

Mélange Spatio-temporel de Patches pour une Réduction d'Artefacts Préservant la Géométrie des Images et des Vidéos

Maxime DAISY, Pierre BUYSSENS, David TSCHUMPERLÉ, Olivier LÉZORAY

GREYC, Équipe Image
6 Bd Maréchal Juin
14000 Caen

maxime.daisy@ensicaen.fr, pierre.buysens@unicaen.fr,
david.tschumperle@ensicaen.fr, olivier.lezoray@unicaen.fr

Résumé – Les méthodes d’inpainting par patch provoquent souvent des artefacts de bloc engendrés par les discontinuités visibles entre les morceaux de patches recollés. Dans cet article nous proposons une méthode de mélange spatio-temporel de patches réduisant ce type d’artefacts. Nous définissons un modèle tensoriel guidant le mélange afin de mieux conserver les textures et les structures de la zone reconstruite. Rapide et générique, cette méthode améliore la qualité visuelle des résultats d’inpainting par patch des images et des vidéos.

Abstract – Patch-based inpainting methods often lead to block-effect artifacts provoked by visible discontinuities between reconstructed patch chunks. In this paper we propose a post-inpainting space-time patch blending algorithm able to lessen these artifacts. We define a tensor model that guides the spatial patch blending process and helps preserving both textures and structures of the reconstructed area as far as possible. Fast to compute and generic, it really improves the visual quality of exemplar-based inpainting results for images and videos.

1 Introduction et Contexte

L’inpainting par patch permet de reconstruire de larges portions d’images contenant des données manquantes ou corrompues. Présent dans certains logiciels d’infographie, il permet aux utilisateurs de traiter leurs photographies personnelles, mais aussi à des professionnels d’enlever des objets parasites dans des images ou des vidéos : des micros ou des fils par exemple. Comme décrit dans [10], les méthodes *basées patch* ont pour origines des travaux sur la synthèse de texture [9, 2]. Celles-ci, basées sur le principe d’auto-similarité, peuvent être classées en trois groupes principaux : 1) Les méthodes *gloutonnes* [3, 5, 6, 15, 19, 13] copient des morceaux de patches depuis la zone connue dans la zone à reconstruire. Montrant de bons résultats pour la synthèse de texture, elle sont moins performante en ce qui concerne les structures, et provoquent des artefacts de bloc. 2) Les méthodes *hybrides* [12, 4] continuent les structures principales avec une méthode basée géométrie, et reconstruisent les textures ensuite. En pratique, ces méthodes sont assez lentes, et utilisent une décomposition cartoon/texture qui est un problème mal posé. 3) Les méthodes basées *energie* minimisent des énergies modélisant la cohérence de l’inpainting, en utilisant la multi-résolution [18] ou des cadres variationnels [1]. D’autres méthodes utilisent des statistiques sur les décalages entre les motifs répétitifs d’une image et complètent l’image avec les décalages les plus fréquents [11]. Les résultats de [18] présentent des déformations de texture dûs à la multi-résolution et peuvent se montrer flou principalement à cause de la combinaison linéaire de plusieurs patches. Pour contrer

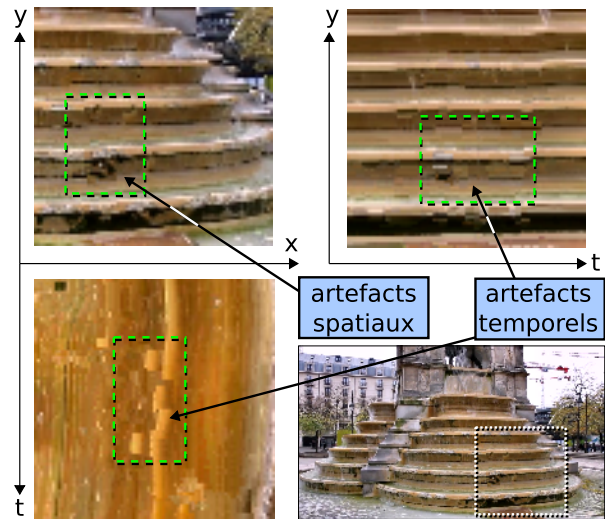


FIGURE 1 – Artefacts de bloc dans une vidéo inpaintée.

ces effets, [1] utilise une variante de Poisson dérivée de [16] pour réduire la jointure entre les morceaux de patches. Dans cet article nous nous focalisons sur le mélange de patches pour la réduction d’artefacts spatio-temporels (voir Fig. 1) dans des images et vidéo déjà inpaintées. Notre première contribution est un modèle tensoriel pour un mélange de patches guidé par la géométrie. Notre deuxième contribution est une extension de ce modèle pour la vidéo. Ces contributions permettent de réduire les artefacts de bloc tout en conservant les structures et les textures présentes dans les résultats.

