

ANALYSE DE MOUVEMENTS GLOBAUX DANS LES SEQUENCES
 D'IMAGES ET APPLICATION AU CODAGE DE TELEVISION

Jean-Luc DUGELAY

et

Henri SANSON

CCETT

4, rue du Clos Courtel, F 35510 - Cesson Sévigné, France

RÉSUMÉ

Cet article présente une approche qui combine méthode différentielle et analyse hiérarchique de l'image pour l'identification de mouvements globaux par zone d'image. Deux algorithmes sont proposés, le premier se situe dans une approche purement paramétrique et réalise l'identification de modèles affines par blocs, le second est plus orienté vers l'analyse du mouvement 3D et réalise conjointement la segmentation et l'identification d'un modèle du deuxième ordre pour chaque région obtenue.

1. INTRODUCTION

L'analyse de mouvement dans les séquences d'images occupe une place importante dans les techniques de compression d'images animées. L'étude présentée dans cet article s'inscrit dans le cadre du codage de la Télévision Stéréoscopique pour une transmission sur un canal de type MAC. La particularité d'une telle application est que la compression requise ne peut être obtenue que par sous-échantillonnage et interpolation temporels d'une part, ou bien par sous-échantillonnage et filtrage spatiaux d'autre part. Ce type de codage confère donc à l'analyse du mouvement un rôle prépondérant. Cependant, l'amélioration des techniques d'estimation et de compensation du mouvement devient un objectif incontournable en codage d'images animées, car l'exploitation de la redondance temporelle permet d'espérer atteindre de très bas débits.

Les approches classiques de l'utilisation du mouvement en codage d'images de télévision consistent à attribuer un vecteur de translation par bloc de pixels dans l'image. Le débit alloué à l'information de mouvement est ainsi limité si la taille des blocs est assez importante (typiquement 8x8 au minimum) et la précision sur les vecteurs faible (1/2 pixel). Par contre, les zones d'image pouvant être reconstruites de façon satisfaisante s'en trouvent réduites. Bien plus que les performances intrinsèques des estimateurs utilisés, c'est le choix des descripteurs du mouvement qui se révèle inadéquat pour de nombreux cas réels:

- le modèle de translation se révèle insuffisant pour prendre en compte certains types de mouvement, sauf si on sur-segmente localement le champ des déplacements, auquel cas le débit alloué à ce champ augmente de façon importante.

- la structure de blocs est en général sans aucun rapport avec le contenu des images et rend difficile la prise en compte des discontinuités du mouvement, aux frontières des objets animés.

Nous proposons ici d'utiliser des modèles de mouvement plus généraux que la translation, dans le but de globaliser la description de celui-ci et d'aboutir à une compensation efficace à l'aide d'un modèle par zone d'image relativement grande. L'objectif est de parvenir à une compensation performante avec peu d'information pour décrire le mouvement.

Le présent article décrit deux algorithmes développés dans cette optique. Le premier réalise l'identification de modèles affines par bloc de taille fixe, mais son extension à l'identification de modèles d'ordre supérieur est immédiate. Le second est plus ambitieux et réalise conjointement l'estimation, à un facteur près, de paramètres du mouvement tri-dimensionnel et la segmentation de l'image en régions où ces paramètres sont valides, dans le cas où les paramètres de calibration intrinsèque de la caméra sont connus.

ABSTRACT

This paper presents an approach which combines a differential method and hierarchical image analysis to identify global motion per image area. Two algorithms are proposed, the first, purely parametric in nature, identifies affine models per block, the second is oriented more specifically towards 3D motion analysis and performs jointly the segmentation and the identification of a physical second order model for each region.

Son emploi s'inscrit particulièrement bien dans un codage orienté objet.

Les deux méthodes reposent sur des idées communes telles que l'emploi de modèles de mouvement par région, les méthodes différentielles pour leur identification, combinées à une analyse hiérarchique de l'image.

