}(t)$ , contenant les valeurs de tous les pixels alignées. Nous rassemblons les  $n$  images d'une séquence dans une matrice  $\mathbf{I}$ , après avoir soustrait l'image moyenne  $\bar{\mathbf{i}}$  :

$$\bar{\mathbf{i}} = \frac{1}{n} \sum_{t=1}^n \mathbf{i}(t) \quad (6.1)$$

$$\mathbf{I} = \left[ (\mathbf{i}(t_1) - \bar{\mathbf{i}})^t, \dots, (\mathbf{i}(t_n) - \bar{\mathbf{i}})^t \right] \quad (6.2)$$

La formulation de l'ACP est alors :

$$\frac{1}{n} \mathbf{I}^T \mathbf{I} \mathbf{E} = \mathbf{E} \mathbf{D} \quad (6.3)$$

Séquence	PC1	PC2	PC3	PC4
Cheval, marche	33.7	23.7	11.4	8.56
Cheval, canter	32.5	14.5	9.17	8.78
Cheval, galop	31.1	19.9	11.0	8.33
Cheval, course	44.7	11.6	9.93	7.79
Tigre, course	15.2	10.5	6.14	4.69
Antilope, marche	21.5	12.2	8.40	6.91
Girafe, marche	42.8	15.8	11.1	5.63

TAB. 6.2 – pourcentage de la variance couverte par chaque composante principale par rapport à la variance totale, pour sept séquences de test.

$$\mathbf{E}^T \mathbf{E} = \mathbf{I} \quad (6.4)$$

Finalement, nous prenons comme vecteurs descriptifs de l’animation les projections sur les composantes principales placées dans les colonnes de  $\mathbf{E}$  et normalisées par les racines carrées des valeurs propres, contenues dans la matrice diagonale  $\mathbf{D}$  :

$$\mathbf{p}(t) = (\mathbf{i}(t) - \bar{\mathbf{i}}) \mathbf{E} \sqrt{\mathbf{D}^{-1}} \quad (6.5)$$

Le Tableau 3.2 récapitule les résultats de l’ACP, c’est-à-dire la part de la variance couverte par chaque composante principale par rapport à la variance totale, pour sept séquences de test.

### 3.3 Interpolation

Notre but est de générer des paramètres d’animation (positions et valeurs d’angles)  $\mathbf{x}(t)$  à partir des coefficients de projection  $\mathbf{p}(t)$  calculés à partir de l’ACP. Pour cela nous interpolons  $m$  exemples de pose 3D  $[\mathbf{x}(t_i)]_{i=1\dots m}$ , correspondant à  $m$  images de la séquence vidéo pour lesquelles nous connaissons les coefficients de projection  $[\mathbf{p}(t_i)]_{i=1\dots m}$  au temps  $t_i$  dans la séquence vidéo. Pour plus de clarté, nous notons  $\mathbf{x}_i$  et  $\mathbf{p}_i$  pour  $\mathbf{x}(t_i)$  et  $\mathbf{p}(t_i)$  respectivement.

Trois méthodes principales pour l’interpolation de données clairsemées sont utilisées dans les méthodes à base d’exemples : l’interpolation linéaire [BLCD02], les fonctions de base radiales (FBR) [LCF00], ou une combinaison des deux [SIC01]. Dans le dernier cas, l’interpolation linéaire permet de garantir une interpolation stable dans le cas où les données d’entrées sont clairsemées. Dans notre cas, les données d’entrée sont les résultats de l’ACP et sont donc très compactes. Les Fonctions de Base Radiales (FBR) sont donc suffisantes dans notre cas. Ce schéma général d’interpolation est formulé comme une combinaison linéaire de fonctions de distance  $h(r)$  (les FBR) depuis  $m$  points d’interpolation dans l’espace d’entrée :

$$\mathbf{x}(\mathbf{p}) = \sum_{k=1}^m h(\|\mathbf{p} - \mathbf{p}_k\|) \mathbf{a}_k \quad (6.6)$$

Où  $\mathbf{p}$  est le vecteur d'entrée et  $\mathbf{x}$  le vecteur prédit.  $h(r)$  est la FBR. Les  $\mathbf{a}_k$  sont des vecteurs inconnus à déterminer. Si les valeurs de FBR  $h(\|\mathbf{p} - \mathbf{p}_k\|)$  sont empilées dans un seul vecteur  $\mathbf{h}(\mathbf{p})$  et les vecteurs inconnus  $\mathbf{a}_k$  dans les lignes d'une matrice  $\mathbf{A}$ , nous avons la formulation :

$$\mathbf{x}(\mathbf{p}) = \mathbf{h}(\mathbf{p})\mathbf{A} \quad (6.7)$$

$$\mathbf{h}(\mathbf{p}) = [h(\|\mathbf{p} - \mathbf{p}_1\|) \dots h(\|\mathbf{p} - \mathbf{p}_m\|)] \quad (6.8)$$

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_m \end{pmatrix} \quad (6.9)$$

Aux points d'interpolation, nous utilisons  $m$  exemples de pose 3D  $\mathbf{x}_i$  et les valeurs des  $m$  paramètres d'entrée associés  $\mathbf{p}_i$  de l'image-clé correspondante. Il faut trouver  $\mathbf{A}$  tel que  $\|\mathbf{x}(\mathbf{p}_i) - \mathbf{x}_i\|$  est minimal. Le minimum au sens des moindres carrés est obtenu de façon classique grâce à la pseudo-inverse :

$$\mathbf{A} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{X} \quad (6.10)$$

Où,

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{pmatrix} \quad (6.11)$$

$$\mathbf{H} = \begin{pmatrix} \mathbf{h}(\mathbf{p}_1) \\ \vdots \\ \mathbf{h}(\mathbf{p}_m) \end{pmatrix} \quad (6.12)$$

La formulation finale est donc :

$$\mathbf{x}(\mathbf{p}) = \mathbf{h}(\mathbf{p}) (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{X} \quad (6.13)$$

Notez qu'on peut reformuler cette équation exactement comme une interpolation des  $\mathbf{x}_i$  :

$$\mathbf{x}(\mathbf{p}) = \sum_{i=1}^m w_i(\mathbf{p}) \mathbf{x}_i \quad (6.14)$$

en extrayant les  $w_i(\mathbf{p})$  de la matrice  $\mathbf{h}(\mathbf{p}) (\mathbf{H}^t \mathbf{H})^{-1} \mathbf{H}^t$

Dans [AM00], Alexa et al. compressent du mouvement 3D à l'aide de composantes principales (CP) apprises sur des séquences connues de mouvement 3D. Dans notre cas, les CP sont apprises à partir d'images et l'animation de données 3D est contrôlée par interpolation.

La valeur de  $\sum_{i=1}^m w_i(\mathbf{p})$  doit rester proche de 1 pour garantir qu'un point  $\mathbf{p}$  dans l'espace d'entrée est assez proche des points d'interpolation, et les  $w_i(\mathbf{p})$  doivent être proches de  $[0, 1]$  pour que  $\mathbf{x}(\mathbf{p})$  reste proche de l'enveloppe convexe des exemples de pose 3D.

Pour  $h(r)$ , on choisit généralement une gaussienne pour ses propriétés de continuité  $C^\infty$  :

$$h(r) = e^{-\alpha r^2} \quad (6.15)$$

Le paramètre  $\alpha$  dans l'Equation 6.15 reste à déterminer. Nous l'évaluons comme suit. Statistiquement, les projections sur les CP sont homogènes à un écart-type. Cela signifie que les données seront réparties approximativement dans chaque direction dans l'intervalle  $[-1; +1]$ , avec des variations selon la nature de la distribution. A supposer que les points d'interpolation sont bien répartis, nous prenons une valeur de 2 comme estimation grossière de la distance entre les points d'interpolation. A égale distance de deux points d'interpolation, nous voulons que leurs influences soient égales, c'est-à-dire que nous voulons  $h(r) = 0.5$  quand  $r = 1$ . Ceci nous donne une estimation de  $\alpha = \ln 2$ .

Dans tous les travaux antérieurs sur l'animation à base d'exemple, c'est l'utilisateur qui doit décider où les exemples de pose 3D doivent être placés [LCF00, SIC01, PKC<sup>+</sup>03]. Dans notre cas, cela signifierait sélectionner des images-clés parmi les milliers d'images d'une séquence vidéo. Nous proposons donc un critère automatique pour sélectionner les images-clés, connaissant le nombre d'images-clés souhaité.

## 4 Sélection des images-clés

### 4.1 Critère automatique de sélection

Comme notre contrôle de l'animation est basé sur les images, nous voulons avoir une fonction de prédiction lisse de l'espace image vers l'espace des positions 3D : Une petite variation dans l'espace image doit induire une petite variation dans l'espace des poses 3D. Notons que le schéma d'interpolation par FBR nécessite l'inversion de la matrice  $\mathbf{H}^T \mathbf{H}$  construite à partir des points d'interpolation, comme cela a été montré dans la section précédente. En conséquence, pour assurer une interpolation stable et donc une animation fluide, nous sélectionnons parmi la séquence les images-clés qui donnent le meilleur conditionnement de la matrice  $\mathbf{H}^T \mathbf{H}$  à inverser. Le conditionnement est le ratio de la plus grande valeur singulière de la matrice à la plus

petite. Un conditionnement élevé indique une matrice presque singulière.

Ce critère est applicable de manière générale à n'importe quelle méthode à base d'exemples. Il peut être utilisé pour sélectionner un certain nombre d'exemples, des images-clés dans notre cas, parmi un grand ensemble de données. Les valeurs singulières de  $\mathbf{H}^T\mathbf{H}$  sont les carrés des valeurs singulières de  $\mathbf{H}$ . La matrice  $\mathbf{H}$  mesure les distances respectives entre les points d'interpolation. Intuitivement, le critère du conditionnement sélectionne les exemples d'entrée qui sont le mieux répartis parmi les données. Choisir des valeurs singulières proches les unes des autres signifie qu'ils sont répartis régulièrement dans toutes les dimensions de l'espace.

Dans la pratique, comme nous le verrons dans les sections 5 et 6, seuls quelques composantes principales et quelques exemples de poses 3D sont nécessaires. Ceci nous permet d'implémenter une approche combinatoire simple pour le critère du conditionnement : pour chaque séquence de  $n$  images, étant donné un nombre  $c$  de composantes principales à considérer et un nombre  $m$  d'images-clés à choisir, nous calculons le conditionnement des  $\binom{n}{m}$  matrices  $\mathbf{H}^T\mathbf{H}$  (La matrice  $\mathbf{H}^T\mathbf{H}$  est carrée de dimension  $m$ ), et nous gardons les  $m$  images-clés parmi toute la séquence qui donne le plus petit conditionnement de  $\mathbf{H}^T\mathbf{H}$ . Les calculs sont rapides en gardant seulement quelques composantes principales. nous avons testé jusqu'à 5 CP, mais l'expérience a montré que 2 CP sont suffisantes comme nous allons le montrer dans les sections suivantes. A titre d'exemple, pour la prédiction de 3 séquences d'animation à partir d'images synthétiques, nous affichons les projections sur les deux premières composantes principales comme un graphe 2D et nous cherchons les quatre meilleurs exemples avec le critère du conditionnement (Dans la prochaine section, nous montrerons que 2 CP et 4 images-clés est la meilleure configuration pour la prédiction de cette démarche spécifique). Dans ce cas le critère du conditionnement a la particularité de sélectionner des exemples approximativement aux positions extrémales des projections sur les deux premières CP. (Figure 6.3).

Intuitivement, plus le nombre d'images-clés utilisées sera important, meilleure sera l'interpolation. Comme l'utilisateur devra fournir un exemple de pose 3D pour chaque image-clé, il faut trouver un compromis entre précision et quantité de travail à fournir. La question du nombre d'images-clés nécessaires doit être examinée au cas-par-cas. D'après nos expériences sur différentes allures, nous avons observé de bons résultats avec 4 exemples de pose pour la course et 8 exemples de pose pour la marche.

## 4.2 Résolution des ambiguïtés 2D avec plusieurs modèles

Telle quelle, notre méthode prédit du mouvement 3D à partir d'images. Pour chaque silhouette, nous prédisons une pose 3D de façon non ambiguë. Cependant, dans certains cas, deux poses 3D différentes vues de côté peuvent donner des silhouettes très similaires (Figure 6.4). C'est très courant dans les mouvements qui consistent en une succession de deux

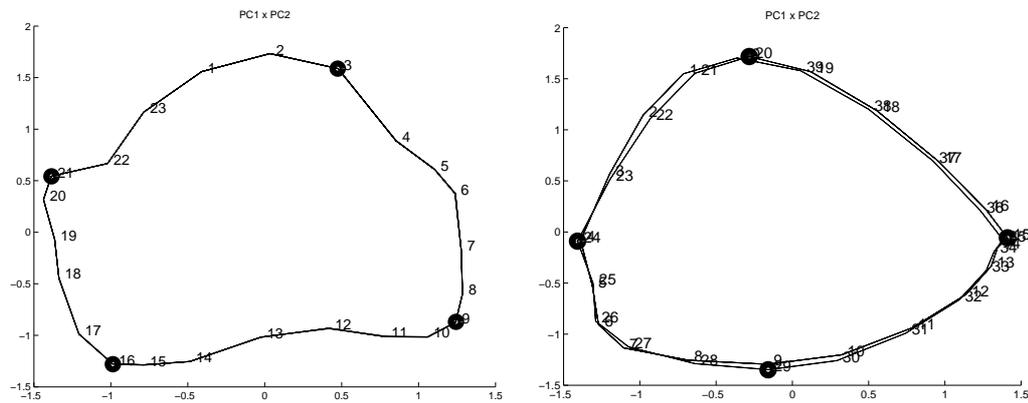


FIG. 6.3 –  $PC1(t) \times PC2(t)$  : Les projections sur les CP au cours du temps pour deux séquences synthétiques du cheval : le canter et la marche. Les images sont numérotées pour un cycle. Les cercles sont les images-clés choisies par le critère du conditionnement.

phases symétriques entre elles, comme la marche des quadrupèdes. Le mouvement prédit par FBR donne toujours de bons résultats mais seulement pour la moitié de la période du mouvement original.

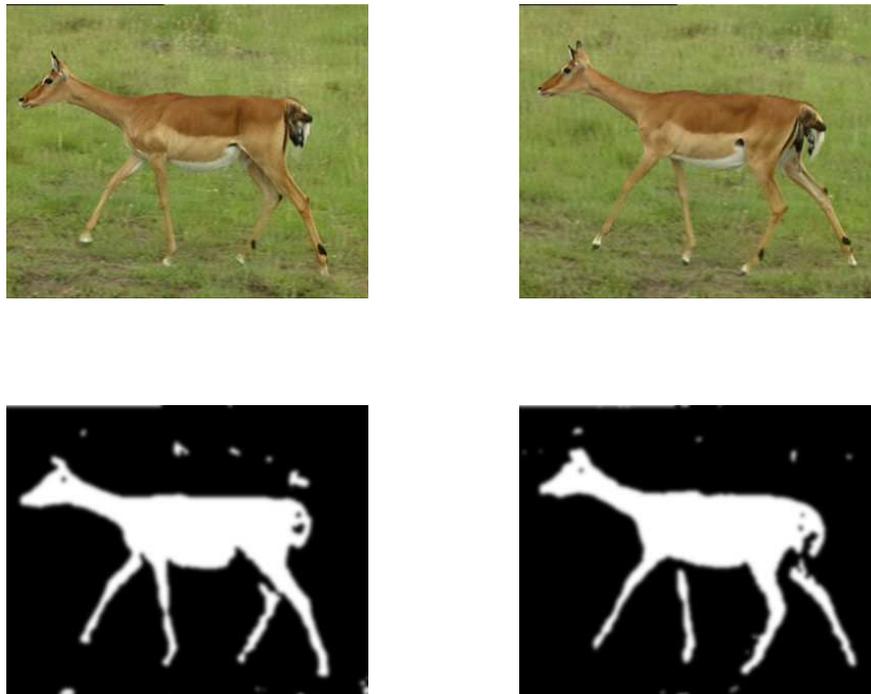


FIG. 6.4 – Deux poses différentes peuvent donner des silhouettes similaires.

Pour éviter ce problème, il est d'abord nécessaire de fournir deux exemples de pose 3D pour

chacune des silhouettes ambiguës, et ensuite de donner une méthode pour choisir correctement entre ces deux poses pendant la génération de l'animation 3D.

Nous résolvons le premier problème à l'aide d'un algorithme simple :

1. Nous sélectionnons  $m$  images-clés initiales à l'aide de la méthode standard et nous construisons l'animation 3D en associant des poses 3D aux images-clés et en appliquant la prédiction par FBR. Si l'utilisateur remarque des problèmes liés aux ambiguïtés de pose, nous allons à l'étape 2.
2. Pour chaque image-clé, nous cherchons automatiquement l'image la plus proche dans l'espace des CP et nous la proposons à l'utilisateur comme une pose 3D alternative pour cette silhouette. Nous contraignons cette image à être à au moins 3 pas de temps de l'image originale pour garantir qu'elle est dans un autre demi-cycle.
3. Quand l'utilisateur valide l'image proposée comme étant l'image-clé correspondant à la même silhouette mais avec une pose différente, nous demandons à l'artiste de donner la pose 3D correspondante.
4. Nous itérons jusqu'à ce que les  $m$  images-clés initiales aient leur image-clé associée correspondant à la pose 3D symétrique.

A la fin de ce processus, nous avons doublé le nombre d'images-clés et d'exemples de pose 3D correspondants. La Figure 6.5 détaille un exemple de cet algorithme pour  $m = 4$  images-clés. Nous sommes à présent capables de générer un cycle complet de mouvement 3D. Pour générer de l'animation, la même méthode de prédiction à partir d'images est utilisée, mais au lieu de garder les mêmes  $m$  exemples de pose 3D, nous changeons parmi  $q$  ensembles de  $m$  exemples de pose 3D au cours de temps, choisis parmi les  $2m$  exemples de pose 3D sélectionnés par l'algorithme précédent. Nous permutons donc entre  $q$  modèles de prédiction au cours du temps. La prédiction des paramètres d'animation est étendue de la façon suivante :

$$\mathbf{x}(\mathbf{p}_t) = \sum_{k=1}^q w_{\sigma_k(s_t)}(\mathbf{p}_t) \mathbf{x}_{\sigma_k(s_t)} \quad (6.16)$$

$$s_t = \text{switch}(\mathbf{p}_t, s_{t-1}) \quad (6.17)$$

Où  $s_t$  est un entier qui représente l'état courant, ou le numéro du modèle à utiliser, les  $\mathbf{x}_i$  sont les  $2m$  exemples de pose 3D, et les  $w_i$  les poids attribués à chaque modèle étant donnés l'image d'entrée et l'état courant. La fonction  $\text{switch}(\mathbf{p}, s)$  indique quel ensemble de  $m$  exemples de pose 3D doit être utilisé dans l'algorithme de prédiction. C'est une variable d'état discrète, incrémentée à chaque fois que nous avons atteint la dernière silhouette 2D dans un ensemble de  $m$  images-clés. Le changement de silhouette dans les images-clés est facilement détecté grâce à une fonction de distance dans l'espace des CP.

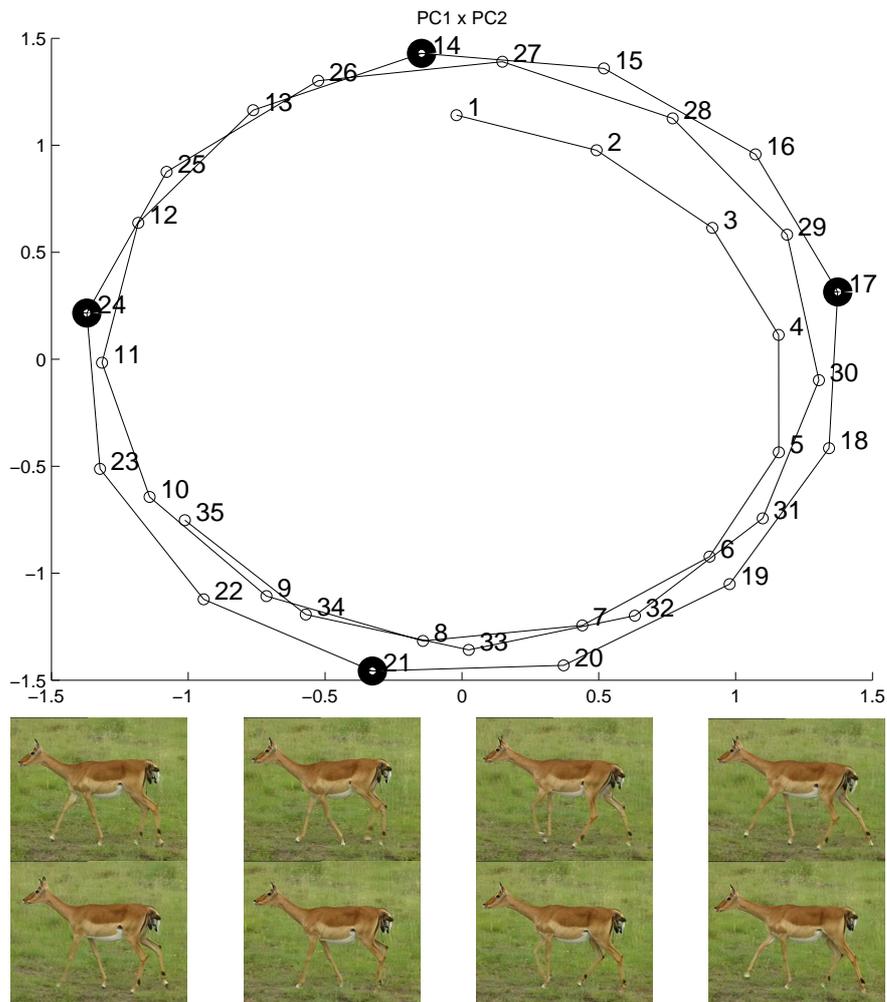


FIG. 6.5 – Sélection d’images-clés supplémentaires pour résoudre les ambiguïtés dans l’espace d’ACP. Les images 14, 17, 21 and 24 sont sélectionnées comme images-clés initiales. Dans une deuxième étape, les images 27, 4, 8 et 11 sont sélectionnées comme candidats pour les poses ambiguës, à partir de leurs coordonnées dans l’espace d’ACP. Première ligne : images 14, 17, 21 24 ; Deuxième ligne : images 27, 4, 8, 11.

