De l'estimation de mouvement pour l'analyse temps réel de vidéos dans le domaine compressé

Marc LENY[1,2]
marc.leny@fr.thalesgroup.com

Didier NICHOLSON[1]
didier.nicholson@fr.thalesgroup.com

Françoise PRÊTEUX[2]
francoise.preteux@int-evry.fr

[1] Thales Communications, Laboratoire MMP
146 Bd de Valmy, 92700 Colombes, France
Fax : +33 1 46 13 25 55

[2]GET / Institut National des Télécommunications, Département ARTEMIS
9, Rue Charles Fourier, 91011 Evry Cedex, France
Fax : +33 1 60 76 43 81

Thèmes : 4.2 Segmentation et estimation de ruptures - 2.2 Codage et compression d'images et de vidéos

Résumé – Analyser des vidéos directement dans le domaine compressé nécessite de disposer de procédures précises d'estimation des vecteurs de mouvement. Notre contribution porte sur la mise au point d'une procédure temps réel de traitement de ces vecteurs faisant suite à une analyse statistique de leur répartition, et aboutissant aux solutions de filtrage adéquates. Les nouveaux algorithmes ont été implantés et leur apport dans le cadre d'un corpus de séquences de vidéo-surveillance démontré avec une accélération du temps de calcul d'un facteur cinq par rapport aux performances décrites dans la littérature.

Abstract *– Directly analysing compressed video sequences requires accurate procedures to evaluate motion vectors. Our work focuses on a real time application processing these vectors. A statistical analysis of their spatial distribution makes it possible to select the appropriate filtering scheme. Then the related algorithms were implemented and validated on a corpus of surveillance sequences. Compared to previous publications, the overall result is a five time faster analysis.*

1. Introduction

Face à la multiplication des systèmes de vidéo-surveillance et au volume d'archives vidéos générées, l'indexation systématique, automatique et en temps réel de ces sources d'information reste un défi scientifique et technologique. Une nouvelle voix de recherche est actuellement explorée à partir de méthodes capables d'analyser les séquences vidéos directement dans le domaine compressé afin d'y identifier principalement des objets mobiles ou les mouvements de caméra [6,7,8,9,13,15]. Toutefois, l'information dans le flux compressé est singulière et nécessite l'application de procédures et traitements spécifiques et originaux [10,11,12,14]. Il ne s'agit plus de traiter des successions de pixels dans le domaine spatial, mais des coefficients dans le domaine fréquentiel et des informations représentatives de l'évolution temporelle d'une image. Cet article présente la procédure temps réel de traitement des vecteurs estimation de mouvement (ci-après VEM) développée. Son originalité repose sur les solutions de filtrage mises en œuvre qui sont dûment justifiées par une analyse statistique de la distribution spatiale desdits vecteurs.

2. Des vidéos compressées aux caractéristiques

Un flux vidéo compressé, selon les standards H.26x et MPEG-1/2/4/AVC, offre deux types d'information que nous nous proposons d'exploiter.

Le premier renvoie aux VEM qui exploitent la redondance temporelle au sein d'une séquence vidéo en copiant pour l'image courante des macroblocs d'images précédemment décodées. Ne sont alors conservés dans le flux que le vecteur de copie correspondant et éventuellement une erreur résiduelle permettant d'affiner ce patchwork. En revanche, si une zone contient un nouvel élément, on la stocke intégralement dans le flux (blocs "Intra", en rouge fig. 1). Ce principe est mis en oeuvre lors du codage selon des critères de minimisation d'erreur sur la valeur des pixels. Dès 1999, R. Wang et T. Huang [2] ont utilisé ces VEM pour détecter la présence d'activité sur une vidéo. Toutefois, il arrive que des pixels ne correspondent pas au mouvement réel d'un objet, comme l'illustrent les problèmes bien connus de surfaces ou murs unis [3] (blank wall ou aperture – fig. 2).

