

Rapport de Stage de Recherche

Animation d'arbres sous le vent à partir de vidéo

Master 2 Recherche Image, Vision, Robotique 2004-2005

Julien Diener

sous l'encadrement de Lionel REVERET

**mené au sein de l'UMR GRAVIR (CNRS, INPG, INRIA, et UJF)
Projet EVASION**

avril 2005 - juin 2005

Table des matières

| | | |
|----------|---|-----------|
| 1 | Avant-Propos | 4 |
| 1.1 | Présentation de l'INRIA Rhône-Alpes | 4 |
| 1.2 | Présentation de l'équipe Evasion | 5 |
| 2 | Introduction | 6 |
| 2.1 | Sujet, motivations | 6 |
| 2.2 | Contribution | 7 |
| 3 | État de l'Art | 8 |
| 3.1 | Modèle Géométrique de Scènes Végétales | 8 |
| 3.2 | Modèle pour l'animation | 10 |
| 3.3 | Analyse vidéo du mouvement | 14 |
| 4 | Méthode proposée | 18 |
| 4.1 | Présentation de la méthode | 18 |
| 4.2 | Outils disponibles | 19 |
| 4.2.1 | Acquisition de données | 19 |
| 4.2.2 | Détection des branches : méthode statistique | 20 |
| 4.3 | Extraction de la hiérarchie du mouvement | 22 |
| 4.3.1 | Choix de la classification hiérarchique | 22 |
| 4.3.2 | Système métrique : distance vs mouvement | 22 |
| 4.3.3 | Une classification cohérente dans le temps | 25 |
| 4.3.4 | De la hiérarchie au modèle d'animation | 27 |
| 4.4 | Toute la chaîne : de la vidéo à la 3D animée | 28 |
| 4.4.1 | Restructuration de la hiérarchie du mouvement en une structure d'arbre réaliste | 29 |
| 4.4.2 | De la structure plane vers le modèle 3D | 30 |
| 5 | Conclusion | 32 |
| 5.1 | Résultats | 32 |
| 5.2 | Perspectives | 35 |
| 5.2.1 | Estimation 3D du sujet filmé | 36 |
| 5.2.2 | Étude de l'environnement filmé | 36 |
| 5.2.3 | Création de cartes d'animation pour une application générique à des modèles 3D | 36 |

Chapitre 1

Avant-Propos

1.1 Présentation de l'INRIA Rhône-Alpes

L'INRIA, Institut National de Recherche en Informatique et en Automatique, est un établissement public à caractère scientifique et technologique, placé sous la double tutelle du Ministère de la Recherche et du Ministère de l'Economie, des Finances et de l'Industrie.

Créée en décembre 1992, l'unité de recherche INRIA Rhône-Alpes regroupe plus de 400 personnes réparties sur trois sites : la Zirst de Meylan-Montbonnot, le campus universitaire de Grenoble et le site technopolitain de Lyon (dont Lyon-Gerland et le domaine scientifique de la Doua).

Quatre pôles de recherche

L'INRIA Rhône-Alpes mène ses activités en étroite collaboration avec les laboratoires de recherche publics et privés, nationaux et internationaux, et elle entretient des liens privilégiés avec l'institut d'Informatique et Mathématiques Appliquées de Grenoble (IMAG). Ces activités sont organisées autour de quatre pôles de recherche :

- Maîtriser les systèmes et réseaux informatiques
- Aider à la conception et à la création
- Percevoir, simuler et agir
- Modéliser les phénomènes complexes

Développement et transfert de technologies

Acteur de la recherche au sein d'une région qui a retenu le pôle du numérique comme l'un des deux grands axes de développement technologique, aux côtés des biotechnologies, l'INRIA Rhône-Alpes mène une politique volontariste de transfert vers le monde économique et social et participe activement à la création d'entreprises innovantes.

Diffusion des connaissances

L'INRIA Rhône-Alpes accueille plus de 130 doctorants, ingénieurs et stagiaires. Ses chercheurs participent à l'enseignement supérieur au sein des universités et grandes écoles de la région Rhône-Alpes (Institut National Polytechnique de Grenoble, université Joseph Fourier, université Pierre Mendès-France, université de Savoie, Ecole Nationale Supérieure de Lyon). Afin d'assurer la diffusion des connaissances et du savoir-faire, l'INRIA Rhône-Alpes

entreprind continuellement des actions en accord avec les objectifs suivants :

- Formation par la recherche
- Information scientifique et technique
- Partenariats internationaux
- Organisation de séminaires et colloques

1.2 Présentation de l'équipe Evasion

L'équipe EVASION (Environnements Virtuels pour l'Animation et la Synthèse d'Images d'Objets Naturels) du laboratoire GRAVIR (CNRS, INPG, INRIA, UJF) a été créée le 1er janvier 2003. Elle regroupe six chercheurs ou enseignants-chercheurs permanents, dix huit étudiants en thèse et un ingénieur expert. Ses travaux de recherche sont dédiés à la modélisation, à l'animation et à la visualisation d'objets et de phénomènes naturels.

Pour cela, deux grands axes de recherche sont privilégiés : D'une part le développement d'outils fondamentaux destinés à la spécification de scènes et d'objets naturels complexes, à la mise au point de modèles alternatifs pour la forme, le mouvement et l'apparence ainsi qu'à la conception d'algorithmes reposant sur un niveau de détail adaptatif pour gérer au mieux la complexité ; D'autre part la validation de ces outils sur des scènes naturelles spécifiques, qui vont du monde minéral (océan, ruisseaux, lave, avalanches, nuages) au monde animal (simulation d'organes, visages, corps et chevelure d'un personnage, mouvements d'animaux), en passant par les scènes végétales (morphogénèse de plantes, animation de prairies et d'arbres).

Chapitre 2

Introduction

2.1 Sujet, motivations

Le but de ce stage est d'utiliser la vidéo de plantes naturelles pour contrôler une animation de plantes 3D virtuelles d'apparence réaliste et en temps réel. Il y a à cela plusieurs motivations.

Tout d'abord, même si une grande quantité de travaux concernant l'animation de végétaux virtuels ont été développés jusqu'à présent, les moyens couramment utilisés dans les applications en temps réel ne sont généralement pas assez réalistes ou bien supposent un ensemble de réglages généralement peu intuitif pour un animateur voulant obtenir un effet particulier. De plus, les modèles utilisés jusqu'à présent sont limités par la puissance de calcul qu'ils demandent.

La vidéo est naturellement un indice visuel très pertinent et donc présente une source de mouvements particulièrement réalistes. Dans une perspective d'animation, cela nous permettra, premièrement, de choisir librement dans une vidéo la séquence qui correspond à une animation intéressante, et deuxièmement, de ne pas avoir à régler des paramètres complexes de contrôle de mouvement a priori pour atteindre directement le résultat désiré.

Nous allons donc essayer de traiter l'information contenue dans la vidéo de manière similaire au système de compréhension qui intervient lors du processus de perception visuelle de l'oeil humain. Il nous faudra pour cela définir une règle qui relie le mouvement observé à la conception automatique d'une modélisation adéquate du sujet. Ceci permettra ensuite d'utiliser ce modèle d'animation pour générer une plante 3D similaire à la plante filmée et qui possède le même comportement.

Nous nous intéresserons tout particulièrement à la construction d'un premier modèle pour valider l'intérêt de cette technique. La difficulté principale consistera à choisir la formulation d'un principe approprié qui permettra la division, sous forme d'une hiérarchique, de l'image de la vidéo en régions suivant une même trajectoire. Puis nous devrons définir un procédé générique de création du modèle 3D basé sur une transformation de la structure de cette hiérarchie.

2.2 Contribution

Nous montrerons comment les informations obtenues par l'étude du flot optique dans l'image peuvent être utilisées pour définir un modèle cohérent du mouvement observé. Notre méthode repose sur la classification hiérarchique des éléments suivis dans l'image. Dans cette optique, il nous faudra définir en particulier un système métrique qui permette une comparaison utile de ces éléments. De plus, nous verrons qu'à partir de là, nous posséderons déjà assez d'informations sur la structure du mouvement pour reconstruire un modèle 3D animé en temps réel.

Le chapitre 3 étudiera d'abord plusieurs travaux par rapport auxquels la technique développée ici se positionne.

Ensuite, le chapitre 4 exposera notre technique en trois parties. Tout d'abord nous commencerons par présenter les outils existants que nous avons utilisés. Puis nous verrons comment, à partir de ceux-ci, il est possible de développer la hiérarchie du mouvement qui supportera l'animation récupérée de la vidéo, et les choix mathématiques nécessaires à son obtention. Enfin, dans la dernière section il sera question de la création d'un modèle 3D animé qui aura la même structure que celle de la hiérarchie.

Pour finir, nous présenterons certains résultats à l'aide de graphiques, ainsi que plusieurs possibilités de développement auxquelles pourraient aboutir notre méthode.

Chapitre 3

État de l'Art

L'animation de scènes végétales réalistes dans les applications temps-réel représente un défi particulier en raison de leur énorme complexité. Avec le développement des technologies informatiques, ce challenge est devenu maintenant à portée de main et nous voyons de plus en plus apparaître des techniques permettant l'affichage de plantes ou de paysages virtuels de grande qualité. Cependant, même si une grande quantité de travaux ont été effectués sur l'animation de tels objets 3D, nous ne pouvons finalement en voir que rarement dans des applications grand public. Il semblerait donc que le rapport coût/qualité des animations soit encore trop haut pour les développeurs.

3.1 Modèle Géométrique de Scènes Végétales

Le réalisme d'une scène dépend d'abord de la qualité des modèles 3D qui y sont présent. Ces modèles utilisés sont généralement créés avec des logiciels de modélisation, ce qui implique un travail artistique considérable. Pour obtenir un meilleur réalisme ou une plus grande précision, de nombreuses techniques existent pour reconstruire des objets réels, en particulier artificiels, à partir d'images ou de vidéos. Seulement, les plantes sont généralement trop complexes et possèdent des formes trop irrégulières pour être reconstruites ainsi.

Reche et al. ont développé une technique [AR04] permettant de reconstruire un modèle d'un arbre réel à partir d'un nombre réduit d'images (20-30) de façon à être rendu de manière interactive.

La construction de l'arbre passe, tout d'abord, par la création d'une carte d'opacité pour chaque image prise (suivant l'algorithme de Ruzon et Tomasi [MR00]). Le modèle consistera en une grille récursive (similaire à un octree, voir fig. 3.1), telle qu'à chacune de ces cases sera défini un taux de transparence en fonction des cartes d'opacité. Finalement un imposteur sera attaché à chaque cellule de la grille dont la texture affichée dépendra de l'interpolation entre la projection d'une partie des images provenant des deux angles de vue les plus proches.

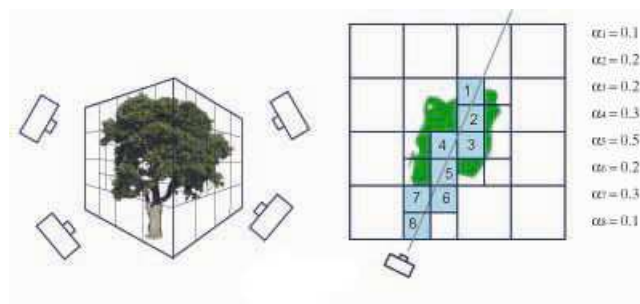


FIG. 3.1 – Gauche : Grille récursive dans laquelle est stocké le modèle récupéré. Droite : Détection des cellules de la grille où sera stockée l'information contenue dans un pixel d'une image.

Nous obtenons finalement un arbre qui est presque l'exacte représentation du sujet (visuellement). Ce modèle, composé d'un ensemble de points (auxquels sont attachés les imposteurs), va finalement afficher un nombre de polygones bien inférieur à ce qui serait normalement nécessaire pour atteindre le même niveau de qualité. Le rendu pourra donc se faire de façon interactive.

Ce modèle n'a pas été développé en particulier pour le support de l'animation. Celle-ci, si elle consiste en l'application d'un mouvement indépendant pour chaque imposteur, se montrerait trop coûteuse en puissance de calcul. Cependant, une partition en volumes englobants (qui serait susceptible d'accélérer l'affichage du modèle statique) pourrait correspondre à un nombre bien plus réduit de contrôleurs pour l'animation.