2. MODELES DE MOUVEMENT

L'attribution d'un vecteur de translation à une région de l'image constitue une approximation d'ordre zéro de son champ de vitesse, qui n'est valide que si cette région correspond à un objet animé d'une translation pure dans un plan de l'espace parallèle au plan image. Lorsque ce n'est pas le cas, l'utilisation d'une approximation d'ordre plus élevée s'avère nécessaire. Ceci constitue une approche purement paramétrique pour une analyse bidimensionnelle, où, pour une région donnée, on cherche le modèle optimal selon une critère donné, l'erreur de compensation par exemple. Le choix de l'ordre du modèle résulte d'un compromis entre la qualité de compensation espérée, la taille des régions et la complexité et le débit résultants.

Une seconde approche, physique, consiste à schématiser les aspects cinématiques et géométriques 3D ainsi que le processus de formation des images par la caméra pour en déduire le modèle de mouvement 2D. On constate que le modèle quadratique du champ de vitesse permet de représenter de nombreux cas réels [1]. Dans ce cas, le champ des vitesses pour une région donnée s'exprime de la façon suivante:

$$\begin{aligned} v_x &= a_x^1 + a_x^2 \cdot x + a_x^3 \cdot y + a_x^4 \cdot x^2 + a_x^5 \cdot x \cdot y + a_x^6 \cdot y^2 \\ v_y &= a_y^1 + a_y^2 \cdot x + a_y^3 \cdot y + a_y^4 \cdot x^2 + a_y^5 \cdot x \cdot y + a_y^6 \cdot y^2 \end{aligned} \quad (1)$$

où v_x et v_y représentent les composantes du vecteur vitesse au point de coordonnées (x,y) .

Cette approche est pleinement justifiée dès que l'on souhaite déterminer les paramètres du mouvement 3D, moyennant la connaissance de la calibration intrinsèque de la caméra. Elle repose, par contre sur quelques hypothèses qui peuvent s'avérer restrictives, telles que celles d'un mouvement rigide et de régions correspondant à des facettes planes. Si ces hypothèses ne sont pas satisfaites, on risque d'aboutir à une sur-segmentation de l'image avec des régions de taille réduite. Dans ce cas, l'emploi des termes du second ordre ne serait plus justifié, car induisant des variations de vitesse négligeables. Elle donne cependant une idée de l'ordre des modèles admissibles.



3. ESTIMATION

Actuellement, 2 approches sont possibles pour l'identification de modèles de mouvement [2].

- l'identification indirecte, ou a posteriori, qui consiste à approcher au mieux un champ dense préalablement estimé par une méthode classique, sur une zone donnée.
- l'identification directe, où les paramètres du modèle sont estimés à partir de la séquence, sans étape intermédiaire.

Seule la seconde est envisagée ici, car les estimateurs de flot optique conduisent en général à un champ irrégulier et peu significatif, rendant délicate une approximation par des modèles d'ordre peu élevé. Le deuxième algorithme présenté ici, utilise cependant une interprétation du flot optique en phase d'initialisation.

Le critère choisi pour évaluer la qualité d'un modèle est l'erreur quadratique moyenne de reconstruction de la région de l'image à l'instant t à partir de l'image précédente, soit:

$$C(R, a_i) = \sum_{\vec{r} \in R} \text{TFD}^2(\vec{r}, a_i), \text{ où } \text{TFD}(\vec{r}, a_i) = I(\vec{r}, t) - I(\vec{r} - \vec{V}(\vec{r}, a_i), t-1)$$