Comme nous avons doublé le nombre initial d’exemples de pose  $m$ , commuter entre  $q = 2$  modèles de  $m$  exemples de pose 3D devrait nous permettre d’explorer tout l’espace de poses 3D souhaité. Cependant, la transition entre deux modèles seulement s’est avérée être instable. Nous résolvons ce problème en introduisant des recouvrement entre modèles. L’utilisation de  $q = 4$  modèles permet des transitions lisses. Dans la pratique nous utilisons  $m = 4$  exemples de pose 3D pour décrire la moitié d’un cycle d’animation. La fonction  $\sigma_k(s)$  donne les indices de 4 exemples de pose avec 4 états, avec 2 poses en commun entre deux états consécutifs :

$$\sigma_i(j) = \sigma_{ij} \quad (6.18)$$

$$(\sigma_{ij}) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 5 & 6 \\ 5 & 6 & 7 & 8 \\ 7 & 8 & 1 & 2 \end{pmatrix} \quad (6.19)$$

Comme chaque modèle a deux poses en commun avec le modèle suivant, le changement de modèle peut se faire de manière lisse lorsque le mouvement est entre ces deux poses.

## 5 Validation sur des images de synthèse

Nous avons validé notre méthode en prenant comme images d'entrée le rendu d'un modèle 3D simple de cheval. Dans le modèle qui nous avons choisi, les joints sont représentés comme des ellipsoïdes (Figure 6.2), afin d'éliminer toutes les perturbations qu'un algorithme de skinning pourrait introduire. Nous donnons des résultats obtenus pour 3 séquences synthétiques de cheval : le galop, le canter et la marche. En utilisant des images de synthèse, nous pouvons tester le pipeline complet présenté dans la section 3. De plus, nous pouvons comparer numériquement l'animation 3D obtenue avec l'animation originale. Cette évaluation nous donne des informations sur le nombre de composantes principales et d'exemples à utiliser.

Nous avons évalué le résultat en testant plusieurs valeurs pour le nombre d'images-clés (à partir de deux) et de composantes principales (à partir de une). Étant donné le nombre d'images-clés à sélectionner à partir de la vidéo, le critère du conditionnement nous dit quelles images-clés sélectionner. Les exemples de pose 3D correspondant sont données par l'animation 3D originale. Nous évaluons les résultats en calculant la moyenne (et l'écart-type) de la différence pour tous les joints et les axes de rotation principaux (perpendiculaires à l'image, 36 valeurs dans notre cas), entre les valeurs angulaires originales et les valeurs prédites.

Sur la Figure 6.6, nous observons immédiatement qu'ajouter la troisième et la quatrième composante introduit du bruit. Ceci suggère qu'elles codent des informations non pertinentes pour le mouvement. Quant au nombre de clés, comme on pouvait s'y attendre, plus on fournit d'exemple, meilleur est le résultat. Cependant choisir 4 clés est un bon compromis. Ajouter une cinquième clé diminue l'erreur angulaire moyenne de moins d'un degré. Deux ou trois exemple, même si ils sont sélectionnés de manière optimale par le critère du conditionnement, ne sont pas suffisants. Avec quatre exemples de pose 3D et deux composantes principales, nous obtenons une très bonne correspondance entre l'animation 3D originale et l'animation prédite.

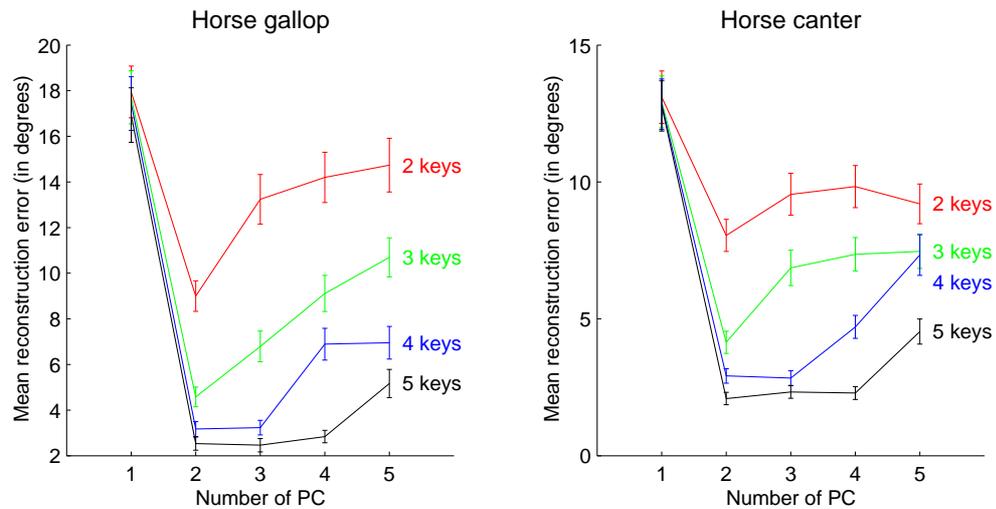


FIG. 6.6 – Évaluation pour les séquences du galop et du canter. Chaque courbe correspond à l'erreur moyenne pour un nombre fixe d'exemples (entre 2 et 5), en fonction du nombre de composantes utilisées en entrée.

## 6 Application à des séquences vidéos réelles

Nous détaillons maintenant comment appliquer notre méthode à des séquences vidéos réelles, parfois annotées avec un dessin rapide de l'animal comme nous l'avons expliqué dans la section 3. Comme nous allons le montrer, en gardant seulement les deux premières composantes principales et en appliquant un filtre passe-bande sur les trajectoires dans l'espace des CP nous pouvons obtenir des résultats aussi bons qu'avec des données synthétiques.

### 6.1 Restriction aux deux premières CP

Lors de nos tests sur des séquences vidéo de synthèse, les deux premières composantes pouvaient être facilement interprétées en terme de mouvement. Par exemple, pour le galop du cheval, la première composante (PC1) code une variation entre une phase de vol, ou aucun pied ne touche le sol, et une phase où le corps de l'animal est groupé. La seconde composante (PC2) correspond à une opposition entre une phase de décollage, quand le cheval prend son appui au sol et saute, et une phase de réception, quand les pattes avant atterrissent (Figure 6.7).

L'évaluation numérique sur les images de synthèse a suggéré que les deux premières CP sont optimales pour obtenir une bonne prédiction. La segmentation d'image et le dessin à la main introduiront naturellement plus de bruit dans les courbes de CP, rendant les CP instables et peu fiables pour prédire du mouvement. Nous décidons donc que seules les deux premières composantes doivent être gardées pour les vidéos réelles et les images esquissées.

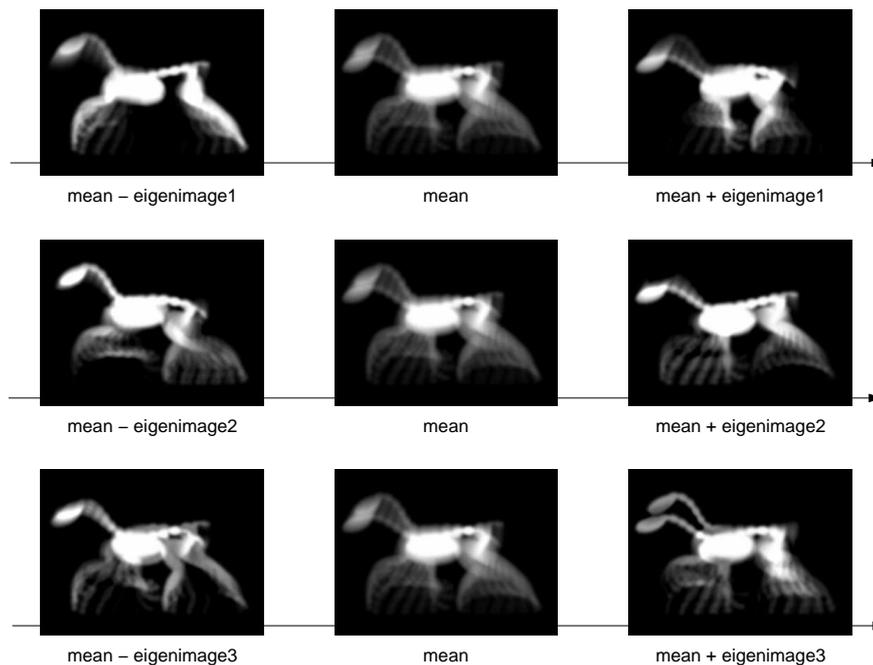


FIG. 6.7 – Les variations capturées par les trois premiers vecteurs propre pour la séquence du cheval au galop. La colonne du milieu est l'image moyenne, chaque ligne correspond à la variation selon un vecteur propre.

Nous confirmons cette hypothèse sur la séquence du guépard où les deux premières CP peuvent être interprétées de manière similaire au cas du cheval de synthèse (Figure 6.8).

## 6.2 Régularisation du spectre

Sur les séquences de synthèse, les trajectoires au cours du temps des deux premières composantes sont déphasées d'un quart de cycle. Ceci produit le motif circulaire visible sur la figure 6.3. Cette configuration s'est reproduite sur tous nos exemples de vidéos de synthèse. En conséquence, nous considérons la configuration des projections sur PC1 et PC2 en motif circulaire comme une caractérisation des séquences vidéo utilisables avec notre méthode. Dans le domaine de Fourier, cette configuration correspond à des pics de fréquence au même endroit pour les projections sur PC1 et PC2, et une différence de phase d'environ  $\frac{\pi}{2}$ .

Les vidéos réelles peuvent donc être considérées comme non utilisables par notre méthode si leurs projections sur PC1 et PC2 ne sont pas contenues dans une certaine bande de fréquence que nous estimons automatiquement. La première composante encode la plupart de la variance et est donc considérée comme la plus représentative du mouvement cyclique général. Son spectre est donc centré autour d'une fréquence correspondant à la période de la variation

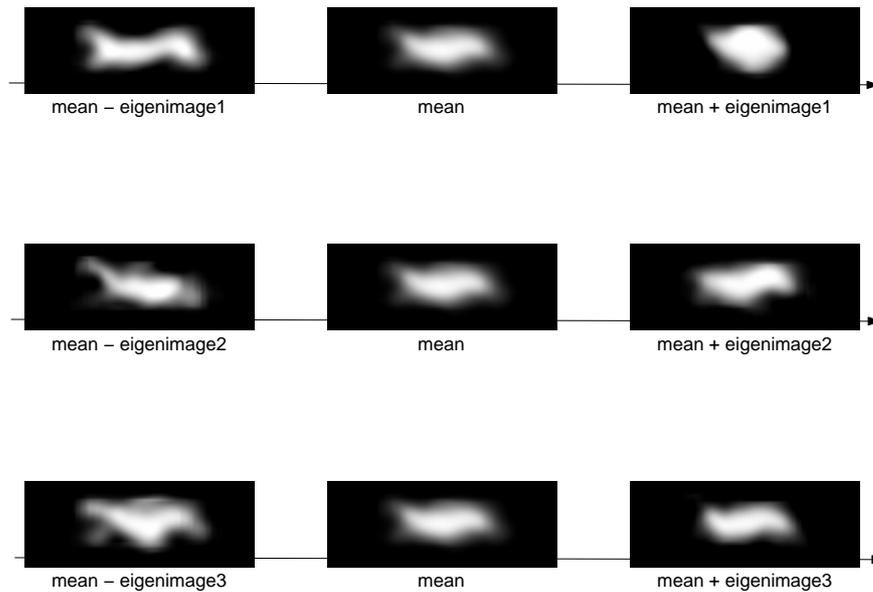


FIG. 6.8 – Les variations encodées par les trois premiers vecteurs propres pour la séquence du guépard.

des images. Tous nos tests confirment cette hypothèse. A partir d'une transformée de Fourier nous obtenons la fréquence de plus grande amplitude. Nous ajustons ensuite une fonction de pic centrée sur cette fréquence de la forme :

$$peak(f) = \frac{1}{1 + \left(\frac{f-f_0}{f_b-f_0}\right)^2} \quad (6.20)$$

Nous choisissons  $f_0$  comme la fréquence de plus grande amplitude et  $f_b$  comme la fréquence la plus proche de  $f_0$  qui a une amplitude de la moitié du maximum. Nous déduisons une bande de fréquence de  $[-3(f_b - f_0); +3(f_b - f_0)]$  (Figure 6.9).

La seconde composante est filtrée à l'aide du même filtre passe-bande. Pour obtenir le motif circulaire, il faudrait que la seconde composante ait un pic similaire. Nous pouvons évaluer à quel point cette hypothèse est vérifiée en regardant si le signal reconstruit après filtrage est proche du signal original. Pour ce faire, nous calculons le coefficient de corrélation entre le signal de la composante principale originale et le signal filtré. Nous avons observé que pour nos séquences de test, les deux premières composantes ont le même pic en commun ( $r \geq .6$ ), alors que les composantes suivantes ne l'ont pas ( $r \leq .3$ ) (table 6.3). Ceci fournit un critère numérique pour évaluer si notre méthode peut être appliquée à une séquence vidéo ou non.

La figure 6.9 montre les résultats dans le domaine de Fourier pour la séquence du guépard (qui est conforme au critère). Quand une séquence vidéo n'est pas conforme à ce critère, nous suggérons d'utiliser les croquis. Si l'approche par croquis n'est pas conforme non plus au

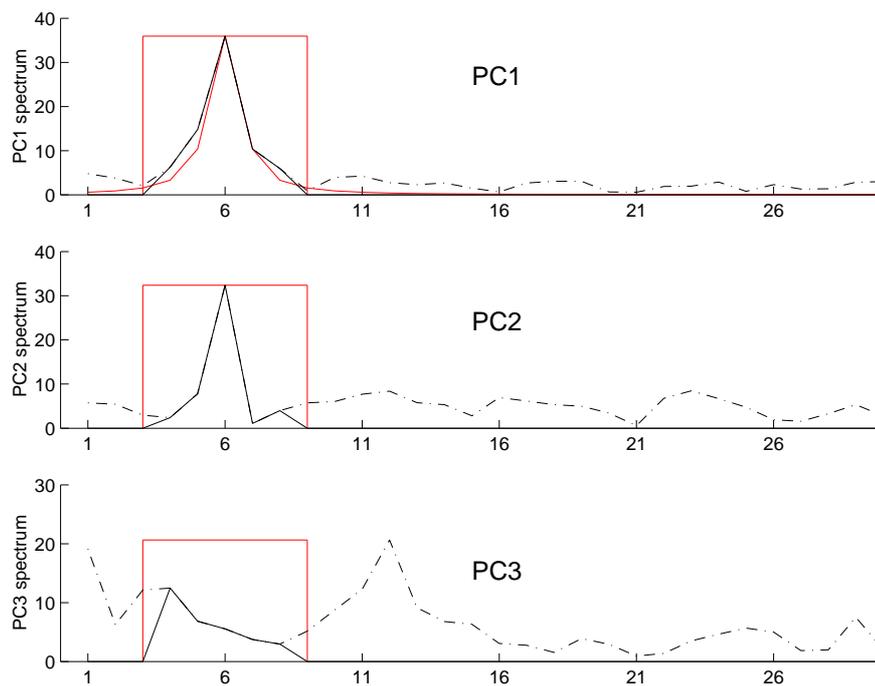


FIG. 6.9 – Le spectre des 2 premières CP du spectre du guépard. La fonction de pic est ajustée sur PC1, et une fenêtre rectangulaire est déduite.

PC :	PC1	PC2	PC3	PC4	PC5
Cheval, gallop	.90	.92	.01	<.01	<.01
Cheval, canter	.94	.89	.08	<.01	.01
Cheval, marche	.99	.94	<.01	.05	<.01
Guépard, course	.91	.61	.14	.21	.25
Antilope, marche	.78	.91	.14	.10	.06
Tigre, course	.87	.72	.26	.18	.23
Girafe, marche	.92	.82	.24	.19	.28

TAB. 6.3 – Coefficients de corrélation entre les signaux filtrés et non filtrés pour les cinq premières composantes, calculés pour quelques séquences de test.

critère, non considérons que notre méthode n'est pas applicable à la vidéo considérée.

### 6.3 Factorisation des mouvements non cycliques

Le critère défini à la section précédente considère une séquence comme utilisable par notre méthode si ses deux premières composantes principales ont les même pic dans le domaine des

fréquences. Cependant, dans certains cas, la première composante n'est pas liée à la locomotion. Par exemple, une de nos séquences de test représente un rhinocéros qui marche. Pendant la marche, le rhinocéros baisse lentement la tête tout au long de la séquence, pour finir par prendre de la nourriture au sol. La figure 6.10 montre les variations encodées par les trois premiers vecteurs propres de cette séquence. La première composante code une opposition entre monter et baisser la tête. Cette information est répartie tout au long de la séquence et n'est pas liée à la locomotion. Elle ne présente pas une évolution cyclique (Figure 6.11), et ne peut donc pas être utilisée par notre méthode. La seconde et la troisième composante, cependant, ont un pic de fréquence en commun et leur graphe a bien la forme circulaire que nous recherchons. Elles encodent une alternance entre grouper et écarter les pattes arrières (pour PC1) et les pattes avant (pour PC2). Notre méthode a pu être appliquée avec succès en utilisant ces composantes. En sélectionnant manuellement les composantes à utiliser, nous avons donc pu éliminer les variations d'images qui ne correspondent pas à la locomotion.

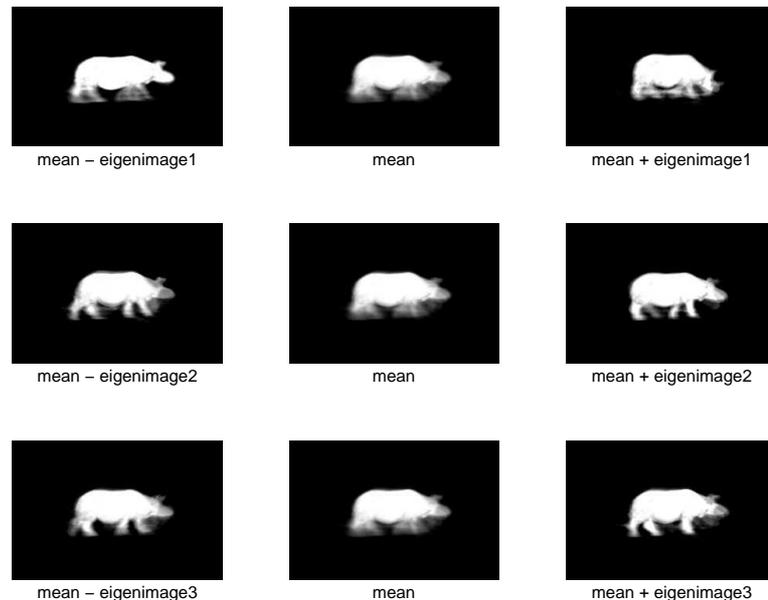


FIG. 6.10 – Les variations encodées par les trois premiers vecteurs propre de la séquence de marche du rhinocéros.