Notre méthode consiste principalement à étudier une vidéo d'un arbre pour en extraire une structure semblable à cette partition tout en récupérant ses déplacements au cours du temps. Ainsi, cela fournirait une animation pré-calculée, et donc qui supporterait un affichage rapide.

Tous les travaux qui apportent des moyens d'affichage d'objets végétaux ont pour but principal d'obtenir une apparence vraisemblable. Dans le cas de scènes extérieures, un grand nombre de plantes est indispensable au réalisme, et cela même pour des applications temps réel. Généralement, la quantité va donc être préférée à la qualité, qui peut être limitée par un nombre très réduit de polygones par objets. Par exemple, pour certaines scènes statiques, comme celles utilisées dans les simulateurs de vol, il est nécessaire d'utiliser des techniques qui permettent l'affichage d'un nombre vraiment important de plantes (ex : instanciation de modèles composés de quelques imposteurs,...).

Dans [PD04], Decaudin et Neyret proposent une technique pour l'affichage en temps réel de terrains recouverts d'une forêt, pouvant contenir plusieurs centaines voir milliers d'arbres, par l'utilisation des capacités des cartes graphiques récentes. Pour cela, des textures 3D sont utilisées sous forme de cellules triangulaires (texcells) qui représentent une zone contenant quelques arbres qui ont été modélisés de façon générale (polygonale). Ils sont composés de plusieurs triangles superposés auxquels sont appliquées des textures semi-transparentes 2D et qui vont recouvrir le terrain ; ils en existent deux types (fig. 3.2) :

- Ceux qui sont découpés en tranches horizontales. ils sont utilisés pour la majorité du terrain car ils sont plus rapides pour l'affichage.
- Ceux orientables, découpés en diagonale de façon à faire face à la caméra, utilisés pour la ligne d'horizon.

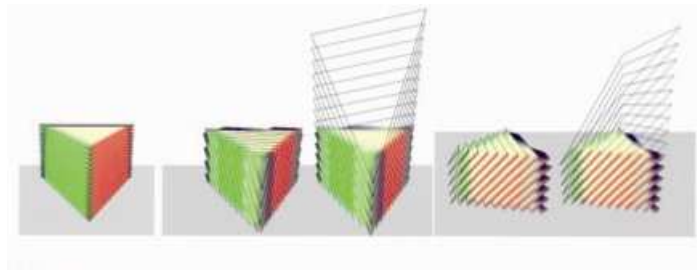


FIG. 3.2 – Gauche : Un texcell horizontal. Milieu : Un texcell de silhouette faisant face à la caméra. Droite : Une vue de coté.

Cette technique supporte plusieurs niveaux de détails qui afficheront un nombre de plus en plus réduit de tranches (donc plus lointaines) proportionnellement à la distance à la caméra. Les textures seront calculées par une formule de mélange de celles du niveau supérieur.

De plus, ce travail propose une répartition aperiodique des texcells [FN99]. L'idée est d'imposer des conditions sur les bords de ces texcells, considérés comme des triangles : Si pour chaque bord, il en existe deux autres en concordance, cela donnera 16 possibilités de texcells. Il sera alors possible d'alterner irrégulièrement ces texcells. Pour sauvegarder la mémoire utilisée, un sous-ensemble de huit peuvent être utilisés qui sera suffisant.

Avec cette technique, il est possible d'afficher une scène particulièrement grande couverte d'arbres modélisés indépendamment et pouvant être d'apparence complexes, mais elle non plus n'a pas été développée pour supporter une animation. Toutefois, il faut noter que pour le type d'applications auxquelles se destine cette méthode, l'animation n'est pas souvent primordiale.

Pour que l'animation que nous nous proposons d'extraire ait d'importantes conséquences visuelles sur une telle scène, il nous faudrait étudier une vidéo d'un paysage adéquat sinon le mouvement risquerait de ne pas être suffisamment apparent sur une vue trop éloignée.

Que se soit par l'utilisation de ce type de techniques avancées ou simplement par une modélisation adéquate, avec le développement de la capacité d'affichage des nouvelles cartes graphiques, il est maintenant possible de garnir une scène virtuelle d'une végétation luxuriante dans les applications temps-réel. Nous pouvons d'ailleurs nous en apercevoir par l'apparition de plus en plus fréquente de telles scènes dans les jeux récents. Cependant, même si dans le cas de forêt vu de loin l'animation n'est peut être pas fondamentale, pour d'autres applications l'immobilité des scènes naturelles représente un manque de réalisme considérable.

3.2 Modèle pour l'animation

Pour augmenter l'impression d'immersion, l'animation de l'environnement est essentielle notamment pour les scènes extérieures. Le mouvement des végétaux est difficile à représenter correctement. Il consiste généralement à simuler, l'application de forces initiales sur différentes parties d'un modèle (qui doit être réaliste), et leurs propagations au long de celui-ci. Puis d'appliquer le mouvement obtenu au modèle 3D à afficher. Le problème majeur pour une plante est qu'il est pratiquement impossible de tenir compte de l'ensemble de sa structure sans provoquer de trop long temps de calcul.

Beaudoin et Keyser ont proposé [JB04] un contrôle de l'animation de végétaux par une simulation du mouvement qui s'effectue par niveaux de détail de façon à réduire le temps de calcul. Pour cela, une simulation physique sur tous les types de branches utilisables (restreinte à un environnement donné, ex : vent maximum) va générer au préalable une table de référence. Ensuite un pré-calcul de l'ensemble des simplifications (fig. 3.3) sera opéré de manière à définir pour chacune un taux d'erreur maximum à l'affichage, proportionnel à la distance de la caméra. Il ne restera plus à l'application qu'à changer de niveau de détail les branches, à l'aide de la table de référence dans la majorité des cas, quand le taux d'erreur sera inférieur à une limite choisie (ex : 1 pixel).

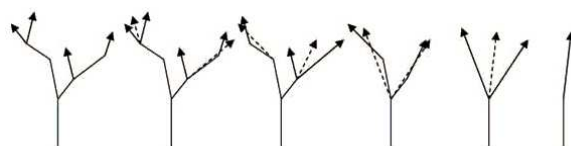


FIG. 3.3 – Simplification par étape d'un modèle d'animation en une structure fondamentale.

Cette technique permet donc de calculer, à moindre coût, le mouvement créé par un moteur de simulation basé sur des lois bien établies de la physique. Pour cela, un pré-traitement permet, d'abord de décider quel niveau de simplification du modèle original utiliser durant la simulation avec une limite de perte de qualité, et ensuite d'accélérer le passage entre ces niveaux. D'un autre côté, si elle permet d'accélérer le calcul, nous sommes encore loin du temps-réel, ou même interactif si le modèle est complexe. Mais cela peut permettre d'obtenir par exemple une animation voulue par des testes assez rapides qui seront ensuite utilisés pour un pré-calcul du mouvement ou des poses du modèle.

L'un des points importants à retenir est la liaison avec la qualité visuelle de ce type de niveaux de détail. [JB04] montre que la restriction du mouvement d'un arbre à un certain niveau de branche n'augmente que peu la différence visuelle avec cet arbre animé à tous les niveaux. Il nous a donc paru avantageux d'essayer d'extraire le mouvement de la vidéo sous une telle forme hiérarchique. Celle-ci aura ainsi l'intérêt de correspondre au type de structure observée et de permettre une application naturelle au modèle 3D de plantes tout en gardant la possibilité d'obtenir plusieurs niveaux de détails utilisables bien adaptés au support de ce type d'animation.

L'une des principales difficultés pour la rapidité d'affichage de plantes virtuelles, et en général de scènes naturelles, est que la complexité des modèles est assez homogène. Pourtant seul la partie proche de la caméra devra tenir compte du plus haut niveau de complexité. Pour obtenir le temps-réel, il sera donc nécessaire d'utiliser des modèles dont non seulement la structure d'animation devra être définie par niveaux de détails, mais aussi la structure géométrique qui lui est attachée.

Dans [FP01], Perbet et Cani présentent une telle méthode qui permet d'afficher et d'animer une prairie en temps réel et de façon réaliste. Pour cela, le terrain recouvert est divisé en carrés réguliers. Tous ceux dans le champ de vision vont appartenir à un des trois niveaux de détails suivants (LOD : Level of Detail) :

1. Le niveau 3D, le plus proche de la caméra, où chaque brin d'herbe est représenté par une chaîne de segments (fig. 3.4).

2. Le niveau 2,5D fait de plusieurs couches, composées de plusieurs bandes de textures semi-transparentes (fig. 3.4).
3. Le niveau 2D, qui est une texture 2D recouvrant le terrain au loin (en fait, juste un peu plus haut que celui-ci)

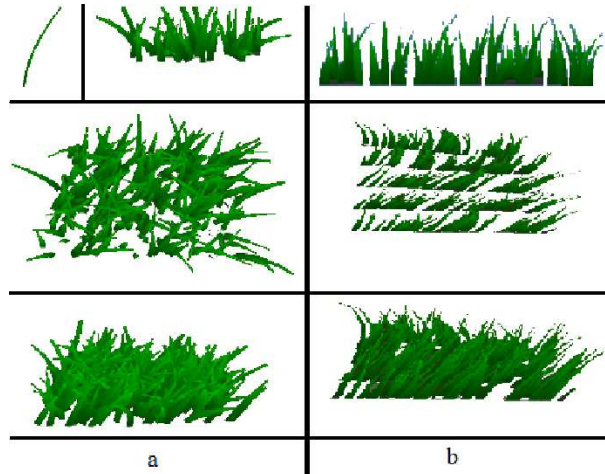


FIG. 3.4 – a) La représentation 3D, chaque brin d’herbe est indépendant. b) La représentation 2,5D, chaque bande représente plusieurs brins d’herbe.

Pour appliquer l’animation, un " receveur " est lié à chacune des primitives animables, c’est-à-dire des deux premiers niveaux, sous la forme d’un masque 2D (fig. 3.5) qui donnera à chacune une direction et un indice représentant le taux de courbure. La valeur de l’indice suivra au cours du temps une fonction qui a été développée par simulation.

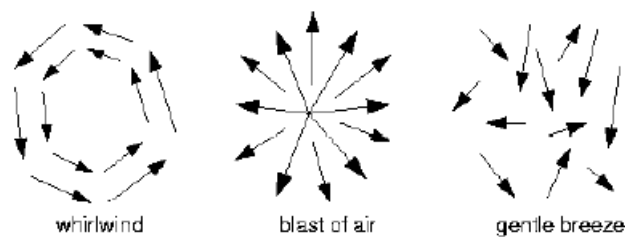


FIG. 3.5 – Différents types de vents sous la forme de masques 2D.

Le passage entre chaque niveau nécessite que, pour chaque patch, ces niveaux représentent les mêmes brins d’herbe. Donc une fois le premier niveau généré aléatoirement (plusieurs brins d’herbe avec différentes orientations), les autres en seront des projections sous différents angles.

Ce type de représentation pour une grande scène végétale peut donc être considérée comme complet pour une application animée en temps-réel, c’est-à-dire qu’elle considère un modèle pour l’animation lié à un autre pour l’affichage. Il nous faudra donc obtenir ce type

d'interaction directe entre ces deux types de modèles si nous voulons obtenir une animation rapide. Cependant, dans le cas d'une prairie, la structure des brins d'herbe est suffisamment simple et peut être identique pour tous. A l'opposé, la structure d'un arbre, plus complexe, dépend du modèle choisi. C'est pourquoi, dans notre cas, il sera préférable d'extraire en premier lieu la hiérarchie du mouvement, puis de l'utiliser ensuite pour générer celui de notre modèle d'affichage.

Dans [FP01], les conséquences visuelles de l'application d'une force sont immédiates grâce à la simplicité de la structure animable, et donc il est aisé de décider quels évènements doivent être enclenchés pour obtenir une certaine animation.