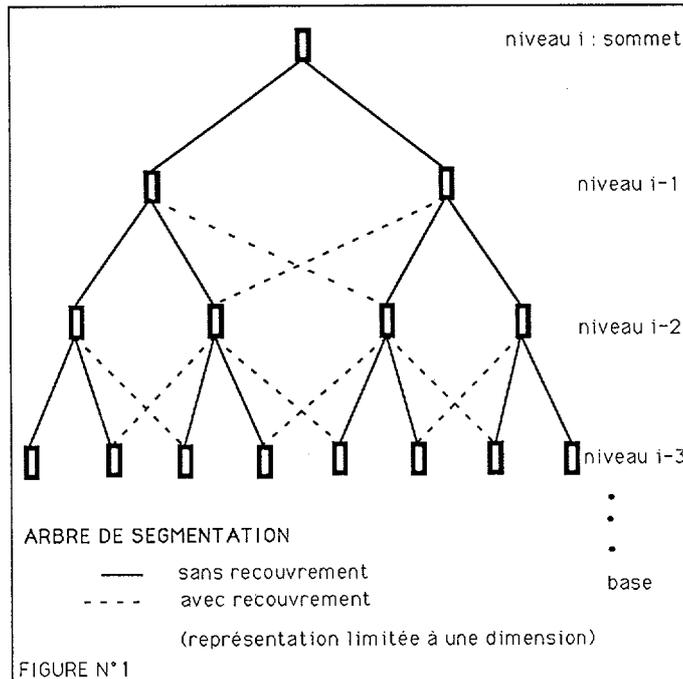
est la différence entre l'image t et l'image t-1 transformée couramment appelée Transformed Frame Difference, et $\vec{V}(\vec{r}, a_i)$ le vecteur vitesse engendré au point \vec{r} par le modèle $\{a_i\}$.

Le modèle optimal est déterminé par une méthode différentielle de type prédiction-correction, utilisant la corrélation entre les gradients spatiaux et temporels de la séquence. Les méthodes, spécifiques des algorithmes, seront détaillées plus loin.

Les déplacements étant a priori réels, la luminance ainsi que ses gradients sont obtenus par une interpolation de Lagrange sur un voisinage 2x2 ou 4x4 autour du point déplacé.

4. ANALYSE HIERARCHIQUE DE L'IMAGE

L'image est analysée de façon hiérarchique selon un "quad-tree" avec recouvrement [Fig.1], soit pour l'estimation seule (algorithme 1), soit pour l'estimation et la segmentation conjointes (algorithme 2). Le terme de recouvrement signifie qu'à un niveau donné de l'arbre, un bloc est mis en relation avec les 4 blocs adjacents du niveau supérieur ainsi qu'avec les 16 blocs adjacents du niveau inférieur. Ces relations sont mises à profit d'une part pour propager des modèles de mouvement d'un niveau à l'autre, et d'autre part pour affiner la segmentation [3].



Nous décrivons maintenant de façon détaillée les deux algorithmes dont les principes généraux viennent d'être décrits.

5.1. Identification de modèles affines par blocs

La notion de modèle n'a de sens que pour une région, ce qui pose implicitement le problème de la segmentation. Une façon de le contourner est d'en fixer une arbitrairement. Utilisant ce qui se fait traditionnellement en codage, un découpage en blocs apparaît comme une solution pratique et présente l'avantage de permettre la comparaison avec des estimateurs existants. Un compromis est à faire entre la taille des blocs et l'ordre du modèle. Des blocs trop grands risquant d'intercepter plusieurs objets mobiles, un compromis acceptable consiste à allouer un modèle affine sur des blocs de dimension réduite, typiquement 8x8 ou 16x16. Pour chaque bloc du "quad-tree", le champ de vitesse est donné par:

$$\vec{V}(\vec{r}) = \vec{T} + A.(\vec{r} - \vec{r}_c) \quad (2)$$

où \vec{r}_c désigne le centre du bloc courant, \vec{T} le déplacement moyen du bloc et A une matrice de déformation du bloc par rapport à son centre.

Ce modèle est déterminé selon une procédure de prédiction-correction qui combine une propagation hiérarchique et un certain nombre d'itérations sur chaque bloc.

