Vers une méthodologie pour l'identification paramétrique robuste du mouvement de régions par optimisation non-linéaire.

Henri SANSON

Centre Commun d'Études de Télédiffusion et Télécommunications
4 rue du Clos Courtel, B.P. 59, 35512 Cesson-Sévigne Cedex, France
Tel: +33 99 12 47 45, E-mail: sanson@ccett.fr

RÉSUMÉ

Cet article met l'accent sur certains problèmes survenant lors de l'identification de modèles paramétriques du mouvement dans les séquences d'images par des méthodes différentielles, entrainant ainsi un mauvais comportement des algorithmes mis en œuvre. Des solutions appropriées sont proposées, rendant ces techniques particulièrement efficaces, même pour l'estimation de mouvements de forte amplitude.

ABSTRACT

This paper highlights some problems related to the identification of parametric motion models in image sequences by the so-called differential methods and that often entail a bad behavior of the algorithms used, leading to poor results. Appropriate solutions are proposed, turning differential algorithms particularly effective for complex motion estimation, even of large magnitude.

1 Introduction.

Ce papier traite le problème de l'identification d'un modèle paramétrique du mouvement apparent (également baptisé flot optique) sur une région de forme arbitraire entre 2 images numériques I_1 et I_2. Nous formulons ici ce problème de façon très classique: soit $\vec{d}(\vec{p})$ le champ dense des déplacements défini sur un domaine $\mathcal{R}$ du plan image. Sous l'hypothèse habituelle que la luminance ne varie pas le long des trajectoires des points de $\mathcal{R}$, le champ $\vec{d}(\vec{p})$ recherché doit minimiser l'énergie d'erreur de prédiction, qui n'est autre que la norme L^2 entre la fonction de luminance sur $\mathcal{R}$ et sa version déplacée:

$$E(\vec{d}(\vec{p})) = \sum_{\vec{p}\in\mathcal{R}} DFD^2(\vec{p}, \vec{d}(\vec{p})) \quad (1)$$
$$\text{avec: } DFD(\vec{p}, \vec{d}(\vec{p})) = I_2(\vec{p}) - I_1(\vec{p} - \vec{d}(\vec{p}))$$

Cette approche de l'estimation de mouvement est particulièrement simpliste, et ne rend pas compte de certains phénomènes physiques réels, notamment sur le plan de la photométrie 3D, mais elle est abondamment utilisée et donne en pratique des résultats satisfisants dans la plupart des situations. Nous supposons de plus ici que le champ de déplacements peut être représenté sous la forme d'un développement linéaire sur une base fonctionnelle donnée (description paramétrique):

$$\vec{d}(\vec{p}, A) = \sum_{k=0}^{D-1} \vec{a}_k f_k(\vec{p}) \quad (2)$$
$$\text{avec: } A = [a_0^x, \cdots, a_{D-1}^x, a_0^y, \cdots, a_{D-1}^y]^T$$

Ainsi formulée, l'estimation de mouvement n'est rien d'autre qu'un problème d'optimisation non-linéaire sans contrainte, de dimension supposée raisonnable en pratique.

2 Choix de modélisations.

Nous nous concentrons ici sur une analyse des points délicats soulevés par l'identification de modèles polynomiaux [1]. Nous sommes cependant persuadés que les enseignements tirés de cette analyse et les solutions proposées peuvent être étendus à d'autres types de modélisations, plus intéressantes pour rendre compte de champs continus les plus variés possible, telles que les Eléments Finis polynomiaux [2]. La complexité essentielle de ces représentations réside surtout dans la grande liberté existant pour la construction du maillage. Un polynôme bivarié est défini par l'utilisation des fonctions de base suivantes:

$$f_{k_{ij}} = (\frac{x-x_0}{\Delta x})^{i-j}(\frac{y-y_0}{\Delta y})^j \text{ ou } (\frac{x-x_0}{\Delta x})^i(\frac{y-y_0}{\Delta y})^j \quad (3)$$

selon le cas, séparable ou non. Les indices sont tels que $0 \leq i, j \leq n$, le degré du polynôme. Un point important à souligner dès à présent est qu'ils constituent une hiérarchie de modèles ordonnée par le degré [3]. On peut de même considérer des Eléments Finis hiérarchiques définis par une succession de maillages emboîtés, à condition de savoir prédéfinir ou construire de manière adaptative une telle hiérarchie.