Les possibilités de contrôles par l'animateur représentent l'une des difficultés principales des modèles animables. En effet, même si un objet virtuel est capable d'être mis en mouvement de façon très réaliste, il est essentiel de pouvoir parvenir à l'effet désiré. Ceci se révèle un problème majeur en particulier pour les animations issues de la simulation : il est difficile d'obtenir un type de mouvement précis seulement par le contrôle de l'application de forces physiques.

Pour la création d'une forêt interactive [TDG00], Di Giacomo et al. ont développé un modèle de construction procédurale d'arbres, avec niveaux de détail, couplé d'une double méthode d'animation. Le mouvement de l'arbre sera calculé par une méthode procédurale pour l'influence du vent, sous la forme de cercles d'influence (fig. 3.6) où en chaque point peut être calculée une force oscillatoire. Celle-ci définit une torsion à appliquer à la racine de chaque branche qui se répercutera sur les branches parentes. Puis, s'il y a interaction avec l'utilisateur, une méthode basée sur des lois physiques est appliquée.

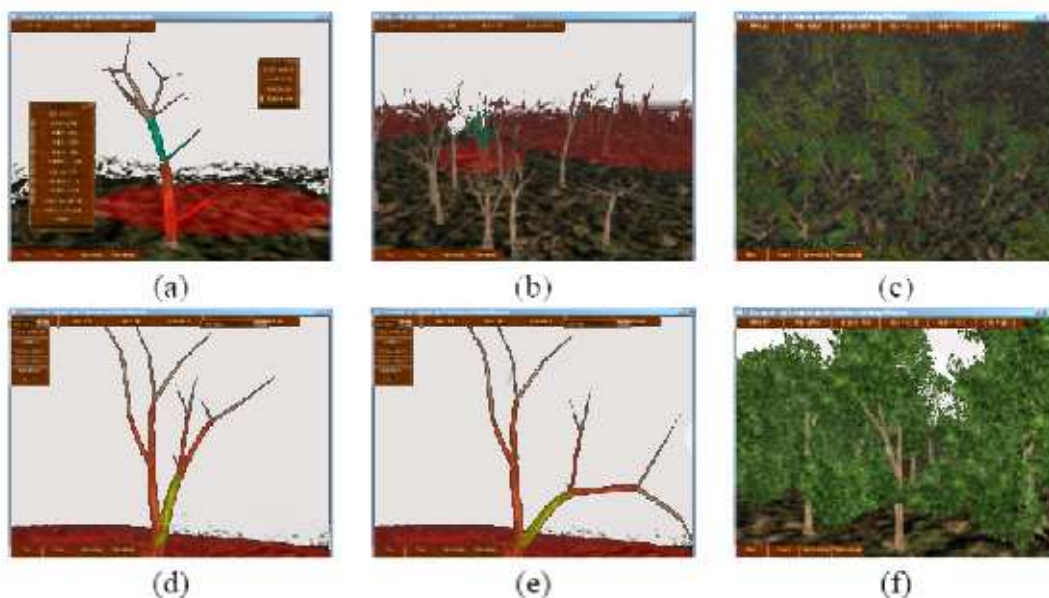


FIG. 3.6 – Le système de représentation des forces (cercle rouge sur le sol) est intuitif, mais comment estimer les conséquences visuelles pour des modèles suffisamment complexes (à partir de 4-5 niveaux de branches).

Comme pour [FP01], cette technique est complète (liaison entre modèle d'animation et d'affichage). Nous pouvons voir ici que pour un arbre, chacun des éléments composant celui-ci (branches, feuilles) réagit à des forces, mais suivant un algorithme de simulation, le mouvement suivi est aussi dépendant des autres éléments qui lui sont liés. Il sera donc difficile d'évaluer quelle forces appliquer pour créer une animation précise (fig. 3.6).

Dans l'extraction d'animation par vidéo, nous obtenons d'abord les mouvements terminaux (des groupes de feuilles) qui, de plus, correspondent à une séquence choisie au préalable. Le calcul des mouvements des plus bas niveaux de la hiérarchie de l'arbre est seulement réalisé ensuite de telle sorte qu'ils supportent de façon réaliste l'animation finale.

3.3 Analyse vidéo du mouvement

La difficulté principale pour arriver à un mouvement vraisemblable provient du fait que nous voyons souvent des plantes animées autour de nous. La vision humaine reconnaît donc instinctivement ce genre de mouvements, et inversement s'aperçoit facilement du manque de réalisme d'une animation imparfaite. Malgré cela, il est difficile de définir un mouvement réaliste, car il est finalement la somme d'effets très complexes (ex : le vent n'est jamais régulier). Il est donc naturel de vouloir obtenir des informations sur les éléments créateurs de ces effets directement à partir de l'observation de la réalité. A partir de là, il paraît alors intéressant d'extraire les paramètres d'un mouvement observé dans une vidéo, qui pourront être en suite appliqués à un modèle suivant un algorithme de simulation.

Un tel travail a été réalisé par Sun et al. [MS03]. Il consiste d'abord en l'extraction d'oscillations harmoniques telles que des vagues agissant sur un bateau ou le vent sur un arbre. Puis ces paramètres serviront à déplacer des objets synthétiques ajoutés dans la vidéo considérée.

Pour l'arbre, l'utilisateur va indiquer une zone de l'image correspondant à la pointe d'une branche principale et sa position au repos. Cette région de l'image sera suivie par une technique de vision et le mouvement récupéré permettra, suivant une solution approximative au problème d'oscillations harmoniques inverses, de trouver les paramètres physiques de la branche observée (amortissement et oscillation naturelle). Puis l'étude du mouvement de la branche, considérée désormais comme un cylindre déformable à un degré de liberté permettra de calculer la force du vent, et donc sa vitesse, au cours du temps (fig. 3.7).

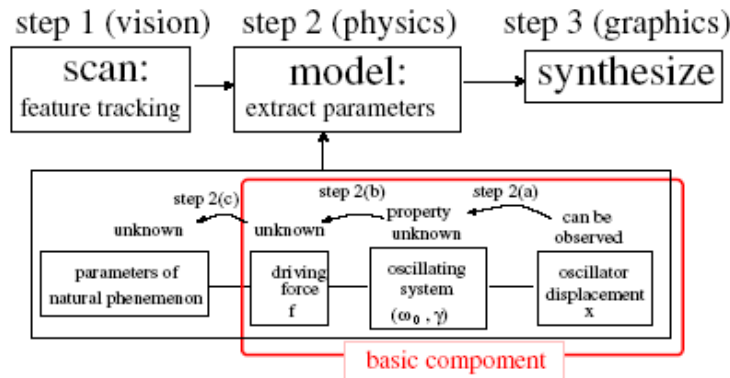


FIG. 3.7 – Procédure suivie dans [MS03].

Remarque : *Comme les informations proviennent d'une vidéo qui n'est pas calibrée, les paramètres extraits seront estimés à un facteur d'échelle près et pour du vent à priori perpendiculaire à la direction de vue de la caméra.*

Cette technique nous montre un moyen d'extraction d'informations sur l'environnement observé, qui pourront par la suite être appliquées dans une scène virtuelle. L'intérêt principal est de pouvoir appliquer ces données à tout élément de la scène à la seule condition de posséder l'algorithme de simulation adéquat. Cela permet donc de compenser la difficulté de modélisation de phénomènes naturels, mais le résultat visuel d'une scène animée de la sorte dépend encore beaucoup de la qualité de ces simulateurs ; qualité qui dépend généralement du temps de calcul que l'on peut lui attribuer.

Nous allons nous aussi étudier le mouvement de certaines parties d'un arbre en mouvement dans une vidéo. Mais, à la différence de [MS03], nous nous restreignons à la définition de mouvement pour un modèle 3D similaire au sujet filmé et, du moins dans le contexte de ce stage, nous ne nous intéresserons ni aux forces créatrices du mouvement ni aux caractéristiques physiques de l'arbre filmé. Dans ce cas, la possibilité de pouvoir définir directement les poses de ce modèle par rapport à l'image filmé aura l'avantage d'éviter l'utilisation de simulateurs, et de pouvoir obtenir un mouvement réaliste à un moindre coût.

Dans [LF04] Favreau et al. ont développé un système d'extraction de mouvements simples d'animaux à partir de vidéos (comme la marche ou la course, dans un documentaire animalier) avec une faible participation de l'utilisateur (fig. 3.8).

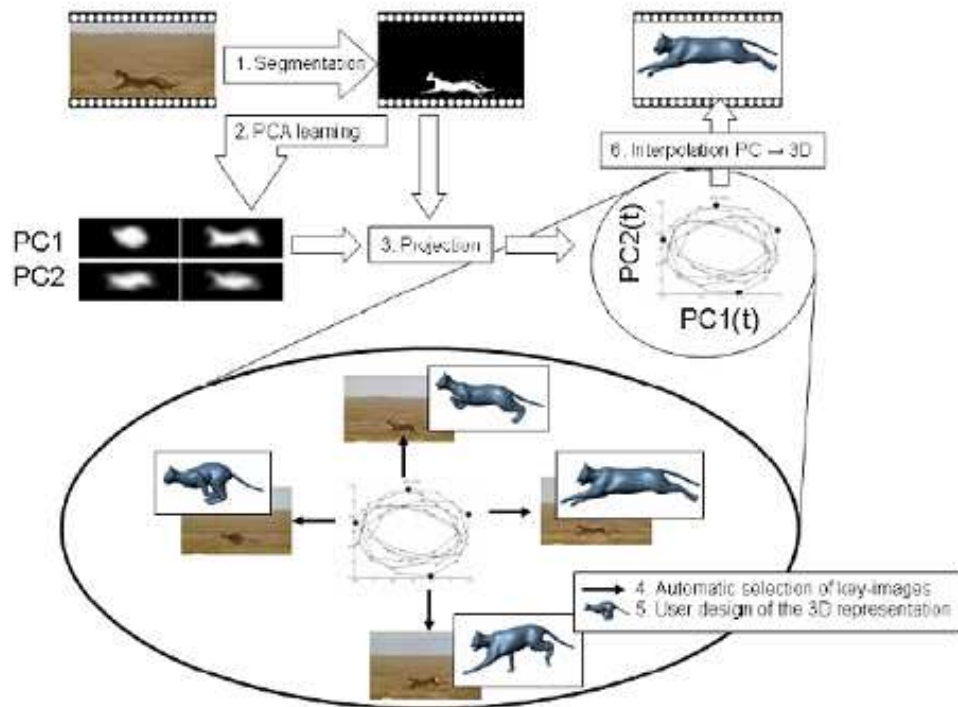


FIG. 3.8 – Etapes suivies par [LF04]

Pour cela, il faut tout d'abord que le sujet soit extrait du fond. Ceci est fait par segmentation automatique de l'image. Ensuite, l'analyse des composants principaux sur la totalité de la vidéo définit deux coefficients par image représentant la position du sujet étudié (pour un mouvement cyclique ces coefficients vont suivre des trajectoires circulaires au cours du temps). Un ensemble d'images clés sera automatiquement sélectionné pour lesquelles l'utilisateur devra indiquer quelles sont les positions correspondantes du modèle 3D. Par la suite, toutes les positions de ce modèle seront interpolées parallèlement à l'interpolation des coefficients des images clés.

Comme pour [LF04], nous ferons au préalable une étude de l'ensemble de la séquence vidéo pour en extraire des valeurs caractéristiques pour chaque image. Celles-ci nous serviront ensuite à définir la hiérarchie du mouvement. Par contre, dans notre cas, la structure de cette hiérarchie sera elle-même suffisante pour créer un modèle 3D supportant l'animation.

Chapitre 4

Méthode proposée

4.1 Présentation de la méthode

Le cadre de travail dans lequel nous nous plaçons, consistant à étudier simplement une vidéo d'une plante animée (caméra fixe non-calibrée), possède quelques contraintes particulières.

Tout d'abord, nous ne pouvons attendre de la vidéo que la structure de l'arbre soit apparente. Mais, sans vraiment posséder de modèle à priori du sujet (comme dans certains travaux de reconnaissance de mouvement d'humains ou d'animaux), nous en connaissons le type de structure, c'est à dire typiquement hiérarchique, composée d'une branche racine qui se subdivise en sous branches, et ce jusqu'aux feuilles.