## 7 Résultats

La figure 6.12 montre l'évolution des poids des exemples de pose 3D pour la séquence du guépard et la séquence du rhinocéros. Aux exemples, nous avons une interpolation exacte. Pour le reste de la séquence, nous observons une généralisation correcte, l'influence de chaque

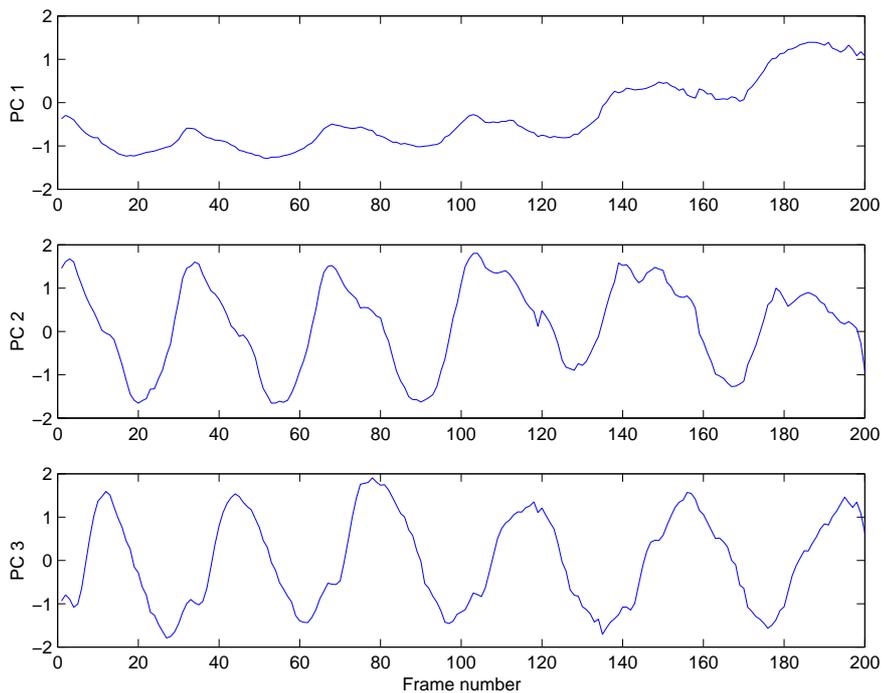


FIG. 6.11 – Évolution au cours du temps des trois premières composantes principales de la séquence de marche du rhinocéros.

exemple apparaît au bon rythme dans un ordre cohérent. Notez que la somme des poids reste toujours proche de 1, ce qui garantit que les paramètres d'entrée sont toujours proche d'un point d'interpolation. Les poids sont parfois en dehors de l'intervalle  $[0, 1]$  car ils ne sont pas contraints dans la formulation par FBR. Ceci laisse la pose prédite quitter l'enveloppe convexe des exemples de pose 3D. Cette flexibilité permet une extrapolation dans l'espace 3D induite par les variations des images au long de la séquence. Un contrôle pourrait facilement être ajouté pour maintenir ces poids dans un certain domaine pour empêcher la génération de poses étranges, trop loin de l'enveloppe convexe des exemples. Jusqu'ici cependant, aucune des séquences testées n'a nécessité un tel contrôle. Dans les sous-sections suivantes, nous effectuons plusieurs tests comparatifs pour évaluer plus finement notre méthode.

### 7.1 Sélection des images-clés : comparaison entre différents animaux

Nous avons appliqué notre méthode à plusieurs espèces de quadrupèdes, allant d'animaux légers comme le chat à des animaux lourds comme l'éléphant. Les animaux lourds ont des démarches plus lentes. Ceci peut être vu sur la figure 6.12, où le nombre d'images entre deux pics successifs dans les poids d'interpolation sont plus grands pour le rhinoceros que pour

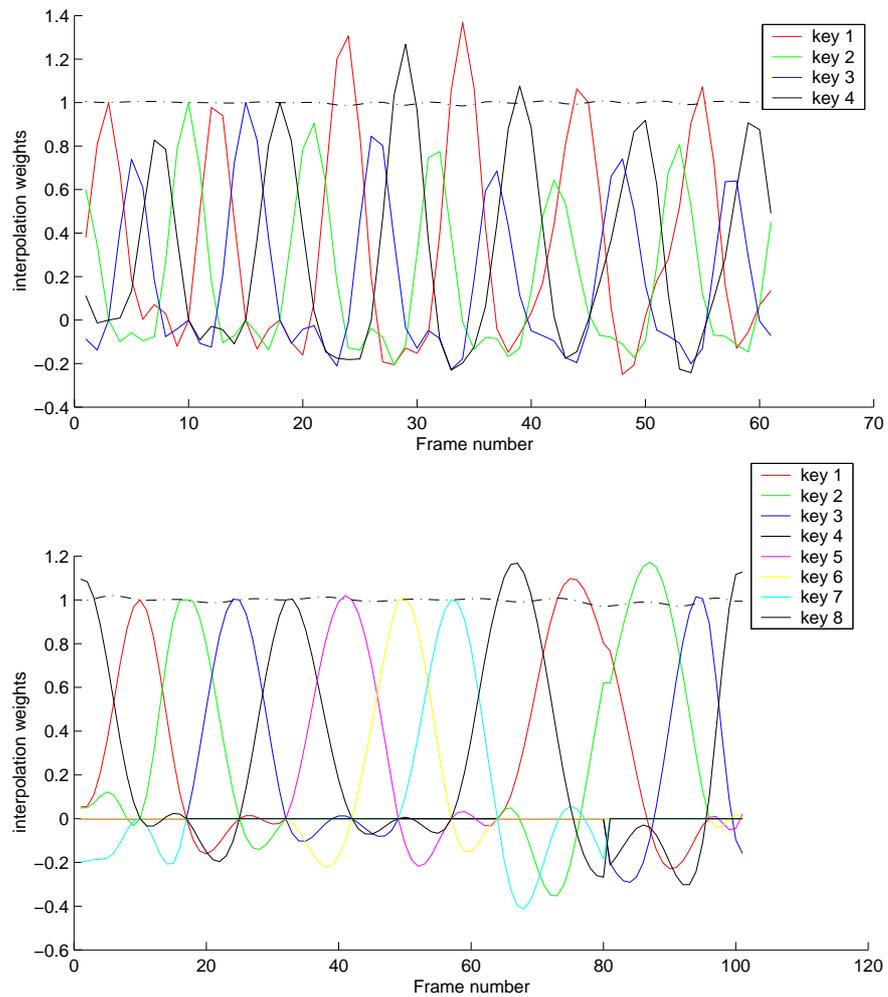


FIG. 6.12 – Evolution des poids des 4 exemples pour la séquence du guépard(haut), et des 8 exemples pour la séquence du rhinocéros (en bas). La ligne en pointillés est la somme des poids.

le guépard. Cependant, les images-clés sélectionnées restent très similaires entre les animaux légers et les animaux lourds.

La figure 6.13 rassemble nos résultats finaux pour la sélection d'images-clés. Elle montre les images-clés détectées automatiquement et les poses 3D associées fournies par l'artiste.

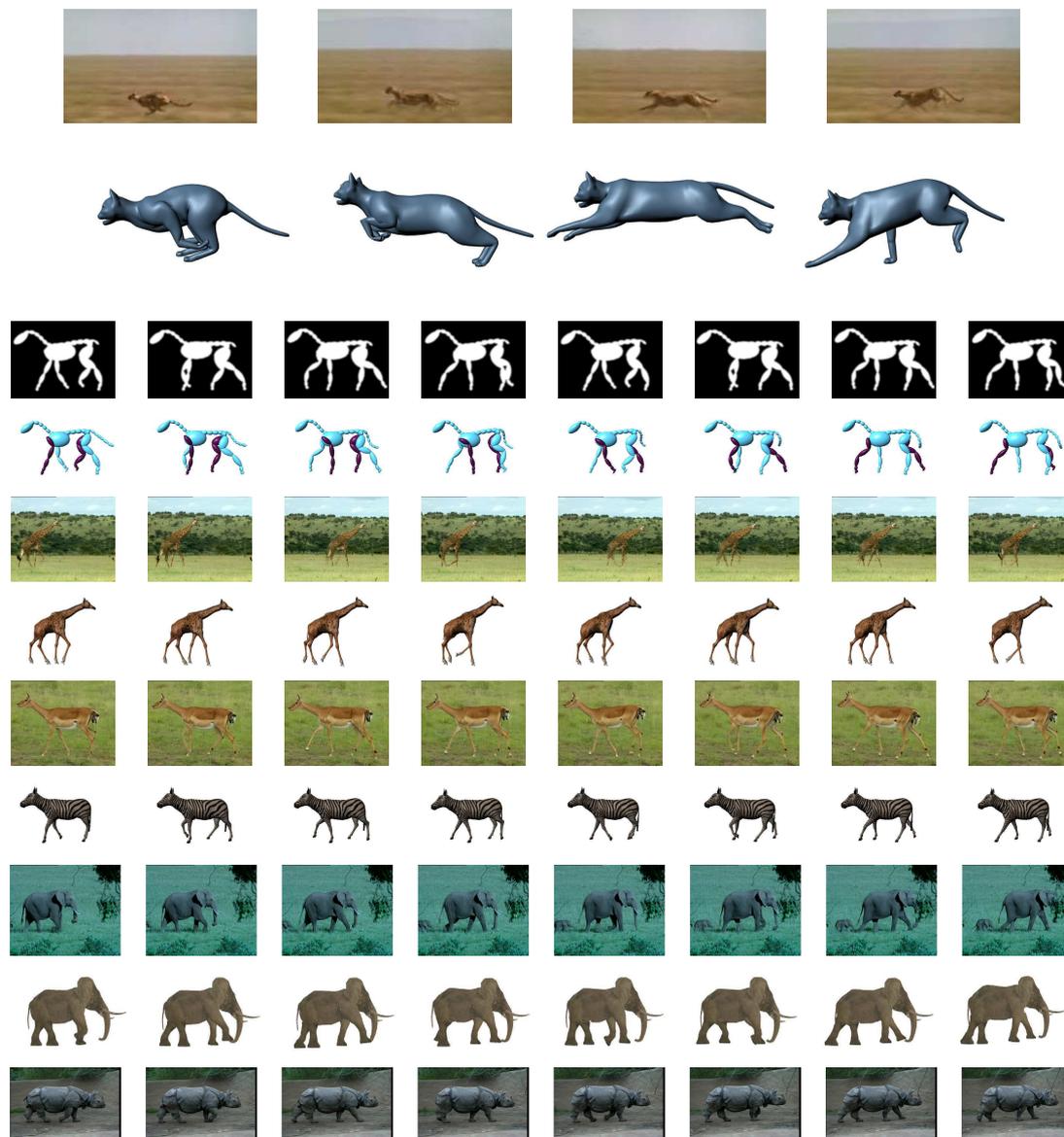


FIG. 6.13 – Sélection d'images-clés pour plusieurs séquences vidéo.

## 7.2 Sélection des images-clés : Comparaison entre images réelles et images de synthèse

Nous avons testé notre méthode sur deux séquences de marche de cheval : une vidéo réelle et une vidéo de synthèse, et nous avons comparé les résultats. Les images-clés sélectionnées dans les deux cas sont très similaires (Figure 6.14). Notez que même si l'apparence du cheval synthétique n'est pas réaliste, son mouvement est bien capturé. C'est du au fait que notre

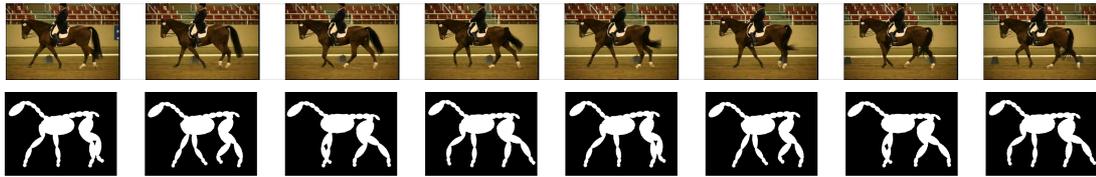


FIG. 6.14 – Les images-clés sélectionnées pour la marche du cheval, d’après une séquence réelle (haut), et une séquence synthétique (bas).

Esquisse	6	12	18	24	30	36	42	48
Segmentation	8	13	18	24	32	37	42	48

TAB. 6.4 – Indices des images-clés sélectionnées pour la séquence de la marche de l’éléphant, en utilisant les croquis et la segmentation d’images.

méthode s’appuie sur la dynamique de la séquence d’images en 2D pour capturer la dynamique d’un mouvement 3D, mais n’a pas besoin de réalisme au niveau d’une seule image.

### 7.3 Sélection des images-clés : Comparaison entre images réelles et croquis

Afin de valider l’utilisation du croquis, nous avons choisi une séquence de marche d’un éléphant où la segmentation peut être appliquée avec succès. Nous avons dessiné par dessus la séquence et comparé les résultats obtenus en utilisant les deux méthodes d’extraction de silhouette : la segmentation d’images et le dessin. Les images-clés sélectionnées sont légèrement différentes (Table 4), mais correspondent à des poses très similaires (Figure 6.15). Ceci montre que les croquis sont une bonne alternative à la segmentation. Ils ne capturent pas exactement la forme de l’animal, mais encodent correctement sa démarche.



FIG. 6.15 – Images-clés sélectionnées pour la séquence de la marche de l’éléphant, en utilisant des croquis dessinés par l’utilisateur sur les images (en haut) et la segmentation directe d’images (en bas). Nous ne montrons pas tous les croquis, mais ils ont été dessinés par-dessus chaque image.

## 8 Bilan

Alors que la capture de mouvement classique avec marqueurs n'est pas faisable sur des sujets non coopératifs comme les animaux sauvages, nous avons montré que les vidéos réelles peuvent fournir de l'information significative sur le mouvement. Dans ce chapitre, nous avons proposé des outils nouveaux et robustes pour analyser des séquences vidéo et les rendre utilisables pour la génération de mouvement 3D.

Nous utilisons l'Analyse en Composantes Principales pour extraire des paramètres à partir d'une version binaire des images d'entrée. Comme nos résultats l'ont montré, un petit nombre de paramètres est suffisant pour les mouvements cycliques d'animaux : nous obtenons de bons résultats avec seulement deux composantes principales. Nous proposons aussi un critère pour sélectionner le meilleur ensemble d'images-clés à partir de la vidéo. Dans notre application, les images-clés sélectionnées peuvent être facilement interprétées comme les images extrémales dans l'espace 2D des composantes principales.

Notre travail montre que l'Analyse en Composantes Principales (ACP) peut être appliquée à une séquence d'images 2D pour contrôler du mouvement 3D. L'ACP sur les images aide à quantifier les changements importants d'apparence sur la vidéo. L'interpolation par FBR des exemples de pose a pour but de transposer la dynamique des variations de la vidéo dans l'espace des poses 3D. La sélection automatique des exemples permet à l'artiste de concentrer son effort sur les images-clés les plus importantes.

Notre méthode pourrait tirer bénéfice de techniques plus avancées d'analyse vidéo. Par exemple, elle est très sensible aux rotations de la caméra, ainsi qu'à sa translation hors du plan image. Ce sont des mouvements non liés au mouvement qui apparaissent dans les composantes principales. Des techniques de compensation de caméra pourraient nous aider à appliquer notre méthode à un plus large ensemble de vidéos.

Dans le futur, nous prévoyons d'explorer les mouvements non cycliques comme les transitions entre allures. L'ajout de contraintes physiques pourrait aider à animer les parties non cycliques du mouvement. Des méthodes de réduction de dimension non linéaire pourraient aussi être utilisées pour mieux visualiser et paramétrer des mouvements complexes.

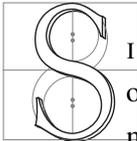
Nous étudions aussi comment réutiliser les données d'ACP déjà obtenues et leurs poses 3D associées pour analyser automatiquement une nouvelle séquence vidéo grâce à une adaptation morphologique dans l'espace image.



---

# Analyse des séquences de posés à partir de vidéos.

---



SI LES MÉTHODES CLASSIQUES de capture de mouvement à l'aide de marqueurs optiques ou magnétiques ne sont pas adaptées aux animaux sauvages, nous avons montré qu'il était possible d'extraire, à partir d'une séquence vidéo représentant un mouvement cyclique de locomotion d'un quadrupède, un petit nombre de paramètres globaux qui permettent de décrire la dynamique du mouvement. En ajoutant de la connaissance a priori sur le mouvement représenté, en l'occurrence un modèle 3D de l'animal à animer, ainsi que quelques poses-clés, il est possible de créer du mouvement 3D à partir de la vidéo. Cependant, les paramètres globaux ne permettent pas de décrire de manière précise certaines contraintes locales, comme les posés des pieds au sol. Si les mouvements capturés au chapitre précédent sont globalement satisfaisants, il n'y a aucune garantie que les pieds resteront fixes lors des contacts avec le sol. Or c'est un critère indispensable pour qu'une animation soit visuellement plausible.

Dans ce chapitre, nous poursuivons l'objectif fixé pour cette partie : extraire, à partir de vidéo, un petit nombre de paramètres permettant de décrire un mouvement de locomotion.

Cependant, nous voulons à présent extraire des paramètres locaux : la succession des posés des différents pieds sur le sol. Nous verrons que ces posés permettent d'utiliser un nouveau formalisme l'*analyse en séquence*, qui permet d'identifier l'allure utilisée par l'animal, ainsi que des éventuelles transitions entre allures. Ce travail a été commencé lors d'une collaboration avec l'Université de Washington dans le cadre d'une bourse EXPLORADOC, et a été poursuivi en collaboration avec le Muséum National d'Histoire Naturelle, dans le cadre du projet ANR Kameleon.

## 1 Contexte

En animation, il existe quelques méthodes permettant d'animer des quadrupèdes à partir de posés. Torkos et al. [TvdP98] utilisent une technique d'animation par optimisation qui calcule d'abord le mouvement d'un squelette très simplifié. Le squelette simplifié consiste en deux masses ponctuelles pour le bassin et les épaules, 5 segments pour les pattes et la colonne vertébrale. Ensuite, mouvement du squelette complet est ensuite reconstruit à l'aide de méthodes d'animation classiques telles que la cinématique inverse. Cette méthode permet d'obtenir des animations de marche/sauts pour un quadrupède (Figure 7.1), mais le mouvement obtenu ne semble pas naturel visuellement. La physique semble avoir un poids trop important et le mouvement ressemble à celui d'un objet lourd qu'on lance en l'air. De manière générale, les méthodes d'animation par optimisation avec des contraintes physiques ne semblent pas efficaces pour animer des mouvements lents comme la marche, où le contrôle exercé par les muscles est beaucoup plus important que la physique.

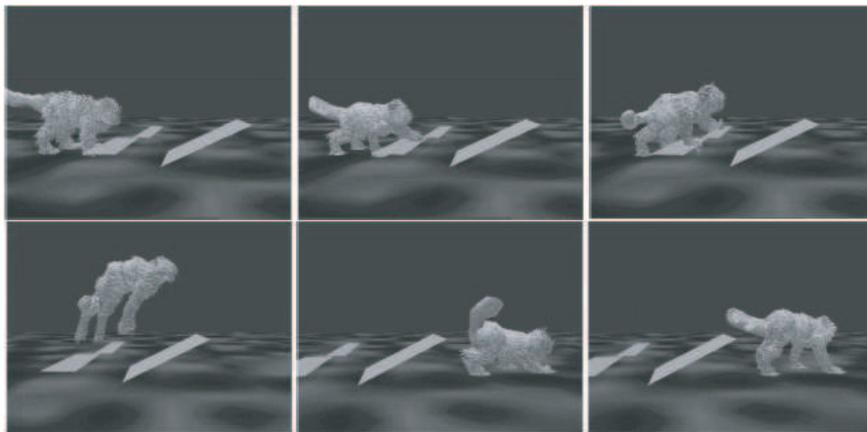


FIG. 7.1 – Torkos et al. [TvdP98] animent des quadrupèdes à partir de données de posés.

En biomécanique, de nombreux travaux se sont intéressés à l'identification des allures. Depuis les travaux de Hildebrand en 1966 [Hil66, Hil89], plusieurs méthodes ont permis d'iden-

tifier les différentes allures, mais aucune ne pouvait suivre toutes les coordinations temporelles entre les membres lors du passage d'une allure symétrique à une allure asymétrique, et inversement. Plus tard, A.W. English [EL82], Z. Afelt et al. [ABD83] et J.W. Blaszczyk [Bla01] se sont intéressés aux coordinations intra- et inter-paires pour comprendre les mécanismes de transition, mais en réalisant une étude plutôt descriptive, sans s'appuyer sur une base théorique. Nous allons nous appuyer sur une nouvelle méthode d'analyse des allures introduite par Abou-rachid [Abo03], qui permet de décrire toutes les allures, symétriques et asymétriques, dans un même formalisme.

En vision par ordinateur, l'analyse des allures à partir de vidéo a principalement concerné l'identification des humains à partir de leur allure ([LG05], voir [MG01] pour un état de l'art complet). Hannuna et al. [HCG05] suivent un ensemble de points sur la vidéo puis, à l'aide d'une classification par ACP, déterminent si il s'agit d'une vidéo de locomotion animale ou non, mais n'identifient pas d'allure. A notre connaissance, notre méthode est la première permettant d'identifier explicitement une allure de quadrupède à partir de vidéo, ainsi que les transitions entre allures successives.

La suite de ce chapitre est organisée comme suit : tout d'abord, nous montrons comment extraire, à partir d'une vidéo de quadrupède, les posés des pieds. Ensuite, nous présentons la méthode d'analyse en séquence et l'utilisation que nous en faisons pour identifier la ou les allures utilisées par le quadrupède à partir de ces posés. Enfin, nous présentons nos résultats : les allures sont correctement identifiées à partir de vidéos de locomotion cyclique, mais aussi sur une vidéo présentant une transition.