3 Identification du modèle.

La formulation variationnelle du problème de l'estimation de mouvement conduit naturellement à envisager l'utilisation des méthodes de l'optimisation mathématique pour le résoudre. L'extension de techniques du type mise en correspondance paraît quant à elle peu réaliste pour des dimensions d'espace de paramètres supérieures à 1, pour des raisons de complexité arithmétique. On considère désormais la classe des algorithmes d'optimisation non-linéaire différentiable sans contrainte admettant la formulation générale suivante:

$$\delta A^k = A^{k+1} - A^k = -\alpha_k M_k \nabla_A E(A^k) \quad (4)$$

où $\nabla_A E(A^k)$ est le gradient de la fonctionnelle de coût E, $\alpha_k > 0$ un pas, M_k une matrice définie positive, de sorte que

$-M_k \nabla_A E(A^k)$ est une direction de descente. La justification du cadre différentiable repose sur le théorème de Shannon: l'échantillonnage du signal image continu suppose celui-ci à bande limitée, donc C^∞. En pratique, la régularité de la fonction luminance en continu dépend de l'interpolation utilisée. Ici, celle-ci est réalisée par polynômes bi-cubiques par morceaux permettant une représentation C^1 sur $\mathbb{R}^2$ et C^∞ sur l'intérieur de chaque carreau élémentaire [4]. Toutes les grandeurs relatives à la fonction de luminance (valeurs, dérivées) sont calculées avec cette interpolation.

La convergence de ce type d'algorithmes, ne serait-ce que vers un minimum local, n'est garantie sous des hypothèses suffisamment générales que pour l'algorithme du gradient à pas optimal [5]. Autrement, la convergence globale n'est assurée que sous des hypothèses de forte convexité de la fonction de coût [6], ce qui n'est pas a priori le cas ici. Intuitivement, on peut espérer se rapprocher de ce cas modèle par l'emploi d'une stratégie multi-résolution, au moins localement. C'est pourquoi nous considérons désormais l'approche couramment utilisée consistant en une procédure d'optimisation non-linéaire intégrée dans un schéma multi-échelle [1], [7]. Malgré cela, il subsiste certains points délicats, rarement évoqués et ayant suscité peu de travaux, que nous allons mettre en évidence dans ce qui suit, et pour lesquels nous allons proposer des solutions. Le but recherché ici est d'éviter des situations difficiles où l'on est sûr du mauvais comportement des algorithmes. Pour autant, nous n'avons pas de preuve de convergence une fois le défaut corrigé.

3.1 Choix de l'algorithme différentiel.

Une comparaison objective et expérimentale a été menée entre 3 méthodes que l'on peut considérer comme représentatives du point de vue de la convergence: un gradient à pas adaptatif (M_k=matrice identité), Newton [5] (M_k=inverse du Hessien de E) et Gauss-Newton [8] (M_k=inverse d'une approximation du Hessien de E). On trouvera dans [1] les résultats détaillés des travaux réalisés. C'est volontairement que les méthodes requérant le calcul d'un pas optimal ont été oubliées. En effet, l'obtention de ce pas, outre qu'elle n'est pas garantie non plus, exige la résolution d'un problème d'optimisation mono-dimensionnelle à chaque itération principale, et donc un certain nombre d'évaluations de la fonction de coût, donc de DFDs. Or, ce sont ces quantités qui sont chères à calculer. Nous avons donc privilégié les techniques ne nécessitant pas de recherche linéaire.