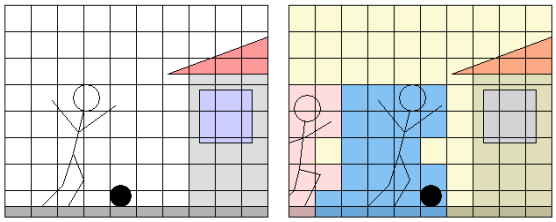


Figure 1 - Principe de la redondance temporelle
(blocs jaunes : inchangés, bleus : prédits via un vecteur, rouges : nouveaux, codés en Intra)

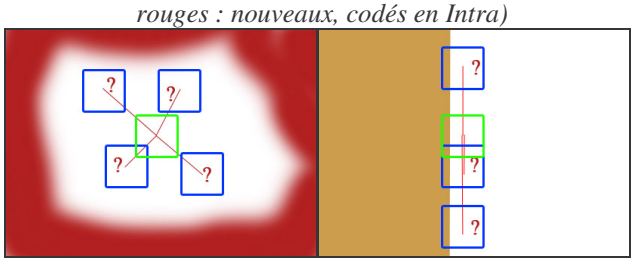


Figure 2 - Phénomènes d'*aperture* et *blank wall* (en vert, bloc actuel à prédire, en bleu, quelques blocs candidats sur l'image de référence)

Le second type d'information provient des blocs DCT (*Discrete Cosine Transform*), qui correspondent au passage dans le domaine fréquentiel d'un sous-élément (8x8 pixels pour du MPEG-1 ou 2, 4x4 pour MPEG-4 AVC) d'une image. Cette transformée permet aussi de quantifier les coefficients associés aux hautes fréquences moins significatives visuellement. B. Shen et I. Sethi en 1996 [1] se sont intéressés aux blocs DCT et ont montré que les coefficients AC_{10} et AC_{01} sont proportionnels aux gradients moyens (vertical et horizontal) sur chaque bloc.

Ronan Coudray [4] a étudié la possibilité d'indexer une vidéo au travers de plans de coupe détectés dans le domaine compressé, à partir de blocs codés en mode intra. Miguel Coimbra [3] a exploité ces coefficients DCT pour détecter et segmenter des piétons dans une station de métro directement à partir du flux vidéo MPEG-2.
Les résultats obtenus par ces différents travaux tout en montrant l'intérêt de traiter les informations portées par les VEM et les coefficients DCT mettent en évidence les limites mêmes de l'état de l'art actuel quant à la robustesse au bruit et une exploitation fiable en temps réel. Nous proposons ici des éléments de solution fondés sur une analyse statistique du bruit sur un flux vidéo compressé.

3. Etude statistique

Pour conduire cette étude, le corpus de séquences de vidéosurveillance détaillé tableau 1 a été utilisé.

TAB. 1 : Distribution des VEM

	Speedway	Caretaker [17] - Métro Turin	WCAM Séquence Outdoor 1	WCAM Séquence Outdoor 2 (de nuit, fortement bruitée)	Séquence statique (plan fixe)
Largeur	720	704	352	352	720
Hauteur	576	288	288	288	576
Entrelacement	✓	×	×	×	✓
Images par seconde	25	6	25	25	25
Référence de la séquence	A	B	C	D	E

Concernant le contenu, la séquence *Speedway* présente des véhicules sur une voie rapide, celle dans le métro de Turin, des piétons dans une station (borne d'achat et tourniquets), les séquences de WCAM [16] en extérieur, des scénarii mêlant voitures et piétons de jour ou de nuit. La séquence statique permet quant à elle une validation sur un plan fixe (l'évolution d'une image à l'autre venant principalement des bruits de capteur).
Les composantes de l'ensemble des vecteurs ont été extraites, puis analysées par un script Matlab agrégeant l'ensemble des vecteurs, puis traçant leur distribution et déterminant leurs caractéristiques (Figure 3, Tableau 2) pour chaque trame individuellement ainsi que sur l'ensemble de la vidéo.

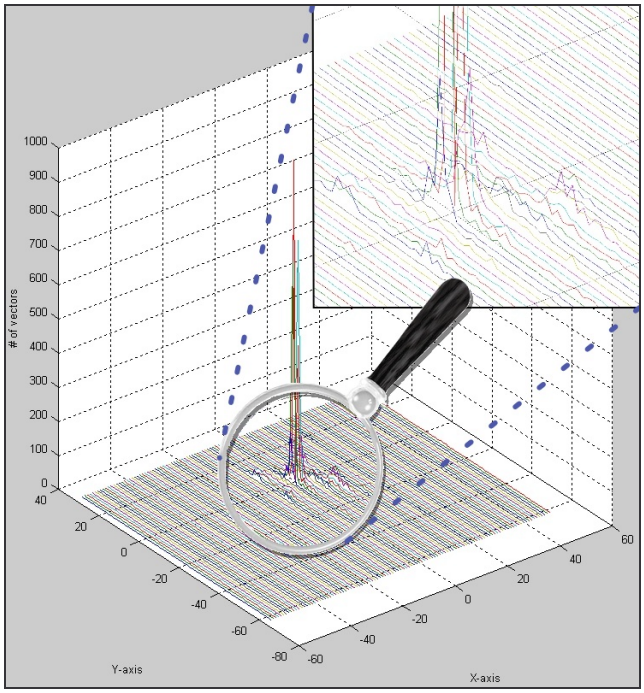


Figure 3 - Distribution des VEM selon leurs composantes sur une séquence (Speedway)