## 2 Extraction des posés des pieds à partir de vidéo

Pour obtenir la séquence des posés, nous nous proposons dans cette section de capturer la position 2D des quatre pieds au cours du temps, à partir d'une vidéo binarisée telle que celles que nous avons utilisées dans la Partie 2. Plus précisément, nous voulons savoir, à chaque pas de temps, quels pieds sont posés au sol et quels pieds ne le sont pas. Notre méthode d'extraction des posés est composée de 3 étapes : la détection de la position 2D des pieds sur chaque image sans les identifier explicitement, la reconstruction des trajectoires 2D de chaque pied, et enfin l'extraction des posés proprement dite.

### 2.1 Détection des pieds sur une vidéo

A cause des nombreuses occlusions subies par les jambes sur une vidéo, un suivi de points d'intérêts tel que la méthode de Lucas-Kanade [LK81] n'est pas applicable au suivi des pieds

sur une vidéo de quadrupède : lorsque les jambes se croisent, les points d'intérêt sont perdus. Pour résoudre ce problème il faudrait réinitialiser manuellement les points à suivre après chaque occlusion. De plus, la forme des pieds sur une vidéo définit une forme qui tourne et se déforme au cours du temps, ce qui provoque des phénomènes de glissement des points d'intérêt. Nous proposons au contraire de détecter, à chaque pas de temps, les zones de l'image qui ont la forme caractéristique d'un pied, qu'on pourrait définir par "l'extrémité d'une forme longiligne". Pour cela nous définissons un filtre qui permet de déterminer, pour chaque position dans l'image courante, si la zone située autour de cette position a la forme recherchée ou non. Une représentation schématique de ce filtre est donnée sur la Figure 7.2.

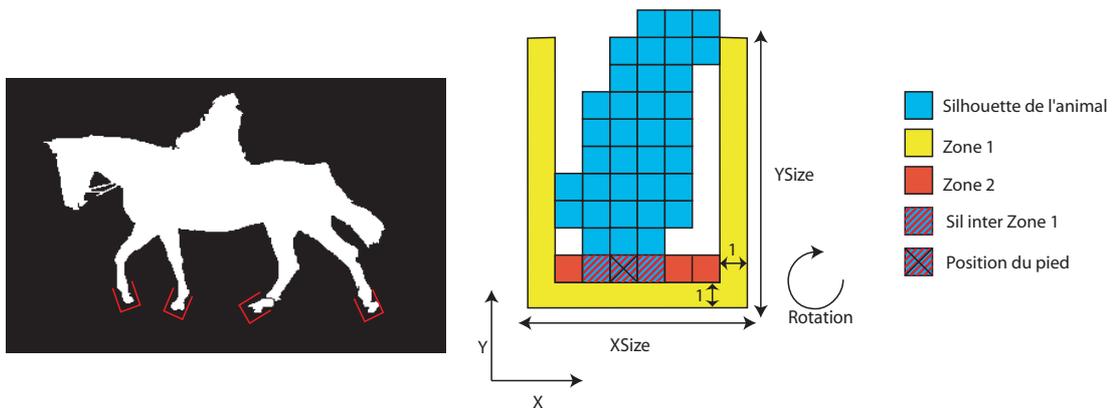


FIG. 7.2 – A gauche : Sur une vidéo binarisée, les pieds exhibent une forme très caractéristique qu'on peut définir par "l'extrémité d'une forme longiligne". A droite : Représentation du filtre utilisé pour détecter des régions susceptibles d'être des pieds. Le filtre détermine si une région de la silhouette de l'animal (en bleu) a la forme caractéristique d'un pied. La réponse est positive si la Zone 1 (en jaune) ne contient aucun pixel de la silhouette de l'animal, et si la Zone 2 (en rouge) en contient au moins un. Les paramètres  $XSize$  et  $YSize$  sont définis manuellement par l'utilisateur pour chaque séquence à traiter. La rotation  $Theta$  ainsi que la position en 2D  $(X, Y)$  du filtre sur l'image sont déterminées lors de la recherche. Les dimensions sont données en pixels. la position du pied estimée est le milieu de l'intersection entre la silhouette de l'animal et la partie rouge du filtre (le pixel marqué d'une croix).

Une position du filtre est définie par trois paramètres  $(X, Y, \Theta)$ , où  $(X, Y)$  est la position 2D sur filtre sur l'image, et  $\Theta$  la valeur de rotation autour de son centre. Chaque image est balayée d'abord de gauche à droite. Pour chaque position en  $X$ , l'image est parcourue de bas en haut jusqu'à ce qu'une réponse positive soit donnée pour une valeur d'angle donnée (Figure 7.3). Dans ce cas, la position courante  $(X, Y, \Theta)$  est marquée comme un *candidat* (une position possible d'un pied). Pour éviter de prendre en compte certains éléments imprévus qui pourraient aussi avoir cette forme (longue queue, éléments du décor mal segmentés), l'utilisateur peut

dessiner un polygone sur l'image pour limiter la zone de recherche.

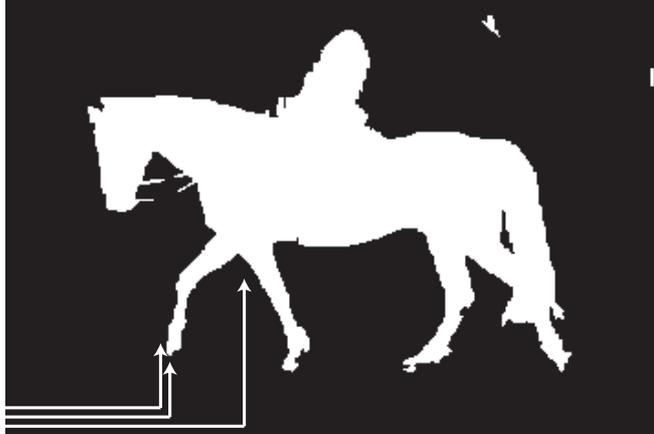


FIG. 7.3 – Les candidats sont recherchés de bas en haut dans chaque colonne de l'image.

Souvent, nous obtenons plusieurs candidats qui correspondent au même pied (Figure 7.4). Nous groupons donc les candidats dont les boîtes englobantes s'intersectent. Pour chaque groupe nous choisissons un représentant et un seul, celui dont la position en 2D est la médiane des positions des membres du groupe. Ensuite, la position précise du pied est choisie comme expliqué sur la Figure 7.2. Au final nous obtenons, pour chaque pas de temps, la position en 2D d'au plus quatre pieds (Figure 7.5). Cependant, à cause des occlusions, un pied peut ne pas être détecté sur certaines images.

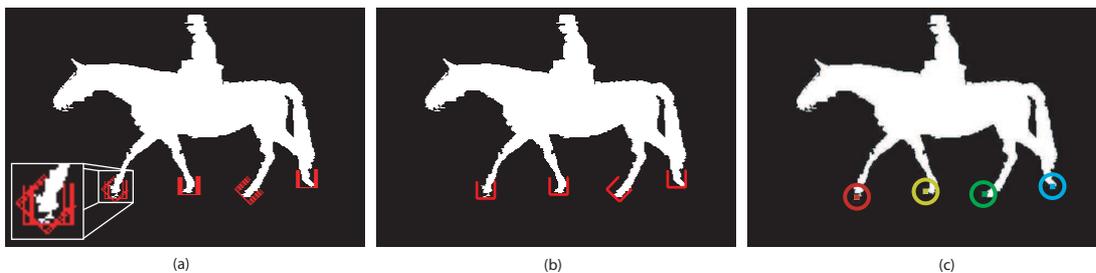


FIG. 7.4 – (a) candidats. Pour un même pied, plusieurs candidats peuvent exister (b) après groupement, un filtre est sélectionné par pied . (c) La position détectée des pieds.

Après cette étape nous avons, à chaque pas de temps, la position en 2D d'au plus quatre pieds, mais nous ne savons pas à quels pieds (avant gauche, avant droit, arrière gauche, arrière droit) ces positions correspondent. Il nous faut donc reconstruire la trajectoire de chaque pied au cours du temps, et les identifier explicitement.

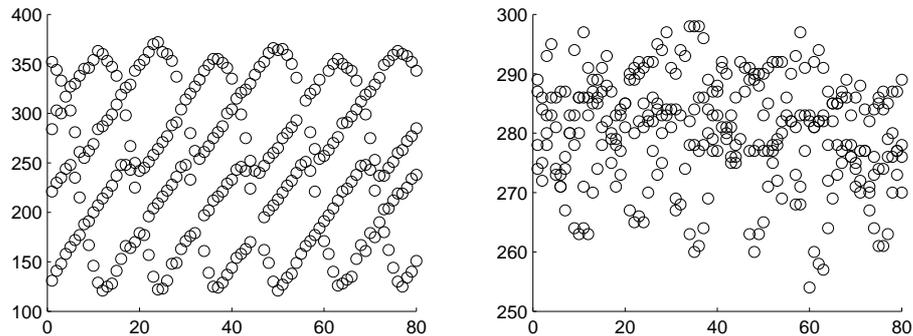


FIG. 7.5 – Les position horizontales(à gauche) et verticales(à droite) des pieds au cours du temps après application du filtre.

## 2.2 Reconstruction des Trajectoires

Les trajectoires sont reconstruites en utilisant une estimation à l'ordre 1 de la trajectoire de chaque pied. La technique de reconstruction est détaillée sur la Figure 7.6. Après reconstruction automatique, l'utilisateur peut valider le résultat ou bien corriger manuellement certaines erreurs.

Enfin, nous demandons à l'utilisateur d'identifier explicitement chaque pied. Il est possible de discriminer automatiquement les pieds avant des pieds arrières à partir de leurs positions et de leur sens de rotation sur l'image, mais distinguer pieds droits et gauches est très difficile, voire impossible, sur des vues proches de la vue de côté. Notamment, il n'est pas possible de distinguer le trot du rack sur des silhouettes vues de profil. La Figure 7.7 montre un exemple de trajectoires 2D obtenues pour les quatre pieds d'un cheval effectuant un mouvement de marche. On remarque que les positions en  $y$  sont beaucoup plus bruitées que les positions en  $x$ . En effet, la segmentation des images en silhouette en noir et blanc ne donne pas des résultats parfaits, et notamment les extrémités des pattes ne sont pas segmentées avec une très grande précision. En conséquence, les longueurs des pattes ne sont pas stables dans le temps. Comme les pattes ont souvent une orientation plus proche de la verticale, les erreurs affectent plus la coordonnée  $y$ . Cependant, nous allons voir que ces informations, même bruitées, permettent de déterminer l'allure utilisée par le quadrupède.

## 2.3 Extraction des posés

Tout d'abord, il nous faut extraire les séquences de posés à partir des trajectoires en 2D des pieds que nous avons obtenues auparavant. Pour cela, les trajectoires de chaque patte sont découpées en cycles à l'aide d'une analyse fréquentielle. La solution la plus évidente serait d'extraire les posés des pieds à partir des positions en  $y$ . Cependant, sur des vidéos natu-

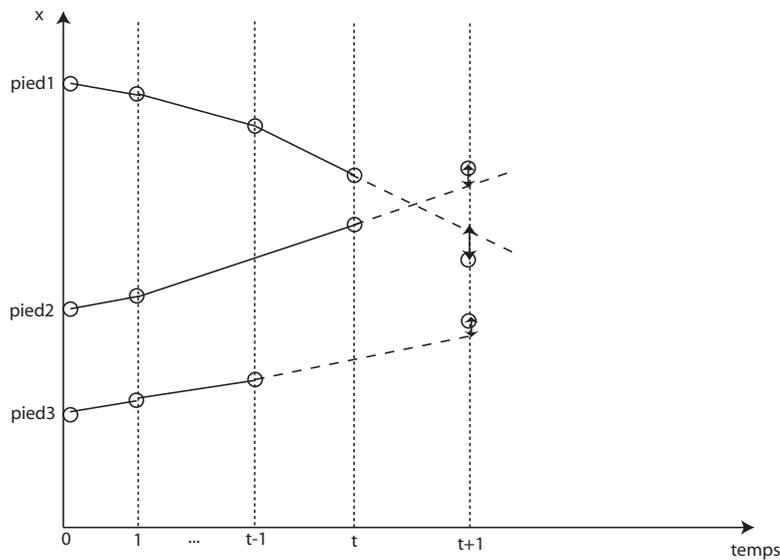


FIG. 7.6 – Illustration de la méthode utilisée pour reconstruire les trajectoires. Notre méthode est appliquée à des trajectoires en 2D, mais elle est représentée ici en 1D pour plus de clarté. Les cercles représentent les positions de pieds issues de l'étape précédente. Les trajectoires sont propagées au cours du temps, en partant du premier pas de temps. La figure représente le fonctionnement de l'algorithme au temps  $t$ . En traits pleins, les trajectoires déjà reconstruites pour l'intervalle de temps  $[0, t]$ . Les trajectoires sont extrapolées linéairement (traits en pointillés) pour estimer les positions des pieds au temps  $t + 1$ . Chaque position mesurée au temps  $t + 1$  est alors rattachée à la trajectoire dont la position estimée est la plus proche. Ensuite on peut passer au pas de temps suivant.

relles, les pieds sont souvent cachés dans l'herbe, et ils sont souvent très difficiles à binariser précisément à partir de leur couleur. Cela induit une erreur sur la longueur du segment extrême de chaque patte. Comme les pattes ont beaucoup plus souvent une orientation proche de la verticale, l'erreur en  $y$  sera plus importante que l'erreur en  $x$ . Nous choisissons donc de détecter les posés à partir des positions extrémales en  $x$  (Figure 7.8). Au final nous savons, pour chaque pas de temps, quels pieds sont au sol. La Figure 7.9 montre quelques exemples de résultats.

L'extension naturelle de ce travail est d'utiliser les données des posés pour ajouter des contraintes locales à notre méthode globale, afin d'assurer que les pieds restent bien au sol au cours du mouvement. Cette extension pourra être réalisée dans le futur. Nous avons utilisé ces posés pour une autre application : l'identification des allures des quadrupèdes.

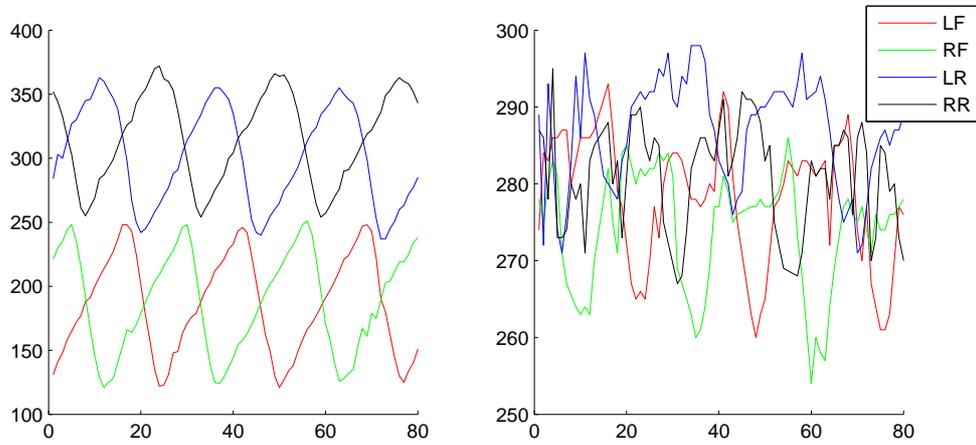
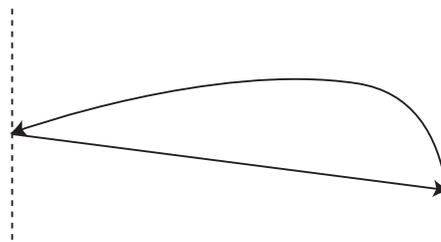


FIG. 7.7 – Les trajectoires de chaque pied après reconstruction, selon la coordonnée  $x$  (à gauche), et  $y$  (à droite).



(a)

FIG. 7.8 – Représentation schématique du mouvement en 2D d'un pied. Comme les positions en  $y$  sont beaucoup plus bruitées que les positions en  $x$ , les posés sont calculés à partir des positions extrémales en  $x$

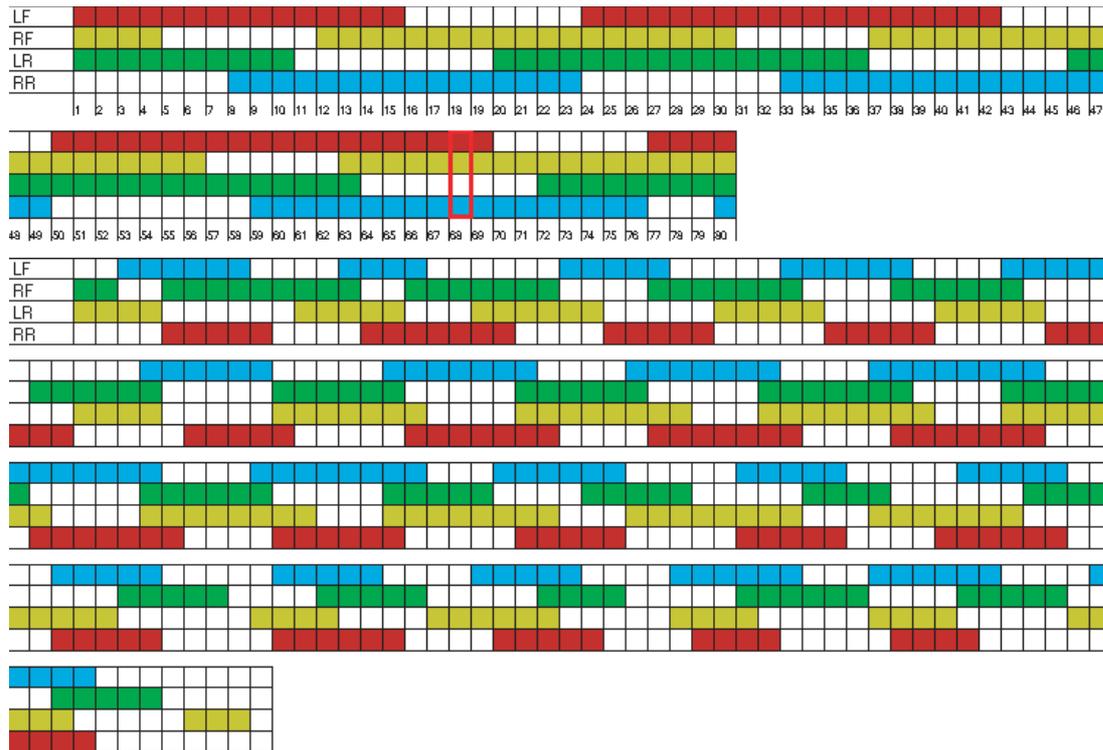


FIG. 7.9 – Séquences de posés obtenues par notre méthode pour une marche (en haut) et pour une séquence de transition (en bas). Chaque colonne correspond à un pas de temps, chaque ligne à un pied (RF : Avant-droit, LF : Avant-gauche, RR : Arrière-droit, LR : Arrière-gauche). Les cases colorées correspondent aux moments où les pieds sont au sol.

### 3 Caractérisation des allures des quadrupèdes par les posés des pieds

Comme nous l'avons déjà vu, la séquence des posés des pieds est une donnée très importante pour décrire un mouvement de locomotion. La séquence des posés des pieds sur le sol détermine l'allure utilisée par l'animal, la présence ou l'absence de phases de suspension au cours du mouvement, sa vitesse. Dans cette section, nous allons d'abord introduire la méthode d'analyse en séquences, qui permet de caractériser les allures des quadrupèdes à l'aide d'un petit nombre de paramètres, appelés les *pair lags*. Ensuite, nous montrons comment déduire ces *pair lags* à partir des posés que nous avons extraits dans la section précédente, et donc identifier les allures et transitions entre allures à partir de séquences vidéos.

#### 3.1 La méthode d'analyse en séquences

La méthode d'analyse en séquences est une approche théorique introduite par Abourachid [Abo03]. Cette approche permet de décrire quantitativement toutes les coordinations locomotrices, en allures constantes comme dans les transitions. Elle part de deux constats. Tout d'abord, les membres d'une même paire (avant ou arrière) sont similaires par leur morphologie et leur fonctionnalité [Hil65, CLS02]. D'autre part, l'influx nerveux moteur se propage d'avant en arrière [GCH<sup>+</sup>00]. D'après ces constats, on déduit que les mouvements locomoteurs s'inscrivent dans une séquence qui débute par les posés des membres antérieurs suivis de ceux des membres postérieurs.

L'analyse en séquences délaisse la durée de cycle d'un membre de référence comme unité locomotrice au profit de la séquence définie comme l'association des cycles consécutifs des deux antérieurs suivis des cycles consécutifs des deux postérieurs (Figure 7.10). La locomotion quadrupède n'est plus considérée sur la base d'une unité temporelle, mais plutôt comme une succession d'actions.

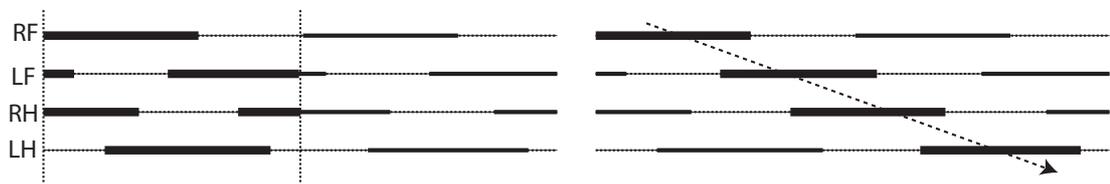


FIG. 7.10 – A gauche : la méthode classique analyse les mouvements en les découpant en intervalles de temps qui commencent par les posés du même pied. A droite : dans l'analyse en séquences, un cycle de mouvement est la succession d'un posé des pattes antérieures puis d'un posé des pattes postérieures.