Soit (\vec{T}_p, A_p) une prédiction pour le bloc courant. Ce modèle engendre un champ de vitesse $\vec{V}_p(\vec{r})$, supposé proche du champ optimal $\vec{V}(\vec{r})$. Sous l'hypothèse d'invariance de la luminance le long des trajectoires ponctuelles, on peut écrire:

$$I(\vec{r}, t) = I(\vec{r} - \vec{V}(\vec{r}), t-1) \quad (3) \text{ avec } \vec{V}(\vec{r}) = \vec{V}_p(\vec{r}) + \delta\vec{V}(\vec{r}) \quad (4).$$

En développant le 2nd membre de (3) au voisinage de la prédiction $\vec{V}_p(\vec{r})$ par rapport au champ $\delta\vec{V}$, on obtient:

$$I(\vec{r}, t) = I(\vec{r} - \vec{V}_p(\vec{r}), t-1) - \delta\vec{V}^T(\vec{r}).\text{grad}I(\vec{r} - \vec{V}_p(\vec{r}), t-1) \quad (5)$$

où l'exposant T désigne le transposé d'un vecteur ou d'une matrice. Une solution particulière pour $\delta\vec{V}$ est donnée par

$$\delta\vec{V}(\vec{r}) = -\text{TFD}(\vec{r}, \vec{T}_p, A_p) \cdot \frac{\text{grad}I(\vec{r} - \vec{V}_p(\vec{r}), t-1)}{\|\text{grad}I(\vec{r} - \vec{V}_p(\vec{r}), t-1)\|^2} \quad (6)$$

Ce champ correctif est alors approché par un modèle $(\delta\vec{T}, \delta A)$, sur critère de moindres carrés pondérés:

$$\sum_{\vec{r} \in R} p(\vec{r}) \cdot \|\delta\vec{V}(\vec{r}) - \delta\vec{T} - \delta A.(\vec{r} - \vec{r}_c)\|^2 \rightarrow \min \quad (7)$$

La formule (6) est d'autant plus fiable que le gradient au point déplacé est important (point contour). L'idée est donc de pondérer la participation des points du bloc dans la sommation (7) par une fonction croissante de l'amplitude de ce gradient. Pour simplifier les calculs, on a choisi $p(\vec{r}) = \|\text{grad}I(\vec{r} - \vec{V}_p(\vec{r}), t-1)\|^2$. La dérivation de (7) par rapport aux paramètres $(\delta\vec{T}, \delta A)$ conduit au système linéaire suivant:

$$\left\{ \begin{array}{l} \delta\vec{T} = \frac{1}{\sum_i p_i} \cdot \left[\sum_i p_i \cdot \vec{d}_i - \delta A \cdot \sum_i p_i \cdot \vec{r}_i \right] \\ \delta A = \left[\sum_i p_i \cdot \vec{d}_i \vec{r}_i^T - \delta\vec{T} \sum_i p_i \cdot \vec{r}_i^T \right] \cdot \left[\sum_i p_i \cdot \vec{r}_i \vec{r}_i^T \right]^{-1} \end{array} \right. \quad (8)$$

où, pour simplifier l'écriture, on a pris l'origine au centre du bloc ($\vec{r}_c = \vec{0}$), indicé les points du bloc par i, et posé:

$$p_i = p(\vec{r}_i) = \|\text{grad}I_{t-1}(\vec{r}_i - \vec{V}_p(\vec{r}_i))\|^2 \text{ et } \vec{d}_i = \delta\vec{V}(\vec{r}_i), \vec{r}_i \in R.$$

En posant :

5. ALGORITHMES

$$R_{dr} = \frac{1}{\sum_{i=1}^N p_i} \cdot \sum_{i=1}^N p_i \cdot \vec{d}_i \cdot \vec{r}_i^T$$

$$R_{rr} = \frac{1}{\sum_{i=1}^N p_i} \cdot \sum_{i=1}^N p_i \cdot \vec{r}_i \cdot \vec{r}_i^T$$

$$\vec{d} = \frac{1}{\sum_{i=1}^N p_i} \cdot \sum_{i=1}^N p_i \cdot \vec{d}_i$$

$$\vec{r} = \frac{1}{\sum_{i=1}^N p_i} \cdot \sum_{i=1}^N p_i \cdot \vec{r}_i$$

on obtient:

$$\begin{cases} \delta \vec{T} + \delta A \cdot \vec{r} = \vec{d} \\ \delta \vec{T} \cdot \vec{r}^T + \delta A \cdot R_{rr} = R_{dr} \end{cases} \quad (9)$$