La conclusion de cette comparaison est qu'une méthode de type Gauss-Newton simplifée, où les composantes x et y sont traitées (presque) séparément à chaque itération, donne les meilleurs résultats, faisant aussi bien que Newton, pour une complexité moindre. Les équations la définissant sont les suivantes:

$$\begin{aligned} \delta A_x^k &= -[R_{ff}^{xx}]^{-1}.R_{df}^x &&= -\tfrac{1}{2}[R_{ff}^{xx}]^{-1}.\nabla_{A_x} E \\ \delta A_y^k &= -[R_{ff}^{yy}]^{-1}.R_{df}^y &&= -\tfrac{1}{2}[R_{ff}^{yy}]^{-1}.\nabla_{A_y} E \end{aligned} \tag{5}$$

avec:

$$\begin{aligned} R_{ff,kl}^{xx} &= \textstyle\sum_{\vec{p}\in\mathcal{R}} I_{1,x}^2.f_k f_l & R_{ff,kl}^{yy} &= \textstyle\sum_{\vec{p}\in\mathcal{R}} I_{1,y}^2.f_k f_l \\ R_{df,k}^{x} &= \textstyle\sum_{\mathcal{R}} I_{1,x}.DFD.f_k & R_{df,k}^{y} &= \textstyle\sum_{\mathcal{R}} I_{1,y}.DFD.f_k \\ I_{1,x} &= \tfrac{\partial I_1}{\partial x}(\vec{p}-\vec{d}(\vec{p},A^k)) & I_{1,y} &= \tfrac{\partial I_1}{\partial y}(\vec{p}-\vec{d}(\vec{p},A^k)) \\ DFD &= DFD(\vec{p},A^k) & f_k &= f_k(\vec{p}) \end{aligned} \tag{6}$$

Essentiellement 2 facteurs vont influer sur le conditionnement de ces matrices symétriques, et semi-définies positives dans le cas général: la texture, ou de façon équivalente les gradients de luminance, et la base retenue pour la modélisation. En ce qui concerne la texture, si le niveau de gris est trop plat (gradient quasi-nuls), les matrices tendront à devenir singulières. On risque d'obtenir des valeurs de paramètres trop grandes, et sans rapport avec la réalité physique. Une façon d'éviter cet écueil est de pénaliser toute variation de paramètres. On aboutit alors à la méthode d'augmentation de Marquardt [8], qui combine en fait les caractéristiques d'une méthode de gradient loin de l'optimum, et celles de Gauss-Newton (rapidité de convergence) près de l'optimum. Les équations (5) deviennent:

$$\begin{aligned} \delta A_x^k &= -[R_{ff}^{xx} + \alpha_x.I_d]^{-1}.R_{df}^x \\ \delta A_y^k &= -[R_{ff}^{yy} + \alpha_y.I_d]^{-1}.R_{df}^y \end{aligned} \tag{7}$$

Pour déterminer des valeurs pertinentes pour α_x et α_y, il faut considérer l'impact d'une variation d'un paramètre sur le champ dense associé, plutôt que cette variation elle-même. Le choix suivant:

$$\begin{aligned} \alpha_x &= \tfrac{1}{\epsilon} max_{ij}(|R_{df,ij}^x|.\parallel f_{ij} \parallel) \\ \alpha_y &= \tfrac{1}{\epsilon} max_{ij}(|R_{df,ij}^y|.\parallel f_{ij} \parallel) \end{aligned} \tag{8}$$

où $\parallel f_{ij} \parallel$ est une norme fonctionnelle, garantit que toute varaition δa_k entrainera une variation correspondante du champ dense de norme inférieure à ϵ. Un choix naturel pour ϵ est $\epsilon = 1$, distance entre 2 pixels sur une même ligne ou une même colonne. La contrepartie à payer est un ralentissement de l'algorithme, donc un nombre plus élevé d'itérations à effectuer.

