

# Utilisation des transformées polynomiales pour l'indexation de documents audiovisuels

C.J. Rivero-Moreno<sup>1</sup> S. Bres<sup>1</sup>

<sup>1</sup> LIRIS, FRE 2672 CNRS, (Lab. d'InfoRmatique en Images et Systèmes d'information)

INSA de Lyon

Bât. Jules Verne, 17 av. Jean Capelle, 69621 Villeurbanne Cedex – France

rivero@rfv.insa-lyon.fr , stephane.bres@liris.cnrs.fr

## Résumé

*Cet article présente une méthode d'indexation des vidéos par combinaison d'informations spatiales et temporelles, extraites séparément des séquences vidéo. Nous utilisons pour ce faire deux familles de filtres qui sont de bons modèles du système visuel humain, à la fois du point de vue de la perception spatiale mais aussi temporelle. Ces filtres sont construits sur des transformées polynomiales qui réalisent une décomposition locale des signaux. Dans la version spatiale, nous utilisons les polynômes de Hermite (en accord avec les modèles de dérivées de Gaussienne) et les polynômes de Laguerre dans la version temporelle, qui préservent la caractéristique de la causalité. Par intégration des deux modèles, nous construisons un système d'indexation spatio-temporel. Une implémentation efficace est réalisée grâce aux versions discrètes de ces polynômes : les polynômes de Krawtchouk et de Meixner.*

*Ce travail est supporté par le Conseil National des Sciences et Technologies (CONACyT) du Mexique, grant 111539, et par le SEP du Mexique.*

## Mots clefs

Signature de vidéos, similarités de vidéos, monitoring, indexation, transformée de Hermite, transformée de Laguerre.

## 1 Introduction

Retrouver des séquences vidéo identiques ou similaires à une séquence requête constitue une problématique qui est au centre de la gestion des grandes bases de documents audiovisuels [2]. La taille de ces bases est en constante augmentation, ce qui nécessite des méthodes d'indexation et de recherche de mieux en mieux adaptées. Ces méthodes peuvent chercher l'exacte réplique de la séquence requête ou un ensemble de séquences similaires

à la requête. Cette dernière pose le problème de la définition de cette notion de similarité. Cependant, c'est ce dernier type de recherche qui semble le mieux répondre aux attentes souvent assez floues des utilisateurs. De plus, compte tenu du volume occupé par les séquences vidéo, il n'est pas envisageable de réaliser des comparaisons entre la requête elle-même et les vidéos de la banque. On utilise à la place une *signature* des vidéos. Ces signatures tentent de reprendre au mieux les caractéristiques essentielles de ces vidéos dans un volume minimal.

Nous proposons dans cet article une nouvelle façon de réaliser la signature d'une séquence vidéo en utilisant des transformations polynomiales, et plus particulièrement, la transformée de Hermite [4] et la transformée de Laguerre [1]. La première est utilisée dans le cadre d'une caractérisation spatiale des images de la séquence vidéo, la seconde pour une caractérisation temporelle.

## 2 Les transformées polynomiales

### 2.1 La transformée de Hermite

La transformée de Hermite [4], définie en deux dimensions, permet de décomposer une image selon un processus qui est un très bon modèle du système visuel humain, au même titre que les filtres de Gabor [5] ou les dérivées de Gaussiennes [8]. Les propriétés des filtres construits sur les polynômes de Hermite ont d'ailleurs des propriétés très voisines des filtres de Gabor [7].

