

CARACTERISATION DU MOUVEMENT DANS LES FILMS D'ANIMATION

B. IONESCU¹, P. LAMBERT², D. COQUIN³, V. BUZULOIU⁴

Acest articol propune o metodă de caracterizare simbolică a mișcării în filmele de animație. Vectorii de mișcare sunt estimați folosind o metodă orientată pe blocuri de pixeli. Utilizând datele obținute, vom construi o matrice de orientări medii și vom calcula un număr de parametri de mișcare. Tipul mișcării este determinat folosind reguli de decizie aplicate parametrilor obținuți. Clasele de mișcare luate în calcul sunt următoarele: mișcarea obiectelor, discontinuitatea mișcării, mișcarea de translație și rotație, mișcarea de tip „zoom”. Folosind rezultatele obținute pentru întregul film, propunem o caracterizare simbolică perceptuală a acestuia, de exemplu: film calm, activitate redusă etc. Ca aplicație analizăm detecția schimbărilor de plan de tip „cut”. O serie de rezultate experimentale sunt prezentate.

This article proposes an approach for the semantic characterization of the animation movie motion content. First, using block-based motion estimation we are computing each frame motion field. Furthermore, the motion is parameterized with the several low-level parameters such as the mean orientation vectors. Each type of motion is then classified using a rule-based approach. The motion types we detect are: the object motion, the motion discontinuity, the booming motion, the dollying motion, the tracking motion, the rotation motion and the zooming motion. Based on these descriptions we propose a global symbolic characterization of the movie content in terms of calmness, activity, etc. Moreover, the motion discontinuity detection is used for the video segmentation by means of cut detection. Several experimental results are presented.

Dans cet article nous proposons une caractérisation symbolique du mouvement présent dans les films d'animation. Pour chaque image du film nous estimons les vecteurs mouvement en utilisant une méthode basée sur les blocs de pixels. A partir des résultats obtenus nous calculons la matrice des orientations moyennes et quelques autres paramètres bas niveau. Le type de mouvement est reconnu en utilisant des règles de décision sur les paramètres obtenus. Les classes de mouvements envisagés sont les suivantes : mouvement d'objets, discontinuité du mouvement, mouvement de translation, rotation et zoom. En analysant le film dans son intégralité une caractérisation symbolique est proposée (par exemple : film

¹ Dr. Ing., Laboratoire de Traitement et Analyse d'Images, Université "POLITEHNICA" Bucarest, Roumanie

² Maître de Conf. HDR, Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance, Université de Savoie, France

³ Maître de Conf., Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance, Université de Savoie, France

⁴ Prof., Laboratoire de Traitement et Analyse d'Images, Université "POLITEHNICA" Bucarest, Roumanie

calme, activité réduite). Nous avons utilisé cette méthode pour évaluer la détection des « cuts ». Enfin nous présentons quelques résultats expérimentaux.

Mots-clés: estimation du mouvement, détection de « cuts », détection du mouvement de caméra, caractérisation symbolique.

1. Introduction

Les systèmes d'indexation (recherche et analyse) d'informations visuelles (vidéo et image) sont entrés dans une nouvelle ère. La première génération de tels systèmes permettait uniquement l'accès au contenu en utilisant des mots dans l'espace textuel. Actuellement, les systèmes d'indexation courants permettent l'accès à l'information visuelle en utilisant des mesures objectives du contenu visuel. Ce type d'analyse permet une interprétation sémantique, très proche de la perception humaine.

La caractérisation du mouvement proposée fait partie intégrante du système de caractérisation symbolique des films d'animation proposé dans le contexte du Festival International du Film d'Animation, qui se déroule depuis 1960, chaque année à Annecy en France (voir [1][2]). La base vidéo des films d'animation créée à la suite de ce festival est très vaste et augmente chaque année. Un système d'analyse qui permet l'accès et l'interprétation du contenu est souhaité.

Le mouvement est un des paramètres principaux de la caractérisation sémantique des films car le mouvement est toujours présent dans ce type de film. L'analyse du mouvement donnera des informations sur l'action et les personnages présents dans le film et fournira des informations sémantiques au niveau du rythme. Il y a deux directions distinctes pour l'analyse du mouvement : le **mouvement de la caméra** et le **mouvement des objets**. La détection et le suivi des objets sont connus dans la littérature par les méthodes "objects tracking" qui consistent en la détection de la direction du déplacement d'un objet spécifique sur plusieurs images (donc dans le temps), et également sa localisation dans un plan. Différentes méthodes sont proposées dans [3][4]. L'analyse du mouvement de la caméra permet d'avoir une caractérisation globale de l'action du film. Par exemple, les changements des plans vidéo utilisent des mouvements de translation, une action importante est souvent suivie d'un mouvement rapide, etc. Différentes techniques d'analyse sont proposées dans [5][6].