TAB. 2 : Distribution des VEM

Séquence	A	B	C	D	E
Moyenne en X	-1,17	-17,9	-8,42	-21,4	-4,23
Ecart type en X	5,56	42,05	20,8	36,0	9,28
Kurtosis en X	14,79	4,39	3,92	5,02	6,47
Moyenne en Y	-1,15	-3,86	-6,87	-8,42	-0,71
Ecart type en Y	8,27	16,7	12,35	18,3	3,29
Kurtosis en Y	16,92	7,28	7,85	8,31	3,25

Il ressort que les répartitions sont loin de suivre un modèle gaussien d'après les valeurs de Kurtosis obtenues ; celles-ci justifient l'utilisation d'un filtre médian. Hormis l'aptitude de ce filtre à réduire le bruit analysé, l'absence de coefficients de lissage par rapport au filtre linéaire permet d'envisager une application plus universelle s'adaptant aux divers types de vidéo (vidéo-surveillance trafic/piéton, analyse d'archive télévisuelle…).

4. Algorithme

4.1 Filtrage spatio-temporel

L'implantation algorithmique exploite le parseur du logiciel de référence MPEG-2 [5] pour l'extraction des VEM et des coefficients de chaque bloc DCT. Dans un premier temps, un vecteur est associé aux blocs codés en mode intra. Ceux-ci sont déterminés par interpolation à

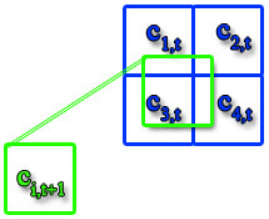
partir des blocs voisins de celui considéré sur les images immédiatement avant et après celle traitée. Un filtrage spatial (médian) est alors appliqué sur chaque trame individuellement.
Le filtrage temporel est ensuite assuré par l'élimination des vecteurs isolés dans un voisinage en 8-connexité pour les trames t-1 et t+1.

4.2 Filtrage orienté contenu

L'étape suivante visant à supprimer les problèmes dits de mur blanc, estime une carte de confiance à partir de l'indice : $Confiance = \sqrt{AC_{10}^{\ 2} + AC_{01}^{\ 2}}$. Cette carte, calculée intégralement sur une trame intra, est ensuite propagée sur l'ensemble de la séquence par l'interpolation suivante (le facteur 8 correspond ici au découpage en blocs 8x8 de MPEG-2) :

$$c_{i,j,t+1} = \frac{1}{64}\begin{bmatrix} c_{i+x/8,\,j+y/8,\,t} \times [\,[8-(x \bmod 8)] \times [8-(y \bmod 8)]\,] \\ + c_{i+x/8+1,\,j+y/8,\,t} \times [\,[x \bmod 8] \times [8-(y \bmod 8)]\,] \\ + c_{i+x/8,\,j+y/8+1,\,t} \times [\,[8-(x \bmod 8)] \times [y \bmod 8]\,] \\ + c_{i+x/8+1,\,j+y/8+1,\,t} \times [\,[x \bmod 8] \times [y \bmod 8]\,] \end{bmatrix}$$

avec t indice temporel de la trame courante, i et j les coordonnées du bloc considéré et x et y les valeurs du VEM associé (en pixels).



Ces traitements sont appliqués pour chacune des images de type P et B. Une option permet de ne pas prendre en compte ce dernier, à des fins de comparaison directe avec la plupart des algorithmes développés dans le domaine décompressé et ne considérant que les images de type I et P. L'ensemble des résultats est synthétisé dans la dernière section de ce document.
La carte de confiance est par ailleurs également remise à jour partiellement en présence de blocs codés intra présents dans les images P et B. Les vecteurs correspondant aux zones de faible confiance, donc de faible gradient - associables à une zone unie - sont finalement éliminés via un seuillage fondé sur l'indice de confiance. Cette dernière opération a pour le moment été déterminée de manière expérimentale à un centième de la confiance maximale de l'image courante, ce qui permet une actualisation de la confiance en continu.

4.3 Segmentation et description

Le seuillage aboutit à une carte binaire contenant les différents blobs représentatifs des objets mobiles, qui sont ensuite comptés sur chaque trame et identifiés par un histogramme de répartition de luminance et couleur, l'ensemble étant mémorisé dans un fichier de descripteurs (en rouge sur la fig. 4). Une première segmentation est donc ainsi proposée directement dans le domaine compressé.