De plus, le fait d'utiliser des vidéos monoculaires comme données d'entrée, limite la dimension des informations obtenues. Nous pourrions remarquer la similitude de ces dernières aux renseignements perçus par un oeil humain. Ces renseignements nous permettent de comprendre en grande partie la géométrie 3D des objets animés qui nous entourent, et ceci grâce au système cognitif lié à la vision : avant toute reconstruction 3D, l'observation monoculaire d'une plante en mouvement nous informe déjà approximativement de la position respective des branches et de leur géométrie. La question se pose donc de comprendre ce que nous percevons quand nous voyons le mouvement d'une plante ou d'un arbre, et ce qui caractérise celui-ci :

- D'abord, les déplacements de chaque élément visible ont un certain type d'amplitude et d'accélération. C'est cette particularité du mouvement que nous voulons extraire.
- Concernant les mouvements d'ensemble (typiquement les feuilles appartenant aux même branches vont suivre des trajectoires similaires) nous tirons deux renseignements essentiels. La relation entre les mouvements de ces groupes est non seulement indispensable au réalisme d'une animation d'arbres, mais aussi suffisante à la reconnaissance (inconsciente) de ce type d'animation.

Cette relation entre le déplacement des branches provient de leur structure hiérarchisée. Nous allons donc étudier les mouvements dans l'image sur un domaine composé d'échantillons répartis uniformément, puis grouper les zones qui restent fortement fixes par rapport à elles-mêmes. Celles-ci correspondront à ces groupes de feuilles qui, intuitivement, sont liées aux mêmes branches. Finalement, nous extrairons leurs mouvements qui sont les caractéristiques de la plantes filmée, sous la forme d'une hiérarchie similaire à la structure réelle de l'arbre par l'uti-

lisation d'une méthode statistique de classification. Cette technique n'est pas une reconstruction exacte du modèle observé, mais elle définit des informations nécessaires sur le mouvement pour contrôler une animation d'apparence vraisemblable dans un environnement virtuel.

4.2 Outils disponibles

4.2.1 Acquisition de données

Sélection d'une vidéo

En premier lieu, nous avons filmé dans un plan fixe plusieurs plantes, animées sous l'effet de déplacements d'air réguliers créés manuellement. Puis l'une des vidéos obtenues a été sélectionnée comme modèle pour la clarté visuelle du mouvement observé. Afin d'utiliser correctement certaines techniques de vision existantes nous avons dû effectuer un rapide pré-traitement de la vidéo sélectionnée : l'arbuste devait être séparé du fond.

Pour cela, nous avons simplement procédé par soustraction de fond, en calculant la moyenne de chaque pixel pour quelques images prises dans le même plan après avoir retiré la plante. La moyenne permet de remédier à l'instabilité inhérente à la caméra et au changement d'éclairage. Puis les pixels de la vidéo qui étaient trop similaires à ceux correspondant dans l'image représentative du fond ont été supprimés. (fig. 4.1)

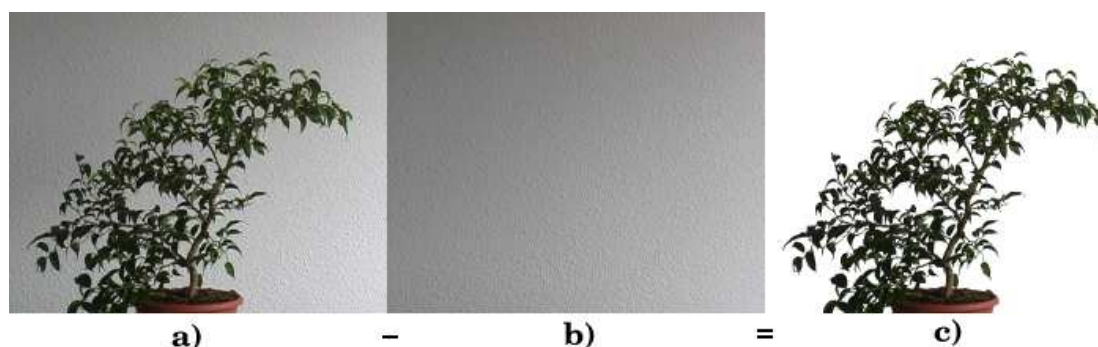


FIG. 4.1 – Gauche : l'image originale. Milieu : l'image représentative du fond. Droite : le fond a été supprimé de la vidéo.

Etude du flot optique

Avant tout traitement de données, il était nécessaire de sélectionner un ensemble de points comme domaine pour l'étude du mouvement dans l'image. L'étude de leurs déplacements nous donnera en effet un échantillonnage du flot optique dans la vidéo.

Ceux-ci devaient être d'abord représentatif, c'est à dire répartis dans toute la partie de l'image où le modèle est présent, mais surtout capable d'être suivis au cours du temps. Nous avons opté pour le nombre de 200 points, de telle sorte que la distance moyenne entre chaque couple de points est d'environ 2 à 3 cm avec un minimum de 0.5 cm si l'on considère que la plante filmée fait 50 cm du bas du tronc à sa feuille la plus haute. Ce nombre s'avérait être suffisant en terme d'échantillonnage, sans trop nous limiter par les temps de calcul qu'il implique. De plus ces points ont été choisis tel qu'ils représentent des maximums locaux du

gradient de l'image, suivant la procédure du choix de points de Shi et Tomasi ('Good Features to Track' [JS94]). Cela permet de réduire la fréquence des erreurs de suivi au long de la séquence (fig. 4.2).



FIG. 4.2 – Les points en vert seront suivis au cours du temps.

La technique de suivi de points que nous avons choisie consiste en une version pyramidale de l'algorithme de Lucas et Kanade [Bou99] :

Pour chaque paire d'images consécutives, on génère une pyramide à plusieurs niveaux de résolution. En commençant par le niveau de moindre résolution (disons le plus bas) de cette pyramide on utilise l'algorithme de Lucas et Kanade [BL81] pour le suivi de points. Puis le résultat est utilisé comme position de recherche initiale du suivi dans les plus hauts niveaux de la pyramide.

Il faut noter ici qu'il n'est pas nécessaire d'avoir des suivis particulièrement précis car nous ne cherchons pas une reconstruction exacte du modèle mais plutôt une récupération de l'aspect du mouvement. Nous verrons par la suite que la technique présentée nécessite seulement qu'ils soient cohérents dans les directions de mouvement. Il est donc admis que quelques suivis ne soient pas parfaits, seul les grosses erreurs, qui sont faciles à repérer, doivent être éliminées.

4.2.2 Détection des branches : méthode statistique

Pour réunir les points échantillons en groupes définissant les branches, nous avons eu recours à des techniques de classification automatique. Il existe de nombreux algorithmes pour cela [LL95] :

Par exemple, la méthode d'agrégation autour de centres mobiles permet de définir une partition de l'ensemble des éléments. Elle consiste à choisir un certain nombre de centre dans la population que l'on étudie (par exemple par tirage aléatoire) puis à grouper tous les points avec le centre qui leur est le plus proche. Ensuite on trouve le nouveau centre des groupes, et on ré-itére l'étape précédente avec ces nouveaux centres jusqu'à une convergence suffisante (plusieurs critères possibles). Cette technique offre l'avantage d'obtenir une partition sur un ensemble volumineux de données à un faible coût, mais ses résultats dépendent des premiers centres choisis et elle impose de fixer à priori le nombre de classes résultantes.

Une autre méthode est celle de la classification hiérarchique qui procède par agglomération progressive ascendante. Il en existe plusieurs versions, mais toutes suivent le même schéma :

l'algorithme consiste à créer à chaque étape, une partition des éléments en agrégeant les deux éléments les plus proches. Les éléments sont à la fois les individus à classer et les regroupements d'individus déjà générés par l'algorithme. Les différentes variantes de cette technique dépendent de la façon de considérer les nouveaux groupes d'éléments agrégés. Comme ils tiennent compte de l'ensemble des données à chaque étape, ces algorithmes de classification vont limiter la quantité classable par le temps de calcul engendré.

Il existe aussi des méthodes dites mixtes, capables d'appliquer une classification hiérarchique sur des ensembles volumineux de données. Leur première étape consiste d'abord à utiliser une classification autour de centres mobiles, pour obtenir un grand nombre de classes mais bien plus réduit que celui d'éléments initiaux. Puis ils appliqueront une classification hiérarchique dont cette partition représenterait les éléments initiaux. Étant donné que les classes obtenues à la première étape sont cohérentes et d'un niveau virtuellement très bas dans la hiérarchie, cela n'influera que peu sur celle-ci pour les niveaux suffisamment supérieurs.

Nous avons choisi la classification hiérarchique pour plusieurs raisons. D'abord car elle est déterministe, donc que la solution est constante pour un même ensemble de point. De plus, le nombre de points à classer ne nous impose pas d'utiliser ni une classification autour de centres mobiles, qui de plus nous contraint à choisir le nombre de classes à l'avance (et donc de branches), ni une classification mixte. Mais la raison principale est la nature même de ce type de classification, c'est à dire qu'elle nous permet d'obtenir une hiérarchie (fig. 4.3) qui est finalement une structure d'arbre.

Nous obtiendrons ainsi, en une seule classification, la structure complète du modèle d'animation. Les points terminaux de cette hiérarchie seront alors les feuilles d'un arbre où chaque union d'éléments représentera une branche. D'autre part, pour chacune de ces branches, quel que soit leur niveau, toutes les feuilles (et sous branches) sont proches physiquement mais avant tout suivent des trajectoires très similaires. Ce fait implique que cette structure soit la plus représentative et la plus adaptée à supporter le mouvement observé.

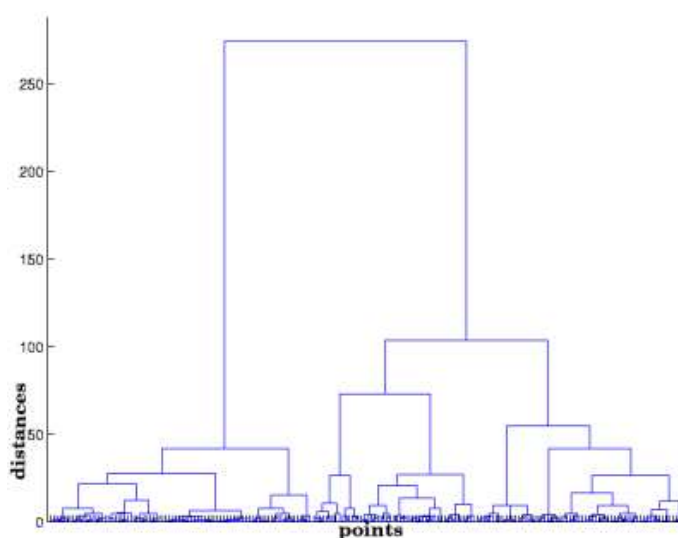


FIG. 4.3 – Exemple de dendrogramme obtenu par la représentation du résultat d'une classification hiérarchique.

4.3 Extraction de la hiérarchie du mouvement

4.3.1 Choix de la classification hiérarchique

Comme nous l'avons vu précédemment il existe plusieurs variantes de la classification hiérarchique. Ces diverses techniques diffèrent par la manière dont elles considèrent les groupes constitués de plusieurs éléments (groupes considérés eux-mêmes comme des éléments) pour les comparaisons de distance [LL95].

Pour une fonction de distance d applicable sur le domaine des éléments à classer, si x, y, z sont trois éléments du domaine et si x et y sont regroupés en l'élément h , nous pourrions par exemple décider que : $d(h, z) = \text{Min}\{d(x, z), d(y, z)\}$. Ce qui constitue le critère d'agrégation appelé *saut minimal*. De même, nous pouvons définir le critère du *saut maximal*, ou bien plus fréquemment employé celle de la *distance moyenne* :

$$d(h, z) = \frac{d(x, z) + d(y, z)}{2}$$

Plus généralement, si x et y désignent des sous-ensembles disjoints de l'ensemble du domaine ayant respectivement n_x et n_y éléments, alors nous aurons.

$$d(h, z) = \frac{n_x d(x, z) + n_y d(y, z)}{n_x + n_y}$$

Nous utiliserons ici une technique plus adaptée dite de *minimisation de la variance*, où le critère d'agrégation considère l'inertie du système :

$$I = I_{inter} + I_{intra} = \sum_q n_q \|g_q - g\|^2 + \sum_q \sum_{i \in q} \|x_i - g_q\|^2$$

A l'étape possédant q groupes (formule simplifiée pour notre cas), x_i sont les éléments initiaux, g est le centre de gravité de l'ensemble des points à classer, g_q le centre de gravité du groupe q et n_q le nombre de points dans le groupe q .