Dans ce papier nous proposons une méthode de caractérisation et d'interprétation sémantique du mouvement appliquée aux films d'animation. La méthode est basée sur les travaux proposés dans [7], où une caractérisation du mouvement de la caméra par les réseaux neuronaux est proposée. Pour l'estimation des vecteurs mouvement nous avons choisi d'utiliser une méthode basée sur les blocs de pixels, méthode efficace et rapide qui est utilisée dans le

codage vidéo. Cette approche estime le déplacement des blocs de pixels en minimisant une fonction de coût. D'autres approches plus lentes utilisent le gradient, des recherches récursives ou le flux optique (voir [8][9]).

Pour chaque image courante, nous proposons un certain nombre de paramètres statistiques, qui concernent la quantité du mouvement dans la scène. Ces paramètres sont calculés à partir des vecteurs de mouvement obtenus précédemment. Ainsi, le mouvement moyen est calculé par des statistiques circulaires en utilisant un partage en blocs 3x3 du champ vectoriel des images analysées. Les paramètres bas niveau obtenus sont également utilisés pour caractériser le mouvement selon les classes suivantes : pas du mouvement, mouvement d'objets, discontinuité du mouvement ou mouvement de la caméra.

Pour la caractérisation sémantique nous proposons un certain nombre de paramètres statistiques calculés sur l'ensemble du film. Ces paramètres permettent de caractériser l'ensemble du film comme : *calme*, *rapide*, *contenant des nombreux changements de point de vue*, etc. Nous avons utilisé la discontinuité du mouvement pour évaluer également la détection des changements abrupts de plans ou « cuts » (voir [10]). Enfin nous présentons certains résultats expérimentaux sur un certain nombre de films d'animation.

2. L'algorithme proposé

Comme nous l'avons déjà mentionné dans l'Introduction, le but de l'analyse proposée est de caractériser les différents mouvements de la caméra par des **paramètres symboliques**. Un exemple d'une telle caractérisation peut être : film calme (pas du mouvement) ou activité élevée (les mouvements prédominants sont des mouvements d'objets). La chaîne de traitements proposés est présentée dans la figure Fig. 1.

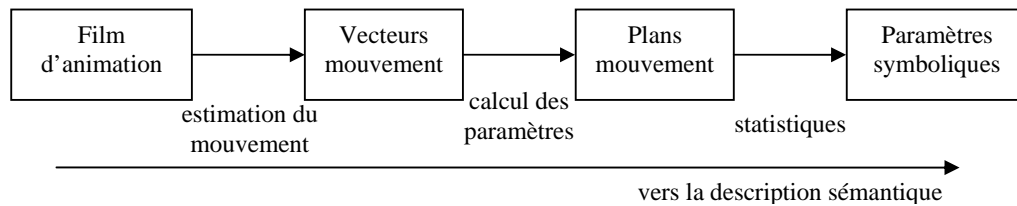


Fig.1. La chaîne de traitement pour la caractérisation symbolique du mouvement.

La première étape d'analyse est *l'estimation des vecteurs mouvement* en utilisant une méthode d'estimation du mouvement basée sur les blocs de pixels. Pour chaque image du film et pour chaque bloc de $n \times n$ pixels (sans recouvrement) on calcule les vecteurs de déplacements (amplitude et orientation, voir Section 2).

La deuxième étape consiste à utiliser ces vecteurs mouvement pour calculer un certain nombre de paramètres qui vont nous permettre de caractériser les différents types de mouvement. Chaque image du film sera caractérisée à partir du mouvement détecté. Les *plans du mouvement* sont obtenus en fusionnant les résultats obtenus, exemple: image0-image30= mouvement de rotation (voir Section 6). Puis, en utilisant les informations obtenues sur la discontinuité du mouvement nous proposons une méthode de détection des « cuts » (voir Section 7). Un « cut » est une transition rapide entre deux plans vidéo. Pour plus d'information sur la segmentation temporelle d'un film voir [10].

La dernière étape est *l'analyse sémantique*. En utilisant des statistiques sur le pourcentage d'apparition des différents mouvements, nous proposons une caractérisation symbolique du film (exemple: film calme, activité élevée, voir Section 7).