Les filtres de Hermite  $d_{n-m,m}(x,y)$  décompose un signal  $l_v(x-p,y-q) = v^2(x-p,y-q) l(x,y)$  à travers une fenêtre Gaussienne  $v(x,y)$  d'écart type  $\sigma$  et d'énergie unité, sur l'ensemble des polynômes orthogonaux de Hermite  $H_{n-m,m}(x/\sigma, y/\sigma)$ . La convolution du signal  $l(x,y)$  par les filtres de Hermite donne les coefficients de cette décomposition. Ces filtres sont séparables ce qui permet des les implémenter efficacement et peuvent s'exprimer par des dérivées de Gaussienne. Ainsi, un filtre  $d_{n-m,m}(x,y)$  correspondant à une dérivation d'ordre  $(n-m)$  en  $x$  et  $m$  en  $y$  se calcule par  $d_{n-m}(x) d_m(y)$ , où chaque filtre 1D vaut :

$$d_n(x) = \left( (-1)^n / (\sqrt{2^n \cdot n!} \sqrt{\pi} \sigma) \right) H_n(x/\sigma) e^{-x^2/\sigma^2}$$

Les polynômes de Hermite  $H_n(x)$  sont définis par :

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2}$$

$H_n(x)$  est le polynôme de Hermite de degré  $n$ , orthogonal aux autres  $H_m(x)$  (pour  $n$  et  $m$  différents) par rapport à la fenêtre de pondération  $\exp(-x^2)$ . En pratique, ces filtres de Hermite sont implémentés à partir des polynômes de Krawtchouk qui en sont l'équivalent discret. La figure 1 donne un exemple de décomposition d'une image par ces filtres.

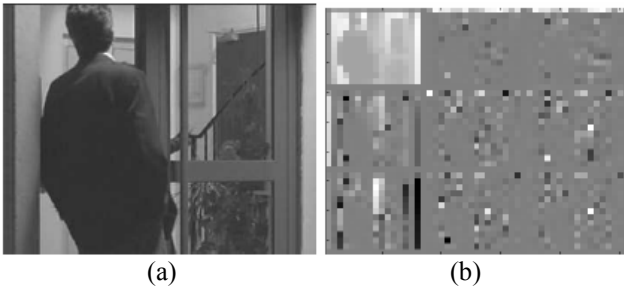


Figure 1 - (a) Image d'origine au format 4 :3 de taille 384x288 (b) Décomposition de Hermite jusqu'à l'ordre  $D=2$  et pour une fenêtre de pondération de longueur  $N = 48$ . Un recouvrement par moitié des fenêtres conduit aux 16 coefficients obtenus.

## 2.2 La transformée de Laguerre

La transformée de Laguerre est une transformée polynomiale au même titre que la transformée précédente. Elle se différencie par le choix de la fenêtre de pondération qui est ici définie sur l'intervalle  $[0, +\infty]$  :

$$v(t) = \sqrt{\sigma} (-\sigma t)^{\alpha/2} e^{-(\sigma t/2)} u(-t)$$

Cela conduit à des polynômes orthogonaux qui sont les polynômes de Laguerre, utilisés eux aussi sur l'intervalle  $[0, +\infty]$ . Nous utilisons cette particularité pour réaliser une décomposition temporelle unilatérale de la séquence vidéo : à partir d'un instant  $t_0$ , seul le passé ou l'avenir est pris en compte. Nous introduisons ainsi la notion de causalité très attachée à la dimension temporelle [3]. De plus, l'aspect de la fenêtre de pondération conduit à une prise en compte importante des événements qui sont proches de l'instant d'analyse (avec un maximum paramétrables à une distance  $\Delta t$ ). La décroissance de la courbe  $v(t)$  permet de prendre en compte des événements plus éloignés de l'instant d'analyse avec une importance de plus en plus faible au fur et à mesure que l'on s'en éloigne. Les polynômes de Laguerre sont définis par la formule de Rodrigues :

$$L_n^{(\alpha)}(t) = \frac{t^{-\alpha} e^t}{n!} \frac{d^n}{dt^n} (t^{n+\alpha} e^{-t})$$

En version discrète, nous utilisons les polynômes de Meixner. La figure 2 montre deux exemples de familles de ces polynômes pour deux séries de paramètres différents.