### 3. CARACTÉRISATION DES ALLURES DES QUADRUPÈDES PAR LES POSÉS DES PIEDS<sup>127</sup>

Pour être rigoureux, il faut choisir un pied de référence, dont le posé déterminera le début d'un cycle. Dans l'analyse en séquence, conformément à un flux nerveux moteur antéro-postérieur, ce membre est un antérieur. Dans le cas des allures symétriques, le choix de l'antérieur droit ou gauche n'a aucune importance puisqu'ils remplissent le même rôle, et on choisit donc l'un ou l'autre indifféremment. En revanche, pour les allures asymétriques, la référence est le pied qui se pose en premier, c'est-à-dire celui qui découvre l'environnement en premier [AHH<sup>+</sup>]. On le nomme *Trailing Antérieur* (TA), de l'anglais *trailer* (remorque). L'autre antérieur est nommé *Leading Antérieur* (LA), de *leader* (premier, meneur), car il est celui des antérieurs qui se pose le plus en avant pour une allure asymétrique. L'homolatéral du TA est le *Trailing Postérieur* (TP) et l'autre postérieur, homolatéral du LA, est appelé *Leading Postérieur* (LP). Ceci permet d'analyser les allures de la même manière selon que le quadrupède pose une patte droite ou une patte gauche en premier.

Les paramètres qui permettent de caractériser une allure sont les déphasages temporels entre différentes pattes, exprimés en pourcentage de la durée d'un cycle du TA. Ces déphasages, appelés *pair lags*, sont :

- le décalage temporel entre les posés des deux antérieurs, ou *fore lag* (FL),
- le décalage temporel entre les posés des deux postérieurs, ou *hind lag* (HL),
- le décalage temporel entre les posés des deux membres trailing, ou *pair lag* (PL).

Le *fore lag* et le *pair lag* sont toujours positifs, le posé du trailing antérieur étant considéré comme le début du cycle. Le *hind lag*, lui, peut être négatif si le leading postérieur est placé avant le trailing postérieur. C'est le cas pour le galop rotatoire.

L'analyse en séquence permet aussi de définir de la même manière les décalages spatiaux, que l'on nomme alors *fore gap*, *hind gap* et *pair gap*. Cependant l'extraction de mesures spatiales à partir de vidéo n'est pas le propos de cette thèse, nous utiliserons donc seulement les décalages temporels. Par cette approche, il est possible d'identifier chaque allure à partir des *pair lags*, comme introduit par [Abo03] (Table 7.1).

#### 3.2 Analyse automatique des allures des quadrupèdes

L'objectif final de ce chapitre est d'appliquer l'analyse en séquences aux posés que nous avons extraits à partir de vidéo. Pour cela, nous procédons en plusieurs étapes (Figure 7.12) :

- Les cycles des paires avant sont trouvés en associant chaque posé du trailing antérieur avec le posé du leading antérieur qui lui succède.
- Les cycles des paires arrières sont trouvés en associant chaque posé du trailing postérieur avec le posé du leading postérieur qui lui est le plus proche dans le temps. Pour éviter les ambiguïtés et les erreurs de mesures pour les allures où le *hind lag* vaut 50% et comme le *hind lag* peut être négatif, on contraint le *hind lag* à être dans l'intervalle  $[-40\%, 60\%]$ .

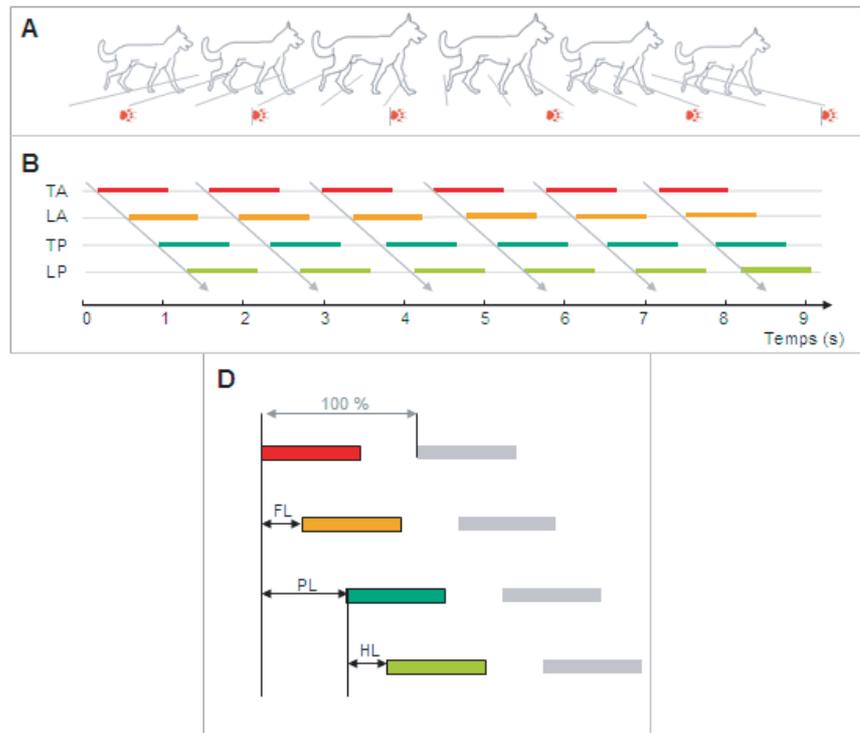


FIG. 7.11 – Analyse en séquences. A : Les débuts des cycles sont les posés du membre *trailing antérieur*, la patte avant qui se pose en premier. B : Diagramme d’allure. L’axe des abscisses représente le temps. Les marques de couleur (orange et rouge pour la paire antérieure et vertes pour la paire postérieure) représentent les temps de contact entre chaque membre et le sol. Chaque séquence est identifiée par les flèches grises. D : Mesure des décalages temporels. Pour chaque séquence, le décalage temporel entre les posés des deux antérieurs (Fore lag, FL), le décalage temporel entre les posés des deux postérieurs (Hind lag, HL) et le décalage temporel entre les posés des deux membres trailing (Pair lag, PL), sont mesurés et exprimés en pourcentage de la durée du cycle du TA.

- On associe les actions des paires en associant un posé du trailing antérieur avec le posé du trailing postérieur qui lui succède.

A la fin de ce processus, le mouvement est découpé en séquences et on connaît les valeurs de pair lags pour chaque cycle. Cela nous permet donc de déduire les allures correspondant à chaque cycle, ainsi que les transitions.

Allure	F lag	H lag	P lag
Amble	50	50	100
Trot	50	50	50
Marche latérale	50	50	> 50
Marche diagonale	50	50	< 50
Galop transverse	< 50	50	dépend de la vitesse
Galop rotatoire	< 50	50	dépend de la vitesse
Demi bond	< 50	0	dépend de la vitesse
Bond	< 50	0	dépend de la vitesse
Saut	0	0	0

TAB. 7.1 – Toutes les allures peuvent être identifiées à l'aide de seulement trois paramètres : le décalage temporel entre les poser des deux antérieurs (Fore lag, F lag), le décalage temporel entre les poser des deux postérieurs (Hind lag, H lag) et le décalage temporel entre les poser de deux membres homolatéraux (Pair lag, P lag). Les décalages temporels sont exprimés en pourcentage de la durée du cycle du TA.

séquence	front lag	hind lag	pair lag	allure trouvée
marche	0.500000	0.500000	0.736842	marche latérale
trot	0.478261	0.478261	0.521739	trot
galop	0.333333	0.166667	0.833333	galop transverse

TAB. 7.2 – lags et allures calculés automatiquement pour quelques séquences synthétiques

## 4 Résultats

Nous avons d'abord appliqué la méthode d'analyse par *pair lags* sur des animations 3D synthétiques, afin de la valider. A partir d'une animation 3D représentant une allure connue, nous extrayons les séquences de posés, puis appliquons l'analyse en séquence. La table 7.2 montre ces résultats : les allures sont identifiées correctement. Ensuite, nous avons confronté notre méthode d'analyse à des vidéos réelles : d'abord des vidéos présentant des allures simples (table 7.3), puis sur une vidéo présentant une transition entre allures (table 7.4). Les allures sont correctement identifiées. La vidéo du cerf qui présente un cerf accélérant progressivement et passant du trot au galop. Une fois encore, les allures sont correctement identifiées et de plus les transitions sont identifiées aux bons endroits. On remarque au passage que la faible résolution temporelle des vidéos (un cycle de course est effectuée en quelques images) induit une faible résolution sur les valeurs de pair lags. Cependant, les allures et les transitions entre allures sont bien identifiées.

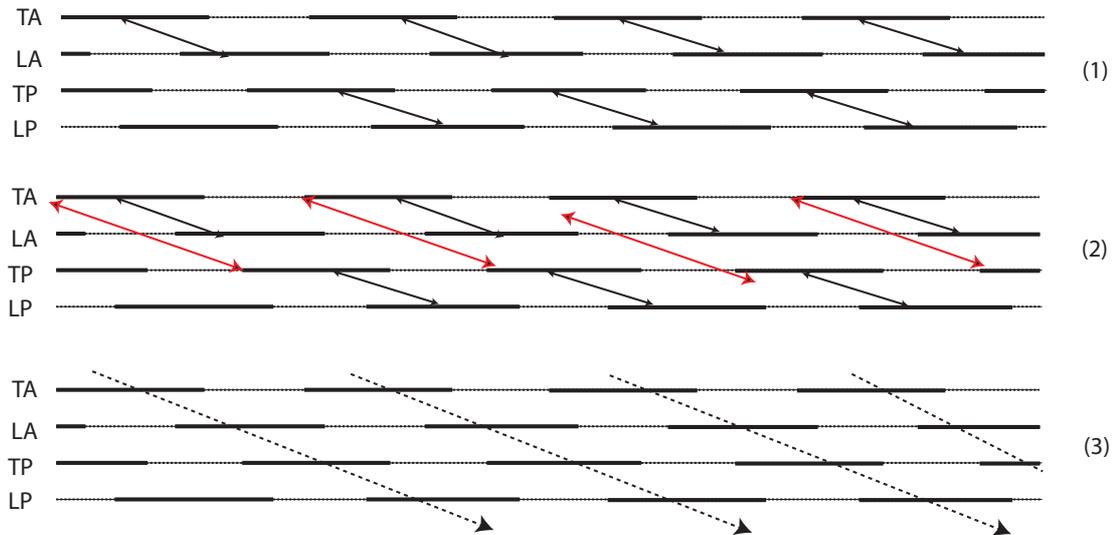


FIG. 7.12 – Calcul des pair lags et recombinaison des séquences à partir de la donnée d’une suite de posés. (1) On regroupe les actions intra-paires. Chaque posé du trailing antérieur est associé avec le posé du leading antérieur qui lui succède, et chaque posé du trailing postérieur est associé avec le posé du leading postérieur qui lui est le plus proche dans le temps. (2) On regroupe les actions des différentes paires. Chaque posé du trailing antérieur est associé avec le posé du trailing postérieur qui lui succède (3) On a décomposé le mouvement en séquences et on peut calculer les pair lags, donc l’allure.

no image	front lag	hind lag	pair lag	allure trouvée
11	0.480000	0.520000	0.840000	marche latérale
36	0.500000	0.500000	0.846154	marche latérale

TAB. 7.3 – lags et allures calculés automatiquement pour la séquence de marche de l’antilope

## 5 Bilan

Si la méthode d’analyse présentée dans le chapitre précédent permettait de contrôler globalement un mouvement de locomotion à l’aide d’un petit nombre de paramètres, elle ne permettait pas d’extraire certaines contraintes locales importantes. Dans ce chapitre, nous avons

no image	front lag	hind lag	pair lag	allure trouvée
20	0.558824	0.500000	0.794118	marche latérale
54	0.531250	0.500000	0.812500	marche latérale

TAB. 7.4 – lags et allures calculés automatiquement pour la séquence de marche du cheval

no image	front lag	hind lag	pair lag	allure
59	0.545455	0.545455	0.454545	trot
70	0.545455	0.545455	0.454545	trot
81	0.545455	0.545455	0.454545	trot
92	0.545455	0.545455	0.454545	trot
103	0.545455	0.545455	0.454545	trot
114	0.545455	0.363636	0.454545	galop transverse
125	0.454545	0.272727	0.454545	galop transverse
136	0.272727	0.272727	0.363636	galop transverse
147	0.181818	0.272727	0.181818	galop transverse
158	0.125000	0.250000	0.125000	galop transverse
166	0.272727	0.272727	0.181818	galop transverse

TAB. 7.5 – lags pour la séquence du cerf. Les transitions sont retrouvées au bon endroit

d’abord montré comment extraire à partir de vidéo et en utilisant peu d’interventions de l’utilisateur, les séquences de posés des pieds. Ces posés donnent des détails locaux très importants très importantes pour le réalisme du mouvement. Ils permettent aussi de déterminer l’allure utilisées par un quadrupède. Cependant, il n’existait jusqu’alors pas de méthodes permettant de décrire formellement toutes les allures, y compris les allures symétriques et asymétriques. En combinant notre approche et la méthode d’analyse en séquence, nous pouvons analyser des vidéos et en déduire précisément les allures utilisées par l’animal, ainsi que le position précise des transitions entre allures dans le temps. Dans le cadre du projet ANR Kameleon, l’analyse semi-automatique des allures permet de gagner beaucoup de temps pour l’étude de la locomotion animale.

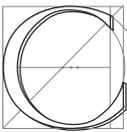
Dans le futur, nous souhaitons construire de contrôleur de mouvement qui combine les indices globaux et les indices locaux. Les indices globaux permettront des décrire la dynamique du mouvement de tout le squelette, tandis que les indices locaux permettront de préserver certaines contraintes d’ordre local. Par exemple, il serait possible d’ajouter des contraintes de cinématique inverse aux bons moments pour éviter les glissements des pieds sur le sol.



---

# Conclusion

---



ETTE THÈSE s'est intéressée à la modélisation du mouvement des quadrupèdes à partir de vidéos. Dans ce dernier chapitre, nous faisons d'abord un résumé de nos contributions, avant d'aborder des pistes envisageables pour des travaux futurs.

Pour finir, nous dressons un bilan général des enseignements que nous tirons de cette thèse.

## Résumé des contributions

Nous avons d'abord traité de la création de squelettes 3D de quadrupèdes. Nous avons validé un protocole de reconstruction 3D de squelette réels à partir d'images. Ensuite, nous avons montré comment construire un modèle déformable de squelette d'animation 3D. Ce modèle permet de créer rapidement des squelettes d'animation de quadrupèdes à partir d'images ou de surfaces 3D.

Ensuite nous avons traité de l'animation de quadrupèdes à partir de séquences vidéo. Dans un premier temps nous avons montré comment animer globalement un mouvement cyclique de locomotion à partir d'une vidéo. En particulier, nous avons proposé un algorithme de sélection automatique des images-clés à partir de vidéo, et proposé un algorithme de résolution des ambiguïtés dues au passage de 2D vers 3D. Ensuite, nous avons proposé une nouvelle méthode d'analyse de vidéos qui permet d'extraire les séquence de posés des pieds à partir d'une vidéo. Nous appliquons cette méthode à l'identification des allures et des transitions entre allures

présentes sur une vidéo. Cette méthode valide une nouvelle méthode de formalisation des allures mise au point au Museum National d'Histoire Naturelle.

## Perspectives

Nos travaux ont porté sur l'animation purement cinématique des quadrupèdes, et nous pensons que la modélisation du mouvement quadrupède pourrait aller plus en avant avec l'utilisation de modèles d'animation physiques. Il existe déjà des travaux qui visent à apprendre des paramètres physiques de personnages à partir de données de mouvement. Utiliser les vidéos pour modéliser des modèles de quadrupèdes dotés de paramètres physique pourrait sans doute apporter beaucoup plus, surtout dans le cas de mouvements extrêmement rapides d'animaux très légers, qui dépendent alors beaucoup des forces et de la gravité.

L'étude des quadrupèdes est extrêmement intéressante qui fait de la richesse de leur mouvement. Le nombre d'allures différentes que les quadrupèdes peuvent utiliser est impressionnant. Cependant, l'étude d'animaux non quadrupèdes est une direction envisageable. Si les travaux sur les humains, donc les bipèdes, abondent, l'étude des animaux à six ou huit apporterait sans doute autant de surprise et d'intérêt que celle des quadrupèdes.

## Bilan

Cette thèse a montré que, même si les vidéos réelles restent un défi à l'exploitation automatique du fait que leur faible résolution et de leurs nombreuses sources de variations, il reste possible, sous certaines conditions, d'en extraire des paramètres pertinent qui permettent de modéliser le mouvement des quadrupèdes et d'animer des modèles 3D de quadrupèdes.

Nos différentes collaborations ont permis d'aborder le problème sous des angles différents. Le fait de collaborer avec une infographiste de métier nous a permis de cerner les nombreuses contraintes qu'impose l'animation de créatures 3D dans un contexte professionnel. Nous avons pu prendre la mesure de la complexité de ce travail et de la place très importante qu'il reste pour développer des outils performants et ergonomiques pour faciliter leur travail. Nous avons du nous fonder dans le mode de travail des infographistes. Une première conclusion est qu'un outil, aussi performant qu'il soit, doit rester orienté vers l'utilisateur et garantir à l'artiste le maximum de contrôle sur le résultat

Notre collaboration avec le Muséum National d'Histoire Naturelle nous a permis de voir à quel point la locomotion animale est un domaine qui reste à étudier. C'est un domaine en plein essor actuellement, et au cours de nos analyse des allures nous avons soulevé de nombreuses questions qui restent du domaine de la recherche aujourd'hui, peut-être à cause de

la difficulté que représente l'analyse de mouvements à partir de la vidéo. Nous espérons que ce projet pourra aller plus en avant dans le futur, et croyons que l'informatique graphique, et particulièrement son application à l'analyse de documents réels et à la modélisation 3D de phénomènes réels, contribuera à faire avancer d'autres sciences dans le futur.