3. L'estimation du mouvement

Pour l'estimation du mouvement nous avons choisi d'utiliser une méthode basée sur l'analyse des blocs de pixels. Ce type d'approche est moins exacte qu'une méthode basée sur le flux optique mais plus efficace et adaptée à nos besoins. Pour la caractérisation globale du mouvement un champ vectoriel moins dense est envisageable. Le principe d'une méthode d'estimation orientée sur les blocs de pixels est de minimiser une fonction de coût pour trouver dans l'image suivante, dans une fenêtre de recherche limitée, la nouvelle position du bloc courant analysé. Les blocs de pixels sont de taille $n \times n$ pixels avec $n = 16$. Nous avons utilisé comme fonction de coût la valeur absolue de l'erreur moyenne (notée MAE) définie par :

$$v = \arg \min_{v \in S} MAE\{f(r, t + dt), f(r - d, t)\} \quad (1)$$

où v est le vecteur mouvement du bloc courant analysé $f(r, t)$, à la position r et au temps t , S est la fenêtre de recherche, $S = (2n + 1)(2n + 1)$ pixels, centrée sur le bloc courant.

Les méthodes par blocs de pixels diffèrent les unes des autres par la méthode de d'optimisation de la recherche dans la fenêtre S . Différents algorithmes ont été proposés: des algorithmes en trois étapes, la recherche logarithmique, la recherche hiérarchique, la recherche hybride, la recherche au demi-pixel et la recherche complète. Afin d'obtenir la méthode la plus exacte possible, nous avons choisi d'utiliser la recherche complète. Nous allons donc chercher le minimum de la fonction MAE dans toute la fenêtre de recherche.

Puis, en utilisant un seuillage des valeurs du MAE on peut trouver des situations pour lesquelles on a une discontinuité du mouvement, c'est à dire que le bloc courant analysé n'a pas forcément une nouvelle position dans l'image

suivante (situation qui correspond aux changements de plans ou mouvements très rapides). Le seuil de discontinuité, $s_{discont}$ a été choisi empiriquement après l'analyse de différents exemples des passages de changement de plan et passages continus. Des résultats expérimentaux sont présentés dans la Section 8. Un exemple de champ vectoriel est présenté dans la figure 2⁵.

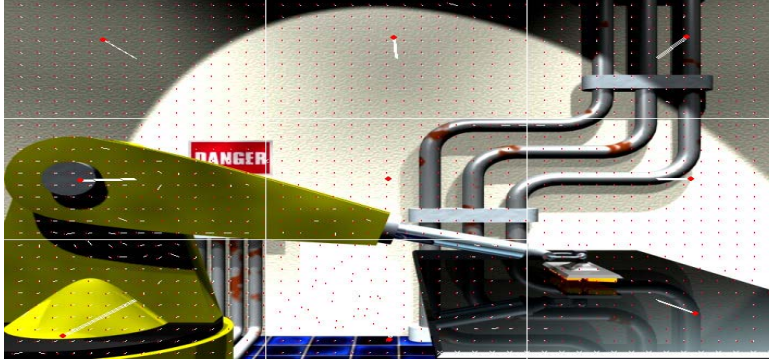


Fig. 2. Exemple d'estimation du mouvement « zoom in » (les vecteurs de mouvement sont représentés par des petites flèches proportionnelles à l'amplitude du déplacement, les orientations moyennes normalisées par la taille de la région sont représentées par de grandes flèches).

4. Les paramètres bas-niveau

Dès que l'estimation du mouvement est finie, chaque image du film est caractérisée par une matrice de vecteurs correspondant à l'amplitude $Ampl(i)(j)$ et à l'orientation $Orient(i)(j)$ du mouvement des blocs de pixels, avec $i = 0..N_{OX}, j = 0..N_{OY}$ où N_{OX}, N_{OY} sont le nombre de blocs sur les axes oX et oY . Nous proposons deux paramètres calculés sur la matrice d'amplitudes $Ampl()$ qui permettront de caractériser globalement le mouvement ou la discontinuité du mouvement. Ces deux paramètres sont définis par:

$$R_{mouvement} = \frac{N_{mouvement}}{N_{total}} \quad \text{et} \quad R_{discont} = \frac{N_{discont}}{N_{total}} \quad (2)$$

où $N_{mouvement}$ est le nombre de blocs de pixels qui ont une valeur de l'amplitude différente de 0, $N_{discont}$ est le nombre de blocs de pixels qui n'ont pas de continuité temporelle du mouvement dans l'image suivante et N_{total} est le nombre total de blocs dans l'image.

