Segmentation de textures dynamiques: une méthode basée sur la transformée en curvelets 3D et une structure d'octree

Sloven DUBOIS[1,2], Renaud PÉTERI[1], Michel MÉNARD[2]

[1]MIA - Laboratoire Mathématiques, Image et Applications, Avenue Michel Crépeau 17042 La Rochelle, France

[2]L3i - Laboratoire Informatique Image et Interaction, Avenue Michel Crépeau 17042 La Rochelle, France

sloven.dubois01@univ-lr.fr, renaud.peteri@univ-lr.fr, michel.menard@univ-lr.fr

Résumé – Ce papier présente une nouvelle approche pour segmenter une séquence vidéo contenant des textures dynamiques. La méthode proposée s'appuie sur une transformée en curvelets 2D+t et une structure d'octree. La transformée en curvelets permet de capturer les structures spatio-temporelles à une échelle et orientation données. La structure en octree permet d'adapter la résolution d'analyse aux phénomènes locaux et améliore donc la segmentation spatio-temporelle. La méthode de segmentation de textures dynamiques est appliquée sur des séquences vidéos de textures dynamiques de synthèse et de scènes réelles. Des perspectives sont finalement exposées.

Abstract – This paper presents a new approach for segmenting a video sequence containing dynamic textures. The proposed method is based on a 2D+t curvelet transform and an octree hierarchical representation. The curvelet transform enables to outline spatio-temporal structures of a given scale and orientation. The octree structure based on motion coherence enables a better spatio-temporal segmentation than a direct application of the 2D+t curvelet transform. Our segmentation method is successfully applied on video sequences of dynamic textures. Future prospects are finally exposed.

1 Introduction

Les textures dynamiques sont un thème de recherche récent dans le domaine de l'analyse de séquences d'images. Celles-ci sont l'extension des textures statiques au domaine temporel. Elles se définissent donc comme un phénomène variant dans le temps et possédant une certaine répétitivité à la fois spatiale et temporelle. Un drapeau dans le vent, les risées à la surface de l'eau, de la fumée, ou un escalator sont autant de textures dynamiques présentes dans des vidéos. Leur étude est un sujet de recherche actif comportant de nombreuses applications comme la synthèse [8, 5], la segmentation [1, 3], la caractérisation [5, 7, 10], ...

Le contexte de nos travaux se situe dans le cadre de l'indexation de textures dynamiques dans des bases vidéos [4]. Il est fréquent qu'une ou plusieurs textures dynamiques apparaissent à divers endroits d'une séquence. Par exemple, dans la vidéo de la figure 1.(a), deux textures dynamiques sont visibles : l'écoulement de la rivière et les herbes au premier plan. Il convient alors de pouvoir segmenter spatio-temporellement ces textures dynamiques, avant d'en extraire des descripteurs qui serviront à l'indexation. De part leur étendue spatiale et temporelle inconnues, segmenter ces textures dynamiques dans une séquence d'images constitue un problème difficile.

Chaque texture dynamique possède ses propres caractéristiques, comme sa stationnarité, sa répétitivité, sa vitesse de propagation, ... En regardant une vidéo de mer (cf. Figure 1.(b)) deux mouvements peuvent être observés : un mouvement haute fréquence (l'écume) porté par un mouvement d'ensemble (les vagues). De nombreuses textures dynamiques peuvent ainsi se décomposer en une onde porteuse et en un phénomène localisé.

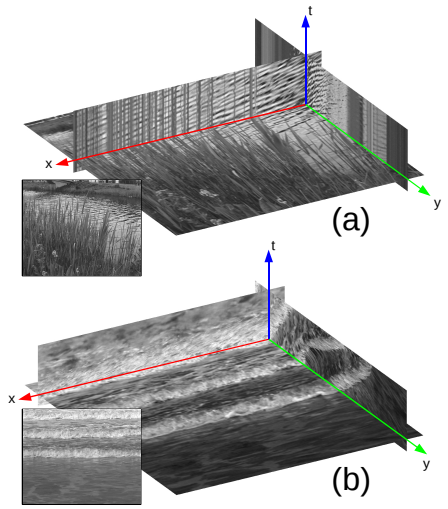

FIG. 1 – Coupes 2D+t de textures dynamiques. Ici, une texture dynamique est vue comme un cube de données qui est coupé au pixel $O(x, y, t)$ afin d'obtenir trois plans $(\overrightarrow{x}O\overrightarrow{y})$, $(\overrightarrow{x}O\overrightarrow{t})$ et $(\overrightarrow{y}O\overrightarrow{t})$.