Le champ des vitesses est donc corrigé de façon indépendante en chaque point, puis cette correction est modélisée de façon affine. Dans [1] et [3], la contrainte de modélisation est prise en compte a priori, i.e. le champ est remplacé par son expression en fonction des paramètres du modèle dans l'équation (3) et le développement est effectué par rapport à ceux-ci. On peut montrer que la solution présentée ici correspond à une simplification de celle de [1] ou [3] concernant l'estimation différentielle proprement dite. Il en résulte un abaissement notable de la complexité des calculs, ne nécessitant plus que l'inversion d'une matrice 2x2 à chaque itération, au lieu d'une matrice 6x6.

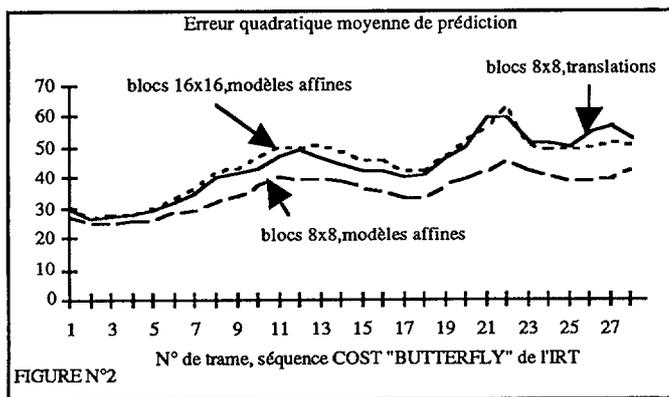
L'estimation est affinée selon un mode hiérarchique descendant, suivant une décomposition en quad-tree sur des blocs de taille décroissante. A un niveau donné de l'arbre, on utilise comme prédiction les modèles obtenus pour les 4 blocs adjacents du niveau inférieur. Pour tenir compte des centres différents, la propagation est donnée par:

$$A_p^1 = A^{-1}$$

$$\vec{T}_p^1 = \vec{T}^{l-1} + A^{-1} \cdot (\vec{r}_c^1 - \vec{r}_c^{l-1}) \quad (10)$$

Une 5^{ème} prédiction, identiquement nulle est ajoutée et permet la réinitialisation de l'algorithme. Après n itérations réalisées à partir de ces prédictions, un choix est effectué a posteriori, sur critère d'erreur de reconstruction.

Résultats



L'algorithme a d'abord été testé sur un mouvement synthétique, appliqué de façon globale sur une image naturelle. Même pour des déformations importantes de l'image, le mouvement est estimé correctement, sauf sur les bords d'image où apparaissent des problèmes d'occlusions. Pour des mouvements importants, il faut cependant recourir à un schéma multirésolution, combiné de façon naturelle avec l'analyse en quad-tree.

Appliqué à des séquences naturelles, on observe les points suivants [Fig 2]:

- à taille de blocs donnée, l'erreur de compensation obtenue avec les modèles affines est inférieure à celle obtenue avec des translations estimées de façon similaire, ceci d'autant plus que les mouvements présents dans la séquence s'éloignent de la translation.

- L'erreur obtenue pour des blocs 16x16 et des modèles affines est de l'ordre de grandeur de celle obtenue avec des blocs 8x8 et

des translations, ce qui permet d'envisager une réduction du débit alloué à l'information de mouvement dans un rapport brut de 3/4. En fait, un codage entropique permet de réduire encore plus ce débit, car la matrice A estimée n'est pas significative dans les zones fixes ou animées d'un mouvement de translation.