## ANNEXE A

---

# Hiérarchie de repères

---

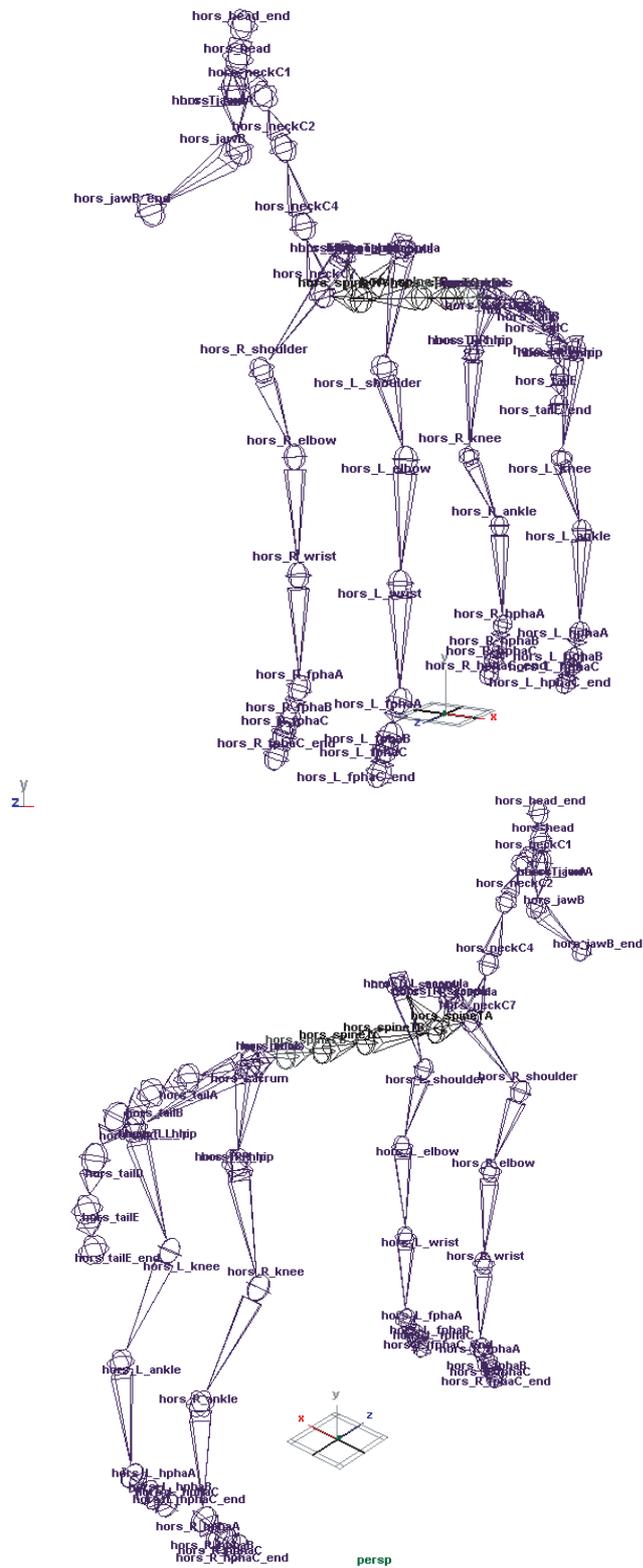


FIG. A.1 – La hiérarchie de repère que nous utilisons dans cette thèse. Comme c'est le cas le plus souvent, le pelvis est choisi comme nœud père

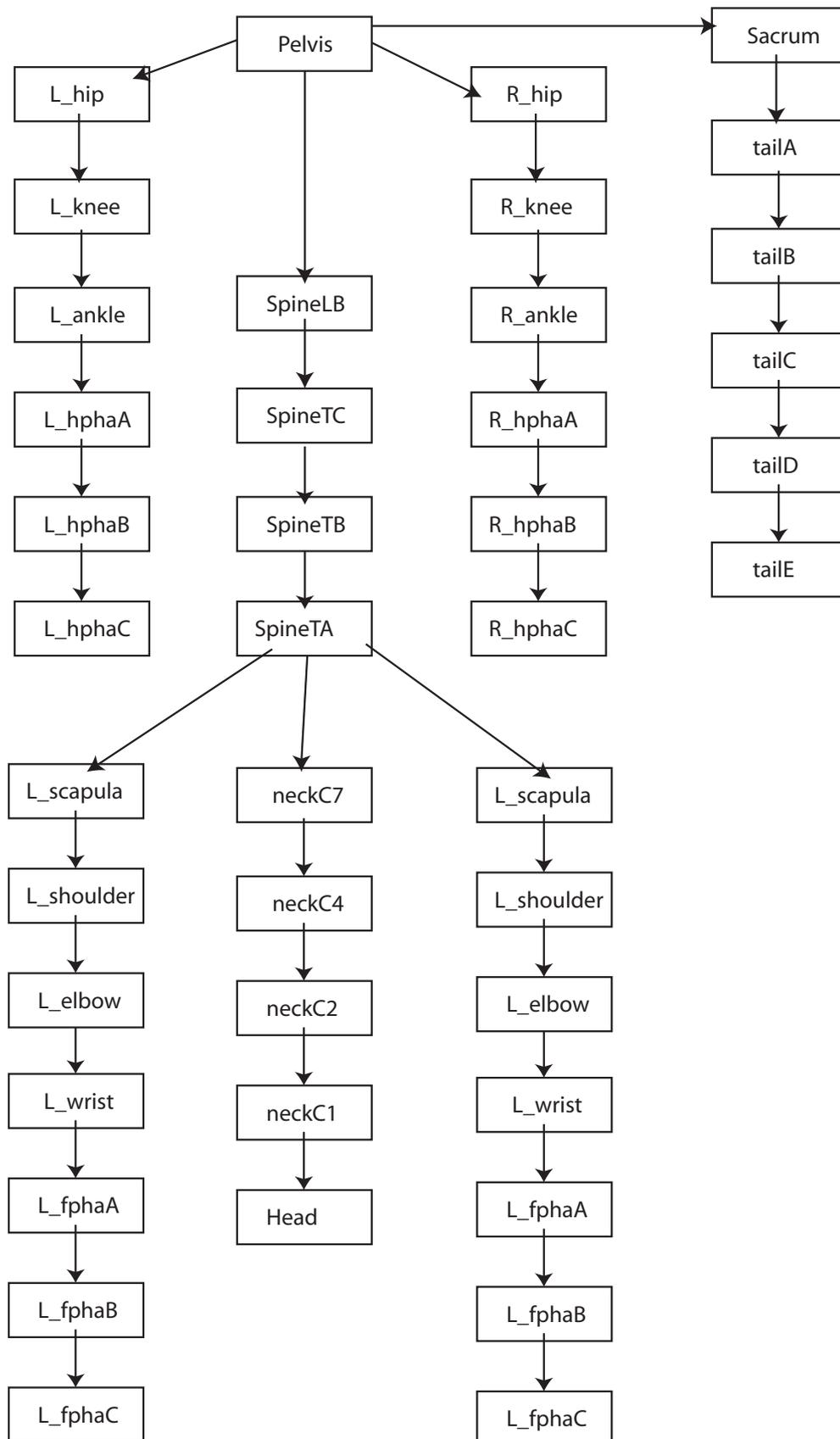


FIG. A.2 – La hiérarchie de repère que nous utilisons dans cette thèse. Comme c'est le cas le plus souvent, le pelvis est choisi comme nœud père

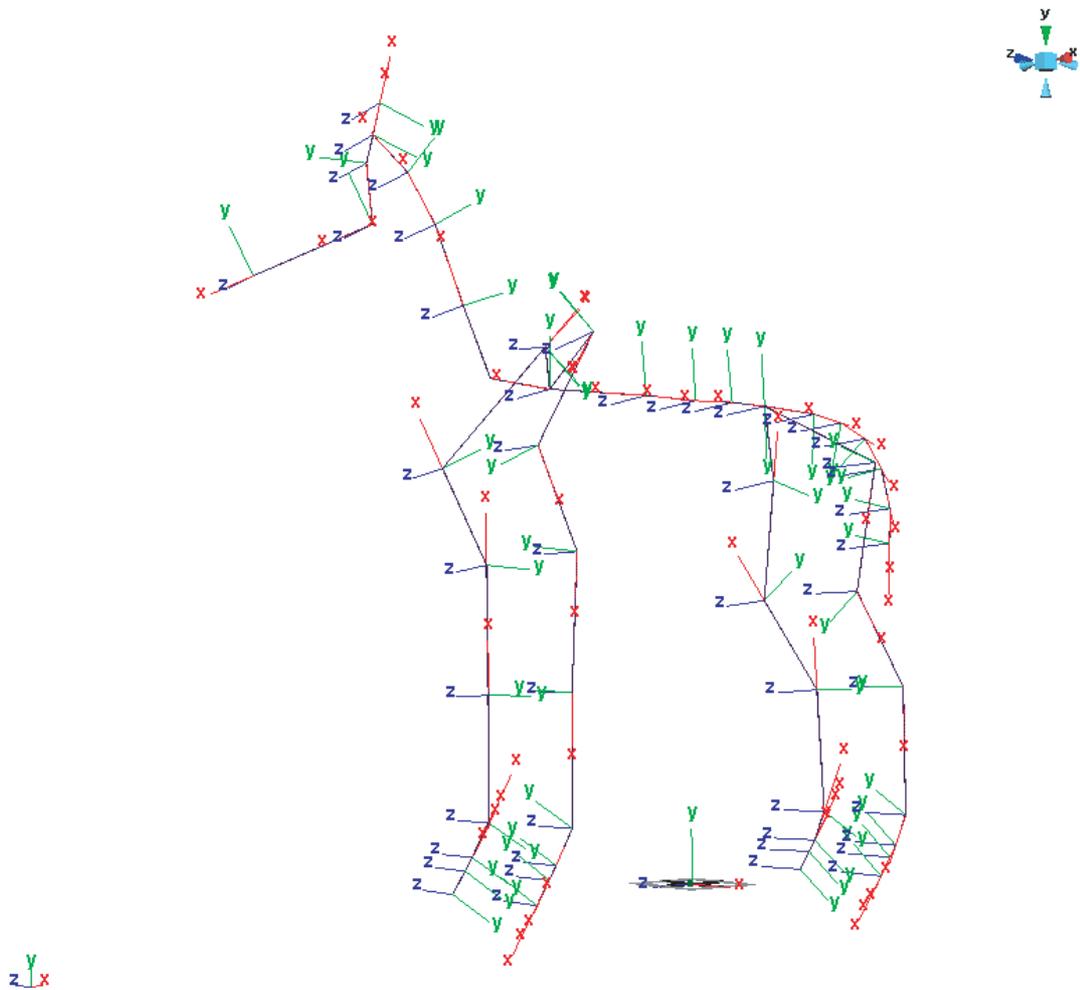


FIG. A.3 – La hiérarchie de repères orientés en 3D obtenue avec nos conventions.

---

## Quaternions

---

Les quaternions sont une extension des nombres complexes. De même qu'un complexe  $c$  est noté  $c = a + bi$  avec  $a^2 = -1$ , on définit un quaternion  $q$  par  $q = w + xi + yj + zk = (w; v)$ , avec :

$$i^2 = j^2 = k^2 = -1$$

$$ij = k, ji = -k$$

$$jk = i; kj = -i$$

$$ki = j; ik = -j$$

Le produit de quaternions (non commutatif) est défini par :

$$q_1 q_2 = (w_1 w_2 - v_1 \cdot v_2, w_1 v_2 + w_2 v_1 + v_1 v_2)$$

Le quaternion conjugué est défini par :

$$\bar{q} = w - xi - yj - zk$$

$$q\bar{q} = w^2 + x^2 + y^2 + z^2$$

Pour représenter les rotations, on utilise les quaternions unitaires, c'est-à-dire les quaternions  $q$  tels que  $q\bar{q} = 1$ . Un vecteur 3D  $(x, y, z)$  est représenté par le quaternion

$$p = (0; x; y; z)$$

, et la rotation d'axe  $\theta$  et d'angle  $u$  ( $(u^2 = 1)$ ) est représentée par :

$$q_{\theta,u} = \left( \cos \frac{\theta}{2}, u_x \sin \frac{\theta}{2}, u_y \sin \frac{\theta}{2}, u_z \sin \frac{\theta}{2} \right)$$

La rotation de la rotation  $q$  au vecteur  $p$  est alors notée :

$$qp\bar{q}$$

A noter qu'on peut convertir un quaternion unitaire en matrice de rotation avec la formule :

$$\begin{pmatrix} 1 * 2y^2 - 2z^2 & 2xy - 2wz & 2xz + 2wy \\ 2xy + 2wz & 1 - 2x^2 - 2z^2 & 2yz - 2wx \\ 2xz - 2wy & 2yz + 2wx & 1 - 2x^2 - 2y^2 \end{pmatrix}$$

Composition des rotations :

$$R_{\alpha,u} R_{\beta,u} \longrightarrow q(\alpha, u) q(\beta, v)$$

Rotation inverse :  $q_{(\theta,u)}^{-1} = q_{(-\theta,u)} = q_{(\theta,-u)} = (-w, v) = (w, -v)$

Conversion  $(w, v) \longrightarrow (\theta, u)$  :

$$\cos\left(\frac{\theta}{2}\right) = w$$

$$\sin\left(\frac{\theta}{2}\right) = \text{norm}(w)$$

$$u = v / \text{norm}(v)$$

	vecteur $v$	quaternion $q_{(\theta,u)} = (w, v)$
composition	$v_1 + v_2 (= v_2 + v_1)$	$q_1 q_2 (\neq q_2 q_1)$
inverse	$v_1^{-1} = -v$	$q_1^{-1} = (-w, v)$
produit avec un scalaire	$kv$	$q^k = q_{(k\theta,u)}$
distance	$\Delta v = v_2 - v_1 = -v_1 + v_2$	$\Delta q = q_1^{-1} q_2$
interpolation	$v = \text{lerp}(v_1, v_2, \alpha) = v_1 + \alpha(-v_1 + v_2)$	$q = \text{slerp}(q_1, q_2, \alpha) = q_1 (q_1^{-1} q_2)^\alpha$
moyenne	$\text{moy}(v_1, v_2) = \text{lerp}(v_1, v_2, 0.5)$	$\text{smoy}(q_1, q_2) = \text{slerp}(q_1, q_2, 0.5)$

---

# TABLE DES MATIÈRES

---

<b>Introduction</b>	<b>7</b>
<b>I Etat de l'art</b>	<b>11</b>
<b>1 Modélisation de quadrupèdes virtuels pour l'animation</b>	<b>13</b>
1 Quadrupèdes : du réel au virtuel . . . . .	14
1.1 Structure anatomique des quadrupèdes . . . . .	14
1.2 Vers un quadrupède virtuel . . . . .	16
2 Modélisation du squelette d'animation . . . . .	17
2.1 Le squelette comme une hiérarchie de repères . . . . .	18
2.2 Représentations des rotations . . . . .	21
2.3 Les cartes exponentielles . . . . .	25
3 Habillage du squelette . . . . .	25
3.1 Interpolation de formes-clés . . . . .	26
3.2 Techniques de skinning . . . . .	28
3.3 Techniques de modélisation anatomique . . . . .	30
4 Bilan . . . . .	32
<b>2 Techniques de contrôle du mouvement</b>	<b>33</b>
1 Caractéristiques du mouvement des quadrupèdes réels . . . . .	33

1.1	Les allures . . . . .	35
2	Animation Cinématique . . . . .	41
2.1	Cinématique directe . . . . .	42
2.2	Cinématique inverse . . . . .	43
3	Animation Physique . . . . .	45
3.1	Animation Dynamique . . . . .	45
3.2	Animation par optimisation . . . . .	46
4	Bilan . . . . .	47
<b>3</b>	<b>Utilisation de données réelles pour l'animation</b>	<b>49</b>
1	Capture de mouvement à l'aide de marqueurs . . . . .	52
2	Extraction de mouvement 3D à partir de vidéos . . . . .	54
2.1	Extraction de formes 3D à partir de vidéos . . . . .	55
2.2	Suivi de mouvement articulé 3D à partir de vidéos . . . . .	56
3	Animation à partir de vidéo . . . . .	59
4	Bilan . . . . .	60
<b>II</b>	<b>Modélisation de squelettes de quadrupèdes</b>	<b>61</b>
<b>4</b>	<b>Reconstruction d'un squelette de quadrupède à partir d'images</b>	<b>63</b>
1	Le modèle de caméra sténopé . . . . .	64
2	Reconstruction d'un squelette 3D de rat à partir d'images . . . . .	66
3	Résultats . . . . .	70
4	Bilan . . . . .	70
<b>5</b>	<b>Construction d'un squelette générique de quadrupèdes à partir d'images</b>	<b>73</b>
1	Contexte . . . . .	74
2	Apprentissage des données . . . . .	75
2.1	Les modèles de squelettes de référence . . . . .	75
2.2	Paramétrisation des données . . . . .	76
2.3	Normalisation . . . . .	77
3	Etude de la dimensionnalité . . . . .	78
4	Paramétrisation géométrique . . . . .	81
5	Résultats et discussion . . . . .	83
5.1	Reconstruction des données . . . . .	83
5.2	Animation d'un nouveau quadrupède . . . . .	86
6	Bilan . . . . .	86

<b>III</b>	<b>Animation de quadrupèdes à partir de vidéos</b>	<b>89</b>
<b>6</b>	<b>Animation de mouvements cycliques à partir de vidéos</b>	<b>91</b>
1	Contexte . . . . .	92
2	Vue d'ensemble de la méthode . . . . .	93
3	Prédiction d'animation 3D à partir d'ACP sur les images. . . . .	94
3.1	Reduction de la variabilité dans des images binaires . . . . .	95
3.2	Description du mouvement par les composantes principales . . . . .	97
3.3	Interpolation . . . . .	98
4	Sélection des images-clés . . . . .	100
4.1	Critère automatique de sélection . . . . .	100
4.2	Résolution des ambiguïtés 2D avec plusieurs modèles . . . . .	101
5	Validation sur des images de synthèse . . . . .	105
6	Application à des séquences vidéos réelles . . . . .	106
6.1	Restriction aux deux premières CP . . . . .	106
6.2	Régularisation du spectre . . . . .	107
6.3	Factorisation des mouvements non cycliques . . . . .	109
7	Résultats . . . . .	110
7.1	Sélection des images-clés : comparaison entre différents animaux . . .	111
7.2	Sélection des images-clés : Comparaison entre images réelles et images de synthèse . . . . .	113
7.3	Sélection des images-clés : Comparaison entre images réelles et croquis	114
8	Bilan . . . . .	115
<b>7</b>	<b>Analyse des séquences de posés à partir de vidéos.</b>	<b>117</b>
1	Contexte . . . . .	118
2	Extraction des posés des pieds à partir de vidéo . . . . .	119
2.1	Détection des pieds sur une vidéo . . . . .	119
2.2	Reconstruction des Trajectoires . . . . .	122
2.3	Extraction des posés . . . . .	122
3	Caractérisation des allures des quadrupèdes par les posés des pieds . . . . .	126
3.1	La méthode d'analyse en séquences . . . . .	126
3.2	Analyse automatique des allures des quadrupèdes . . . . .	127
4	Résultats . . . . .	129
5	Bilan . . . . .	130
	<b>Conclusion</b>	<b>133</b>

<b>A Hiérarchie de repères</b>	<b>137</b>
<b>B Quaternions</b>	<b>141</b>

---

# TABLE DES FIGURES

---

1	Images extraites de <i>l'âge de glace</i> et du <i>monde de Narnia</i> . Que ce soit dans les films d'animation ou les effets spéciaux, les animaux virtuels sont de plus en plus présents. . . . .	8
1.1	A gauche : le megazostrodon, mammifère primitif. A droite : le cheval est historiquement le quadrupède le plus étudié. . . . .	14
1.2	Schéma général de construction d'un mammifère (d'après [Ren94]) . . . . .	15
1.3	Schéma de deux types d'orientation des membres (d'après [Ren94]). Les membres sont de type transversal chez les quadrupèdes plus primitifs comme les amphibiens, et de type parasagittal chez les mammifères. . . . .	16
1.4	exemples de plantigrades (le rat) (a), de digitigrade (le tigre) (b) et d'ongulé (le cheval)(c) (d'après [EDB56], modifié). Notez les positions respectives des articulations du genou (1), du tarse (2), et de la première phalange (3). Ces positions peuvent donner l'illusion que les membres ont un segment de plus que des membres d'humains. Il n'en est rien, seule la longueur et la position des segments change. . . . .	17
1.5	Structure de base d'un quadrupède 3D : Un squelette d'animation (en jaune) entouré d'une surface 3D (en gris). . . . .	18

1.6	Articulations réelles et articulations idéales : (a)-lorsque deux membres d'une articulation réelle sont en mouvement relatif, les centrodes (c'est-à-dire les lieux des centres instantanés de rotation des deux objets) montrent le glissement qui intervient entre les deux membres (image extraite de [KG83]); (b)- en négligeant les translations, l'articulation peut être ramenée à une rotule (3 degrés de liberté) ou à un modèle plus contraint (images extraites de [Mau99].)	19
1.7	Les positions des repères sont relatives. En conséquence, changer l'orientation d'un repère déplacera tous ces repères fils. . . . .	19
1.8	L'organisation en joints, présentée en 2D pour plus de clarté. Un joint est défini par plusieurs transformations dans un ordre précis : d'abord la translation (en vert) , puis une rotation fixe qui définit l'orientation naturelle du joints(en rouge), puis une rotation qui varie au cours du temps et définit le mouvement du joint (en noir) . . . . .	20
1.9	Les limites induites sur l'amplitudes de quelques articulations par la morphologie du squelette. . . . .	22
1.10	Les angles d'Euler : exemple de rotation selon la convention XYZ. Trois rotations successives sont effectuées : d'abord le repère $(x, y, z)$ est tourné d'un angle $\alpha$ autour de l'axe $x$ . On obtient le repère $(x, y', z')$ . Ensuite on tourne d'un angle $\beta$ autour de l'axe $y'$ et on obtient le repère $(x', y', Z)$ . Enfin on tourne d'un angle $\gamma$ autour de l'axe $Z$ et on obtient le nouveau repère $(X, Y, Z)$ . . . . .	23
1.11	. Toute rotation 3D peut être représentée par une seule rotation d'angle $\theta$ et d'axe $u$ , où $u$ est un vecteur 3D unitaire. . . . .	24
1.12	Habillage d'un squelette d'animation par une chaîne de solides. . . . .	26
1.13	La texture est une image plaquée sur le maillage 3D. A chaque sommet 3D du maillage est associé une position dans l'image, ou coordonnée de texture. . . .	27
1.14	En haut : des poses-clés choisies pour un maillage de lion. En bas : un champ de déformation appris à partir des exemples permet d'interpoler entre les formes-clés pour déformer de façon réaliste le maillage en n'en déplaçant que certains sommets (d'après [SZGP05]). . . . .	28
1.15	La technique du skinning rigide : chaque point du maillage se déplace avec le joint auquel il est attaché. Le maillage est étiré d'un côté de l'articulation et les morceaux s'intersectent de l'autre côté. Si le maillage a été découpé en membres, des trous peuvent même apparaître (image extraite de [Blo02]). . . .	28