2 Un Modèle Tensoriel pour un Mélange de Patches Guidé par Géométrie

Le mélange de patches décrit dans [8] permet de diminuer l'effet des artefacts de bloc dans une image inpaintée par patch, dans toutes les directions de la même façon. Cependant, cette méthode endommage les structures et les textures de l'image (voir Fig. 2, milieu). La méthode proposée définit un modèle tensoriel aidant le mélange à préserver les textures et les structures locales de l'image (voir Fig. 2, droite). Pour ce faire, la largeur de mélange doit être localement adaptée à la géométrie locale de l'image. Aussi, le mélange doit être orienté selon la direction des structures de l'image afin qu'elles soient préservées. Comme les amplitudes et les orientations du mélange sont aisément représentable par les valeurs propres et vecteurs propres de tenseurs, ce modèle est un des plus adaptés pour construire le notre. Les tenseurs de structure donnant une bonne description de la géométrie locale d'une image, nous les utilisons comme base pour notre modèle tensoriel de mélange. Un champs de tenseurs *complet* peut être reconstruit en utilisant la carte de correspondance générée par inpainting (plus d'information Section 4).



FIGURE 2 – Comparaison entre le mélange isotrope (milieu) et anisotrope (droite) sur un résultat d'inpainting avec [7] sans mélange spatial de patches (gauche).

Pour des tenseurs de structure \mathbf{S} , la valeur et la direction du changement d'intensité local sont représentés par leurs valeurs propres $\lambda_{\mathbf{S}\{1,2\}}$ et vecteurs propres $\mathbf{e}_{\mathbf{S}\{1,2\}}$. De façon similaire, les propriétés locales du mélange sont codées dans des tenseurs de mélange \mathbf{B} . Leurs valeurs/vecteurs propres $\lambda_{\mathbf{B}\{1,2\}}$ $\mathbf{e}_{\mathbf{B}\{1,2\}}$ représentent respectivement la largeur de bande et la direction du mélange. Une correspondance peut être faite entre les différentes configurations des tenseurs de structure et des tenseurs de mélange. Dans les zones homogènes, alors que le mélange est omnidirectionnel et assez fort pour enlever les petite variations, les tenseurs de mélange sont isotropes mais plus gros que les tenseurs de structure. Dans les zones de texture, les tenseurs de mélange sont de la même forme que les tenseurs de structure, c-à-d petits et isotropes. Pour les zones de fortes structures, alors que les petites craquelures doivent être effacées, les structures doivent rester intactes. Les tenseurs de mélange sont donc orientés selon la direction de ces structures, orthogonalement aux tenseurs de structure. À partir de cette description, nous proposons une suite d'opérations sur les tenseurs de structure afin de construire les tenseurs de mélange. Les valeurs propres des tenseurs de structure sont directe-

ment dépendantes du gradient, et donc de l'intervalle des valeurs possibles pour un pixel. Celles des tenseurs de mélange en revanche, ne le sont pas. La première étape est donc de normaliser les valeurs propres des tenseurs de structure :

$$\hat{\lambda}_{\mathbf{S}(p)i} = \frac{\lambda_{\mathbf{S}(p)i}}{\max_{q \in \mathcal{I}} \lambda_{\mathbf{S}(q)i}}$$

Le ratio entre les valeurs propres d'un tenseur de structure donne le type de géométrie local dans des images couleurs. Plus ce ratio est petit, plus le tenseur est anisotrope. Les valeurs propres d'un tenseur de mélange représentant la largeur du mélange dépend aussi de ce ratio. Nous modifions les valeur propres

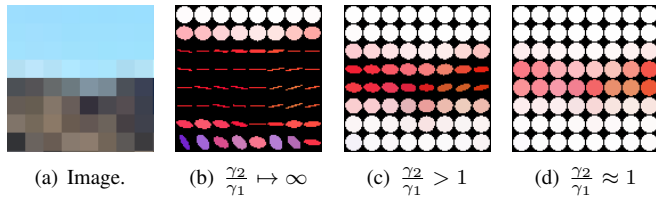


FIGURE 3 – Illustration de l'effet des paramètres γ_i .