3.2 Choix de la base.

Le comportement d'un algorithme différentiel dépend fortement de la base utilisée, sur le plan mathématique ou numérique. Les algorithmes de gradient ne sont pas invariants par changement de base, par exemple. L'évaluation des conditionnements matriciels pour différents choix de bases (canonique avec différentes normalisations, Bernstein) ont montré qu'il était important que la base retenue soit normée sur la région, ou plus simplement sur le rectangle circonscrit à celle-ci (voir figure 1):

Ainsi, la contribution des différents paramètres est équilibrée, et on améliore l'isotropie de la fonction de coût. Pour les polynômes, ceci est obtenu en choisissant comme coefficients de normalisation:

$$\begin{aligned} x_0^s &= \tfrac{x_{min}^s + x_{max}^s}{2} & y_0^s &= \tfrac{y_{min}^s + y_{max}^s}{2} \\ \Delta x^s &= \tfrac{x_{max}^s - x_{min}^s}{2} & \Delta y^s &= \tfrac{y_{max}^s - y_{min}^s}{2} \end{aligned} \tag{9}$$

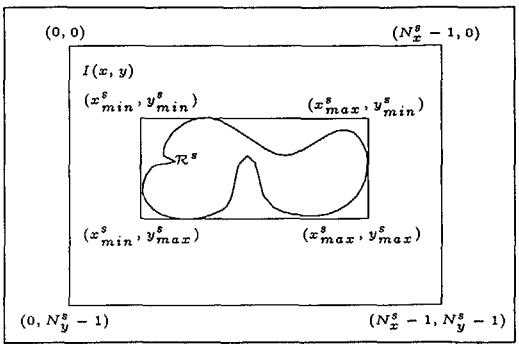

FIG. 1 - *Rectangle circonscrit à une région, à une échelle s.*

Pour les Eléments Finis sous la forme usuelle de définition par vecteurs de contrôle nodaux, cette propriété est naturellement satisfaite pour un interpolation sur l'élément par polynômes de Bernstein, et approximativement pour les polynômes de Lagrange (pour les bas degrés) [2].

3.3 Construction des pyramides multi-échelles d'images.

Les approches multi-échelles descendantes sont souvent utilisées pour l'estimation de forts mouvements. La plupart du temps, une pyramide de type Gaussienne (de Burt) est employée. En fait, il n'y a pas de raison évidente, à notre connaissance, pour préférer une pyramide dyadique, si ce n'est pour des aspects pratiques et de réduction de complexité arithmétique. De même, le choix du filtre passe-bas utilisé pour générer ces pyramides devrait se faire sur la base des caractéristiques de la méthode d'optimisation. Des simulations effectuées sur des situations plus ou moins critiques, il ressort que les résultats sont relativement indifférents au choix du filtre, sauf dans quelques cas difficiles, tels celui de la figure (??). La méthode de Gauss-Newton étant basée sur une hypothèse de linéarité locale des fonctions de coût marginales, donc des images, il importe de minimiser en un certain sens les dérivées d'ordre supérieur à 1 des images après filtrage. Si de plus, on impose une solution séparable, à phase linéaire, et la plus critique possible vis-à-vis du sous - échantillonnage (il faut conserver le maximum de hautes fréquences d'après le paragraphe précédent), on arrive au filtre dont le noyau par composante x ou y est le moyenneur sur $[-2, +2]$ (il suffit d'écrire la dérivée d'un produit de convolution). L'approximation par B-splines, minimisant l'énergie de courbure, peut aussi être envisagée, mais le coût opératoire est nettement supérieur. Nous pensons cependant qu'une étude plus rigoureuse et approfondie de l'impact de l'analyse multi-échelle sur la convergence de ce type d'algorithmes, par exemple en partant des espaces échelles continus générés par des opérateurs d'évolution aux dérivées partielles, apportera des réponses de portée plus générale.

Quant au nombre de niveaux de pyramide à utiliser, un raisonnement simple, consistant à supposer une borne supérieure δd_{max} sur la correction de chaque paramètre pour une résolution donnée conduit à la relation suivante entre mouvement maximal accessible et nombre de niveaux, où expérimentalement $2 \leq \delta d_{max} < 3$:

$$\| \vec{d} \|_{max} \leq \delta d_{max} \cdot (2^{n_s} - 1) = \delta d_{max} \sum_{s=0}^{n_s - 1} 2^s \qquad (10)$$