Ces ondes porteuses ont principalement deux caractéristiques : une direction spatio-temporelle et une périodicité. La transformée en curvelets semble être un outil adapté pour caractériser cette onde. En effet, elle a été introduite afin de pallier les li-

mites rencontrées par la transformée en ondelettes : cette dernière capture très bien les singularités mono-dimensionnelles, alors que les curvelets peuvent détecter des singularitées d'ordre supérieur (des surfaces dans le cas de vidéos).

Dans cet article, après une présentation succinte de la transformée en curvelet 3D, nous verrons comment celle-ci, à l'aide d'une structure d'octree, permet de segmenter des vidéos au sens de la texture dynamique. Nous présenterons tout d'abord des résultats de synthèse afin de valider la démarche puis nous montrerons l'application sur des vidéos réelles issues de DynTex [6].

2 Transformée en curvelet 3D

La transformée en curvelets du volume $f(\mathbf{x}), \mathbf{x} = (x_1, x_2, x_3)$ correspondant aux coordonnées 3D, donne une collection de coefficients $c(j, \ell, k)$ définis par le produit scalaire suivant :

$$c(j,\ell,k) := \langle f, \varphi_{j,\ell,k} \rangle = \int_{\mathbb{R}^3} f(\mathbf{x}) \overline{\varphi_{j,\ell,k}(\mathbf{x})} d\mathbf{x} \qquad (1)$$

où $\varphi_{j,\ell,k}$ est l'atome de curvelet à l'échelle $j \in \mathbb{Z}$, dans la direction $\ell \in \mathbb{Z}$ et à la position $k = (k_1, k_2, k_3)$.

L'équation (1) peut s'exprimer dans le domaine des fréquences de la manière suivante :

$$c(j,\ell,k) := \frac{1}{(2\pi)^2} \int \widehat{f}(\boldsymbol{\omega}) \overline{\widehat{\varphi}_{j,\ell,k}(\boldsymbol{\omega})} d\boldsymbol{\omega} \qquad (2)$$

où $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3)$ est la variable du domaine fréquentiel.

Dans cet article, nous utilisons l'implémentation discrète de la transformée en curvelets [2].

$\widehat{\varphi}_{j,\ell,k}$ est définie dans le domaine fréquentiel par :

$$\widehat{\varphi}_{j,\ell,k}(\boldsymbol{\omega}) = U_{j,l}(\boldsymbol{\omega}) e^{i\langle x_k^{(j,\ell)}, \boldsymbol{\omega} \rangle} \qquad (3)$$

où $U_{j,l}(\boldsymbol{\omega})$ est la fenêtre fréquentielle discrète qui isole les fréquences à l'échelle j et selon la direction ℓ. $e^{i\langle x_k^{(j,\ell)}, \boldsymbol{\omega} \rangle}$ représente la translation de l'atome de curvelet à la position k. Cette fenêtre fréquentielle s'exprime par :

$$U_{j,l}(\boldsymbol{\omega}) = W_j(\boldsymbol{\omega}) V_{j,\ell}(\boldsymbol{\omega}) \qquad (4)$$

où $W_j(\boldsymbol{\omega})$ et $V_{j,\ell}(\boldsymbol{\omega})$ correspondent respectivement à la fenêtre fréquentielle radiale et angulaire.

La fenêtre radiale à l'échelle j est donnée par :

$$W_j(\boldsymbol{\omega}) = \sqrt{\Phi_{j+1}^2(\boldsymbol{\omega}) - \Phi_j^2(\boldsymbol{\omega})} \qquad (5)$$

où Φ est définie par le produit de fenêtres passe-bas de dimension 1 : $\Phi_j(\omega_1, \omega_2, \omega_3) = \phi(2^{-j}\omega_1)\phi(2^{-j}\omega_2)\phi(2^{-j}\omega_3)$. La fonction ϕ respecte $0 \leqslant \phi \leqslant 1$, est égale à 1 sur $[-1; 1]$ et à 0 sur $]-\infty; 2]$ et sur $[2; +\infty[$. Cette fenêtre est représentée par un gris moyen sur la figure 2.

La fenêtre angulaire se définit à partir de la face d'un cube. Par exemple, $V_{j,\ell}(\boldsymbol{\omega})$ est définie relativement à l'axe ω_1 par :

$$V_{j,l}(\boldsymbol{\omega}) = \phi\left(2^{j/2}\frac{\omega_2 - \alpha_l\omega_1}{\omega_1}\right)\phi\left(2^{j/2}\frac{\omega_3 - \beta_l\omega_1}{\omega_1}\right) \qquad (6)$$

Pour les autres faces du cube, la définition est similaire en échangeant les rôles de ω_1, ω_2 et ω_3. Cette fenêtre est représentée par un gris léger sur la figure 2.