normalisées $\hat{\lambda}_{\mathbf{S}(p)i}$ en utilisant ce ratio pour calculer les valeurs propres des tenseurs de mélange $\lambda_{\mathbf{B}i}$. La fonction utilisée est inspirée par les EDP de diffusion [17], et est définie par :

$$\lambda_{\mathbf{B}i} = \frac{1}{(1 + \hat{\lambda}_{\mathbf{S}1} + \hat{\lambda}_{\mathbf{S}2})^{\gamma_i}} \quad (1)$$

où γ_i ($i \in \{1, 2\}$) contrôle l'anisotropie des tenseurs de mélange. Des exemples de différentes configurations de γ_i sont fournis Fig. 3. Finalement, comme les tenseurs de mélange sont des matrices symétriques définies positives, ils peuvent être construits de la manière suivante :

$$\mathbf{B} = \lambda_{\sigma \mathbf{B}1} \mathbf{e}_{\mathbf{S}1}^{\perp} \mathbf{e}_{\mathbf{S}1}^{\perp T} + \lambda_{\sigma \mathbf{B}2} \mathbf{e}_{\mathbf{S}2}^{\perp} \mathbf{e}_{\mathbf{S}2}^{\perp T} \quad (2)$$

où $\lambda_{\sigma \mathbf{B}i} = \sigma_{\mathbf{B}} \lambda_{\mathbf{B}i}$, $\sigma_{\mathbf{B}}$ étant la largeur maximale du mélange. L'application de ce modèle au mélange de patches est fait en utilisant l'algorithme de [8] avec la formule suivante :

$$J(p) = \frac{\sum_{\psi_q \in \Psi_p} w_{\mathbf{B}}(q,p) \psi_q(p-q)}{\varepsilon + \sum_{\psi_q \in \Psi_p} w_{\mathbf{B}}(q,p)} \quad (3)$$

avec

$$w_{\mathbf{B}}(p, q) = e^{-\frac{X^T \mathbf{B}(p) X}{2\sigma_{\mathbf{B}}}} \quad \text{with } X = q - p \quad (4)$$

une fonction de poids Gaussienne anisotropique. Table 1 résume l'effet local du mélange anisotrope de patches. Ces propriétés

Zone d'image	Force du mélange	Direction
homogène	fort	toutes
texture	très faible	toutes
structure	faible	structure

TABLE 1 – Configurations locales des tenseurs de mélange en fonction de la géométrie locale de l'image.

forcent le mélange à respecter les structures et les textures tout en réduisant les artefacts. Dans la Fig. 2 le mélange isotrope (milieu) endommage les structures. En revanche, le mélange anisotrope (droite) réduit les artefacts dans les zones texturées et laisse les structure intactes entre les pierres.

3 Extension au Mélange Spatio-temporel de Patches pour la Vidéo

Notre seconde contribution est l'adaptation du modèle décrit Section 2 pour les vidéos. Certains articles [18, 14] ont montré qu'utiliser la dimension temporelle comme une dimension spatiale fournit de bons résultats pour l'inpainting par exemple, en utilisant des patches tridimensionnels par exemple. Les tenseurs de mélange sont donc adaptables à la vidéo en ajoutant une composante temporelle comme suit :

$$\mathbf{B} = \lambda_{\sigma_{\mathbf{B}1}} \mathbf{e}_{\mathbf{S}1}^{\perp} \mathbf{e}_{\mathbf{S}1}^{\perp T} + \lambda_{\sigma_{\mathbf{B}2}} \mathbf{e}_{\mathbf{S}2}^{\perp} \mathbf{e}_{\mathbf{S}2}^{\perp T} + \lambda_{\sigma_t} \mathbf{e}_t \mathbf{e}_t^T \quad (5)$$