L'inertie I du système est constante, et le critère utilisé ici (appelé *critère de Ward généralisé*) consiste à minimiser l'inertie intra-classes (I_{intra}) qui représente l'homogénéité des classes, ou de la même manière, à maximiser l'inertie inter-classes (I_{inter}). Pour cela, la *relation de Huygens* est utilisée pour calculer, à chaque étape et pour chaque couple d'éléments, le déplacement d'inertie ΔI_{ij} qui résulterait de la association. Ce critère d'agrégation va finalement choisir la combinaison qui minimisera ce nouvel indice de disparité (appelé aussi *indice de niveau*), à la place de la distance entre chaque paire d'éléments.

4.3.2 Système métrique : distance vs mouvement

Toutes les techniques de classification ont pour but de grouper les points en fonction de leurs distances respectives, elles sont donc particulièrement liées au choix de la métrique utilisée.

Nous nous sommes rapidement aperçus que les distances entre les seules positions des points ne nous permettraient pas d'obtenir des groupes particulièrement significatifs pour représenter le mouvement observé dans une vidéo. En effet, celle-ci ne dépendrait que trop du

choix initial des points, c'est-à-dire de la répartition des pixels de fort contraste. Ce n'est pas la position des feuilles sur l'arbre qui nous renseignera sur la position des branches, du moins pas obligatoirement. Par contre, il nous faudra remarquer que ces groupes (fig. 4.4) ont l'intérêt d'être réguliers dans l'espace (formes et positions relatives).

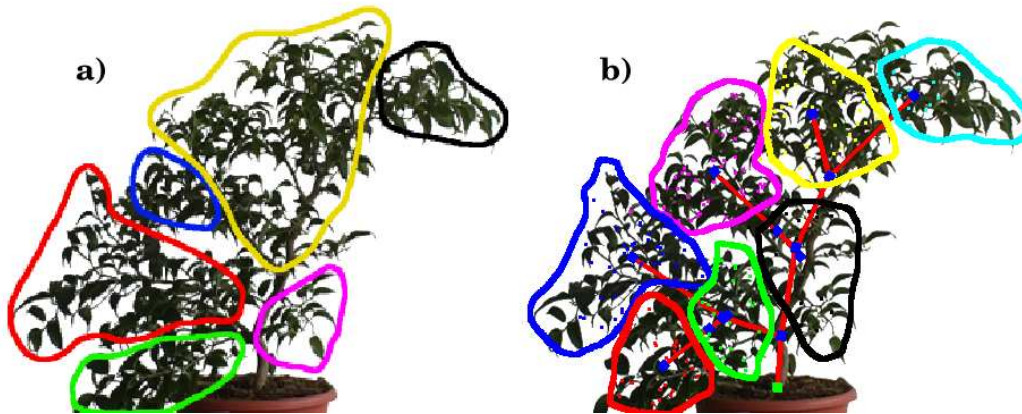


FIG. 4.4 – a) Véritable classification obtenue par l'observation du mouvement des branches. b) Hiérarchie obtenue en utilisant les distances de positions pour la classification.

Étant donné que nous étudions le mouvement, il semble donc normal de recourir aux informations sur leurs vitesses au cours du temps pour regrouper nos points. Comme nous connaissons les coordonnées des points dans chaque image, les vitesses sont les différences de positions entre deux images consécutives. Ces vitesses seront considérées comme instantanées, et si nous estimons qu'un groupe de feuilles, considéré donc comme une branche, effectue par exemple une rotation, alors les vitesses des échantillons de ce groupe pourront être assimilées à des translations sans perdre une trop grande précision.

Nous avons décidé de ne pas utiliser la vitesse à proprement parler, c'est à dire la norme des vecteurs vitesses. La raison principale est que ces valeurs ne sont pas une caractéristique pertinente pour la classification. Par exemple, dans le cadre d'une rotation d'un groupe (branche), les vitesses instantanées de tous les échantillons du groupe, exprimées sous forme de translations, seront similaires pour les directions mais pas nécessairement pour les vitesses qui seront plus rapide pour les feuilles extérieures (fig. 4.5). Les coordonnées des vitesses ont donc été converties en coordonnées polaires, puis les seules directions de mouvement ont été retenues. La métrique obtenue calcule donc des distances entre ces directions (que nous appellerons *distance de mouvement*).

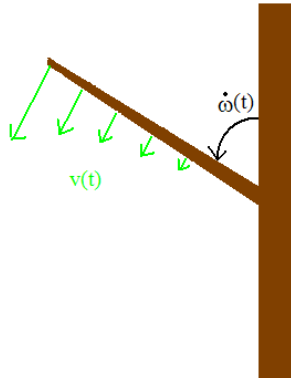


FIG. 4.5 – Sur une branche en rotation, les vitesses instantanées sont de même direction mais pas de même norme.

Une nouvelle hiérarchie est obtenue par cette méthode, différente par la forme des groupes et par sa structure hiérarchique de celle basée sur les distances euclidiennes (fig. 4.6, pour les distances de positions, les groupes n’avaient pas de rapport avec le mouvement alors que pour les distances de mouvement, deux groupes se détache : celui de droite attaché à la branche principale et celui de gauche dont les oscillations sont plus rapides). Cette nouvelle hiérarchie est clairement représentative du mouvement, mais les groupes obtenus ont des aspects et une répartition trop irrégulières pour être utilisables de façon aisée (ex : certaines branches ont des sous-branches entrelacées). C’est pourquoi notre choix s’est porté sur un compromis entre ces deux systèmes métriques qui passe par la multiplication des fonctions de distances respectives (l’unité de cette distance sera donc le pixel-radian). Simplement, pour éviter que la distance de position ne prenne trop d’importance, il faudra s’assurer que les points échantillons soient bien répartis.

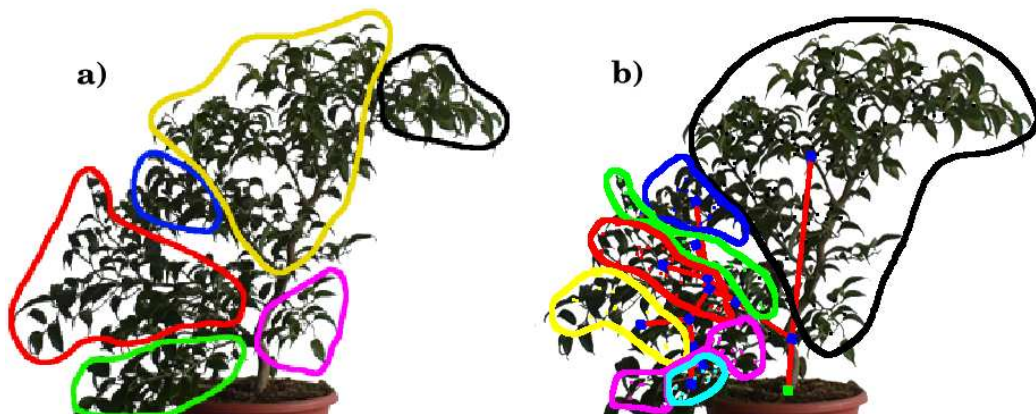


FIG. 4.6 – a) Véritable classification obtenue par l’observation du mouvement des branches. b) La hiérarchie obtenue en utilisant, pour la classification, des distances proportionnelles à l’écart entre les directions du mouvement. Le groupe noir reste très stable dans la vidéo utilisée ; les autres groupes sont de forme allongée perpendiculairement à l’axe principal du mouvement (instantané).

La structure obtenue avec la classification hiérarchique utilisant cette mesure aura non seulement un aspect régulier donc visuellement significatif, mais surtout caractéristique de l'animation recherchée : si nous extrayons le mouvement des groupes mais gardons ceux-ci fixe par rapport à eux-même, nous garderons l'aspect essentiel de l'animation filmée. (fig. 4.7)

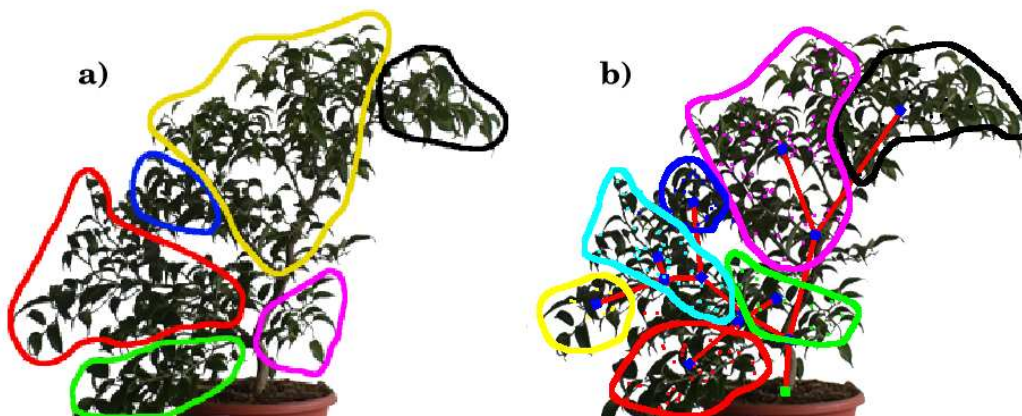


FIG. 4.7 – La classification hiérarchique obtenue en utilisant les distances de position et les distances de mouvement, est très proches de celle que nous aurions choisie intuitivement : significative du mouvement visionné et de structure plus réaliste.

4.3.3 Une classification cohérente dans le temps

Jusqu'à présent, nous avons défini une fonction de distance utilisable pour chaque point à un instant correspondant à une image. L'appliquons d'une classification hiérarchique sur le domaine des échantillons suivis parvient à générer un ensemble de groupes suffisamment représentatif de chaque image. Seulement ceux-ci ne sont représentatifs que pour l'image où ils ont été calculés, et ne sont pas particulièrement cohérents d'un début de la séquence à l'autre. De plus certaines images, car les mouvements des points sont trop désordonnés ou trop homogènes (fig. 4.8), ne propose qu'une hiérarchie très peu représentative.

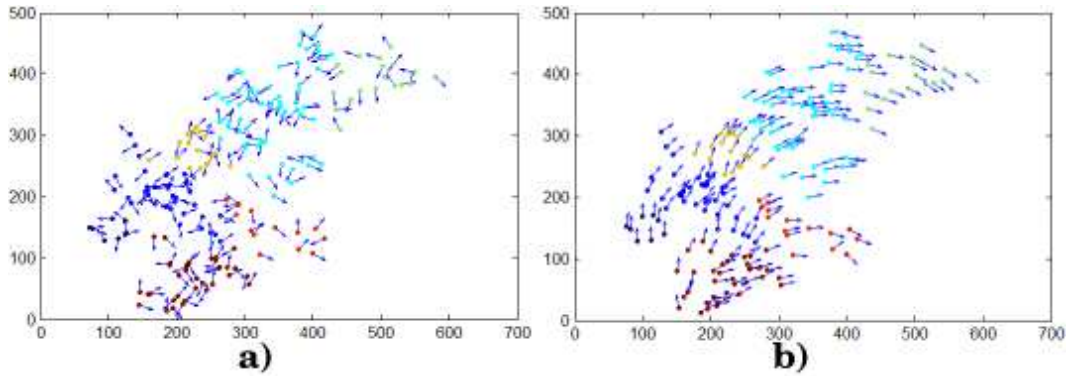


FIG. 4.8 – a) le mouvement général des points est trop désordonné. b) le mouvement est trop ordonné. Dans les deux cas, l'influence de cette image sur la hiérarchie sera négative car elle va tendre à rapprocher des points sans qu'ils soient sur une branche proche. (ne pas tenir compte des couleurs des points dans ce graphique)

C'est pourquoi nous devons tenir compte du plus grand nombre d'images, mais exclure celles qui montrent des caractéristiques trop extrêmes. Pour cela, à chaque image, un coefficient représentant le taux de désordre est calculé tel que, si la fonction d_{mvt} représente la distance de mouvement, les x_j sont les points et N est leur nombre :

$$E_i = \frac{2}{N(N-1)} \sum_{k=1}^{N-1} \sum_{l=k+1}^N d_{mvt}(x_k, x_l)$$

qui est en fait la moyenne de ces distances, pour tous les couples de points. Puis, si E est l'ensemble des E_i conservés, nous excluons toutes les images telles que nous ne gardions qu'un ensemble sans aucun élément trop différent :

$$\forall E_i (E_i \in E) \Rightarrow (E_i \in [\mu(E) - 2 \times \sigma(E); \mu(E) + 2 \times \sigma(E)])$$

où $\mu(E)$ est la moyenne des E_i et $\sigma(E)$ leur écart-type.