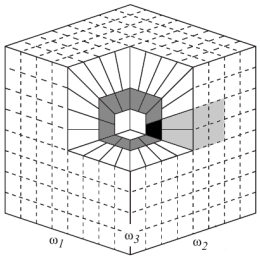

FIG. 2 – Découpage du domaine fréquentiel (adapté de [2]). Les gris léger et moyen représentent respectivement les fenêtres $V_{j,\ell}(\overrightarrow{\omega})$ et $W_j(\overrightarrow{\omega})$. La composition des deux fenêtres $U_{j,\ell}(\overrightarrow{\omega})$ est colorée en noir.

Pour plus d'information sur l'algorithme de la transformée en curvelets 3D, se référer à [2, 9]. Celui-ci se résume de la manière suivante :

- calculer la transformée de Fourier 3D de f ;
- pour chaque échelle et chaque direction, obtenir le produit $U_{j,\ell}(\boldsymbol{\omega})\widehat{f}(\boldsymbol{\omega})$;
- "déformer" le produit autour de l'origine pour obtenir $\mathcal{W}(U_{j,\ell}\widehat{f})(\boldsymbol{\omega})$. $\mathcal{W}$ est l'opération de "déformation" ;
- prendre l'inverse de la transformée de Fourier 3D de $\mathcal{W}(U_{j,\ell}\widehat{f})$ pour collecter les coefficients discrets $c(j, \ell, k)$.

La transformée en curvelets 3D a été conçue pour des données à trois dimensions (x,y,z). Pour pouvoir l'appliquer sur des séquences d'images, la relation $z = c.t$ est utilisée. c est une constante permettant de rendre homogènes les variables spatiales et temporelle. c devra s'adapter à la séquence d'images.

Dans la section suivante, nous utiliserons la transformée en curvelets pour segmenter spatio-temporellement une séquence vidéo.

3 Segmentation spatio-temporelle au sens des textures dynamiques

Comme cela a été précisée, une texture dynamique est souvent constituée d'une onde porteuse se propageant dans une direction spatio-temporelle. Si une vidéo possède plusieurs textures dynamiques, différentes directions vont alors apparaître à des échelles éventuellement différentes. La première méthode mise en place utilise la transformée en curvelet 3D afin de séparer ces différentes orientations.

Le principe de cette méthode est le suivant :

- calcul de la transformée en curvelet 3D sur toute la vidéo ;
- calcul de l'énergie pour chaque orientation ℓ à chaque échelle j ;
- segmentation des différentes directions spatio-temporelles selon des critères d'orientation et d'échelle des coefficients de curvelets.

Cette première méthode fournit des résultats intéressants quant à la détection des textures dynamiques, mais n'a pas permis une localisation précise des frontières des différents phénomènes.

Le principe de la seconde méthode est de ne pas appliquer la transformée en curvelet sur toute la vidéo, mais au contraire, afin d'améliorer la segmentation, d'effectuer une analyse locale des ondes porteuses. L'approche repose sur une décomposition de la vidéo à l'aide d'un octree. Celui-ci est utilisé pour partitionner un espace tridimensionnel, en le subdivisant récursivement en huit sous-espaces, tant que celui-ci n'est pas homogène (cf. Figure 3).

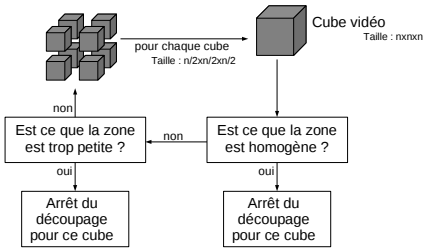

FIG. 3 – Principe général de l'octree

Dans le cadre des textures dynamiques, le critère d'homogénéité fonctionnera donc de la manière suivante :

- calcul de la transformée en curvelet sur une vidéo de taille (t_x, t_y, t_t)
- calcul de l'énergie pour chaque orientation ℓ de chaque échelle j
- normalisation des energies par une fonction pénalisant les directions spatiales
- **Si** plus d'une direction est détectée **alors**
  - recommencer avec huit sous-cubes de taille $(\frac{t_x}{2}, \frac{t_y}{2}, \frac{t_t}{2})$
- **Sinon**
  - arrêter
- **Fin Si**

A l'aide de la structure en octree, un arbre est obtenu. Celui-ci est constitué d'un ensemble de noeuds et de feuilles. Un noeud représente une subdivision d'une zone de la vidéo en 8 sous-cubes, et une feuille définit une zone de la vidéo homogène en terme de direction. Afin d'établir une segmentation, il faut parcourir l'arbre de décomposition et comptabiliser les différentes directions spatio-temporelles. Puis un algorithme de détection des maxima locaux est utilisé pour extraire les principales directions et ainsi déterminer les différentes textures dynamiques présentes dans la vidéo.