où λ_{σ_t} et \mathbf{e}_t sont la largeur mélange et les vecteurs propres temporels des tenseurs de mélange. Comme les propriétés spatiales des tenseurs de mélange pour la vidéo sont les mêmes que pour les images, nous discutons exclusivement de l'aspect temporel par la suite. Nous nous proposons de définir le comportement des tenseurs de mélange en énumérant les différentes configurations qui apparaissent avec l'ajout de la composante temporelle. Dans le cas de mouvements d'amplitude *très faibles* dans la vidéo, le gradient temporel d'une image I est $\frac{\partial I}{\partial t} \rightarrow 0$. La meilleure configuration dans ce cas est une valeur de λ_{σ_t} très élevée : la dimension temporelle est mélangée très fortement pour enlever les petites imperfections. Pour des mouvements d'amplitude *modérés*, le mélange doit être appliqué avec précautions de manière à ne pas laisser apparaître des rémanences d'objets dans les images où il ne sont pas supposés se trouver à un temps t . La meilleure manière d'appliquer le mélange dans ce cas est de prendre \mathbf{e}_t orienté selon l'axe du mouvement. Enfin, pour des mouvements d'amplitude *élevée*, $\frac{\partial I}{\partial t} \rightarrow \infty$, le meilleur choix est d'utiliser $\lambda_{\sigma_t} \rightarrow 0$. De cette manière, les tenseurs de mélange sont orientés orthogonalement à l'axe temporel. Comme elle n'endommage que très peu la dimension temporelle, c'est la configuration à choisir par défaut quand le mouvement n'est pas connu. Pour construire les tenseurs de mélange spatio-temporel, nous estimons la géométrie spatio-temporelle de la vidéo en utilisant un gradient temporel : $\nabla_t I = \frac{\partial I}{\partial t}$. D'autres outils peuvent être utilisés pour estimer cette géométrie, tel que le flot optique par exemple.

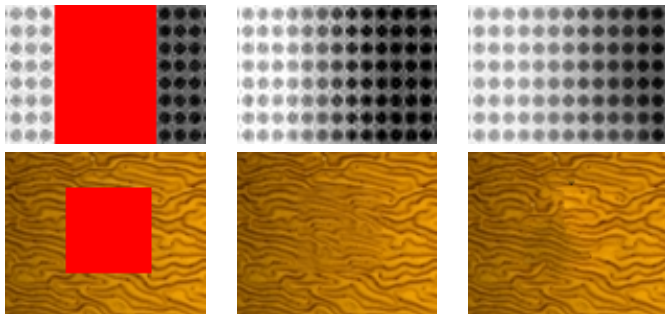


FIGURE 4 – Comparaisons (zoomées) entre la méthode de l'état de l'art [1] (milieu) et la méthode proposée (droite).

4 Algorithme

Algorithme 1 fournit un pseudo-code de la méthode proposée. \mathbf{S}_i est calculé en utilisant des schémas de différences finies selon les données connues localement. Aux lignes 4 – 6, le mélange isotrope des tenseurs de structure réduit les discontinuités dans le modèle géométriques. Aussi, le champs de tenseurs \mathbf{B} est régularisé par sa convolution par un noyau Gaussien, rendant le modèle plus robuste aux variations locales.

Entrées : V_i : vidéo, ϕ : carte de correspondance d'inpainting,
 Ω : masque, Ψ : ensemble des sources des patches reconstruits

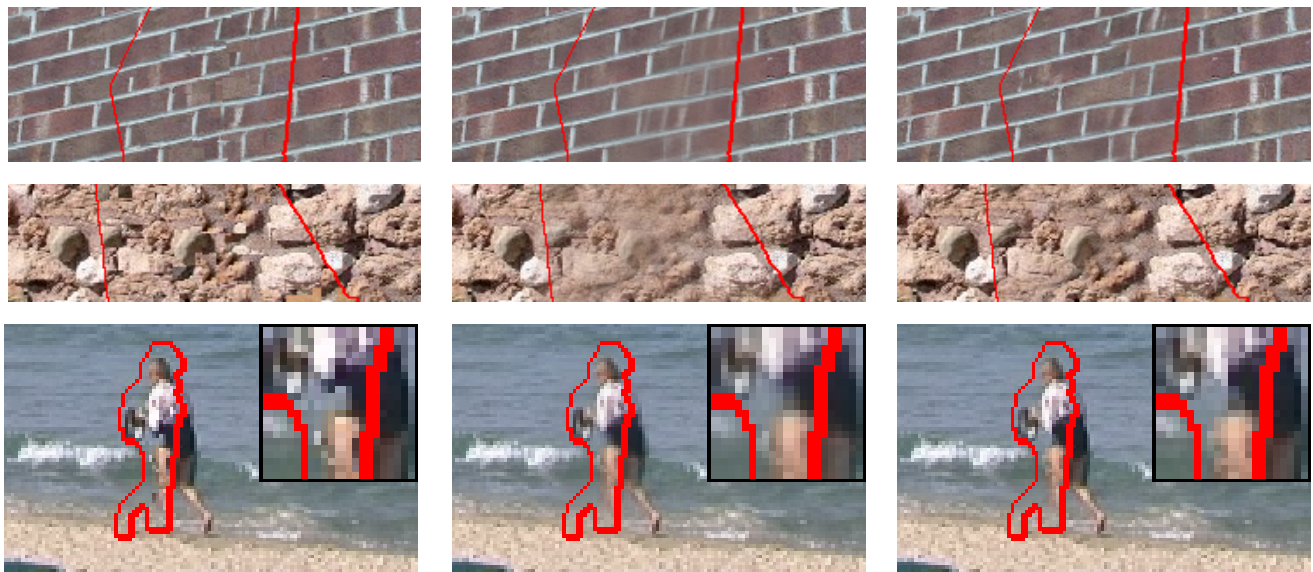
Sorties : V_o : vidéo avec artefacts réduits $\leftarrow V_i$