Ceci donne des résultats satisfaisant pour notre séquence, notamment car le mouvement représenté est répétitif. Mais il est possible que l'étude de ce coefficient de désordre du mouvement puisse aussi nous permettre de diviser une séquence composée de plusieurs type de mouvement (vent faible - vent fort,...). Dans le cas de la vidéo sélectionnée, qui dure cinq secondes donc 125 images, 14 d'entre elles sont ainsi exclues.

Ensuite, la distance d utilisée pour la classification calculée de façon à ce que, si les d_i sont les distances entre points pour chaque image i gardée, si N est leur nombre et, si x et y sont deux points échantillons, alors :

$$d(x, y) = \left(\prod_{i=1}^N d_i(x, y) \right)^{\frac{1}{N}}$$

Nous avons utilisé une moyenne géométrique pour deux raisons : D'abord, l'usage d'une moyenne arithmétique impliquerait de choisir une proportion relative pour les valeurs des distances, particulièrement entre distance de position et distance de mouvement ; De plus cette

moyenne est (intuitivement) inversement proportionnelle à une fonction qui définit la probabilité d'appartenance de toutes paires de points x et y à un même groupe G_q :

$$P(x \in G_q \wedge y \in G_q) = \frac{M - d(x,y)}{M}$$

avec M la valeur maximale que peut prendre la distance d .

Typiquement, si deux points se suivent souvent, ils auront une petite distance les séparant et donc une grande probabilité d'appartenir au même groupe, et inversement. Pour cela, chaque image ajoutée à la séquence va ou bien rapprocher ou bien les éloigner chaque couple de points dans la hiérarchie en, respectivement, réduisant ou augmentant la valeur de la fonction $P(x \in G_q \wedge y \in G_q)$.

Remarque : Actuellement, pour des raisons de gestion de la mémoire durant le calcul du produit, le nombre d'images utilisées est restreint à 40, qui sont choisies comme ayant les plus grosses moyennes des vitesses des points, donc plus représentatives du mouvement que les autres. Cette restriction pourrait être évitée en choisissant une échelle de valeurs appropriée pour les d_i , telle que l'on puisse stocker leur produit dans une seule variable.

4.3.4 De la hiérarchie au modèle d'animation

La hiérarchie obtenue ainsi est sous la forme d'un arbre binaire dont les noeuds terminaux sont les points suivis dans l'image. Comme dans [JB04], pour réduire le coût de l'animation (en gardant un meilleur rapport coût/qualité), on ne calculera le mouvement que jusqu'à un certain niveau de l'arbre. La première étape consiste donc à couper cet arbre pour ne garder qu'un certain nombre de groupes de feuilles (restreintes désormais au mouvement de leur groupe) qui en deviendront les nouveaux noeuds terminaux. De la même manière, il est possible de choisir plusieurs coupes qui serviront ensuite de niveaux de détail pour l'animation. Le choix du nombre de branches peut être fait de façon arbitraire (nous avons opté pour sept groupes), quel que soit ce nombre nous obtiendrons automatiquement la meilleure qualité pour ce niveau de complexité. Mais la hiérarchie peut aussi nous informer sur les distances entre groupes, et celles-ci peuvent être utilisées pour déterminer un choix optimal de découpe (les plus représentatifs des différents niveaux de qualité d'animation possible) par rapport à l'irrégularité des distances sur la séquence des subdivisions de l'arbre (fig. 4.9).

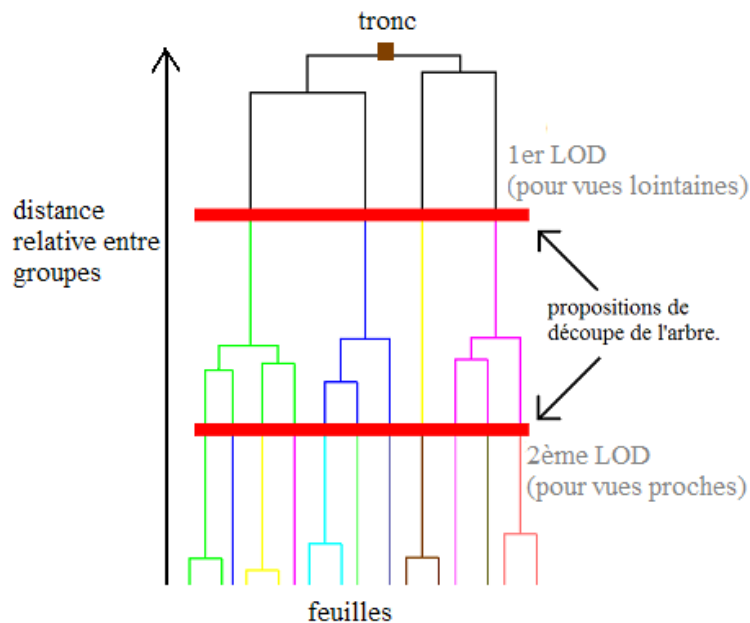


FIG. 4.9 – Exemple de hiérarchie particulièrement bien adaptée pour le choix des niveaux de détails (découpe en 4 et en 11 branches à animer).

Nous obtenons ainsi un modèle composé d’au maximum N groupes de feuilles, qui aura $2(N - 1)$ branches, et la possibilité de choisir plusieurs niveaux (N) de détails en animant seulement les groupes correspondant à une découpe plus haute de la hiérarchie.

4.4 Toute la chaîne : de la vidéo à la 3D animée

Le but de toute cette technique d’extraction de mouvement est, finalement, d’obtenir une animation temps-réel et réaliste d’un modèle 3D. Jusqu’ici nous avons développé un moyen de création d’une structure décrivant particulièrement bien le mouvement par rapport au niveau de sa complexité. Cette structure se présente sous la forme d’une hiérarchie composée de branches reliant un tronc unique et un ensemble de feuilles (les points suivis dans la vidéo).

Nous avons donc décidé d’utiliser cet arbre comme base initiale pour la création de notre modèle 3D. L’une des grandes possibilités de cette structure (comme expliqué dans le paragraphe précédent) est de pouvoir la découper à différents niveaux de complexité représentant différentes qualités d’animation. Ainsi, pour un choix de N groupes terminaux, donc d’objets virtuels représentant des ensembles de feuilles (ex : quelques polygones fixes entre eux), il y aura seulement $2(N - 1)$ rotations pour animer l’arbre de façon réaliste, c’est à dire le nombre des branches de plus hauts niveaux dans la hiérarchie. Dans notre cas, nous avons gardé 7 branches terminales, et donc l’animation se calcule avec 12 rotations par image affichée.

Cependant il existe deux difficultés à surmonter avant de posséder un modèle d’arbre d’apparence réaliste : en premier lieu, même si elle en possède la structure, la hiérarchie du mouvement n’a pas tout à fait l’apparence d’un arbre (fig. 4.10), de plus elle n’est pour l’instant définie qu’en deux dimensions seulement puisqu’elle est liée à la projection dans le plan de la caméra.

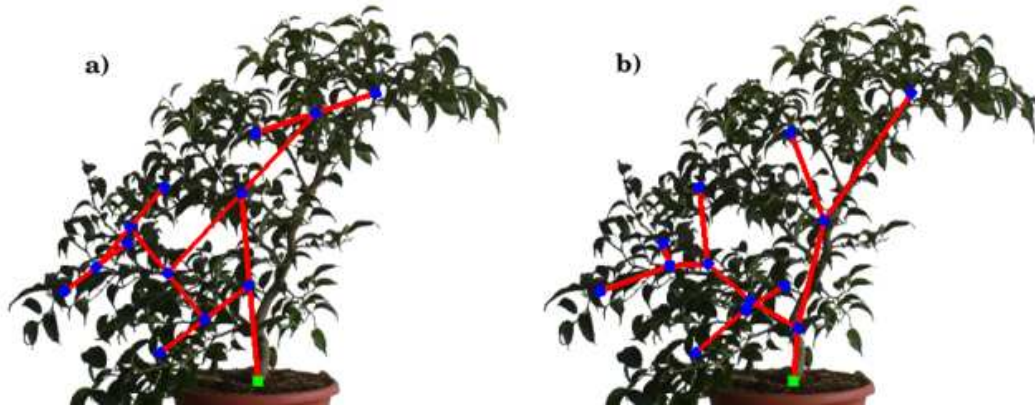


FIG. 4.10 – a) Hiérarchie obtenue telle quelle par la classification. b) Structure d’arbre plus ordinaire et plus proche de la réalité (hiérarchie 2D actuelle).

4.4.1 Restructuration de la hiérarchie du mouvement en une structure d’arbre réaliste

Le premier problème provient de la manière dont les centres des groupes sont calculés, c’est à dire la moyenne des positions de leurs points. Il en découle donc que pour les groupes du dernier niveau (les groupes feuilles ou noeuds terminaux de l’arbre, qui dépendent de la limite choisie dans la hiérarchie) aurons des centres raisonnables. Par contre les autres groupes (les branches) posséderont leur centre exactement entre leur deux fils. De plus, il faudra transférer les mouvements des groupes qui pour l’instant sont stockés sous forme de translation, en rotation des branches filles par rapport aux branches mères et cette position de groupes n’est pas idéale pour supporter ces rotations telle que le mouvement soit visuellement juste.

Pour résoudre cela, il suffit en fait de rapprocher le centre de chaque branche (donc les noeuds de l’arbre qui ne sont pas terminaux) vers leur noeud parent (fig. 4.11). Pour notre modèle, nous avons choisi de déplacer ces noeuds de la moitié de la longueur de la branche. En fait il serait aussi possible de décider d’un rapport de longueur proportionnel au nombre de points que possède chacune des deux branches filles Une autre solution plus exacte consisterait à définir les centres de rotation (noeuds) en optimisant les coefficients des matrices de rotation appliquées aux points qui en dépendent.

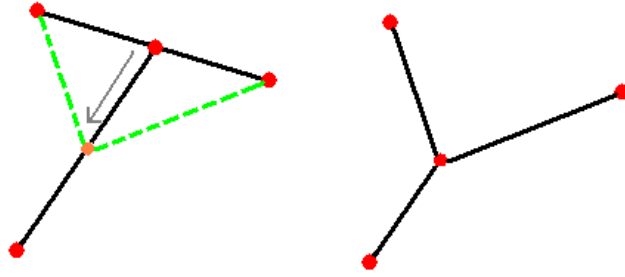


FIG. 4.11 – Transfert d'un noeud (non-terminal) de la hiérarchie.

4.4.2 De la structure plane vers le modèle 3D

Afin de transférer ce modèle 2D vers la 3D, plusieurs techniques se sont révélées possibles. Nous avons choisi une procédure qui utilisait, de façon automatique, les données déjà en notre possession.

Volume des ensembles de feuilles

Tout d'abord, un petit ensemble de trois feuilles est affiché à la place de chaque point utilisé pour le suivi, sous la forme d'un motif texturé (fig. 4.12). Puis une valeur pour la troisième coordonnée (z si on considère (x,y) comme le plan de l'image) pour chacun de ces points, est définie en fonction de leur place dans l'arbre.