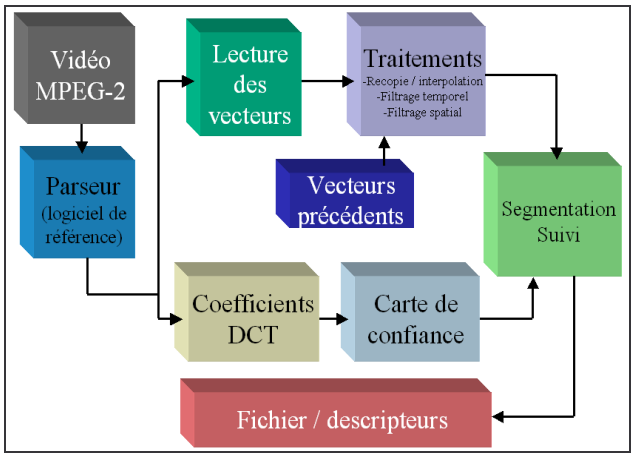


Figure 4 - Schéma bloc de l'application

5. Résultats et perspectives

Une fois l'ensemble de ces filtrages effectués, le bruit est considérablement réduit (fig. 5 et tab 3). Les valeurs extrêmes ont disparu, réduisant globalement de moitié l'écart type (selon les séquences), la valeur de Kurtosis est maintenant pour la séquence Speedway par exemple de 3.62 et 4.87 en X et Y. Dans tous les autres cas, la valeur de Kurtosis est réduite et permet d'assimiler la distribution à une loi normale.

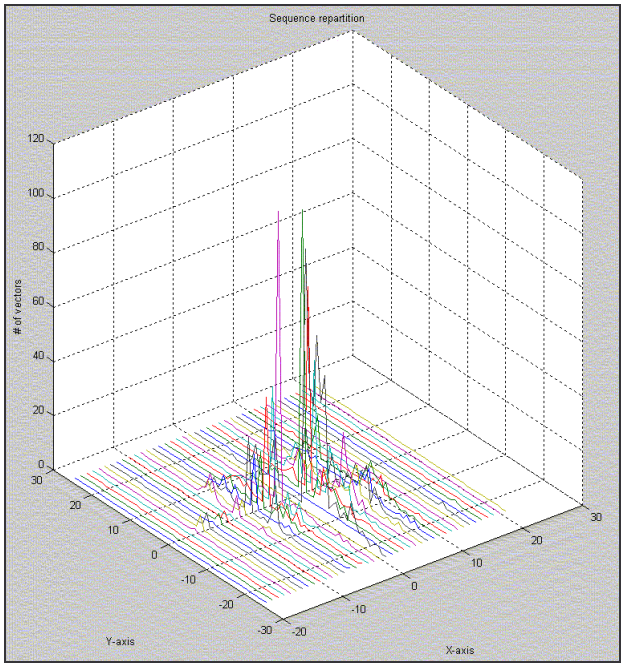


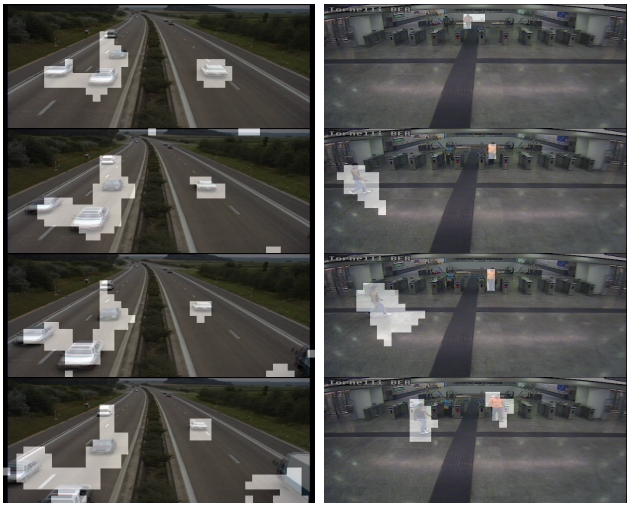
Figure 5 - Répartition des VEM après filtrages selon leurs composantes sur la séquence Speedway

Cette procédure a été appliquée au corpus détaillé précédemment. Le programme développé a été testé sur un Pentium 4 3GHz avec pour tâche le filtrage de l'ensemble des vecteurs, le comptage pour chaque trame du nombre de blobs attachés aux objets mobiles détectés ainsi que la détermination des histogrammes associés. La deuxième

partie du tableau 3 reprend l'ensemble des résultats sur les différentes séquences.