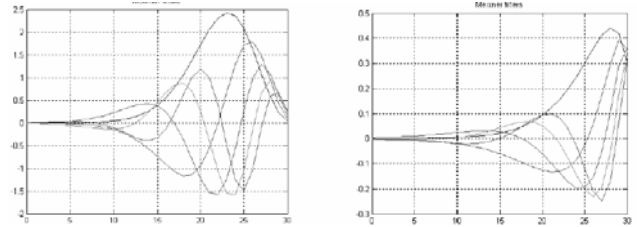


Figure 2 - Exemple de filtres de Meixner, jusqu'à l'ordre 4. On fait varier la position du maximum par rapport à l'instant d'analyse (borne à droite).

## 3 Signatures spatio-temporelles

A partir de ces décompositions par transformées polynomiales, nous pouvons obtenir des caractéristiques complémentaires sur les dimensions spatiales et sur la dimension temporelle :

- La transformée de Hermite nous permet d'analyser spatialement chaque image de la vidéo. La signature spatio-temporelle sera déduite de l'évolution de ces mesures avec le temps (sur chaque image de la séquence).
- La transformée de Laguerre nous permet d'avoir une analyse de l'évolution temporelle d'une zone donnée. La signature spatio-temporelle sera constituée de ces évolutions temporelles sur les différentes zones spatiales considérées.

### 3.1 Signature spatiale (Hermite)

Pour construire la signature spatiale d'une vidéo, nous calculons, sur la luminance de chaque image, une décomposition de Hermite sur la partie au format (4/3) centrale. Cette décomposition donne une analyse à plusieurs niveaux en fréquence, selon l'ordre des dérivées en  $x$  et en  $y$  considérées. Cela conduit à la représentation en cadrans de la figure 1b, résultat de la transformée de Hermite de la figure 1a. Les caractéristiques images mises en évidence sont du type contours, lignes, bandes ... Ce sont principalement ces caractéristiques avec orientations verticales et horizontales qui sont extraites. Afin de réduire la quantité d'informations contenues dans la signature, nous choisissons les paramètres de la transformée (longueur de la fenêtre gaussienne d'analyse et importance du recouvrement entre fenêtres) pour avoir 16x12 coefficients dans ces cadrans.

Nous calculons ensuite les histogrammes obtenus par projections sur l'horizontale (16 valeurs) et sur la verticale

(12 valeurs) des coefficients obtenus pour les cadrans (1,0) et (0,1) qui correspondent aux polynômes d'ordre 1 dans une direction (x ou y) et 0 dans l'autre. Cela s'apparente à l'utilisation de dérivées premières de Gaussiennes. Nous obtenons donc une information de type gradients vertical et horizontal respectivement. Nous ajoutons à la signature, le calcul de moments sur les autres cadrans. Les valeurs de ces histogrammes sont quantifiées sur 4 bits (16 valeurs).

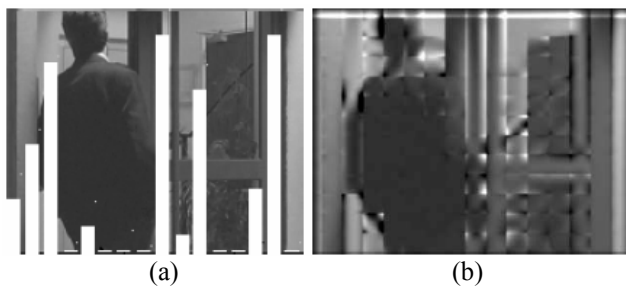


Figure 3 - (a) Exemple d'histogramme-signature (b) Image reconstruite jusqu'à l'ordre 2.

Un exemple de ce type d'histogramme est présenté sur la figure 3a. Il est calculé par projection du cadran (1,0). La signature spatio-temporelle finale, déduite de cette signature spatiale, est construite par agglomération temporelle des différentes mesures calculées à différents instants.