Une autre caractérisation proposée est la matrice des vecteurs moyens du mouvement. L'image courante analysée est découpée en 3×3 blocs. En utilisant des statistiques circulaires (voir [11]) sur chacun de 9 blocs considérés, on calcule

⁵ pour la version couleur voir la section des publications sur <http://alpha.imag.pub.ro>

l'orientation et l'amplitude moyenne des vecteurs de mouvement contenus dans chaque région de l'image. Etant donné n valeurs angulaires θ_i avec $i=1..n$, la direction moyenne est la direction du vecteur résultant de la somme des vecteurs unitaires dans les n directions θ_i . La direction du vecteur résultant est définie par $\arctan\left(\frac{B}{A}\right)$, ou A et B sont définis par :

$$A = \sum_i \cos(\theta_i) \quad \text{et} \quad B = \sum_i \sin(\theta_i) \quad (3)$$

Pour chaque image on obtient deux matrices de dimensions 3×3 qui correspondent aux orientations et aux amplitudes moyennes de chaque région: $O_m(k,l)$ et $A_m(k,l)$ avec $k,l \in \{0,1,2\}$ (voir figure Fig. 2).

5. Les classes du mouvement à détecter

Parmi les différents mouvements possibles de la caméra, l'algorithme proposé permet de prendre en compte les situations suivantes : « **pas de mouvement** » - donc absence de mouvement; « **discontinuité du mouvement** » - correspond aux changements de plans vidéo, aux transitions et au mouvement très rapide de la caméra; « **mouvement d'objets** » - déplacements des régions de pixels ou mouvement très faible; « **zoom in/out** » - agrandissement/rétrécissement d'image ; « **déplacement vers la gauche/droite** »; « **déplacement vers le haut ou vers le bas** » et « **rotation dans le sens horaire/anti-horaire** ».

6. Les règles de décision

En analysant les paramètres bas niveau proposés ($R_{mouvement}$, $R_{discont}$ et $O_m(k,l)$, $A_m(k,l)$ avec $k,l \in \{0,1,2\}$) nous avons défini les règles de classification pour les mouvements énumérés dans la section précédente. En ce qui concerne les mouvements de caméra, les conditions sur les paramètres bas niveau reposent sur les travaux proposés dans [7]. Chaque situation de mouvement est caractérisée par un ensemble de règles disjointes:

- « **pas de mouvement** »: condition : $R_{mouvement} \leq 0.01$.
- « **discontinuité du mouvement** »: condition $R_{discont} > 0.5$.
- « **mouvement d'objets** »: condition $R_{mouvement} > 0.01$ ET $R_{mouvement} \leq 0.5$ ⁶.

Si $R_{mouvement} > 0.5$, alors on se trouve dans la situation d'un mouvement de caméra. Les distributions des vecteurs moyens du mouvement (définis par O_m et

⁶ Les valeurs de ces seuils ont été déterminées empiriquement en analysant plusieurs séquences.

A_m) pour les différents types de mouvements de caméra à détecter sont présentés par les figures Fig. 4 et Fig. 5 présentées ci-dessous.

Pour vérifier l'appartenance à une des situations présentées ci-dessus on analyse la matrice $O_m(k,l)$ où $k,l \in \{0,1,2\}$. Seuls les vecteurs qui ont une amplitude supérieure à un pixel sont pris en compte, c'est-à-dire $A_m(k,l) > 0$. Les angles sont comparés en utilisant un intervalle de tolérance, $d = \pm 45$ degrés pour les mouvements de translation et les situations 2 et 3 du « zoom » et $d = \pm 22$ degrés pour les mouvements de rotation et la première situation du « zoom » (voir figure Fig. 4.a. et Fig. 4.b.). Le principe est présenté par la figure Fig. 6.

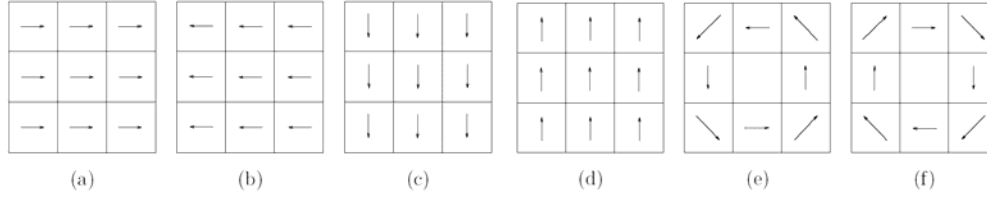
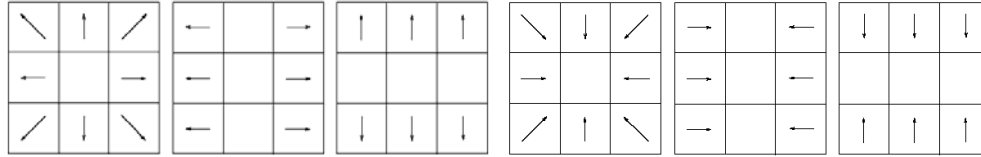


Fig. 5. Orientations des vecteurs moyens pour: (a)/(b) « déplacement vers la gauche/droite », (c)/(d) « déplacement vers le haut/bas », (e)/(f) « rotation horaire/anti-horaire ».