1.16	Illustration de la technique du skinning avec deux repères. Le point obtenu $P'$ est une combinaison linéaire des positions qu'il aurait si il était influencé seulement par le premier repère ( $P_{vert}$ ) ou par le second ( $P_{rouge}$ ). $P'$ se trouve obligatoirement sur la droite reliant $P_{vert}$ à $P_{rouge}$ . . . . .	29
1.17	Problèmes liés à la technique du skinning : (a)-aplatissement au niveau du coude ; (b)-amincissement au niveau de l'articulation quand le bras est vrillé. . . . .	30
1.18	Le modèle anatomique de Wilhelms et al. comprend le squelette (en blanc), les muscles (en rouge), des tissus non musculaires (en violet) et enfin la peau (en bas à droite). . . . .	31
1.19	Pratscher et al.[PCLS05] adaptent un modèle générique de muscles à l'intérieur d'un maillage 3D muni d'un squelette d'animation. . . . .	31
2.1	A gauche : ordre des posés pour la séquence latérale et la séquence diagonale. A droite : lorsque la patte la plus en avant (celle marquée d'un double cercle) vient d'être posée, la séquence latérale donne un meilleur équilibre statique que la séquence diagonale. Les pieds sont notés RF pour right front (avant droit), LF pour left front (avant gauche), RR pour right rear (arrière droit), et LR pour left rear (arrière gauche). . . . .	35
2.2	Exemples de représentations d'une allure de manière visuelle. Une représentation doit inclure la coordination entre les membres, donc les successions leur levés et posés. <i>En haut</i> : On représente l'allure comme une succession de phases d'appui ou de suspension. A chaque phase, une combinaison différente de membres est en contact avec le sol. La flèche représente la tête de l'animal. Pour chaque phase, les disques noirs désignent les pattes qui sont au sol, et les phases se succèdent dans le temps de gauche à droite. <i>En bas</i> : Le temps se déroule de gauche à droite. Pour chaque membre, les segments en traits pleins désignent les moments où le membre est au sol et les segments en pointillés désignent les moments où les membres sont levés. A noter que cette notation est plus riche car elle permet d'exprimer la longueur des phases. Les segments en gras désignent les posés qui font partie d'une même <i>séquence</i> , comme nous le définirons au chapitre 7. . . . .	37
2.3	Allure symétrique : le décalage $\Delta$ entre les posés des pattes avant est le même qu'entre les posés des pattes arrières. . . . .	37
2.4	La marche. . . . .	38
2.5	L'amble. . . . .	38
2.6	Le trot. . . . .	38
2.7	Le rack. . . . .	39

2.8	Allure asymétrique : le décalage entre les posés des pattes avant n'est pas le même qu'entre les posés des pattes arrières. . . . .	39
2.9	Le canter. . . . .	40
2.10	Le galop transverse. . . . .	40
2.11	Le galop rotatoire. . . . .	40
2.12	Le demi-bond. . . . .	41
2.13	Le bond. . . . .	41
2.14	Cinématique directe : on définit le positionnement des éléments d'une chaîne articulée en réglant les orientations de chaque articulation. Amener l'extrémité de la chaîne à un endroit précis est difficile. . . . .	42
2.15	Animation par position-clés. Les valeurs des trois rotations du joint sont définies à quelques positions dans le temps (les carrés sur les courbes sont les clés). Ensuite toutes les valeurs intermédiaires sont interpolées. . . . .	43
2.16	Cinématique inverse : Le mouvement de la chaîne articulée est contrôlé par les positions de ses extrémités. Les valeurs de rotations de toutes les articulations intermédiaires sont calculées en fonctions de ces positions. . . . .	44
2.17	. Le squelette normalisé introduit par Kulpa et al. [KMA05] . . . . .	45
2.18	Illustration de la méthode de Liu & Popović [LP02]. En haut : l'animation simplifiée donnée par l'utilisateur. En bas : le résultat. . . . .	47
3.1	Une des planches réalisées par Muybridge : Plusieurs appareils photo sont déclenchés par des fils tendus sur le chemin de l'animal. La phase de suspension est clairement visible sur les premières images. . . . .	50
3.2	Quelques planches réalisées par Marey : Avec un sujet clair sur fond foncé, la même plaque est exposée plusieurs fois. . . . .	51
3.3	Le Rotoscope (Image extraite du brevet de 1914). . . . .	51
3.4	Un acteur muni de marqueurs. . . . .	52
3.5	Un animal muni de marqueurs. . . . .	55
3.6	D'après[BHB99]. A partir d'un suivi de points, la forme 3D de la girafe est extraite(au centre), ainsi que les modes principaux de variation de la forme (à droite) . . . . .	56
3.7	Ramanan et al. [RF03] reconstruisent des modèles articulés simples d'animaux à partir de vidéo. . . . .	56
3.8	Suivi 3D à l'aide d'un modèle 3D a priori : à chaque instant, on cherche la pose du modèle 3D qui maximise la corrélation avec l'image (D'après [ST03], modifié). . . . .	57

3.9	Représentation schématique d'un espace de recherche. Le présence de minima locaux rend la recherche du minimum global difficile. . . . .	58
3.10	Le flot optique donne le mouvement local en 2D pour chaque position sur l'image. . . . .	59
3.11	D'après Bregler et al.[BLCD02]. Chaque image de la vidéo est décrite comme une combinaison d'images-clés. L'utilisateur donne une pose-clé 3D qui correspond à chacune des images-clés, et les poses-clés sont alors interpolées pour créer de l'animation 3D à partir de vidéo. . . . .	60
4.1	Le modèle de caméra sténopé, présenté ici en 2D. Tous les rayons convergent vers le foyer $F$ . Contrairement au appareils sténopé réels, les points ne sont pas projetés sur un écran $E_r$ situé derrière le foyer, mais sur une écran virtuel $E$ situé entre le foyer et la scène, à une distance du foyer égale à la distance focale $f_x$ est la distance focale de la caméra. Le point de coordonnées $(x, z)$ est projeté sur l'écran $E$ en un point 1D de coordonnées homogènes $(u, w)$ . . . . .	65
4.2	Une mire de calibration. . . . .	66
4.3	Une photographie du squelette réel de rat. . . . .	67
4.4	La donnée de la projection du point 3D par une caméra contraint le point 3D à être sur une droite : la droite reliant le foyer de la caméra à la projection du point 3D sur l'écran. A partir de la donnée de 2 projections, on peut estimer la position du point 3D comme étant la position 3D la plus proche des droites. . . . .	68
4.5	La caméra est immobile, et le rat est placé sur un plateau qui peut pivoter sur lui-même. . . . .	70
4.6	Le modèle 3D de rat reconstruit, vu depuis une caméra virtuelle correspondant aux paramètres de caméra extraits des images. Une fois le modèle superposé aux images d'origine, on peut apprécier visuellement la qualité de la reconstruction. . . . .	71
5.1	Construction des données d'apprentissage - l'exemple de la vache . . . . .	76
5.2	Distribution des composantes principales pour les quatre paramétrisations . . . . .	78
5.3	Les trois premiers modes de variation de l'ACP sur les squelettes. Étonnamment, ces modes sont faciles à identifier intuitivement comme la hauteur de l'animal au repos, l'inclinaison de la colonne vertébrale, et un paramètre ongulé/plantigrande. . . . .	80
5.4	Les trois mesures géométriques qui contrôlent notre modèle déformable. . . . .	81
5.5	Les trois premiers modes de variation du modèle déformable, contrôlés par la hauteur de l'animal, l'inclinaison de la colonne vertébrale, et le paramètre ongulé/plantigrade (angle du pied). . . . .	83

5.6	Les trois premiers modes de variation du modèle déformable, contrôle par la hauteur de l'animal, l'inclinaison de la colonne vertébrale, et la longueur du cou.	84
5.7	La moyenne et l'écart-type de l'erreur de prédiction pour les 4 conditions : translation (à gauche) et rotation (à droite) en utilisant le modèle contrôlé par ACP. . . . .	85
5.8	La moyenne et l'écart-type de l'erreur de prédiction pour les 4 conditions : translation (à gauche) et rotation (à droite) en utilisant notre modèle déformable.	85
5.9	Comparaison entre les conditions $QR \times LC$ (à gauche) et $QR \times GC$ (à droite) dans le cas de l'éléphant. La référence est en violet et le squelette prédit est en bleu. . . . .	86
5.10	Animation d'un nouveau quadrupède : ajustement du squelette, habillage, rendu final. . . . .	87
6.1	Vue d'ensemble de la méthode . . . . .	94
6.2	Les résultats de segmentation obtenus pour nos trois sources de données : vidéo, croquis et images de synthèse. . . . .	95
6.3	PC1(t)xPC2(t) : Les projections sur les CP au cours du temps pour deux séquences synthétiques du cheval : le canter et la marche. Les images sont numérotées pour un cycle. Les cercles sont les images-clés choisies par le critère du conditionnement. . . . .	102
6.4	Deux poses différentes peuvent donner des silhouettes similaires. . . . .	102
6.5	Sélection d'images-clés supplémentaires pour résoudre les ambiguïtés dans l'espace d'ACP. Les images 14, 17, 21 and 24 sont sélectionnées comme images-clés initiales. Dans une deuxième étape, les images 27, 4, 8 et 11 sont sélectionnées comme candidats pour les poses ambiguës, à partir de leurs coordonnées dans l'espace d'ACP. Première ligne : images 14, 17, 21 24 ; Deuxième ligne : images 27, 4, 8, 11. . . . .	104
6.6	Évaluation pour les séquences du galop et du canter. Chaque courbe correspond à l'erreur moyenne pour un nombre fixe d'exemples (entre 2 et 5), en fonction du nombre de composantes utilisées en entrée. . . . .	106
6.7	Les variations capturées par les trois premiers vecteurs propre pour la séquence du cheval au galop. La colonne du milieu est l'image moyenne, chaque ligne correspond à la variation selon un vecteur propre. . . . .	107
6.8	Les variations encodées par les trois premiers vecteurs propres pour la séquence du guépard. . . . .	108

6.9	Le spectre des 2 premières CP du spectre du guépard. La fonction de pic est ajustée sur PC1, et une fenêtre rectangulaire est déduite. . . . .	109
6.10	Les variations encodées par les trois premiers vecteurs propre de la séquence de marche du rhinocéros. . . . .	110
6.11	Évolution au cours du temps des trois premières composantes principales de la séquence de marche du rhinocéros. . . . .	111
6.12	Evolution des poids des 4 exemples pour la séquence du guépard(haut), et des 8 exemples pour la séquence du rhinocéros (en bas). La ligne en pointillés est la somme des poids. . . . .	112
6.13	Selection d'images-clés pour plusieurs séquences vidéo. . . . .	113
6.14	Les images-clés sélectionnées pour la marche du cheval, d'après une séquence réelle (haut), et une séquence synthétique (bas). . . . .	114
6.15	Images-clés sélectionnées pour la séquence de la marche de l'éléphant, en utilisant des croquis dessinés par l'utilisateur sur les images (en haut) et la segmentation directe d'images (en bas). Nous ne montrons pas tous les croquis, mais ils ont été dessiné par-dessus chaque image. . . . .	114
7.1	Torkos et al. [TvdP98] animent des quadrupèdes à partir de données de posés. .	118
7.2	A gauche : Sur une vidéo binarisée, les pieds exhibent une forme très caractéristique qu'on peut définir par "l'extrémité d'une forme longiligne". A droite : Représentation du filtre utilisé pour détecter des régions susceptibles d'être des pieds. Le filtre détermine si une région de la silhouette de l'animal (en bleu) à la forme caractéristique d'un pied. La réponse est positive si la Zone 1 (en jaune) ne contient aucun pixel de la silhouette de l'animal, et si la Zone 2 (en rouge) en contient au moins un. Les paramètres <i>XSize</i> et <i>YSize</i> sont définis manuellement par l'utilisateur pour chaque séquence à traiter. La rotation <i>Theta</i> ainsi que la position en 2D ( <i>X,Y</i> ) du filtre sur l'image sont déterminées lors de la recherche. Les dimensions sont données en pixels. la position du pied estimée est le milieu de l'intersection entre la silhouette de l'animal et la partie rouge du filtre (le pixel marqué d'une croix). . . . .	120
7.3	Les candidats sont recherchés de bas en haut dans chaque colonne de l'image. .	121
7.4	(a) candidats. Pour un même pied, plusieurs candidats peuvent exister (b) après groupement, un filtre est sélectionné par pied . (c) La position détectée des pieds.	121
7.5	Les position horizontales(à gauche) et verticales(à droite) des pieds au cours du temps après application du filtre. . . . .	122

- 7.6 Illustration de la méthode utilisée pour reconstruire les trajectoires. Notre méthode est appliquée à des trajectoires en 2D, mais elle est représentée ici en 1D pour plus de clarté. Les cercles représentent les position de pieds issues de l'étape précédente. Les trajectoires sont propagées au cours du temps, en partant du premier pas de temps. La figure représente le fonctionnement de l'algorithme au temps  $t$ . En traits pleins, les trajectoires déjà reconstruites pour l'intervalle de temps  $[0, t]$ . Les trajectoires sont extrapolées linéairement (traits en pointillés) pour estimer les positions des pieds au temps  $t + 1$ . Chaque position mesurée au temps  $t + 1$  est alors rattachée à la trajectoire dont la position estimée est la plus proche. Ensuite on peut passer au pas de temps suivant. . . . . 123
- 7.7 Les trajectoires de chaque pied après reconstruction, selon la coordonnée  $x$  (à gauche), et  $y$  (à droite). . . . . 124
- 7.8 Représentation schématique du mouvement en 2D d'un pied. Comme les positions en  $y$  sont beaucoup plus bruitées que les positions en  $x$ , les posés sont calculés à partir des positions extrémales en  $x$  . . . . . 124
- 7.9 Séquences de posés obtenues par notre méthode pour une marche (en haut) et pour une séquence de transition (en bas). Chaque colonne correspond à un pas de temps, chaque ligne à un pied (RF : Avant-droit, LF : Avant-gauche, RR : Arrière-droit, LR : Arrière-gauche). Les cases colorées correspondent aux moments où les pieds sont au sol. . . . . 125
- 7.10 A gauche : la méthode classique analyse les mouvements en les découpant en intervalles de temps qui commencent par les posés du même pied. A droite : dans l'analyse en séquences, un cycle de mouvement est la succession d'un posé des pattes antérieures puis d'un posé des pattes postérieures. . . . . 126
- 7.11 Analyse en séquences. A : Les débuts des cycles sont les posés du membre *trailing antérieur*, la patte avant qui se pose en premier. B : Diagramme d'allure. L'axe des abscisses représente le temps. Les marques de couleur (orange et rouge pour la paire antérieure et vertes pour la paire postérieure) représentent les temps de contact entre chaque membre et le sol. Chaque séquence est identifiée par les flèches grises. D : Mesure des décalages temporels. Pour chaque séquence, le décalage temporel entre les poser des deux antérieurs (Fore lag, FL), le décalage temporel entre les poser des deux postérieurs (Hind lag, HL) et le décalage temporel entre les poser des deux membres trailing (Pair lag, PL), sont mesurés et exprimés en pourcentage de la durée du cycle du TA. . . . 128

7.12	Calcul des pair lags et recombinaison des séquences à partir de la donnée d'une suite de posés. (1) On regroupe les actions intra-paires. Chaque posé du trailing antérieur est associé avec le posé du leading antérieur qui lui succède, et chaque posé du trailing postérieur est associé avec le posé du leading postérieur qui lui est le plus proche dans le temps. (2) On regroupe les actions des différentes paires. Chaque posé du trailing antérieur est associé avec le posé du trailing postérieur qui lui succède (3) On a décomposé le mouvement en séquences et on peut calculer les pair lags, donc l'allure. . . . .	130
A.1	La hiérarchie de repère que nous utilisons dans cette thèse. Comme c'est le cas le plus souvent, le pelvis est choisi comme nœud père . . . . .	138
A.2	La hiérarchie de repère que nous utilisons dans cette thèse. Comme c'est le cas le plus souvent, le pelvis est choisi comme nœud père . . . . .	139
A.3	La hiérarchie de repères orientés en 3D obtenue avec nos conventions. . . . .	140



---

# LISTE DES TABLEAUX

---

5.1	Les écarts-types les plus grands rencontrés dans nos exemples, pour tous les paramètres de rotations et de translation et pour les 4 paramétrisations testées. . . . .	77
6.1	Nombre d'images et taille des images pour quelques séquences de test. . . . .	97
6.2	pourcentage de la variance couverte par chaque composante principale par rapport à la variance totale, pour sept séquences de test. . . . .	98
6.3	Coefficients de corrélation entre les signaux filtrés et non filtrés pour les cinq premières composantes, calculés pour quelques séquences de test. . . . .	109
6.4	Indices des images-clés sélectionnées pour la séquence de la marche de l'éléphant, en utilisant les croquis et la segmentation d'images. . . . .	114
7.1	Toutes les allures peuvent être identifiées à l'aide de seulement trois paramètres : le décalage temporel entre les poser des deux antérieurs (Fore lag, F lag), le décalage temporel entre les poser des deux postérieurs (Hind lag, H lag) et le décalage temporel entre les poser de deux membres homolatéraux (Pair lag, P lag). Les décalages temporels sont exprimés en pourcentage de la durée du cycle du TA. . . . .	129
7.2	lags et allures calculés automatiquement pour quelques séquences synthétiques	129
7.3	lags et allures calculés automatiquement pour la séquence de marche de l'antilope	130
7.4	lags et allures calculés automatiquement pour la séquence de marche du cheval	130
7.5	lags pour la séquence du cerf. Les transitions sont retrouvées au bon endroit . . .	131



---

# BIBLIOGRAPHIE

---

- [ABD83] Z. Afelt, J.W. Blaszczyk, et Cz. Dobrzecka. Speed control in animal locomotion : transitions between symmetrical and asymmetrical gaits in the dog. *Acta Neurobiologiae Experimentalis*, 45 :233–250, 1983.
- [Abo03] A. Abourachid. A new way of analysing symmetrical and asymmetrical gaits in quadrupeds. *C. R. Biologies*, 326 :625–630, 2003.
- [ACP03] B. Allen, B. Curless, et Z. Popović. The space of all body shapes : reconstruction and parameterization from range scans. *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH 2003)*, 22(3) :587–594, 2003.
- [AG85] W.W. Armstrong et M. Green. The dynamics of articulated rigid bodies for the purpose of animation, 1985.
- [AHH<sup>+</sup>] A. Abourachid, M. Herbin, R. Hackert, L. Maes, et V. Martin. Experimental study of coordination patterns during unsteady locomotion in mammals, soumis au j. expt. biol.
- [AM00] M. Alexa et W. Müller. Representing animation by principal components. In *Proceeding of EUROGRAPHICS 2000*, volume 19, 2000.
- [ASBP00] J.V. Albro, G.A. Sohl, J.E. Bobrow, et F.C. Park. On the computation of optimal high-dives. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '00)*, volume 4, pages 3958–3963, 2000.
- [Bae01] P. Baerlocher. *Inverse Kinematics Techniques for the Interactive Posture Control of Articulated Figures*. PhD thesis, Ecole Polytechnique Fédérale de Lausanne, 2001.
- [BHB99] C. Bregler, A. Hertzmann, et H. Biermann. Recovering non-rigid 3d shape from image streams. In *CVPR99*, 1999.

- [BKK<sup>+</sup>85] N.I. Badler, J.D. Korein, J.U. Korein, G.M. Radack, et L.S. Brotman. Positioning and animating human figures in a task-oriented environment. *The Visual Computer*, 1(4) :212–220, Décembre 1985.
- [Bla01] J. W. Blaszczyk. Gait transitions during unrestrained locomotion in dogs. *Equine Veterinary Journal. Supplement.*, 33 :112–115, 2001.
- [BLCD02] C. Bregler, L. Loeb, E. Chuang, et H. Deshpande. Turning to the masters : motion capturing cartoons. In *Proceedings of SIGGRAPH'02*, pages 399–407, 2002.
- [Blo02] J. Bloomenthal. Medial-based vertex deformation. In *Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation (SCA '02)*, pages 147–151, 2002.
- [BM98] C. Bregler et J. Malik. Tracking people with twists and exponential maps. In *Proceedings of Conf. Computer Vision and Pattern Recognition 1998*, pages 8–15, 1998.
- [BMP] C. Bregler, J. Malik, et K. Pullen. Twist based acquisition and tracking of animal and human kinematics. *Int. Journal of Computer Vision*, 56(3).
- [BMW87] N.I. Badler, K.H. Manoochehri, et G. Walters. Articulated figure positioning by multiple constraints. *IEEE Computer Graphics and Applications*, 7(6) :28–38, Juin 1987.
- [BV99] V. Blanz et T. Vetter. A morphable model for the synthesis of 3d faces. In *Proceedings of SIGGRAPH '99*, pages 187–194, 1999.
- [Cal75] W.F. Calderon. *Animal painting and anatomy*. Dover Publications, Inc., New-York, 1975.
- [CB04] B. Le Calennec et R. Boulic. Interactive motion deformation with prioritized constraints. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation (SCA '04)*, pages 163–171, 2004.
- [CGC<sup>+</sup>02] S. Capell, S. Green, B. Curless, T. Duchamp, et Z. Popović. Interactive skeleton-driven dynamic deformations. In *SIGGRAPH '02 : Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, 2002.
- [CK98] J.P. Costeira et T. Kanade. A multibody factorization method for independently moving objects. *Int. J. Comput. Vision*, 29(3) :159–179, 1998.
- [CLS02] M. Cartmill, P. Lemelin, et D. Schmitt. Support polygons and symmetrical gaits in mammals. *Zoological Journal of the Linnean Society*, (136) :401–420, 2002.
- [CS95] L.S. Crawford et S.S. Sastry. Biological motor control approaches for a planar driver. In *Proceedings of IEEE Conference on Decision and Control*, pages 3881–3886, 1995.
- [CZ92] David T. Chen et David Zeltzer. Pump it up : computer animation of a biomechanically based model of muscle using the finite element method. In *SIGGRAPH '92 : Proceedings of the 19th annual conference on Computer graphics and interactive techniques*, 1992.
- [DAC<sup>+</sup>03] J. Davis, M. Agrawala, E. Chuang, Z. Popović, et D. Salesin. A sketching interface for articulated figure animation. In *Proceedings of Eurographics/SIGGRAPH Symposium of Computer Animation (SCA'03)*, 2003.

