

Décomposition de séquences d'images vidéo.

Mathieu LUGIEZ^{1,2}, Michel MÉNARD¹, Abdallah EL-HAMIDI²

¹Laboratoire L3i

Avenue Michel Crépeau, 17042 La Rochelle, France

²Laboratoire Mathématique Image et Applications

Avenue Michel Crépeau, 17042 La Rochelle, France

mathieu.lugiez02@univ-lr.fr, michel.menard@univ-lr.fr
abdallah.el-hamidi@univ-lr.fr

Résumé – La caractérisation de textures présentes dans les séquences vidéo représente un des thèmes majeurs de l'analyse d'images. Elle est particulièrement utile dans le cas de l'indexation et de l'inpainting vidéo. Suivant des travaux récents, initiés par Y.Meyer[1], les modèles de décomposition apparaissent comme l'une des voies les plus prometteuses pour extraire les textures au sein d'une séquence d'images. Les textures dynamiques se distinguent des textures statiques par le fait qu'elles présentent des variations temporelles caractéristiques. Cependant, à notre connaissance, peu de travaux traitent actuellement de cet aspect pourtant inhérent aux séquences d'images [2, 3]. Dans cet article nous proposons d'étendre le modèle de la décomposition d'images couleur proposé par Aujol et Chambolle [4] à l'aspect temporel. Le but est d'extraire la composante dynamique de la texture. L'algorithme repose également sur le schéma de projection proposé par Chambolle [5].

Abstract – Following recent work, initiate by Y.Meyer [1], decomposition approach appears as the best way to extract meaningful texture information present in image sequences. Indeed, dynamic textures present characteristic variations along the temporal dimension which can be catch in a decomposition model. However, only very few work in literature is available to deal with spatiotemporal decomposition. In this way, we propose in the present paper, to extend color decomposition model to time aspect through an efficient scheme based on projected gradient algorithm. Therefore we provide a useful and innovative way to decompose color image sequences and thus to extract dynamic texture component independently of noise and geometrical information.

1 Introduction

La décomposition d'images en composantes d'intérêt apparaît comme un des objectifs majeurs dans le domaine de l'analyse d'images. Cette approche a d'abord été développée dans un but de restauration d'images bruitées. Ce problème mal posé se formalise de la façon suivante : soient f une image dégradée, A un opérateur linéaire et w un bruit additif. Alors la restauration d'images consiste à estimer l'image u telle que $f = Au + w$. L'approche classique de résolution repose sur l'introduction d'un terme de régularisation. Afin de mieux prendre en compte les discontinuités et les contours des images, Rudin, Osher et Fatemi on proposé dans leurs travaux un terme de régularisation reposant sur la Variation Totale (TV) et utilisant l'espace BV^1 [6]. Ils proposent de minimiser la fonctionnelle suivante :

$$\inf_{u \in BV} F^{ROF}(u) \left(J(u) + \frac{1}{2\lambda} \|f - u\|_{L^2}^2 \right) \quad (1)$$

avec λ le paramètre contrôlant la norme L_2 de la partie résiduelle $f - u$.

Cependant, Y. Meyer a prouvé que ce modèle ne prenait pas

¹ $BV(\Omega)$ est le sous espace des fonctions $u \in L^1(\Omega)$ telle que la quantité suivante, appelée variation totale de u , soit finie : $J(u) = \sup \{ \int_{\Omega} u(x) \operatorname{div}(\xi(x)) dx \}$ tel que $\xi \in C_c^1(\Omega, \mathbb{R}^2)$, $\|\xi\|_{L^\infty(\Omega)} \leq 1$

en compte les oscillations de l'image et n'intégrait pas la texture dans la composante bruit. Il a donc proposé un nouvel espace G^2 , proche de l'espace dual de BV pour modéliser ces oscillations. Il a ainsi proposé un modèle de décomposition d'image en composantes géométrie et texture (cf. Figure 1 pour un exemple de décomposition d'image). Cette approche de la décomposition a conduit au développement de nombreux travaux dans les domaines de la déconvolution [7], de la classification [8], de la segmentation [9] et de l'inpainting [10, 11, 12].