FIG. 4.12 – Texture semi-transparente utilisée pour le modèle 3D (agrandie).

Cette valeur est calculée tel que, si $d_i(x)$ est la distance de Mahalanobis du point x par rapport au centre de l'ensemble de points X_i (distance définie sur X_i), X_{all} est l'ensemble de tout les points suivi dans l'image, X_q est le groupe terminal $\subset X_{all}$ auquel appartient x , et si :

$$M_i = \mu(d_i(\{x \in X_i\})) \quad \text{et} \quad S_i = \sigma(d_i(\{x \in X_i\}))$$

$$Max_{all} = M_{all} + 2S_{all} \quad \text{et} \quad Max_q = M_q + S_q$$

Alors la profondeur dans l'espace du point x sera :

$$x(z) = z_{all} + z_q = \sigma_{all} \times \frac{Max_{all} - d_{all}(x)}{Max_{all}} + 2\sigma_q \times \frac{Max_q - d_q(x)}{Max_q}$$

Ce qui revient, en fait, à donner une valeur z inversement proportionnel à : la distance du point par rapport à l'arbre (z_{all}), et la distance par rapport à son groupe (z_q), voir fig. 4.13. Puis une copie de ces points, symétriques au premiers ($x'(z) = -x(z)$), ont été ajouté à la structure, de façon à couvrir la face opposée à la caméra d'origine.

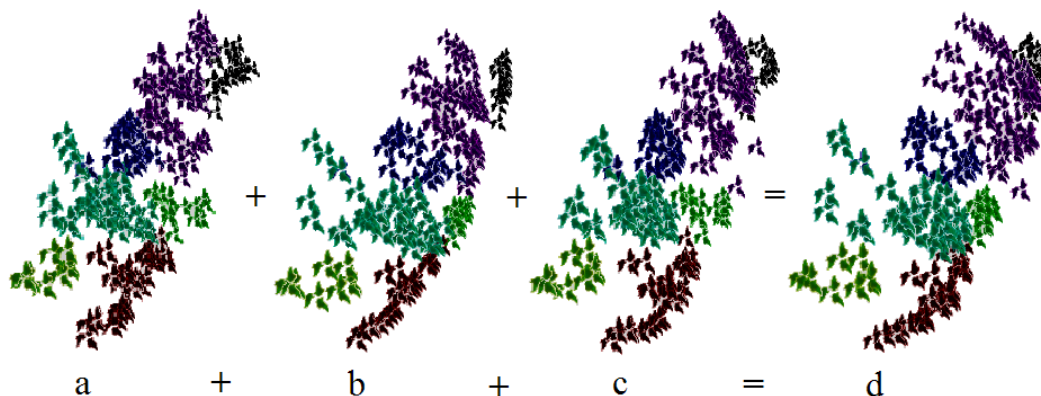


FIG. 4.13 – Calcul de la profondeur des points du modèle 3D (les couleurs représentent les groupes) : *a*) les points d'origine ($z = 0$); *b*) hauteur par rapport à l'ensemble des points (z_{all}), une 'bosse' pour tout l'arbre ; *c*) par rapport à chacun des groupes (z_q), une 'bosse' par groupe ; *d*) la position finale des points ($x(z) = z_{all} + z_q$).

Définition asymétrique du modèle

Nous avons donné du volume au modèle 3D, mais celui-ci reste symétrique par rapport au plan de l'image ($x - y$). Ceci provient du fait que la hiérarchie de l'arbre est toujours plane. La structure est un squelette articulé tel que toute l'animation se fait par des rotations autour de l'axe des z des repères de chaque branche, les autres rotations restent nulles tout au long de la séquence. Nous avons donc, par une rapide édition, appliqué de petites rotations à chacune des branches de l'arbre tel que sa structure ne soit plus seulement définie dans un plan (fig. 4.14).

Ainsi, les groupes de feuilles du modèle auront des positions irrégulières quel que soit l'angle de vue, tout en gardant une animation réaliste et équivalente à celle filmée.

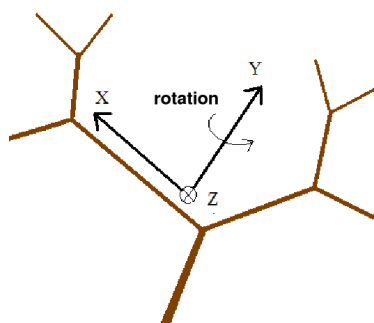


FIG. 4.14 – Rotation aléatoire pour réduire la symétrie de la plante.

Chapitre 5

Conclusion

A partir d'une séquence vidéo, nous avons montré que l'étude du mouvement observé peut permettre de définir de manière automatique une hiérarchie similaire au système de branches de la plante visionnée. De plus la forme sous laquelle se présente celle-ci pourra, par la suite, permettre de choisir les niveaux de détail appropriés pour l'application de destination.

Cette hiérarchie, plus qu'une tentative de reconstruction du modèle observé, est en fait la structure de son mouvement, c'est à dire les ensembles de la plante qui posséderont des mouvements similaires (ce qui est intuitivement une conséquence de l'appartenance à une même branche). Elle se présente donc comme particulièrement bien adaptée à supporter le contrôle de ce mouvement sur un modèle 3D, de façon simplifiée mais fortement réaliste, par l'utilisation d'un squelette animé composé d'objets rigides.

Finalement, nous avons proposé un premier modèle de la sorte qui confirme ces présomptions, nous obtenons bien un mouvement qui est visuellement presque identique à celui filmé, et cela animé en temps-réel.

5.1 Résultats

Nous avons défini une hiérarchie de groupes animés suivant le mouvement d'un modèle sur une séquence vidéo de cinq secondes (fig. 5.1 et 5.2). Ces groupes sont définis de telle manière à ce qu'ils possèdent la caractéristique de se mouvoir de façon rigide (fig. 5.3) et à posséder des formes compactes.

Ensuite, la création du modèle 3D grâce à cette hiérarchie (fig. 5.4) nous a permis de lui appliquer l'animation observée sur la plante filmée de manière immédiate, pour en fin de compte obtenir une animation de plante virtuelle réaliste et en temps-réel.

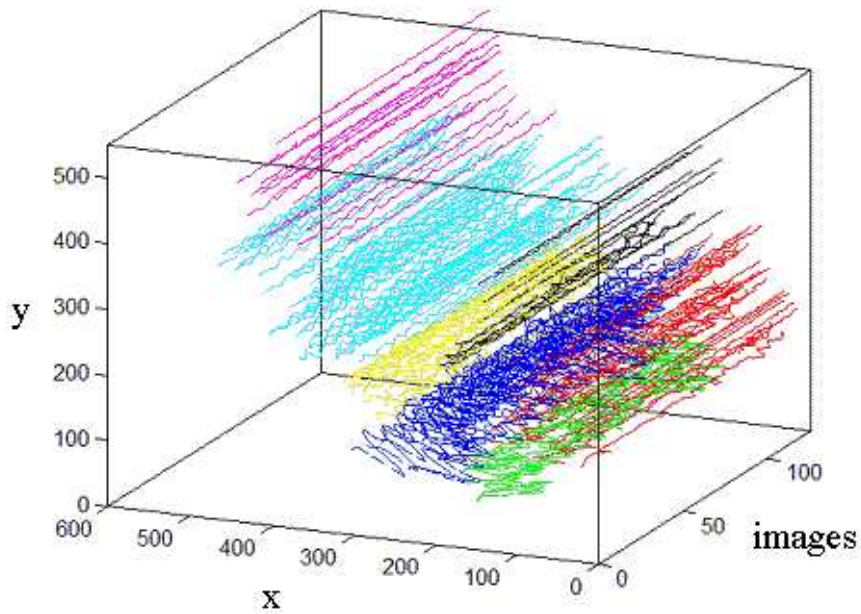


FIG. 5.1 – Regroupement des points en fonction de leur mouvement au cours du temps. Les coordonnées des 200 points dans chaque image de la séquence que nous avons choisies sont considérées comme des échantillons d'un champ vectoriel 3D (deux dimensions pour l'image et une pour le temps) représentant un courant coulant dans la direction du temps.

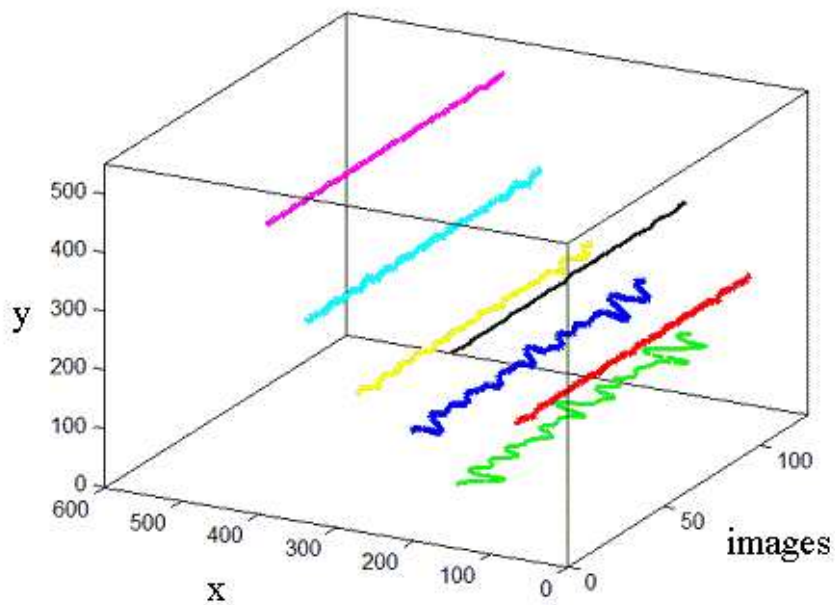


FIG. 5.2 – Le mouvement de chacun des groupes terminaux de la hiérarchie extraite.