3.4 Propagation des modèles entre 2 échelles consécutives.

En plus de la conversion d'échelle systématique (multiplication par 2 des paramètres), nous proposons ici une propagation sélective dans le but d'accroître la robustesse du schéma. En effet, il peut arriver que l'optimum trouvé à l'échelle précédente ne constitue pas un bon point de départ pour l'échelle courante, par exemple lorsque le filtrage a éliminé trop d'information de texture. Dans ce cas, il est intéressant de réinitialiser l'estimation avec $\hat{A}^0_{0 \to s}$. Autrement dit, si $E(\hat{A}_{s+1 \to s}) > E(\hat{A}^0_{0 \to s})$, $\hat{A}^0_{0 \to s}$ est choisi, sinon $\hat{A}_{s+1 \to s}$ est retenu. En général, $\hat{A}^0_0 = 0$. Il s'agit d'un choix a priori, peu gourmand en calculs.

3.5 Représentation d'une région discrète à differentes résolutions.

Ce point est rarement abordé, bien qu'il s'agisse d'une question assez épineuse. Pour y répondre, on impose une taille minimale à la région à chaque échelle. Les simulations ont montré que ceci permettait une estimation fiable du mouvement, même pour des amplitudes supérieures au diamètre de la région originale. Soit $L_{\mathcal{R}}(x, y) = 1 \Leftrightarrow (x, y) \in \mathcal{R}$ le masque de la région à l'échelle s. La pyramide des masques est construite classiquement par simple sous - échantillonnage:

$$L^{s+1}_{\mathcal{R}}(x, y) = L^s_{\mathcal{R}}(2.x, 2.y) \qquad (11)$$

Cependant, s'il existe une portion filaire à une échelle donnée (figure 2), la région peut être déconnectée, voire supprimée. Pour éviter cela, une dilatation morphologique conditionnelle est opérée au niveau de ces points filaires, avnt sous-échantillonnage. D'autre part, si à un niveau donné (y compris l'original), la région ne satisfait pas le critère de taille minimale, une succession de dilatations par un élément structurant 3×3 lui est appliquée, jusqu'à atteindre la taille κ.

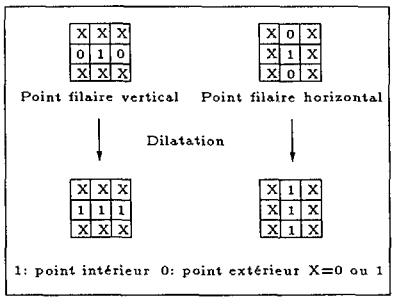

FIG. 2 - *Définition et dilatation d'un point filaire.*

3.6 Adaptation de la complexité du modèle.

Il s'agit ici de choisir parmi la hiérarchie de modèles disponibles celui qui convient à la région courante et en parti-

culier à sa taille. Soit $|\mathcal{R}^s|$ cette taille et $D(n)$ le nombre de paramètres (par composante x and y) du n^{me} modèle de la hiérarchie (pour les polynômes, n est simplement le degré), alors on impose la condition suivante:

$$|\mathcal{R}^s| > \kappa.D(n) \tag{12}$$

Pour être cohérent avec le point précédent, il est impératif que le modèle d'ordre 0 corresponde à la translation. Pour les Eléments Finis, la situation est moins simple, car outre la grande richesse de maillages emboîtés possibles, il faudrait plutôt considérer le nombre de pixels pris en compte par point de contrôle nodal.

4 Résultats expérimentaux.

Des simulations ont été réalisées sur des zones d'images physiquement homogènes en mouvement, animées de façon plus ou moins complexe: translations, rotations, zooms, déformations, régions petites ou fines, ...Celles-ci ont montré le bien-fondé des solutions avancées. Sur la figure (??), on constate l'aptitude de la méthode à estimer un fort mouvement sur une petite région, tandis que sur la figure (3), on observe la capacité à identifier une déformation importante (un zoom en l'occurence), ceci malgré le flou introduit par le zoom de la caméra.