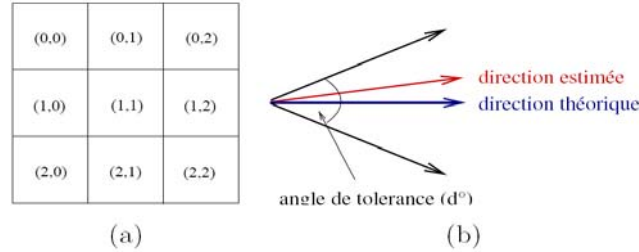


Fig. 6. (a) Les index pour les vecteurs moyens du mouvement, (b) La tolérance des valeurs des angles (la direction estimée et théorique)

En utilisant les notations de la figure Fig. 6.a. nous avons défini par décisions majoritaires les règles de classification suivantes (où N,E,S,O sont les 4 points cardinaux):

- « **zoom in** » : plusieurs situations sont possibles.

Situation 1 : (voir première image de la figure Fig. 4.a.) :

$$\text{Card} \left\{ \begin{array}{l} O_m(0,0) = NO; O_m(0,1) = N; O_m(0,2) = NE; O_m(1,0) = O; \\ O_m(1,2) = E; O_m(2,0) = SO; O_m(2,1) = S; O_m(2,2) = SE \end{array} \right\} \geq 7 \quad (4)$$

Situation 2 : (voir deuxième image de la figure Fig. 4.a.) :

$$\begin{aligned} \text{Card}\{O_m(i,0) = O; O_m(i,2) = E / i = 0..2\} &\geq 4 \text{ ET} \\ \text{Card}\{O_m(i,1) = O / i = 0..2\} &< 2 \text{ ET } \text{Card}\{O_m(i,1) = E / i = 0..2\} < 2 \end{aligned} \quad (5)$$

Situation 3 : (voir troisième image de la figure Fig. 4.a.) :

$$\begin{aligned} \text{Card}\{O_m(0,j) = N; O_m(2,j) = S / j = 0..2\} &\geq 4 \text{ ET} \\ \text{Card}\{O_m(1,j) = N / j = 0..2\} &< 2 \text{ ET } \text{Card}\{O_m(1,j) = S / j = 0..2\} < 2 \end{aligned} \quad (6)$$

Dans la deuxième et troisième situation de « zoom in » l'orientation des vecteurs moyens est similaire à l'orientation des vecteurs moyens correspondant à un mouvement de translation, ce qui justifie l'ajout des deux dernières règles.

- « **zoom out** » : similaire au « zoom in », les conditions sont :

Situation 1 : (voir figure Fig. 4.b. image 1) :

$$\text{Card}\left\{ \begin{array}{l} O_m(0,0) = SE; O_m(0,1) = N; O_m(0,2) = SO; O_m(1,0) = E; \\ O_m(1,2) = O; O_m(2,0) = NE; O_m(2,1) = N; O_m(2,2) = NE \end{array} \right\} \geq 7 \quad (7)$$

Situation 2 : (voir figure Fig. 4.b. image 2) :

$$\begin{aligned} \text{Card}\{O_m(i,0) = E; O_m(i,2) = O / i = 0..2\} &\geq 4 \text{ ET} \\ \text{Card}\{O_m(i,1) = E / i = 0..2\} &< 2 \text{ ET } \text{Card}\{O_m(i,1) = O / i = 0..2\} < 2 \end{aligned} \quad (8)$$

Situation 3 : (voir figure Fig. 4.b. image 3) :

$$\begin{aligned} \text{Card}\{O_m(0,j) = S; O_m(2,j) = N / j = 0..2\} &\geq 4 \text{ ET} \\ \text{Card}\{O_m(1,j) = S / j = 0..2\} &< 2 \text{ ET } \text{Card}\{O_m(1,j) = N / j = 0..2\} < 2 \end{aligned} \quad (9)$$