4 Résultats

L'approche décrite précédement a été validée dans un premier temps sur des vidéos de synthèse et dans un second temps sur des vidéos de la base DynTex [6].

Parmi ces différents tests, nous décrivons dans cet article les résultats obtenus à partir de la vidéo présentée sur la Figure 4.(a). Cette dernière est composée de deux textures dynamiques identiques du point de vue spatial mais possédant des directions spatio-temporelles différentes à une même échelle.

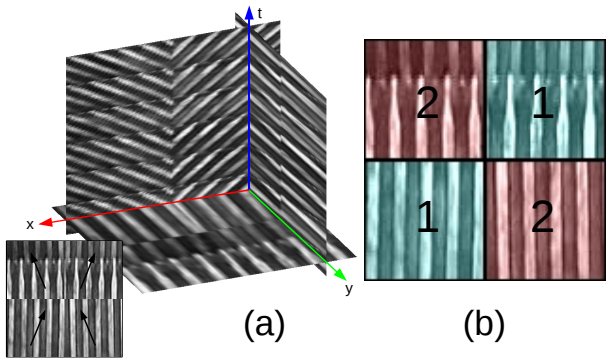

FIG. 4 – Résultat de la segmentation (b) sur une vidéo de synthèse (a). Chaque couleur (bleue et rouge, labélisées respectivement par 1 et 2) représente une région différente.

Le résultat de la segmentation est donné sur la figure 4.(b). Chaque couleur représente une texture dynamique différente. Malgré la ressemblance spatiale très grande, la segmentation distingue les différentes plages. Cette séparation est dû au fait que la méthode discrimine les directions spatio-temporelles en pénalisant l'information spatiale. En effet, lors de la recherche de maxima locaux, une fonction de normalisation est utilisée afin de valoriser l'information spatio-temporelle.

Nous présentons maintenant un résultat obtenu sur une vidéo réelle (cf. Figure 5) issue de DynTex [6].

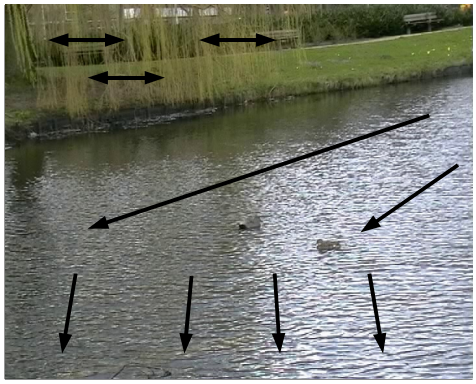

FIG. 5 – Vidéo originale. Les principales directions spatio-temporelles sont symbolisées ici par des flèches.

La Figure 6 montre le résultat de la segmentation de la vidéo de la Figure 5.

Tout d'abord, nous constatons que la rivière, qui est composée de deux directions spatio-temporelles différentes (cf. Figure 5, est bien divisée en deux régions disctinctes. La première (en rouge), au premier plan de la vidéo, contient la texture dynamique de la rivière se propageant vers le bas de l'image. La deuxième (en vert), au milieu de la vidéo, contient la même texture dynamique mais se propageant cette fois ci de droite à

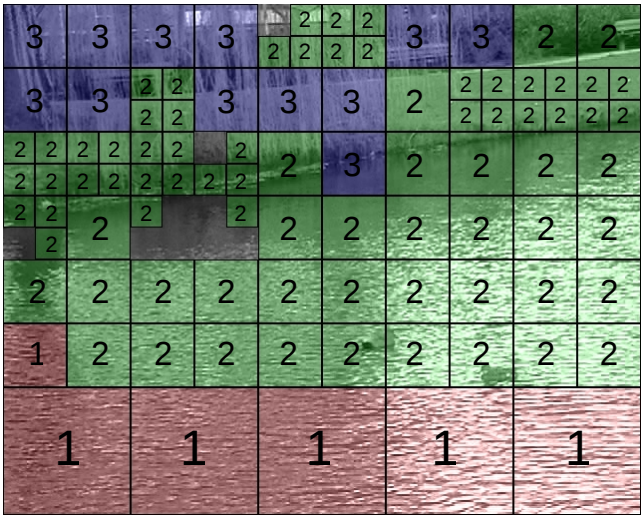

FIG. 6 – Résultat de la segmentation de la vidéo de la figure 5. Chaque couleur (rouge, verte et bleue, labélisées respectivement par 1, 2 et 3) représente une région différente. Une zone dépourvue de couleur correspond à une zone d'ambiguitée. Les traits noirs représentent les frontières des sous-cubes de la structure d'octree.

gauche.