- [dGITT03] V. di Gesù, F. Isgro, D. Tegolo, et E. Trucco. Finding essential features for tracking star fish in a video sequence. In *Proceedings of CIAP03*, pages 504–509, 2003.
- [DN99] A. Dasgupta et Y. Nakamura. Making feasible walking motion of humanoid robots from human motion capture data. In *Proceedings of the IEEE International Conference on Robotics and Automation*, 1999.
- [DRE06] J. Diener, L. Reveret, et E. Fiume. Hierarchical retargetting of 2d motion fields to the animation of 3d plant models. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA'06)*, 2006.
- [DSJ06] K.G. Der, R.W. Sumner, et J. Popović. Inverse kinematics for reduced deformable models. *ACM Transactions on Graphics*, 25(3), 2006.
- [Dun02] F. Dunn. *3D Math Primer for Graphics and Game Development*. Wordware Publishing, Inc., 1st edition, 2002.
- [EDB56] W. Ellenberg, H. Dittich, et H. Baum. *An atlas of animal anatomy for artists*. Lewis S. Brown (eds), Dover Publications, Inc., New-York, 1956.
- [EL82] A.W. English et P.R. Lennard. Interlimb coordination during stepping in the cat : in-phase stepping and gait transitions. *Brain research*, 245 :353–364, 1982.
- [Eva03] F.H. Evans. Detecting fish in underwater video using the em algorithm. In *Proceedings of ICIP03*, pages III : 1029–1032, 2003.
- [FB02] R. Fablet et M.J. Black. Automatic detection and tracking of human motion with a view-based representation. In *ECCV*, volume 1, pages 476–491, 2002.
- [Feh00] G. Feher. *Grand Cours d'Anatomie Artistique*. Könemann, 2000.
- [FP03] Anthony C. Fang et Nancy S. Pollard. Efficient synthesis of physically valid human motion. *ACM Trans. Graph.*, 22(3), 2003.
- [FRDC04] L. Favreau, L. Reveret, C. Depraz, et M.-P. Cani. Animal gaits from video. In *Proceedings of EG/SIGGRAPH Symposium on Computer Animation (SCA'04)*, 2004.
- [FRDC06] L. Favreau, L. Reveret, C. Depraz, et M.-P. Cani. Animal gaits from video : comparative studies. *Graphical Models (GMOD)*, 68(2) :212–234, Mars 2006.
- [GCH<sup>+</sup>00] S. Grillner, L. Cangiano, G.-Y. Hu, R. Thompson, R. Hill, et P. Wallén. The intrinsic function of a motor system - from ion channels to networks and behaviour. *Brain research*, 886 :224–236, 2000.
- [GF02] M. Gleicher et N. Ferrier. Evaluating video-based motion capture. In *Proceedings of Computer Animation 2002*, Juin 2002.
- [GG03] R. Green et L. Guan. Quantifying and recognizing human movement patterns from monocular video images - part i : A new framework for modeling human motion. *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image and Video-Based Biometrics*, Novembre 2003.
- [GL93] F. Glover et M. Laguna. Tabu search. In *Modern Heuristic Techniques for Combinatorial Problems*, 1993.

- [Gle97] M. Gleicher. Motion editing with spacetime constraints. In *Proceedings of the 1997 symposium on Interactive 3D graphics*, pages 139–ff., 1997.
- [Gle98] M. Gleicher. Retargetting motion to new characters. In *Proceedings of SIGGRAPH '98*, pages 33–42, 1998.
- [GM85] M. Girard et A.A. Maciejewski. Computational modeling for the computer animation of legged figures. *Computer Graphics*, 19(3) :263–270, Juillet 1985.
- [GMHP04] K. Grochow, S. L. Martin, A. Hertzmann, et Z. Popović. Style-based inverse kinematics. *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH 2004)*, 2004.
- [Gra98] F. S. Grassia. Practical parameterization of rotations using the exponential map. *J. Graph. Tools*, 3(3) :29–48, 1998.
- [HCG05] S.L. Hannuna, N.W. Campbell, et D.P. Gibson. Identifying quadruped gait in wildlife video. In *Proceedings of ICIP05*, volume I, pages 713–716, 2005.
- [Hil65] M. Hildebrand. Symmetrical gaits of horses. *Science*, 5 :701–708, 1965.
- [Hil66] M. Hildebrand. Analysis of the symmetrical gaits of tetrapods. *Folia Biotheoretica*, 6 :9–22, 1966.
- [Hil89] M. Hildebrand. The quadrupedal gaits of vertebrates. *Bioscience*, 39(11) :766–775, 1989.
- [HML00] G. Huang, D. Metaxas, et J. Lo. Human motion planning based on recursive dynamics and optimal control techniques. In *Proceedings of the International Conference on Computer Graphics (CGI '00)*, page 19, 2000.
- [Hod96] J.K. Hodgins. Three-dimensional human running. In *Proceedings of IEEE International Conference on Robotics and Automation*, 1996.
- [HP97] R. Hartley et P. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2) :146–157, 1997.
- [HS81] B.K.P. Horn et B.G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1-3) :185–203, 1981.
- [HWBO95] J.K. Hodgins, W.L. Wooten, D.C. Brogan, et J.F. O'Brien. Animating human athletics. In *Proceedings of SIGGRAPH '95*, pages 71–78, 1995.
- [IC87] P.M. Isaacs et M.F. Cohen. Controlling dynamic simulation with kinematic constraints. In *Proceedings of SIGGRAPH '87*, pages 215–224, 1987.
- [ICC05] A.J. Ijspeert, A. Crespi, et J.M. Cabelguen. Simulation and robotics studies of salamander locomotion : applying neurobiological principles to the control of locomotion in robots. *Neuroinformatics*, 3(3) :171–195, 2005.
- [JG97] J. Wilhelms et A. Van Gelder. Anatomically based modeling. In *Computer Graphics (ACM Siggraph Proceedings)*, 1997.
- [JP02] D.L. James et D.K. Pai. Dyr : dynamic response textures for real time deformation simulation with graphics hardware. In *SIGGRAPH '02 : Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, 2002.

- [KG83] G. L. Kinzel et L. J. Gutkowski. Joint models, degrees of freedom, and anatomical motion measurement. *Journal of Biomechanical Engineering*, 105 :55–61, 1983.
- [KG04] L. Kovar et M. Gleicher. Automated extraction and parameterization of motions in large data sets. *ACM Trans. Graph.*, 23(3) :559–568, 2004.
- [KGP02] L. Kovar, M. Gleicher, et F. Pighin. Motion graphs. In *Proceedings of SIGGRAPH '02*, pages 473–482, 2002.
- [KGV83] S. Kirkpatrick, C. D. Gelatt, et M. P. Vecchi. Optimization by simulated annealing. *Science*, Number 4598, 13 May 1983, 220, 4598 :671–680, 1983.
- [KJP02] P.G. Kry, D.L. James, et D.K. Pai. Eigenskin : real time large deformation character skinning in hardware. In *Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation (SCA '02)*, pages 153–159, 2002.
- [KMA05] R. Kulpa, F. Multon, et B. Arnaldi. Morphology-independent representation of motions for interactive human-like animation. *Computer Graphics Forum*, 3(24) :343–352, 2005.
- [Ku105] R. Kulpa. *Adaptation interactive et performante des mouvements d'humanoïdes synthétiques : aspects cinématique, cinétique et dynamique*. PhD thesis, SIAMES - INSA de Rennes, 2005.
- [LC94] Z. Liu et M. Cohen. Decomposition of linked figure motion : Diving. In *5th Eurographics Workshop on Animation and Simulation*, 1994.
- [LC95] Z. Liu et M.F. Cohen. Keyframe motion optimization by relaxing speed and timing. In *Proceedings of Computer Animation and Simulation '95*, pages 144–153, 1995.
- [LCF00] J.P. Lewis, M. Cordner, et N. Fong. Pose space deformation : A unified approach to shape interpolation and skeleton-driven deformation. In *Proceedings of SIGGRAPH'00*, pages 165–172, 2000.
- [Lef05] S. Lefebvre. *Modèles d'habillage de surface pour la synthèse d'images*. PhD thesis, Université Joseph Fourier, 2005.
- [LG05] Z. Liu et P. Grother. The humanoid gait challenge problem : Data sets, performance, and analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(2) :162–177, 2005.
- [LK81] B.D. Lucas et T. Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI81*, pages 674–679, 1981.
- [LP02] C.K. Liu et Z. Popović. Synthesis of complex dynamic character motion from simple animations. In *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH 2002)*, pages 408–416, 2002.
- [LrN98] T. Lauthelier et M. rc Neveu. Mélange de formes pour l'animation de personnages. application en production. In *6èmes journées du groupe de travail animation, simulation, systèmes dynamiques*, pages 38–51, Octobre 1998.
- [LTW95] Y. Lee, D. Terzopoulos, et K. Walters. Realistic modeling for facial animation. *Computer Graphics (Proceedings of SIGGRAPH'95)*, 29 :55–62, 1995.

- [Mau99] W. Maurel. *3D Modeling of the Human Upper Limb Including the Biomechanics of Joints, Muscles and Soft Tissues*. PhD thesis, Ecole Polytechnique Fédérale de Lausanne, 1999.
- [MBBT00] J.-S. Monzani, P. Baerlocher, R. Boulic, et D. Thalmann. Using an intermediate skeleton and inverse kinematics for motion retargeting. *Computer Graphics Forum*, 19(3), 2000.
- [MG01] T.B. Moeslund et E. Granum. A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding : CVIU*, 81(3) :231–268, 2001.
- [MK85] A.A. Maciejewski et C.A. Klein. Obstacle avoidance for kinematically redundant manipulators in dynamically varying environments. *The International Journal of Robotics Research*, 4(3) :109–117, 1985.
- [MM99] Alberto Menache et Alberto Manache. *Understanding Motion Capture for Computer Animation and Video Games*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999.
- [Muy57] E. Muybridge. *Animals in Motion*. Dover Publications, 1957.
- [OSBH00] D. Ormoneit, H. Sidenbladh, M.J. Black, et T. Hastie. Learning and tracking cyclic human motion. In *NIPS*, pages 894–900, 2000.
- [PCLS05] M. Pratscher, P. Coleman, J. Laszlo, et K. Singh. Outside-in anatomy based character rigging. In *SCA '05 : Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation*, 2005.
- [Per01] P. Perner. Motion tracking of animals for behavior analysis. In *Proceedings of VF01*, page 779 ff., 2001.
- [PKC<sup>+</sup>03] H. Pyun, Y. Kim, W. Chae, H.W. Kang, et S.Y. Shin. An example-based approach for facial expression cloning. In *Proceedings of Eurographics/SIGGRAPH Symposium on Computer Animation 2003*, pages 167–176, 2003.
- [PR01] N.S. Pollard et P.S.A. Reitsma. Animation of humanlike characters : Dynamic motion filtering with a physically plausible contact model. In *Yale Workshop on Adaptive and Learning Systems*, 2001.
- [PW99] Z. Popović et A. Witkin. Physically based motion transformation. In *proceedings of SIGGRAPH 99*, pages 11–20, 1999.
- [RBC98] C. Rose, B. Bodenheimer, et M.F. Cohen. Verbs and adverbs : Multidimensional motion interpolation using radial basis functions. *IEEE Computer Graphics and Applications*, 18(5) :32–40, Septembre 1998.
- [RE01] L. Reveret et I. Essa. Visual coding and tracking of speech related facial motion. In *Proc. of the IEEE CVPR International Workshop on Cues in Communication*, 2001.
- [Ren94] S. Renous. *Locomotion*. Dunod, 1994.
- [RF03] D. Ramanan et D. Forsyth. Using temporal coherence to build models of animals. In *ICCV*, 2003.
- [RFDC05] L. Reveret, L. Favreau, C. Depraz, et M.-P. Cani. Morphable model of quadrupeds skeletons for animating 3d animals. In *ACM-SIGGRAPH/EG Symposium on Computer Animation (SCA)*, Juiller 2005.

- [RFK06] D. Ramanan, D.A. Forsyth, et K. Barnard. Building models of animals from video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8) :1319–1334, Août 2006.
- [RGBC96] C. Rose, B. Guenter, B. Bodenheimer, et M.F. Cohen. Efficient generation of motion transitions using spacetime constraints. In *Proceedings of SIGGRAPH '96*, pages 147–154, 1996.
- [SBF00] H. Sidenbladh, M.J. Black, et D.J. Fleet. Stochastic tracking of 3d human figures using 2d image motion. In *ECCV (2)*, pages 702–718, 2000.
- [SHP04] A. Safonova, J.K. Hodgins, et N.S. Pollard. Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces. *ACM Trans. Graph.*, 23(3) :514–521, 2004.
- [SIC01] P.-P. J. Sloan, C.F. Rose III, et M.F. Cohen. Shape by example. In *Proceedings of I3D'01*, pages 135–143, 2001.
- [SJ04] R.W. Sumner et J. Popović. Deformation transfer for triangle meshes. *ACM Transactions on Graphics (SIGGRAPH'04)*, 23(3), Août 2004.
- [SM00] J. Shi et J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8) :888–905, 2000.
- [SM01] H.C. Sun et D.N. Metaxas. Automating gait generation. In *SIGGRAPH '01 : Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, 2001.
- [SS88] L. Sciavicco et B. Siciliano. A solution algorithm to the inverse kinematic problem for redundant manipulators. *IEEE Journal of Robotics and Automation*, 4 :403–410, 1988.
- [ST03] C. Sminchisescu et B. TRIGGS. Estimating articulated human motion with covariance scaled sampling. *Int. J. Robotics Research*, 22(6) :371–391, Juin 2003.
- [ST05] C. Sminchisescu et B. Triggs. Building roadmaps of minima and transitions in visual models. *Int. J. Comput. Vision*, 61(1) :81–101, 2005.
- [SWG02] M. Simmons, J. Wilhelms, et A. Van Gelder. Model-based reconstruction for creature animation. In *Proceedings of the ACM SIGGRAPH symposium on Computer animation*, pages 139–146, 2002.
- [SZGP05] R.W. Sumner, M. Zwicker, C. Gotsman, et J. Popović. Mesh-based inverse kinematics. *ACM Trans. Graph.*, 24(3) :488–495, 2005.
- [TBHF03] J. Teran, S. Blemker, V. Ng Thow Hing, et R. Fedkiw. Finite volume methods for the simulation of skeletal muscle. In *SCA '03 : Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*, 2003.
- [TGB00] D. Tolani, A. Goswami, et N.I. Badler. Real-time inverse kinematics techniques for anthropomorphic limbs. *Graphical Models*, 62(5) :353–388, Septembre 2000.
- [TK02] D. Tsutsumi et Y. Kita. Motion tracking of cattle with a constrained deformable model. In *Proceedings of ICPR02*, 2002.
- [TP91] M. Turk et A. Pentland. Eigen faces for recognition. *Journal of Cognitive Neuroscience*, 3(1), 1991.

- [TSB<sup>+</sup>05] J. Teran, E. Sifakis, S. Blemker, V. Ng Thow Hing, C. Lau, et R. Fedkiw. Creating and simulating skeletal muscle from the visible human data set. *IEEE Transactions on Visualization and Computer Graphics*, (11), 2005.
- [TT92] C. Tomasi et T.Kanade. Shape and motion from image streams under orthography : a factorization method. *Int. J. of Computer Vision*, 9(2) :137–154, 1992.
- [TvdP98] N. Torkos et M. van de Panne. Footprint-based quadruped motion synthesis. In *Graphics Interface*, pages 151–160, 1998.
- [TYAB01] L. Torresani, D. Yang, E. Alexander, et C. Bregler. Tracking and modeling non-rigid objects with rank constraints. In *CVPR*, 2001.
- [Wat99] A. Watt. *3D Computer Graphics*. Addison Wesley, 3rd edition, 1999.
- [WBC<sup>+</sup>05] J. Wang, P. Bhat, A. Colburn, M. Agrawala, et M.F. Cohen. Interactive video cutout. In *Proceedings of SIGGRAPH 2005*, 2005.
- [WE84] W.A. Wolovich et H. Elliot. A computational technique for inverse kinematics. In *Proceedings of the 32nd Conference on Decision and Control*, pages 1359–1362, Décembre 1984.
- [WF97] M. Walter et A. Fournier. Growing and animating polygonal models of animals. *Computer Graphics Forum (Proceedings of EUROGRAPHICS'97)*, 16(3), 1997.
- [WFM01] M. Walter, A. Fournier, et D. Meneveaux. Integrating shape and pattern in mammalian models. In *Proceedings of SIGGRAPH 2001*, pages 317–326, 2001.
- [WG03] J. Wilhelms et A. Van Gelder. Combining vision and computer graphics for video motion capture. *The Visual Computer*, 19(6) :360–376, Octobre 2003.
- [WK88] A. Witkin et M. Kass. Spacetime constraints. In *Proceedings of SIGGRAPH '88*, pages 159–168, 1988.
- [WP95] A. Witkin et Z. Popović. Motion warping. *Journal of Computer Graphics*, 29(Annual Conference Series) :105–108, 1995.
- [WP00] L. Wade et R. E. Parent. Fast, fully-automated generation of control skeletons for use in animation. In *Proceedings of of Computer Animation*, pages 164–169, 2000.
- [YN03a] K. Yamane et Y. Nakamura. Dynamics filter - concept and implementation of online motion generator for human figures. In *IEEE Transactions on Robotics and Automation*, volume 19, pages 421–432, Juin 2003.
- [YN03b] K. Yamane et Y. Nakamura. Natural motion animation through constraining and deconstraining at will. *IEEE Transactions on Visualization and Computer Graphics*, 9(3) :352–360, Juillet 2003.
- [ZB94] J. Zhao et N.I. Badler. Inverse kinematics positioning using nonlinear programming for highly articulated figures. *ACM Transactions on Graphics*, 13(4) :313–336, Octobre 1994.
- [ZC06] J. Zhou et C.M. Clark. Autonomous fish tracking by roV using monocular camera. In *Proceedings of CRV06*, pages 68–68, 2006.

- [Zha99] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1999.
- [ZHDM05] J.B. Zurn, D. Hohmann, S.I. Dworkin, et Y. Motai. A real-time rodent tracking system for both light and dark cycle behavior analysis. In *Proceedings of WACV05*, volume I, pages 87–92, 2005.
- [ZOTX03] Y. Zhao, H.Y. Ong, T.-S. Tan, et Y. Xiao. Interactive control of component-based morphing. In *SCA '03 : Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*, 2003.





## **Modélisation du mouvement des quadrupèdes à partir de la vidéo**

Les animaux sont de plus en plus présents dans les effets spéciaux au cinéma et dans les films d'animation 3D. Les quadrupèdes en particulier, de par la richesse et de l'esthétique de leurs mouvements, sont souvent l'objet de création de personnages virtuels.

L'objectif de cette thèse est d'utiliser la vidéo, et surtout des documentaires animaliers, pour créer des animaux virtuels capables de mouvements réalistes.

Dans un premier temps, nous proposons de construire des squelettes d'animation de quadrupèdes à partir d'images. Nous montrons tout d'abord comment reconstruire un squelette réel à partir de plusieurs images. Ensuite, nous construisons un modèle déformable de squelette de quadrupède qui permet de créer un squelette d'animation rapidement à partir d'une image ou d'un modèle 3D de l'animal à animer.

Dans un second temps, nous proposons des méthodes permettant d'animer des quadrupèdes à partir de séquences vidéo. Nous présentons tout d'abord une technique d'animation de mouvements cycliques à partir de vidéo, qui permet notamment de choisir automatiquement les images-clés à partir d'une vidéo. Ensuite, nous proposons une nouvelle méthode d'analyse de vidéos qui permet d'identifier les différentes allures utilisées par l'animal, ainsi que les transitions entre allures.

**Mots-clefs** : animation, analyse d'images, quadrupèdes, locomotion, prédiction de mouvement

## **Modeling quadruped motion from video**

Animals are more and more used in special effects for the movie industry or in 3D feature films. 3D models of quadrupeds in particular are used very often due to the variety and aesthetical appeal of quadrupedal motion.

This thesis aims at using video, and more specifically wild life documentaries, for the creation of virtual quadrupeds models exhibiting realistic motion.

We first develop methods to construct 3D animation skeletons of quadrupeds from images. We show how to reconstruct skeletons of quadrupeds from several images. Then we build a morphable model of quadruped skeleton that can be easily fitted onto an image or a 3D surface of the animal to animate.

Secondly, we propose new methods to animate quadrupeds from video sequences. We first present a method to animate cyclic motion from video. This method includes automatic detection of key-frames from a video sequence. Then we propose a new method for indentifying the gaits used on a video sequence, as well as the transitions between gaits.

**Keywords** : animation, image analysis, quadrupeds, locomotion, motion prediction