- « **déplacement vers la droite** » (voir figure Fig. 5 image b) :

$$\text{Card}\{O_m(i,j) = O / i, j = 0..2\} \geq 6 \quad (10)$$

- « **déplacement vers la gauche** » (voir figure Fig. 5 image a) :

$$\text{Card}\{O_m(i,j) = E / i, j = 0..2\} \geq 6 \quad (11)$$

- « **déplacement vers le haut** » (voir figure Fig. 5 image c) :

$$\text{Card}\{O_m(i,j) = S / i, j = 0..2\} \geq 6 \quad (12)$$

- « **déplacement vers le bas** » (voir figure Fig. 5 image d) :

$$\text{Card}\{O_m(i,j) = N / i, j = 0..2\} \geq 6 \quad (13)$$

- « **rotation dans le sens anti-horaire** » (voir figure Fig. 5 image e) :

$$\text{Card}\left\{ \begin{array}{l} O_m(0,0) = NE; O_m(0,1) = E; O_m(0,2) = SE; O_m(1,0) = N; \\ O_m(1,2) = S; O_m(2,0) = NO; O_m(2,1) = O; O_m(2,2) = SO \end{array} \right\} \geq 6 \quad (14)$$

- « **rotation dans le sens horaire** » (voir figure Fig. 5 image f) :

$$\text{Card}\left\{ \begin{array}{l} O_m(0,0) = SO; O_m(0,1) = O; O_m(0,2) = NO; O_m(1,0) = S; \\ O_m(1,2) = N; O_m(2,0) = SE; O_m(2,1) = E; O_m(2,2) = NE \end{array} \right\} \geq 6 \quad (15)$$

- « **autre mouvement** » : aucune des situations ci-dessus n'est valable.

7. Fusion des résultats : les plans du mouvement

Dès que la classification du mouvement pour le film est finie, chaque image du film est caractérisée par un "label" caractérisant le type de mouvement. Les plans du mouvement sont obtenus en regroupant les mouvements similaires. Dans de nombreuses situations, on rencontre le fait que les plans du mouvement sont artificiellement divisés par de fausses détections de courte durée ; à l'intérieur d'un plan on trouvera d'autres types de mouvement. Pour corriger les résultats nous proposons une étape de fusion. Pour le plan du mouvement courant analysé, plan_i et mouvement B, on cherche si le même type de mouvement est présent dans un plan voisin (dans un intervalle de temps limité, noté T_r) en suivant la chronologie du film. Si le même type de mouvement B, a été rencontré dans le plan_j à l'intérieur de l'intervalle T_r , tous les plans compris entre plan_i et plan_j sont regroupés en un plan de mouvement B qui devient le plan courant et l'algorithme est appliqué de la même façon pour la suite du film. Un cas particulier est le mouvement de type « **discontinuité du mouvement** » pour lequel on ne change rien.

8. Les paramètres symboliques proposés

Pour avoir la description sémantique du mouvement nous proposons un certain nombre de descripteurs bas niveau calculés en utilisant les plans du mouvement obtenus pour tout le film. Les deux premiers paramètres proposés estiment le pourcentage des passages du film sans mouvement et les passages du film dans lesquels on rencontre des mouvements d'objets :

$$P_{\text{pas de mouvement}} = \frac{N_{\text{pas de mouvement}}}{N_{\text{images}}}, P_{\text{mouvement d'objets}} = \frac{N_{\text{mouvement d'objets}}}{N_{\text{images}}} \quad (16)$$

où $N_{\text{pas de mouvement}}$ est le nombre total d'images sans mouvement, $N_{\text{mouvement d'objets}}$ est le nombre total d'images qui comportent un mouvement d'objets et N_{images} est le nombre total des images du film. Le pourcentage des passages du film avec un mouvement de translation est calculé comme :

$$P_{\text{translation}} = \frac{N_{\text{dépl.droite}} + N_{\text{dépl.gauche}} + N_{\text{dépl.haut}} + N_{\text{dépl.bas}}}{N_{\text{images}}} \quad (17)$$

où $N_{\text{dépl.X}}$ est le nombre total d'images qui comportent un mouvement de type déplacement vers X. Les pourcentages de mouvements de rotation et « zoom » sont définis comme :

$$P_{\text{rotation}} = \frac{N_{\text{rot.horaire}} + N_{\text{rot.anti.horaire}}}{N_{\text{images}}}, P_{\text{zoom}} = \frac{N_{\text{zoom in}} + N_{\text{zoom out}}}{N_{\text{images}}} \quad (18)$$

où $N_{\text{rot.X}}$ est le nombre total d'images qui comportent un mouvement de rotation dans le sens X, $N_{\text{zoom in}}$ et $N_{\text{zoom out}}$ sont respectivement le nombre total d'images qui comportent un mouvement de type « zoom in » et « zoom out ».