De plus, les effets de blocs sur les images compensées sont atténués par l'emploi de modèles affines.

5.2. Identification de modèles tridimensionnels par régions

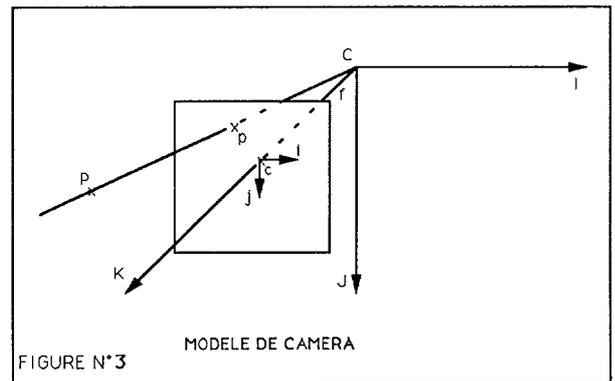
Dans le cadre d'un système de télévision en relief, l'objectif est d'accéder aux mouvements réels et aux structures des objets afin de définir des paramètres scéniques permettant de réaliser un codage spatio-temporel aussi bien pour la vue droite que pour la vue gauche (dans le cas d'une acquisition binoculaire).

L'algorithme complet comprend une première étape d'analyse monoculaire de chaque séquence prise séparément et une seconde étape d'analyse stéréo; seule la première étape est présentée ici.

5.2.1 Le modèle...

5.2.1.a...de caméra

La transformation opérée par la caméra est une projection perspective; dans le repère R(C,I,J,K) défini [Fig.3], le projeté p dans le plan image-caméra d'un point P(X, Y, Z) de la scène a pour coordonnées p(f.X/Z, f.Y/Z, f), soit p(x=f.X/Z, y=f.Y/Z) dans R(c,i,j). Dans ce qui suit et pour simplifier certaines écritures, f est supposée égale à 1, et c être le centre de l'image.



5.2.1.b...de mouvement

Le mouvement, relatif à la caméra d'un objet O de la scène peut se dissocier en deux composantes : vitesse instantanée de translation $T(t_x, t_y, t_z)$ et vitesse instantanée de rotation $W(w_x, w_y, w_z)$.

Soit p_i , le point de l'image (plan caméra) associé à $P_i(X_i, Y_i, Z_i)$, avec (x_i, y_i) ses coordonnées et (u_i, v_i) sa vitesse,

- et sous les hypothèses suivantes :
- les objets sont rigides.
 - les paramètres de rotation sont faibles.
 - la translation en Z (axe de visée) est faible par rapport à la distance caméra/objet.
 - un angle de vue relativement faible.

Les équations de base, reliant le mouvement de p_i et de P_i sont :

$$\begin{cases} u_i = +y_i \cdot w_z - (1 + (x_i)^2) \cdot w_y + x_i y_i \cdot w_x + 1/Z_i \cdot t_x - x_i/Z_i \cdot t_z \\ v_i = -x_i \cdot w_z + (1 + (y_i)^2) \cdot w_x - x_i y_i \cdot w_y + 1/Z_i \cdot t_y - y_i/Z_i \cdot t_z \end{cases} \quad (11)$$

avec Z_i , la profondeur (ie: distance caméra à objets).

5.2.1.c...de structure et de profondeur

L'analyse monoculaire d'une séquence d'images permet de déterminer les paramètres de translation à un facteur de profondeur près, en se fixant à priori un modèle de structure pour chaque région r de l'image.