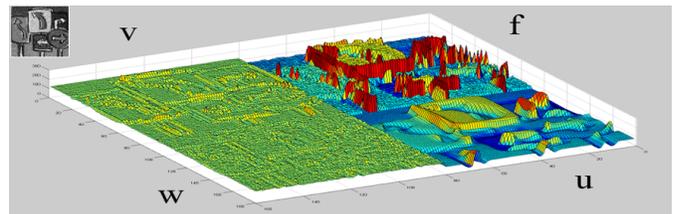


FIG. 1 – Décomposition d'image en niveaux de gris où v est la composante texture, f l'image originale, w le bruit et u la partie régulière.

² G , est l'espace de Banach, composé de toutes les distributions $v : v = \partial_1 g_x + \partial_2 g_y = \operatorname{div}(g)$ with g_x et $g_y \in L^\infty(\mathbb{R}^2)$ doté de la norme suivante $\|v\|_G = \inf \{ \|g\|_{L^\infty(\Omega, \mathbb{R}^2)} / v = \operatorname{div}(g), g = (g_x, g_y), |g(x)| = \sqrt{(|g_x|^2 + |g_y|^2)}(x) \}$

2 L'approche spatio-temporelle

La caractérisation de textures présentes dans des séquences vidéo nécessite la capture de la dynamique de cette texture. Cette analyse conduit naturellement à définir la notion de texture dynamique. Qu'elle soit dynamique ou purement temporelle une texture présente des répétitivités spatiales ainsi qu'une évolution temporelle. L'idée principale est qu'une telle séquence présente des variations le long de l'axe temporel qui doivent être prises en compte dans processus de décomposition. Ces variations peuvent être purement temporelles (clignotement lumineux), spatiales en mouvement, périodiques, pseudo-périodiques ou spatio-temporelles (Voir Fig. 2).

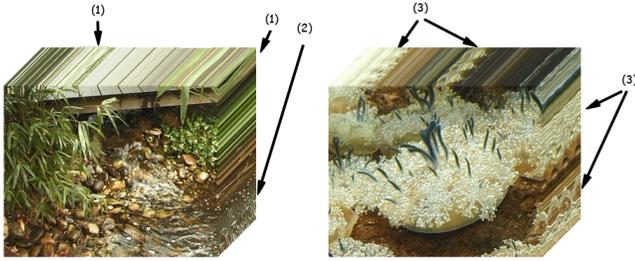


FIG. 2 – Bloc vidéo 2D+t de deux textures dynamiques. Ici, la texture dynamique est vue comme un volume. On peut distinctement voir en (1) les feuilles qui oscillent sous l'effet du vent (une texture spatiale en mouvement), en (2) l'eau s'écoulant et formant des vaguelettes (une texture spatio-temporelle) et en (3) les basses fréquences spatio-temporelles liées aux contractions périodiques de l'anémone de mer.

Pour se faire, un processus de décomposition reposant sur la Variation Totale apparaît comme un outil idéal pour prendre en compte de telles oscillations. Nous considérons une vidéo comme un cube image 3-D [13] (i.e un volume) afin d'appliquer les algorithmes 2-D correctement étendus à cette troisième dimension pour le cas 2D+t.

Afin d'extraire la composante texture dynamique pure, la structure géométrique et le bruit, nous proposons d'étendre le modèle de décomposition d'image couleur aux séquences d'images. Pour se faire, le modèle proposé par Aujol et Chambolle [4], reposant sur les normes duales des espaces BV , G et E^3 , apparaît comme robuste et parfaitement adapté :

$$\inf_{(u,v,w) \in X^3 / f=u+v+w} F^{AC}(u,v,w) = \underbrace{J(u)}_{\text{Régularisation : Variation Totale}} + \underbrace{J^*\left(\frac{v}{\mu}\right)}_{\text{Extraction de la texture}} + \underbrace{B^*\left(\frac{w}{\delta}\right)}_{\text{Extraction du bruit par seuillage}} + \underbrace{\frac{1}{2\lambda} \|f - u - v - w\|_X^2}_{\text{Partie résiduelle}} \quad (2)$$

avec X l'espace Euclidien $\mathbb{R}^{N \times N}$. J^* et B^* , sont respective-

³ Soit $\dot{B}_{1,1}^1$ l'espace dual de l'espace homogène de Besov et l'espace dual de $\dot{B}_{1,1}^1$ est l'espace de Banach $E = \dot{B}_{-1,\infty}^\infty$

ment les transformées de Legendre-Fenchel de J^4 et B . Le paramètre λ contrôle la norme L_2 de la partie résiduelle $f - (u + v + w)$.