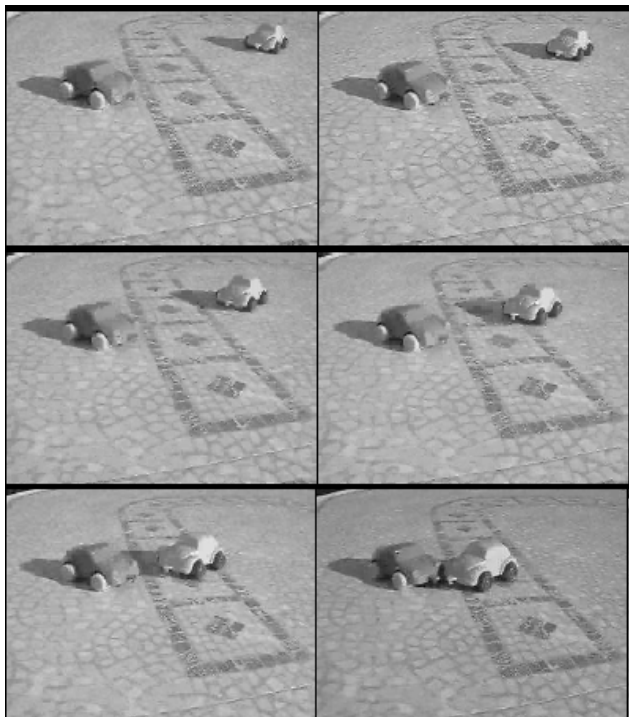


Figure 4 - Quelques images d'une vidéo exemple de  $N=30$  images.

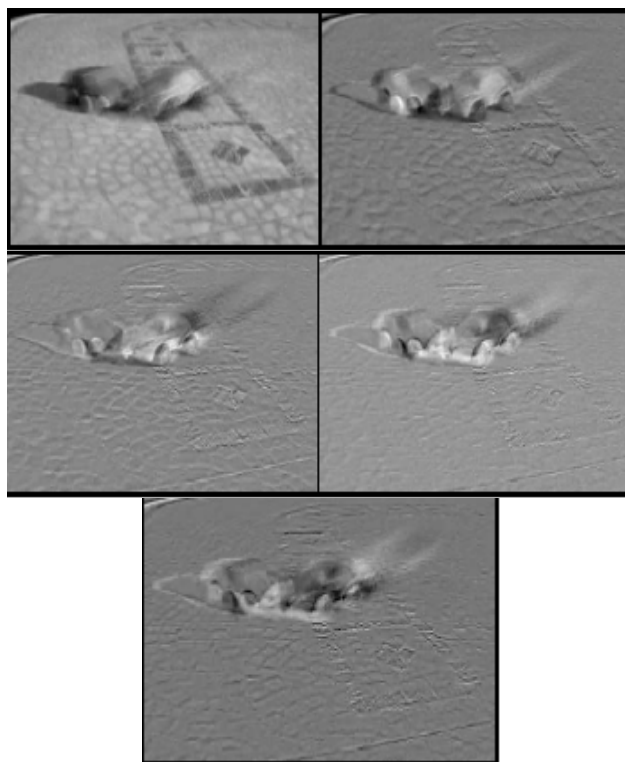


Figure 5 - Coefficients de Laguerre calculés jusqu'à l'ordre  $D=4$  sur la vidéo présentée à la figure 4. Les coefficients ont été calculés ici pour toutes les positions spatiales.



Figure 6 - Quelques images de la séquence de la figure 4 reconstruite à partir des coefficients de la figure 5 ( $D=4$ ).

Notre signature conduit donc à des similarités perceptuelles liées à la position des contours des formes dans les images des séquences vidéo. Pour visualiser l'information que l'on analyse en ne considérant que les coefficients de la décomposition de Hermite jusqu'à l'ordre  $D=2$ , nous avons affiché, sur la figure 3b, le résultat de la reconstruction de l'image de la figure 1a, à partir des coefficients conservés. On constate que l'essentiel de l'image est retrouvée.

### 3.2 Signature temporelle (Laguerre)

Nous construisons une signature temporelle de la séquence à partir de la décomposition par les polynômes de Laguerre, pour un sous ensemble de  $16 \times 12$  pixels issus de la partie centrale de l'image au format (4/3). La signature reprend les coefficients de cette décomposition. Les variations temporelles des pixels choisis sont d'autant mieux représentées qu'elles sont proches de l'instant  $t_0$ . Le nombre de frames prises en compte est paramétrable.