En supposant évoluer dans un univers "facettes", la profondeur s'exprime pour chaque région r sous la forme :

$$1/Z_i = k_x x_i + k_y y_i + k_z \quad (12)$$



Les équations (11) se réécrivent alors [5] :

$$\begin{cases} u_i = a_1 + a_2 \cdot x_i + a_3 \cdot y_i + a_7 \cdot x_i y_i + a_8 \cdot x_i^2 \\ v_i = a_4 + a_5 \cdot x_i + a_6 \cdot y_i + a_8 \cdot x_i y_i + a_7 \cdot y_i^2 \end{cases} \quad (13)$$

avec,

$$\begin{cases} a_1 = -w_y + k_z t_x \\ a_2 = k_x t_x + -k_z t_z \\ a_3 = w_z + k_y t_x \\ a_4 = w_x + k_z t_y \\ a_5 = -w_z + k_x t_y \\ a_6 = k_y t_y + -k_z t_z \\ a_7 = w_x + -k_y t_z \\ a_8 = -w_y + -k_x t_z \end{cases} \quad (14)$$

5.2.2 Estimation du mouvement et de la structure

La méthode différentielle d'estimation de mouvement tridimensionnel apparent initialement proposée par Netravali[6], est une généralisation de l'algorithme de Walker-Rao[7], algorithme basé sur l'observation locale (spatiale et temporelle) de la fonction d'intensité du point.

Cette technique procède par minimisation itérative de la TFD² (différence d'intensité entre le point étudié et ce même point déplacé dans l'image précédente), sous l'hypothèse d'invariance temporelle de la luminosité.

Le terme de "généralisation" est justifié par les deux points suivants :

- L'observation ne concerne plus chaque point de l'image pris séparément, mais concerne des régions regroupant un nombre minimal de points, supposées respecter un modèle de structure.
- Le vecteur mouvement estimé comporte, non plus deux mais six composantes.

Pour les k points d'une région r en mouvement, l'algorithme est défini par l'équation :

$$A_r^n = A_r^{n-1} - [\text{gain} \cdot \sum_{i=0}^k (\text{TFD}_i^{n-1} \cdot \Phi_i^{n-1})] \quad (15)$$

Avec :

- i , l'indice du point-image,
- n , le numéro d'itération,
- (x, y, t) le point courant,
- $(x-u, y-v, t-1)$ le point déplacé (image précédente).

$$A_r = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \end{pmatrix} \quad \Phi_i = \begin{pmatrix} g_x \\ g_x \cdot x_i \\ g_x \cdot y_i \\ g_y \\ g_y \cdot x_i \\ g_y \cdot y_i \\ g_x \cdot x_i y_i + g_y \cdot y_i^2 \\ g_x \cdot x_i^2 + g_y \cdot x_i y_i \end{pmatrix}$$

A_r , le vecteur de paramètres tridimensionnels défini par le système (14), associé à la région r .

Φ_i , le "vecteur gradient" au point i , selon chacune des composantes du vecteur A_r .

$$\text{TFD}_i^{n-1} = I(x_i, y_i, t) - I(x_i - u_i^{n-1}, y_i - v_i^{n-1}, t-1)$$

$$G = (g_x, g_y)^T \equiv G(x_i - u_i^{n-1}, y_i - v_i^{n-1}, t-1)$$

G , le gradient au point déplacé (image précédente).

$$\text{gain} = 1.0 / \sum_{i=1}^k (\Phi_i^{n-1})^2$$

Un schéma multi-prédictions (construit par interprétation du flot optique, par propagation spatiale, par suivi temporel...) est mis en oeuvre afin de garantir une bonne initialisation (C^0 et Z^0) de l'algorithme (15). A^0 est calculé en substituant les valeurs C^0 et Z^0 dans (14).

Après affinage des paramètres de A^0 en A^1 , C^1 et Z^1 se déduisent du vecteur A^1 par linéarisation du système (14) (méthode de

Newton), et en initialisant le système par C^0 et Z^0 (ce qui nous assure d'être proche de la solution). Ce système est ramené à huit équations à huit inconnues en fixant k_z comme constante (la translation n'est estimée qu'à un facteur de profondeur près) :

$$\begin{cases} t_x^1 = t_x^0 + h_5 \\ t_y^1 = t_y^0 + h_6 \\ t_z^1 = t_z^0 + h_4 \\ w_x^1 = w_x^0 + h_3 \\ w_y^1 = w_y^0 + h_2 \\ w_z^1 = w_z^0 + h_1 \\ k_x^1 = k_x^0 + h_8 \\ k_y^1 = k_y^0 + h_7 \\ k_z^1 = k_z^0 \end{cases}$$