3 Extension spatio-temporelle

Soit une séquence d'images originale et bruitée $\mathbf{f} \in L^2(\Omega)$, où Ω est un domaine ouvert borné de \mathbb{R}^3 , avec les conditions de Lipschitz aux bords. Afin d'extraire \mathbf{u} , \mathbf{v} et \mathbf{w} à partir de \mathbf{f} , nous proposons :

- une version spatio-temporelle du vecteur gradient $|\nabla u_{xyt}|$ discrétisé :

$$(\nabla \mathbf{u})_{i,j,k} = ((\nabla \mathbf{u})_{i,j,k}^x, (\nabla \mathbf{u})_{i,j,k}^y, (\nabla \mathbf{u})_{i,j,k}^t), \quad (3)$$

$$\text{où } (\nabla \mathbf{u})_{i,j,k}^t = \begin{cases} \mathbf{u}_{i,j,k+1} - \mathbf{u}_{i,j,k-1} & \text{si } k < N \\ 0 & \text{si } k = N \end{cases}$$

- une redéfinition de la variation totale discrète prenant en compte l'évolution conjointe des plans couleurs inspirée par [14] :

$$J(\mathbf{u}) = \int_{\Omega} \sqrt{\sum_{c=R,G,B} \sum_{i=x,y,t} |((\nabla \mathbf{u})_{i,j,k}^c)|^2} dx dy dt \quad (4)$$

- une version vectorielle de l'espace G adaptée à l'espace couleur RGB et étendu à la dimension temporelle : soit G l'espace de Banach composé de toute les fonctions vectorielles $\vec{\mathbf{v}}(x, y, t) = (\mathbf{v}_R(x, y, t), \mathbf{v}_G(x, y, t), \mathbf{v}_B(x, y, t))$ qui peuvent s'écrire :

$$\vec{\mathbf{v}}(x, y, t) = (\text{div} \vec{\mathbf{g}}_R, \text{div} \vec{\mathbf{g}}_G, \text{div} \vec{\mathbf{g}}_B), \quad (5)$$

$$\text{avec } \mathbf{g}_{x,c}, \mathbf{g}_{y,c}, \mathbf{g}_{t,c} \in L^\infty(\mathbb{R}^3), c = R, G, B$$

induite par la norme $\|\mathbf{v}\|_*$ définie comme la plus petite borne de toutes les normes L^∞ des fonctions $|\vec{\mathbf{g}}|$, où :

$$\begin{aligned} |\vec{\mathbf{g}}| &= \sqrt{|\vec{\mathbf{g}}_R|^2 + |\vec{\mathbf{g}}_G|^2 + |\vec{\mathbf{g}}_B|^2} \\ &= \sqrt{\sum_{c=R,G,B} ((\mathbf{g}_{x,c})^2 + (\mathbf{g}_{y,c})^2 + (\mathbf{g}_{t,c})^2)} \end{aligned}$$

et où l'infimum est calculé parmi toutes les décompositions (6) de $\vec{\mathbf{v}}$.

- une nouvelle règle de seuillage par ondelettes, inspirée par Weickert [15] et adaptée dans Dubois et al. [16] pour le cas couleur. Tous les plans couleurs sont alors pris en compte conjointement par la fonction de seuillage soft, $W_{ST}(S_\theta(w_i^j, \delta))$, qui repose sur la mesure S_θ des coefficients d'ondelettes w_i^j :

$$S_\theta(w_i^j) = w_i^j *$$

$$\left(1 - 39\theta \left(\sqrt{\sum_i c * (w_{R,i}^j)^2 + c * (w_{G,i}^j)^2 + c * (w_{B,i}^j)^2} \right)^{-1} \right) \quad (6)$$

⁴ J est la fonction indicatrice sur $G_\mu = v \in G / \|v\|_G \leq \mu$ (i.e : $J^*(\frac{v}{\mu}) = 0$ si $v \in G_\mu$, $+\infty$ sinon.)