TAB. 3 : Distribution des VEM après filtrage

Séquence	A	B	C	D	E
Analyse statistique					
Moyenne en X	-2,42	-16,1	-6,77	-12,4	-1,12
Ecart type en X	2,92	13,5	7,24	12,2	2,43
Kurtosis en X	3,62	2,64	2,39	2,56	2,17
Moyenne en Y	-1,58	-2,23	-5,12	-3,48	-0,53
Ecart type en Y	6,83	3,11	3,19	4,30	2,31
Kurtosis en Y	4,87	2,43	2,56	4,09	1,25
Nombre d'images traitées par seconde					
Traitement I-P-B	125	220	410	405	140
Traitement I-P	160	265	520	510	170



Séquence Speedway Séquence Métro Turin

Figure 6 - Première segmentation – Suivi sur un GOP

La figure 6 montre que la segmentation obtenue correspond bien aux véhicules pour la séquence Speedway, aux piétons pour la séquence dans le métro de Turin même si la résolution dans le domaine compressé reste grossière (un macrobloc correspondant à une zone de 16x16 pixels). On peut également apprécier l'évolution de la segmentation au cours du temps.

Sur le plan algorithmique, la suite de nos travaux portera tout d'abord sur l'introduction d'un post-traitement spécifique pour isoler les voitures de la route par soustraction du fond. Puis, un suivi temporel sera développé, pour permettre une comparaison avec des méthodes équivalentes dans le domaine décompressé.

En parallèle, les algorithmes seront portés sur une implantation prenant en charge le format MPEG-4 pour intégration et contribution au projet européen Caretaker [17]. De même le corpus vidéo du projet Infom@gic du pôle de compétitivité Cap Digital sera utilisé. Cela permettra une validation à grande échelle (sur un corpus de plusieurs dizaines d'heures) dans le cadre d'une application de vidéo-surveillance de type observation de lieux publics et urbains.

Remerciements

Les auteurs adressent leurs remerciements au Pr. Benoît Macq de l'Université catholique de Louvain La Neuve pour le droit d'usage de la séquence *Speedway*.

Références

[1] B. Shen, I. Sethi, "Direct feature extraction from compressed images", *SPIE Storage and Retrieval for Image and Databases IV 2670*, 1996.

[2] R. Wang, T. Huang, "Fast camera motion analysis in MPEG domain", *ICIP*, 1999.

[3] Miguel Tavares Coimbra, "Compressed Domain Video Processing with Applications to Surveillance", *PhD Thesis, Department of Electronic Engineering, Queen Mary, University of London*, 2004.

[4] R. Coudray, Thèse : "Réutilisation des informations de compensation de mouvement d'un flux MPEG : évaluation qualitative et applications possibles", *Doctorat de l'université de La Rochelle*, 24 novembre 2005.

[5] Reference software MPEG-2 ISO/IEC 13818-5.

[6] R. Coudray, B. Besserer, "Global motion estimation for MPEG-encoded streams", Proc of IEEE Int. Conf. on Image Processing, 2004.

[7] R. Coudray, B. Besserer, "Motion based segmentation using MPEG streams and watershed method", Int. Symposium on Visual Computing - Lecture Notes in Computer Sciences, 2005.

[8] J. Heuer, A. Kaup, "Global motion estimation in image sequences using robust motion vector field segmentation", ACM Multimedia, p 261-264, November 1999.

[9] S. Porter, M. Mirmehdi, B. Thomas, "Video indexing using motion estimation", The British Machine Vision Conference, 2003.

[10] Lan Dong, Imad Zoghlami, Schwartz, "Object tracking in compressed video with confidence measure", *ICME*, 2006

[11] Duan, Yu, Tian, Sun, "Face pose analysis from MPEG compressed video for surveillance applications", *ITRE*, 2003

[12] Li, Jiang, Hashimah, "Dominant color extraction in dct domain", *VIE*, 2003

[13] Creusere, Dahman, "Object detection and localization in compressed video", *SSC*, 2001

[14] Fernando, Canagarajah, Bull, "Statistical feature extraction from compressed video sequences, *ICIP*, 2000

[15] Kim, Choi, Lee, "Fast scene change detection using direct feature extraction from MPEG compressed video", *ICPR*, 2000

[16] Projet européen WCAM http://wcam.epfl.ch, séquences disponibles sur http://wcam.epfl.ch/seq_video_surv/seq_video_surv.html

[17] Projet européen Caretaker, www.caretaker.org