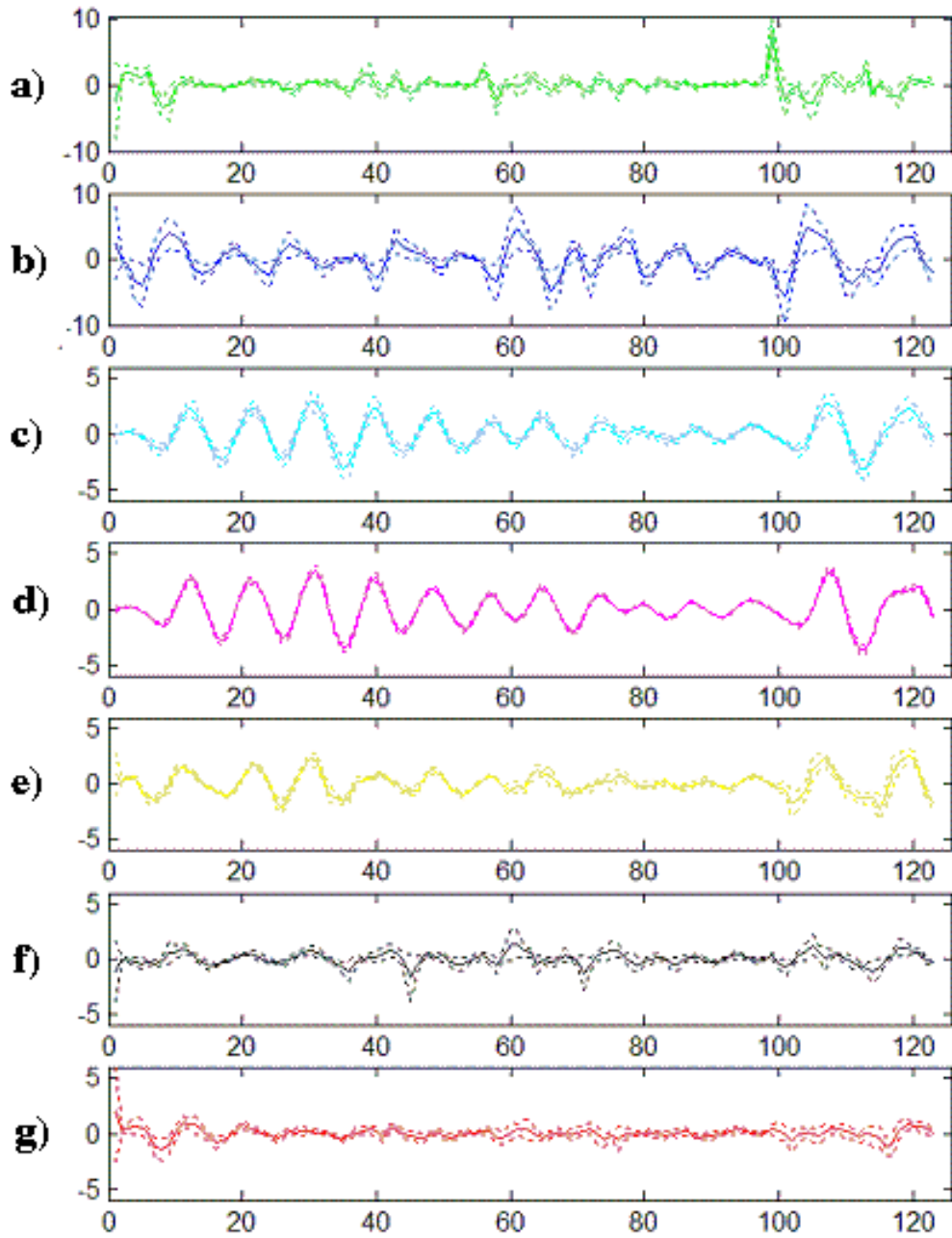


FIG. 5.3 – déplacement (sur l'axe des x de l'image) des groupes terminaux de la hiérarchie (un graphique des valeurs du mouvement sur l'axe des y serait similaire). Les courbes en pointillés dans chaque graphique représentent la valeur moyenne des différences de l'ensemble des déplacements des points appartenant à un groupe avec le mouvement de celui-ci. Ces déplacements sont comparables aux déplacements étudiés par [MS03].

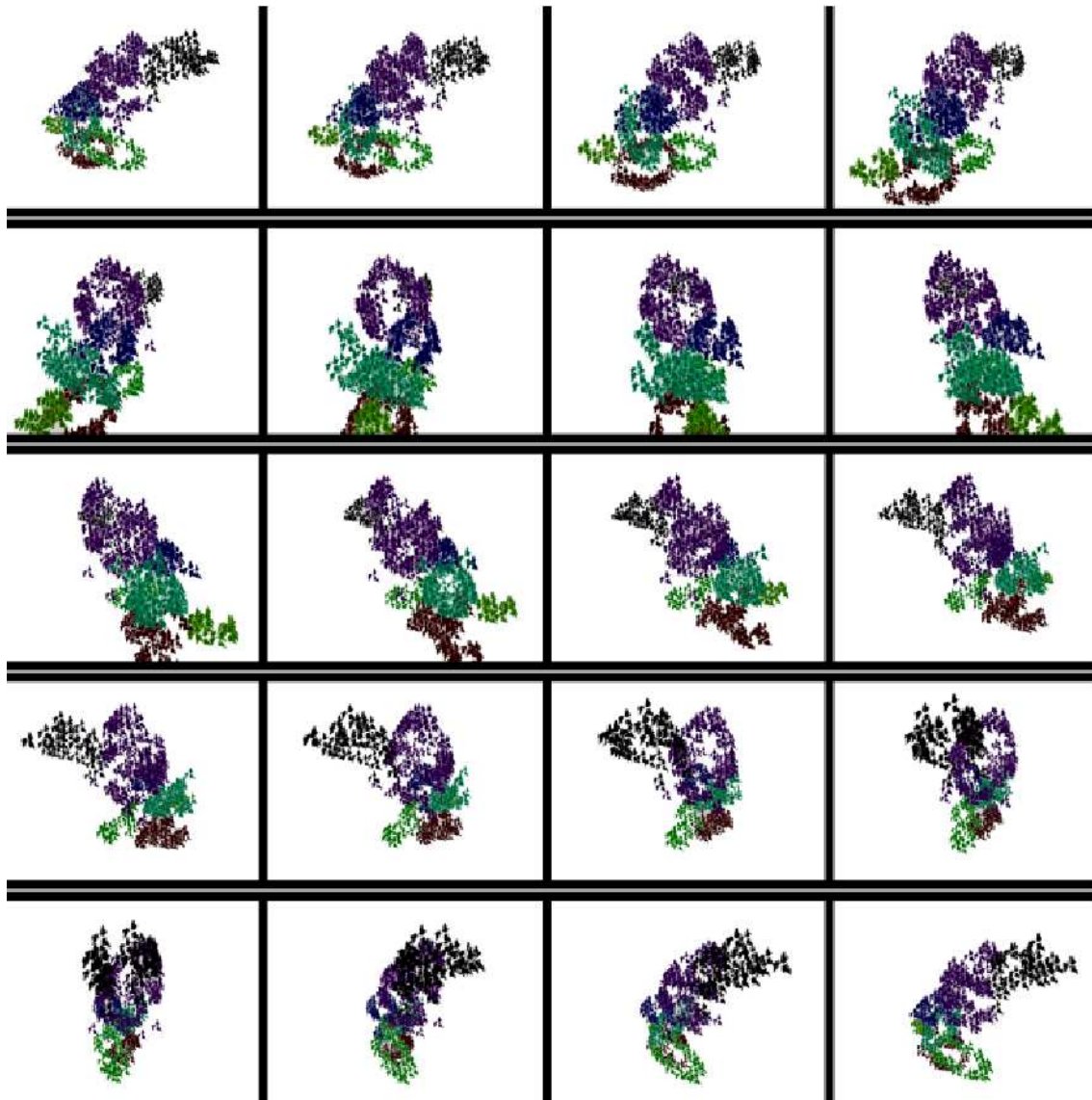


FIG. 5.4 – Modèle 3D final sous plusieurs angles de vue. Les couleurs représentent les groupes.

5.2 Perspectives

Par son côté innovant, la technique que nous avons développée ici, révèle de nombreuses possibilités d'amélioration et d'utilisation dans le cadre de projets plus conséquents.

Tout d'abord, la technique présentée ici représente une réalisation initiale, c'est à dire qu'elle doit être rendue plus robuste pour supporter son application sur d'autres plantes. Ceci passera en premier par la généralisation des techniques de vision utilisées pour le suivi de points. Puis certains choix dans la construction du modèle 3D (en particulier pour sa restructuration , voir 4.4) ont été suffisants dans notre cas, mais pourrait être améliorés.

Ensuite, cette méthode pourrait s'inscrire dans le développement de procédés plus aboutis pour combler la carence actuelle en animation d'arbres dans les applications temps-réel.

5.2.1 Estimation 3D du sujet filmé

Ici, nous avons examiné le mouvement de la vidéo seulement comme un mouvement planaire. Cependant, il serait possible d'estimer une forme 3D du modèle filmé pour obtenir des informations sur son animation plus proche de la réalité et qui permettrait la création d'une structure de modèle 3D de meilleure qualité.

Divers moyens de la sorte seraient réalisables, particulièrement avec une aide simple de l'utilisateur. Nous pourrions, par exemple, définir une forme ovoïde à partir de l'approximation de la silhouette de la plante filmée par une ellipse (moyenne, écart-type). De même, si l'arbre étudié est suffisamment symétrique, une surface de révolution qui lui serait comparable serait aisée à déterminer.

Une fois cette surface spécifiée, la projection des mouvements observés sur elle nous donnerait des informations plus pertinentes sur la position respective des branches et leur structure. Ainsi, la définition de la structure hiérarchique serait naturellement 3D, ce qui augmenterait la qualité du modèle animé généré.

5.2.2 Étude de l'environnement filmé

Notre méthode pourrait servir de pré-traitement des données vidéo dans le cadre d'une étude similaire à [MS03], qui se base sur le suivi du mouvement d'une seule partie de l'arbre indiquée par l'utilisateur.

Grâce à la hiérarchie définie ici, il serait par exemple possible de choisir de façon automatique les zones de l'image à suivre. D'autre part, il serait simplement suffisant de se référer au mouvement des groupes de points qui, étant donné qu'ils sont une moyenne des mouvements d'un nombre possiblement important de points, tendent à supprimer le bruit inhérent des données obtenues par suivi de points. De plus, la figure 5.3 montre bien que les mouvements de ces groupes seront adaptés aux types d'études de courbes utilisées dans [MS03].

Ainsi, nous pourrions obtenir davantage d'informations sur le modèle filmé, telles que la résistance et l'oscillation naturelle de ces branches, et sur les forces qui lui sont appliquées.

5.2.3 Création de cartes d'animation pour une application générique à des modèles 3D

L'estimation, possiblement avec l'aide de l'utilisateur, d'une surface approximative (et minimale) englobant la plante filmée (comme expliqué dans 5.2.1), suivi d'une découpe de cette surface en différentes régions grâce à la méthode présentée ici, nous donnera une carte d'animation composée d'une hiérarchie de points de contrôle (les centres des groupes) liés à une zone d'influence (relative à la taille de la branches représenté par le groupe).

En considérant le modèle 3D comme un volume, un découpage de celui-ci en boîtes englobantes (coordonnées polaires) qui bornent chacune des branches et la probabilité de présence de ces filles, pourra être défini.

Si on enveloppe ce modèle avec la carte d'animation et en prenant en compte certaines caractéristique de l'arbre (ex : axe du tronc, angle moyen entre une branche fille et sa branche mère), il sera alors possible de définir des boîtes englobantes similaires, représentant un échantillonnage du mouvement dans le volume de l'arbre.

Enfin, les mouvements des branches sont pré-calculés par comparaison entre les volumes des boîtes englobantes du modèle et celles obtenues à partir de la carte d'animation : le calcul des

intersections définiraient des coefficients pour une interpolation entre les mouvements des points de contrôle, ou pour l'utilisation de lois probabilistes sur le suivi de ces points.

Bibliographie

- [AR04] G. Drettakis A. Reche, I. Martin. Volumetric reconstruction and interactive rendering of trees from photographs. In *ACM Transactions on Graphics (vol. 23)*, 2004.
- [BL81] T. Kanade B.D. Lucas. An iterative image registration technique with an application to stereo vision. In *Proc. of IJCAI*, pp 674-679, 1981.
- [Bou99] J.-Y. Bouguet. Pyramidal implementation of the lucas kanade feature tracker, description of the algorithm. In *Intel Corporation, Microprocessor Research Labs.*, 1999.
- [FN99] M.-P. Cani F. Neyret. Pattern-based texturing revisited. In *Siggraph International Conference on Computer Graphics and Interactive Techniques*, 1999.
- [FP01] M.-P. Cani F. Perbet. Animating prairies in real-time. In *Symposium on Interactive 3D Graphics*, 2001.
- [JB04] J. Keyser J. Beaudoin. Simulation levels of detail for plant motion. In *ACM Siggraph/Eurographics Symposium on Computer Animation*, 2004.
- [JS94] C. Tomasi J. Shi. Good features to track. In *Proc. of the Conference on Computer Vision and Pattern Recognition*, 1994.
- [LF04] C. Depraz M.-P. Cani L. Favreau, L. Reveret. Animal gaits from video. In *ACM Siggraph/Eurographics Symposium on Computer Animation*, 2004.
- [LL95] M. Piron L. Lebart, A. Morineau. *Statistique Exploratoire Multidimensionnelle*, pp155-176. Dunod, 1995.
- [MR00] C. Tomasi M. Ruzon. Alpha estimation in natural images. In *Proc. of the Conference on Computer Vision and Pattern Recognition*, 2000.
- [MS03] E. Fiume M. Sun, A.D. Jepson. Video input driven animation. In *Ninth IEEE International Conference on Computer Vision*, 2003.
- [PD04] F. Neyret P. Decaudin. Rendering forest scenes in real-time. In *Eurographics Symposium on Rendering (2004)*, 2004.
- [TDG00] F. Faure T. Di Giacomo, S. Capo. An interactive forest. In *Proc. of the Eurographics Workshop on Computer Graphics and Simulation*, 2000.