où w_j^i est le coefficient d'ondelette, j la résolution et où $c = 1$ si $i \in \{x, y, t\}$, $c = 2$ si $i \in \{xy, xt, yt\}$ et $c = 4$ si $i = xyt$. La fonction de seuillage soft par ondelettes W_{ST} pour un coefficient d'ondelette donné, avec le seuil δ est donnée par :

$$W_{ST}(S_\theta(w_j^i), \delta) = \begin{cases} S_\theta(w_j^i) - \delta & \text{if } S_\theta(w_j^i) \geq \delta \\ 0 & \text{if } |S_\theta(w_j^i)| \leq \delta \\ S_\theta(w_j^i) + \delta & \text{if } S_\theta(w_j^i) \leq -\delta \end{cases} \quad (7)$$

4 Implémentation numérique

La projection de Chambolle [5], récemment portée au cas couleur [14], est une méthode robuste et rapide pour résoudre numériquement les problèmes de minimisation induits par la fonctionnelle (2). La projection de Chambolle P sur l'espace λB_G de f est notée $P_{\lambda B_G}(f)$, et la solution est donnée itérativement pour chaque pixels (i, j, k) et à chaque itération $n + 1$ par :

$$\begin{aligned} \mathbf{P}^0 &= 0, \\ \mathbf{P}_{i,j,k}^{n+1} &= \frac{\mathbf{P}_{i,j,k}^n + \tau(\nabla(\text{div}(\mathbf{P}^n) - \frac{\mathbf{f}}{\lambda}))_{i,j,k}}{\max(1, |\mathbf{P}^n + \tau(\nabla(\text{div}(\mathbf{P}^n) - \frac{\mathbf{f}}{\lambda}))_{i,j,k}|_2)} \end{aligned} \quad (8)$$

avec $\tau \leq \frac{1}{8}$ afin d'assurer la convergence [4, 14]. Ainsi pour résoudre le problème de minimisation (2), nous utilisons l'algorithme proposé dans [4] :

1. Initialisation : $\mathbf{u}_0 = 0, \mathbf{v}_0 = 0, \mathbf{w}_0 = 0$
2. Itérations :

$$\begin{aligned} \tilde{\mathbf{w}} &= P_{\delta B_E}(\mathbf{f} - \mathbf{u} - \mathbf{v}) \\ &= \mathbf{f} - \mathbf{u} - \mathbf{v} - W_{ST}(\mathbf{f} - \mathbf{u} - \mathbf{v}, \delta) \\ \tilde{\mathbf{v}} &= P_{G_n}(\mathbf{f} - \mathbf{u} - \tilde{\mathbf{w}}) \\ \tilde{\mathbf{u}} &= \mathbf{f} - \mathbf{u} - \tilde{\mathbf{v}} - \tilde{\mathbf{w}} - P_{G_\lambda}(\mathbf{f} - \tilde{\mathbf{v}} - \tilde{\mathbf{w}}) \end{aligned}$$

3. Critère d'arrêt :

$$\max(|\tilde{\mathbf{u}} - \mathbf{u}|, |\tilde{\mathbf{v}} - \mathbf{v}|, |\tilde{\mathbf{w}} - \mathbf{w}|) \leq \epsilon$$

5 Résultats

Les différentes séquences proposées dans cette section sont issues de la base de donnée Dyntex [17] qui propose une large variété de séquences naturelles présentant toutes des textures dynamiques.

5.1 Extraction de la dynamique

Nous obtenons avec l'approche présentée précédemment la composante texture dynamique pure, les éléments structurels de la séquence régularisée spatio-temporellement ainsi que le bruit. La Figure 3 montre clairement l'apport et l'impact temporel de la décomposition : les vaguelettes situées aux abords

de la fontaine sont totalement intégrées dans la séquence texture dynamique, et le pourtour de la fontaine semble ainsi complètement gelé dans la séquence \mathbf{u} , régularisée dynamiquement. La taille des vaguelettes prise en compte est liée au paramètre d'échelle λ . Comme pour le cas statique, ces résultats sont une étape préliminaire à la caractérisation dans des séquences d'images de la texture dynamique qui représente donc une aide à l'indexation de séquences vidéo.