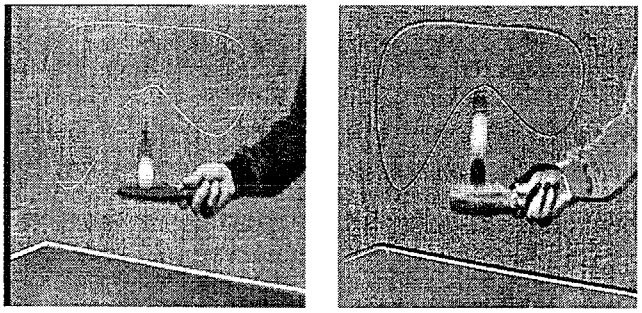

FIG. 3 - *Séquence "Table Tennis" , image 8, trame 1, définition de la région "Tapisserie" (gauche) et image d' Erreur de Prédiction après compensation du mouvement (droite), modèles affines, EQM = 333.*

5 Conclusion.

Dans cet article, nous nous sommes attachés à mettre en évidence un certains nombre de problèmes liés à l'emploi des méthodes différentielles multi-résolutions pour l'estimation paramétrique du mouvement, pour lesquels nous avons proposé des solutions, en nous basant sur les principes généraux de l'optimisation mathématique. Même si les preuves de convergence n'ont pu être apportées, les expériences réalisées ont montrée une robustesse accrue du processus d'identification.

6 Remerciements.

Ce travail a été partiellement financé par le projet RACE II R2072-MAVT.

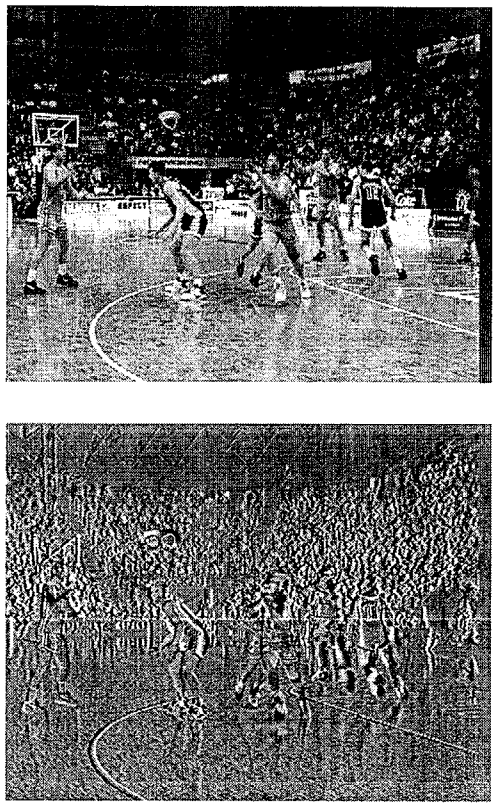

FIG. 4 - *Séquence "Basket-Ball-2" , image 8, trame 1, définition de la région "Ballon" (gauche) et image d' Erreur de Prédiction après compensation du mouvement (droite), modèles affines, EQM = 9.*

Références

[1] J.-L. Dugelay and H. Sanson. Differential methods for the identification of 2D and 3D motion models in image sequences. *Signal Processing: Image Communication*, 7:105–127, 1995.

[2] C. Zienkiewicz. *"The Finite Element method"*. McGraw-Hill Book Company (UK) Limited, Maidenhead, Berkshire, England, 3 edition.

[3] H. Nicolas. *"Hiérarchie de modèles de mouvement et méthodes d'estimation associées. Application au codage de séquences d'images"*. PhD thesis, Université Rennes I, Campus de Beaulieu, 35000 Rennes, September 1992.

[4] D.P. Mitchell & A.N. Netravali. "Reconstruction filters in computer graphics". *Computer Graphics*, 22(4):221–228, August 1988.

[5] D. G. Luenberger. *"Introduction to linear and nonlinear programming"*. Addison-Wesley Publishing Company, Reading, Massachusetts, 1972.

[6] J.-C. Culioli. *"Introduction à l'Optimisation"*. Ellipse, 1994.

[7] J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani. "Hierarchical model-based motion estimation". In *Proc. ECCV'92*, pages 237–252, 1992.

[8] W. Murray. *"Numerical methods for unconstrained optimization"*. Academic Press, London and New-York, 1972.