En notant les termes d'exposant ⁰ en caractères *italiques*,

$$\begin{pmatrix} -1 & 0 & 0 & 0 & -k_y & 0 & -t_x & 0 \\ 0 & 1 & 0 & k_x & 0 & 0 & 0 & t_z \\ 0 & 0 & -1 & k_y & 0 & 0 & t_z & 0 \\ 0 & 0 & 0 & k_z & -k_x & 0 & 0 & -t_x \\ 0 & 1 & 0 & 0 & -k_z & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & -k_z & 0 & 0 \\ 0 & 0 & 0 & k_z & 0 & -k_y & -t_y & 0 \\ 1 & 0 & 0 & 0 & 0 & -k_x & 0 & -t_y \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \\ h_7 \\ h_8 \end{pmatrix} = \begin{pmatrix} -a_3 + w_z + k_y t_x \\ -a_8 - w_y - k_x t_z \\ -a_7 + w_x - k_y t_z \\ -a_2 + k_x t_x - k_z t_z \\ -a_1 - w_y + k_z t_x \\ -a_4 + w_x + k_z t_y \\ -a_6 + k_y t_y - k_z t_z \\ -a_5 - w_z + k_x t_y \end{pmatrix}$$

Après itérations de l'algorithme (15), les résultats obtenus en terme d'erreur d'EQM(TFD) pour un nombre maximum de régions fixé à 256 (niveau n°5 de la pyramide de segmentation d'une image 512x512, séquence "CAMPAGNE"¹) sont du même ordre (ie: 15 à 20) que ceux obtenus à partir d'un flot optique dense type Walker-Rao.

6. CONCLUSION

Dans cet article, nous avons présenté deux algorithmes pour l'identification de modèles décrivant le mouvement global d'une région de l'image. Le premier algorithme se situe dans un cadre paramétrique où est recherché avant tout une description du mouvement plus efficace en terme de qualité de compensation et de débit. Les résultats obtenus sur les séquences testées montrent que cet objectif est atteint. Le second algorithme réalise de façon coopérative la segmentation de la scène en objets animés et l'identification pour chacun d'eux d'un modèle (quasi) quadratique de mouvement et de structure, permettant (moyennant certaines hypothèses et la connaissance de la calibration de la caméra) de remonter aux paramètres du mouvement 3D. Cette dernière approche est particulièrement bien adaptée à un codage orienté objet.

REFERENCES BIBLIOGRAPHIQUES

- [1] M. HOETTER, "Differential estimation of the global motion parameters zoom and pan", Signal Processing, Vol. 16, 1989, pp. 249-265.
- [2] S.F. WU and J. KITTLER, "2D motion parameter estimation: a survey", 6th Scandinavian Conference on Image Analysis, Oulu, Finland, June 19-22, 1989, pp. 1043-1050.
- [3] R. HARTLEY, "Segmentation of optical flow fields by pyramid linking", Pattern Recognition Letters 3 (1985) 253-262.
- [4] S.F. WU and J. KITTLER, "A differential method for simultaneous estimation of rotation, change of scale and translation", Signal Processing: Image Communication 2, 1990, pp. 69-80.
- [5] G. ADIV, "Determining three-dimensional motion and structure from optical flow generated by several moving objects", IEEE Vol.pami-7, n°4, July, 1989.
- [6] A.N. NETRAVALI and J. SALZ, "Algorithms for estimation of three-dimensional motion", AT&T Journal Vol.64, n°2, Feb 1985.
- [7] D.R. WALKER and K.R. RAO, "New techniques in pel-recursive motion compensation", Proc. ICC, Amsterdam, may 14-17, p 703-706, 1984.

¹Dr. Enkelmann (IITB, Karlsruhe, FRG).