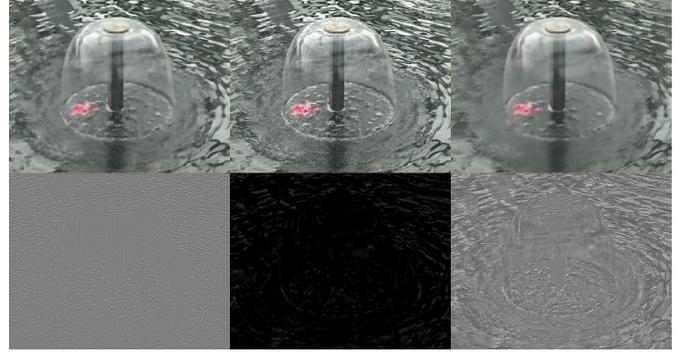


FIG. 3 – Haut : au centre, l'image originale, à gauche, la composante u obtenue par décomposition statique, à droite la composante \mathbf{u} obtenue par décomposition spatio-temporelle (le pourtour de la fontaine semble complètement gelé). Bas : à gauche, la composante texture statique v obtenue par décomposition statique. A droite : la séquence texture dynamique \mathbf{v} obtenue par décomposition spatio-temporelle. Les deux décompositions ont été calculées avec les mêmes paramètres (régularisation forte).

5.2 Débruitage vidéo

Afin de comparer l'efficacité de la décomposition spatio-temporelle avec l'approche statique, on se propose de tester leur capacité dans le cas de la restauration d'une séquence bruitée. Le protocole est le suivant : un bruit gaussien de moyenne nulle et d'écart type σ est ajouté sur chaque image de la séquence. La restauration est effectuée soit image par image pour le cas usuel dit statique (DS) soit sur le volume complet de la séquence à l'aide de la décomposition spatio-temporelle (DST). Puis on calcule le rapport signal sur bruit à partir de la séquence originale non-bruitée. On obtient ainsi les résultats suivants :

N° Séquence	SNR	ROF_{DS}	ROF_{DST}	AC_{DS}	AC_{DST}
54pd210 $\sigma=20$	13.54	17.57	18.55	17.99	19.07
6481g10 $\sigma=20$	14.46	17.58	19.17	17.66	19.73
64cc610 $\sigma=20$	14.53	23.03	29.09	25.10	30.24
64cc610 $\sigma=50$	10.85	16.40	21.04	20.58	26.91

Une vue détaillée des capacités de restauration, ciblée sur une zone d'intérêt présentant de forts contours est présentée sur la Figure 4. On observe clairement que la prise en compte de la dynamique permet d'obtenir de meilleurs résultats que la régularisation statique quelque soit le modèle initial.

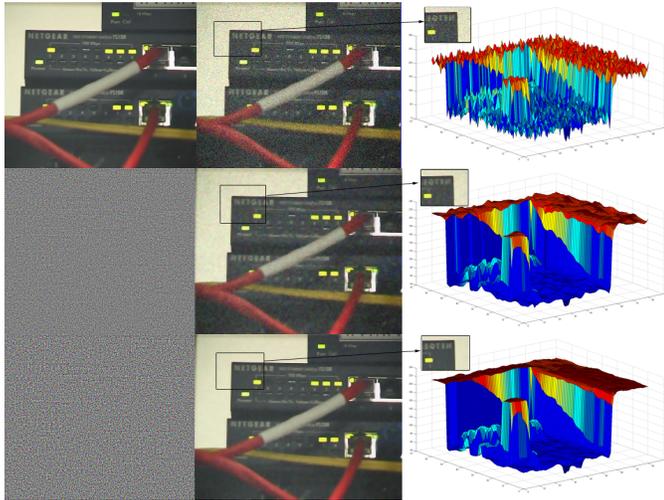


FIG. 4 – De gauche à droite, en haut : l'image originale (issue de la séquence 64cc610 de Dyntex), sa version dégradée par un bruit gaussien ($\sigma = 20$) et une vue de la zone d'intérêt (la led jaune formant le pic en avant plan). Au milieu : la composante bruit (w) issue de la décomposition spatiale (DS), sa composante régularisée spatialement u et une vue de la zone d'intérêt. En bas : la composante bruit (w) issue de la décomposition spatio-temporelle (DST), sa composante régularisée u et une vue de la zone d'intérêt. On peut observer que les zones franches, les aplats, sont mieux restaurés dans le modèle dynamique que dans le modèle statique.