En utilisant les paramètres obtenus nous proposons les descripteurs symboliques suivants : caractérisation d'un film **calme** (lié à $P_{\text{pas de mouvement}}$) ; **l'activité** contenue dans le film (lié à $P_{\text{mouvement d'objets}}$) ; les **changements de point de vue** effectués en utilisant des translations ou des mouvements 3D de la caméra (lié à $P_{\text{translation}}$) ; les **effets de prise de vue** de la caméra (lié à P_{rotation}) ; les **détails** sur des différents objets ou régions d'intérêt du film (lié à P_{zoom}). Des résultats expérimentaux sont présentés dans la Section 8.

9. Résultats expérimentaux

Pour valider la méthode proposée concernant la classification des différents types de mouvements, nous avons utilisé un certain nombre de séquences de synthèse générées à partir d'un logiciel de graphique 3D (3 exemples pour chaque type de mouvement, images de taille 720x486 pixels). Les taux de bonne détection obtenus sont présentés dans la figure Fig. 7. Pour les tests nous avons utilisé comme seuils $R_{\text{discont}} = 0.6$ et $s_{\text{discont}} = 20000$ (sections 2 et 5).

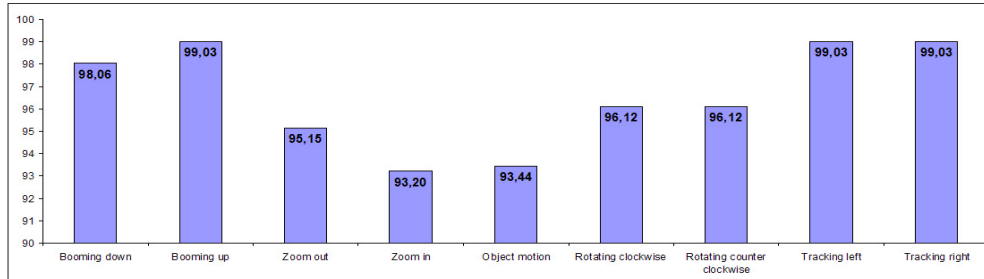


Fig. 7. Les taux de bonne détection pour chaque mouvement.

Nous avons obtenu des taux de détection supérieure à 93% pour tous les mouvements. Les fausses détections sont liées à la non-linéarité du mouvement ; il y a des passages avec un mouvement trop faible ou des situations pour lesquelles les vecteurs mouvement sont nuls à cause de la similarité des textures dans l'image. Le plus mauvais taux de détection est obtenu pour le déplacement d'objet, difficile à détecter précisément avec une approche par blocs. La matrice de confusion obtenue est présentée dans le tableau Table 1 (d.b/d.h=déplacement vers le bas/haut, z.o/z.i.=zoom out/in, m.o.=mouvement d'objets, r.h./r.ah.=rotation horaire/anti-horaire, d.g./d.d.=déplacement vers la gauche/droite et p.m.=pas de mouvement, les chiffres expriment le nombre d'images).

La détection du mouvement permet aussi la détection des changements de plans vidéo. Un changement abrupt ou « cut » est représenté par un plan du mouvement de durée égale à une image, de type « discontinuité du mouvement ».

Table 1.

La matrice de confusion pour la détection du mouvement sur les séquences de synthèse

	d.b.	d.h.	z.o.	z.i.	m.o.	r.h.	r.ah.	d.g.	d.d.	p.m.
Nb.img	103	103	103	103	122	103	103	103	103	
d.b.	101	0	0	0	2	0	0	0	0	0
d.h.	0	102	0	0	1	0	0	0	0	0
z.o.	0	0	98	0	5	0	0	0	0	0
z.i.	0	0	0	96	6	0	0	0	0	1
m.o.	0	0	0	0	114	0	0	0	0	8
r.h.	0	0	0	0	4	99	0	0	0	0
r.ah.	0	0	0	0	4	0	99	0	0	0
d.g.	0	0	0	0	1	0	0	102	0	0
d.d.	0	0	0	0	1	0	0	0	102	0

Pour tester les performances nous avons également utilisé deux longs métrages du même studio de production, avec un temps total de 158 minutes et un nombre de 3166 « cuts ». Les films ont été analysés manuellement pour effectuer la vérité terrain. Les erreurs de détection sont exprimées en utilisant la précision et le rappel définis comme :

$$precision = \frac{BD}{BD + FD}, rappel = \frac{BD}{Nb} \quad (18)$$

où BD=bonnes détections, FD=fausses détections et Nb=nombre total de « cuts ». Les résultats obtenus sont présentés dans le tableau Table 2 ($s_{discont} = 10000$). La précision globale obtenue est de **89.39%** et le rappel de **94.53%**. En utilisant les mêmes films que précédemment, nous avons testé un algorithme de détection des « cuts » basé sur l'approche classique sur les histogrammes couleur [10], et avons obtenu une précision de **93.37%** et un rappel de **88.63%**.