Nous prenons, dans les exemples qui suivent, la valeur  $N=30$ . Nous utilisons par ailleurs la décomposition de Laguerre jusqu'à l'ordre  $D=4$ . Ces valeurs ont été déterminées expérimentalement, et donnent un bon compromis entre la réduction d'information obtenue et la qualité de ce que l'on reconstruit à partir de cette l'information. La position du maximum de la fenêtre de pondération, qui permet de choisir l'instant qui contribue le plus à la signature, est choisie à  $\Delta t=0.8s$  de l'instant d'analyse (vers le futur ou vers le passé selon le sens que l'on choisit).

La signature spatio-temporelle finale est alors construite à partir du regroupement spatial des différentes décompositions de Laguerre, tronquées à l'ordre  $D=4$ .

## 4 Conclusion et perspectives

Nous avons présenté deux façons de construire une signature spatio-temporelle. La première approche consiste à extraire des mesures caractéristiques sur chaque image de la séquence vidéo. Ces caractéristiques images sont calculées ici à partir de la décomposition polynomiale de Hermite, et regroupe à la fois des histogrammes et des moments. C'est la juxtaposition de ces mesures qui permet d'observer leur évolution au cours de la séquence et qui introduit l'aspect temporel dans la signature.

A l'inverse, la deuxième approche est une approche temporelle pure dans un premier temps puisque l'on réalise la décomposition par transformée de Laguerre de l'évolution temporelle de la luminance de chaque point (après sous échantillonnage). L'aspect spatial de la signature spatio-temporelle est introduit après, en organisant, dans la signature, les évolutions temporelles de chaque point en fonction de sa position dans l'image.

Ces deux approches conduisent à deux signatures complémentaires que nous utilisons pour la recherche d'une séquence dans une base de données vidéo.

Nous testons actuellement nos signatures spatiale et temporelle sur une partie de notre base de vidéos de plus de 50Go. Les résultats obtenus jusqu'à maintenant sont encourageants et nous permettent de retrouver les séquences requêtes et les séquences considérées comme similaires d'un point de vue perceptuel : position spatiale des contours similaires et/ou évolution temporelle des intensités lumineuses similaires. Ces deux aspects se pondèrent pour favoriser un aspect plutôt que l'autre.

## Références

- [1] A.C. Brinken, J.A.J. Roufs, Evidence for a generalized Laguerre transform of temporal events by the visual system. *Biological Cybernetics*. Vol. 67 (1992) 395-402.
- [2] S. Dagtas, W. Al-Khatib, A. Ghafoor, R.L. Kashyap, Models for motion-based video indexing and retrieval. *IEEE Trans. Image Processing*, Vol. 9 1 (2000) 88-101.
- [3] T. Lindeberg, D. Fagerstorm, Scale-space with causal time direction. In: *Proc. European Conference on Computer Vision*. Vol. 1064 (1996) 229-240.
- [4] J.B. Martens, The Hermite transform - Theory. *IEEE Trans. Acoust. Speech Signal Processing*. Vol. 38 9 (1990) 1595-1606.
- [5] M. Porat, Y.Y. Zeevi, The generalized Gabor scheme of image representation in biological and machine vision. *IEEE trans. Pattern Analysis Mach. Intell.* Vol. 10 (1998) 452-468.
- [6] J. Puzicha, T. Hofmann, J.M. Buhmann, Non-parametric similarity measures for unsupervised texture segmentation and image retrieval. In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition (1997)* 267-272.
- [7] C.J. Rivero-Moreno, S. Bres, Conditions of similarity between Hermite and Gabor filters as models of the human visual system, N. Petkov and M.A. Westenberg (Eds.), *Computer Analysis of Images and Patterns*, Lectures Notes in Computer Science, vol. 2756, Springer-Verlag Berlin Heidelberg, pp. 762-769, 2003.
- [8] R.A. Young, R.M. Lesperance, W.W. Meyer, The Gaussian derivative model for spatial-temporal vision. *Spatial Vision*. Vol. 14 3,4 (2001) 261-319.