Dans la région représentée en vert dans la Figure 6, une partie du rivage est présente, alors qu'il s'agit d'une partie statique de la séquence vidéo. Ce phénomène s'observe au moment de la fusion des différentes feuilles de l'arbre. En effet, le critère de fusion repose uniquement sur les directions spatio-temporelles, hors une zone statique n'est pas constituée, par définition, d'ondes porteuses. Pour éviter ce problème, il faudrait effectuer une fusion des différentes feuilles en regardant d'autres critères (couleurs, informations spatiales, ...).

La texture dynamique détectée dans la région bleue présente des particularités complexes à analyser : mouvement oscillant en superposition à un fond présentant une texture statique. L'algorithme, même dans un cas difficile, extrait correctement la composante dynamique de la texture.

Les régions dont la texture dynamique ne présente pas de direction particulière sont classées comme régions ambigües et sont laissées en niveaux de gris.

5 Conclusion

Ce papier étudie l'utilisation de la transformée en curvelet 3D pour l'analyse des vidéos. Une méthode pour segmenter spatialement et temporellement des séquences vidéos de textures dynamiques a été présentée. Celle-ci a été appliquée avec succès dans un premier temps sur des cas de synthèse, puis sur des vidéos réelles issues de DynTex [6].

Des travaux sont en ce moment effectués pour améliorer le critère d'homogénéité afin d'augmenter la précision de découpage de l'octree. D'autre part, il est également envisagé d'améliorer la fusion des différentes régions lorsque l'on parcours l'arbre de l'octree. En effet, pour l'instant le regroupement est uniquement effectué sur les directions spatio-temporelles, mais il est possible de regarder d'autres critères pour répondre à certaines ambiguïtés.

La segmentation obtenue peut servir à différencier les différentes textures dynamiques présentes dans une séquence vidéo. Ainsi pour chaque région, on peut extraire un vecteur de caractéristiques pouvant servir dans le cadre de l'indexation de textures dynamiques [4].

Références

[1] T. Amiaz, S. Fazekas, D. Chetverikov, and N. Kiryati. Detecting regions of dynamic texture. In Lecture Notes in Computer Science, editor, *1st International Conference on Scale Space and Variational Methods in Computer Vision (SSVM'07)*, volume 4485, pages 848–859, 2007.

[2] E. Candes, L. Demanet, D. Donoho, and L. Ying. Fast discrete curvelet transforms. Technical report, California Institute of Technology, March 2006.

[3] G. Doretto, D. Cremers, P. Favaro, and S. Soatto. Dynamic texture segmentation. In *Proceedings of Ninth IEEE International Conference on Computer Vision (ICCV'03)*, volume 2, pages 1236–1242, 2003.

[4] S. Dubois, R. Péteri, and M Ménard. A comparison of wavelet based spatio-temporal decomposition methods for dynamic texture recognition. In *4th Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA'09)*, volume 5524, pages 314–321, Povoa de Varzim, Portugal, 2009.

[5] J. Filip, M. Haindl, and D. Chetverikov. Fast synthesis of dynamic colour textures. In *Proceedings of the 18th IAPR Int. Conf. on Pattern Recognition (ICPR'06)*, pages 25–28, Hong Kong, 2006.

[6] R. Péteri, M. Huiskes, and S. Fazekas. Dyntex : A comprehensive database of dynamic textures. http ://www.cwi.nl/projects/dyntex/.

[7] P. Saisan, G. Doretto, Y. N. Wu, and S. Soatto. Dynamic texture recognition. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'01)*, volume 2, pages 58–63, Kauai, Hawaii, December 2001.

[8] M. Szummer and R. W. Picard. Temporal texture modeling. In *Proceedings of IEEE International Conference on Image Processing (ICIP'96)*, volume 3, pages 823–826, 1996.

[9] L. Ying, L. Demanet, and E. Candes. 3d discrete curvelet transform. In *Proceedings of the International Society for Optical Engineering (SPIE)*, 2005.

[10] G. Zhao and M. Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence journal (TPAMI'07)*, 6(29) :915–928, 2007.