6 Conclusion

Nous avons pu voir que la prise en compte des variations spatio-temporelles dans un modèle de décomposition permet d'extraire la composante texture dynamique au sein des séquences vidéo. Cette extension offre un bon outil de débruitage vidéo. En effet le modèle dynamique dépasse clairement le modèle statique et renforce la capacité de débruitage des modèles basés sur la Variation Totale. Cette méthode permet d'étendre également le processus de régularisation aux images volumiques à travers une régularisation tri-dimensionnelle afin de tenir compte de la structure 3D des objets.

Références

[1] Meyer, Y. : Oscillating patterns in image processing and nonlinear evolution equations. The fifteenth dean jacqueline B. Lewis Memorial Lectures (2001)

[2] Lugiez, M., Dubois, S., Ménard, M., El-Hamidi, A. : Spatiotemporal extension of color decomposition model and dynamic color structure-texture extraction. In : 4th European Conference on Colour in Graphics, Imaging, and Vision, Barcelona : Espagne (2008). (2008)

[3] Lugiez, M., Ménard, M., El-Hamidi, A. : Dynamic color texture modeling and color video decomposition using bounded variation and oscillatory functions. In :

ICISP '08 : Proceedings of the 3rd international conference on Image and Signal Processing, Berlin, Heidelberg, Springer-Verlag (2008) 29–37

[4] Aujol, J.F., Chambolle, A. : Dual norms and image decomposition models. International Journal of Computer Vision **63**(1) (2005) 85–104

[5] Chambolle, A. : An algorithm for total variation minimization and its applications. JMIV **20** (2004) 89–97

[6] L.Rudin, S.Osher, E.Fatemi : Nonlinear total variation based noise removal. Physica D **60** (1992) 259–269

[7] Bresson, X., Chan., T. : Fast minimization of the vectorial total variation norm and applications to color image processing. In : SIAM Journal on Imaging Sciences (SIIMS). (submitted 2007)

[8] Aujol, J.F., Chan, T.F. : Combining geometrical and textured information to perform image classification. J. Visual Communication and Image Representation **17**(5) (2006) 1004–1023

[9] Aujol, J.F., Aubert, G., Blanc-Féraud, L., Chambolle, A. : Image decomposition application to SAR images. In Griffin, L.D., Lillholm, M., eds. : Scale-Space. Volume 2695 of Lecture Notes in Computer Science., Springer (2003) 297–312

[10] Bertalmío, M., Vese, L.A., Sapiro, G., Osher, S. : Simultaneous structure and texture image inpainting. In : CVPR, IEEE Computer Society (2003) 707–712

[11] Ballester, C., Caselles, V., Verdera, J., Bertalmio, M., Sapiro, G. : A variational model for Filling-In gray level and color images. In : Proceedings of the Eighth International Conference On Computer Vision (ICCV-01), Los Alamitos, CA, IEEE Computer Society (2001) 10–16

[12] Verdera, J., Caselles, V., Bertalmío, M., Sapiro, G. : Inpainting surface holes. In : ICIP (2). (2003) 903–906

[13] Aubert, G., Kornprobst., P. : Mathematical Problems in Image Processing : Partial Differential Equations and the Calculus of Variations. Volume 147. Springer Verlag (2001)

[14] Duval, V., Aujol, J.F., Vese, L. : A projected gradient algorithm for color image decomposition. Technical report, CMLA Preprint 2008-21 (2008)

[15] Welk, M., Weickert, J., Steidl, G. : A four-pixel scheme for singular differential equations. In Kimmel, R., Sochen, N.A., Weickert, J., eds. : Scale-Space. Volume 3459 of Lecture Notes in Computer Science., Springer (2005) 610–621

[16] Dubois, S., Lugiez, M., Péteri, R., Ménard, M. : Adding a noise component to a color decomposition model for improving color texture extraction. In : 4th European Conference on Colour in Graphics, Imaging, and Vision, Barcelona : Espagne (2008). (2008)

[17] Péteri, R., Huiskes, M., Fazekas, S. : (Dyntex : A comprehensive database of dynamic textures) <http://www.cwi.nl/projects/dyntex/>.