Table 2.

Les résultats de la détection des « cuts » en utilisant le mouvement

	Nb. « cuts »	BD	FD	Précision	Rappel
Film. 1	1597	1492	134	91.76%	93.43%
Film. 2	1569	1501	221	87.17%	95.66%

Nous pouvons conclure que l'utilisation du mouvement donne une amélioration de **6%** pour le rappel mais une réduction de la précision de **4%**.

En utilisant ces mêmes films nous avons calculé les paramètres symboliques proposés dans la Section 7. Nous avons obtenu des résultats similaires. Il semble donc que l'on puisse, grâce à ces outils, caractériser le style propre à un réalisateur, une sorte de « signature » qui traduit l'utilisation de techniques de mouvement identiques dans des films différents (voir figure Fig. 8).

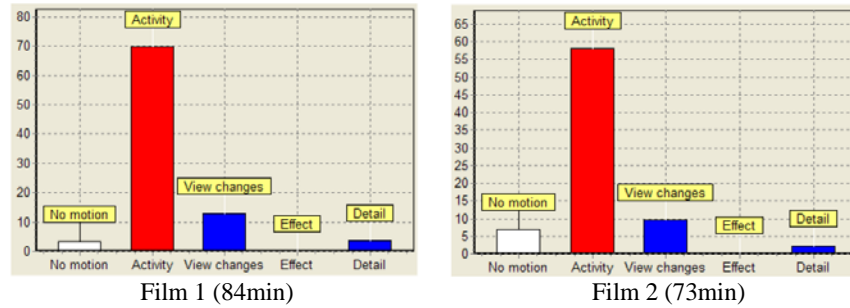


Fig. 8. Exemple de paramètres symboliques pour deux films d'animation du même producteur.

10. Conclusions

Dans cet article nous avons proposé une méthode de caractérisation symbolique appliquée aux films d'animation. Des descripteurs symboliques sont calculés à partir des paramètres bas niveau. Nous avons proposé également une méthode d'analyse du mouvement de caméra en utilisant l'estimation du mouvement. Cette caractérisation du mouvement nous permet de plus, de proposer une méthode de détection des « cuts ». Des résultats expérimentaux, sur plusieurs films et séquences d'animation, viennent confirmer l'intérêt de ces méthodes.

REFERENCES

- [1] *Centre International du Cinéma d'Animation*: <http://www.annecy.org>
- [2] B. Ionescu, D. Coquin, P. Lambert, V. Buzuloiu: Analyse et Caractérisation de Séquences de Films d'Animation, ORASIS journées francophones des jeunes chercheurs en vision par ordinateur, Fourniois - Puy-de-Dôme, France, 24-27 mai 2005.
- [3] Y. Luo, T.-D. Wu, J.-N. Hwang: Object-Based Analysis and Interpretation of Human Motion in Sports Video Sequences by Dynamic Bayesian Networks, *Computer Vision and Image Understanding*, **92**, 2003, pp 196-216.
- [4] H. Denman, N. Rea, A. Kokaram: Content-based analysis for video from snooker broadcasts, *Computer Vision and Image Understanding*, **92**, 2003, pp 176-195.
- [5] C. Colombo, A. Del Bimbo, P. Pala: Semantics in Visual Information Retrieval, *IEEE MultiMedia*, **6**(3), 1999, pp 38-53.
- [6] A. Del Bimbo: Visual Information Retrieval, 1999, pp 203-259.
- [7] M.-S. Lee, Y.-M. Yang, S.-W. Lee: Automatic Video Parsing Using Shot Boundary Detection and Camera Operation Analysis, *Pattern Recognition*, **34**, 2001, pp 711-719.
- [8] F.J. Hampson, J.-C. Pesquet: Motion Estimation in the Presence of Illumination Variations, *Signal Processing: Image Communication*, **16**, 1998, pp 373-381.
- [9] M. Accame, F.G.B. De Natale, D.D. Giusto, High Performance Hierarchical Block-based Motion Estimation for Real-Time Video Coding, *Real-Time Imaging*, **4**, 1998, pp 67-79.
- [10] C.G.M. Snoek, M. Worring: Multimodal Video Indexing: A Review of the State-of-the-art, *Multimedia Tools and Applications*, 2005, sous-presse.
- [11] N.I. Fisher: Statistical Analysis of Circular Data, Cambridge University Press, 1993.