- 1 A : accumulateur de poids $\leftarrow \mathbb{1}_{\bar{\Omega}}$;
- 2 \mathbf{S}_i : tenseurs de structure de V_i , $\forall p \in \bar{\Omega}$;
- 3 \mathbf{S} : reconstruction des tenseurs de structure de \mathbf{S}_i avec Ω et ϕ ;
- 4 appliquer le mélange isotrope [8] sur \mathbf{S}
- 5 \mathbf{B} : calcul des tenseurs de mélange à partir de \mathbf{S} ;
- 6 régulariser \mathbf{B}
- 7 **Pour chaque** patch source de reconstruction $\psi_{\bar{p}} \in \Psi$ **faire**
- 8 G : poids Gaussiens anisotropes utilisant $\mathbf{B}(p)$;
- 9 dessiner G en p dans A de manière *additive*;
- 10 ψ_G : multiplication point-à-point de $\psi_{\bar{p}}$ par G ;
- 11 dessiner ψ_G en p dans V_o de manière *additive*;
- 12 **fin**
- 13 division point-à-point de V_o par A ;

Algorithme 1 : Pseudo-code haut niveau du mélange spatio-temporel de patches pour la vidéo.

5 Résultats et Discussions

Fig. 4, compare notre méthode avec la méthode de l'état de l'art [1]. Alors que la première ligne montre que l'effet de notre méthode (droite) est similaire à celui de [1] (milieu), la seconde montre que notre résultat est moins flou. Fig. 5 compare la méthode proposée avec [8]. De manière générale, dans les résultats avec [8], un effet de superposition se produit, qui n'apparaît pas dans les notes. Aussi, les structures de brique et de roche (première et deuxième ligne) sont mieux préservées avec notre méthode qu'avec [8] qui présente des textures *fantômes*. Nos expériences utilisent une version vidéo de l'algorithme [7] utilisant des patches et des fenêtres de recherche spatio-temporels. Cet algorithme permet un recherche de patches efficace. Dans le résultat vidéo Fig. 5 (dernière ligne), nous pouvons remarquer que le mélange anisotrope (droite) endommage beaucoup moins la vidéo que la méthode isotrope (milieu). Les résultats que nous fournissons montrent clairement que notre méthode améliore les résultats de mélange de patches en réduisant le flou des textures présent dans [8]. Adaptable pour différentes méthodes d'inpainting par patch, notre méthode est facilement reproductible grâce au pseudo code fourni. Dans des prochains travaux, nous nous intéresserons à l'utilisation d'autres outils pour estimer plus précisément le mouvement et améliorer notre modèle de mélange pour les vidéos.



(a) Sans mélange de patch.

(b) Mélange isotrope [8].

(c) Mélange anisotrope proposé.

FIGURE 5 – Comparaison des résultats du mélange de patches proposé dans cet article avec celui de [8]. Les lignes rouges correspondent aux frontières du masque utilisé par le processus d’inpainting.

Références

- [1] P. Arias, G. Facciolo, V. Caselles, and G. Sapiro. A variational framework for exemplar-based image inpainting. *International Journal of Computer Vision*, 93(3) :319–347, jul 2011. 1, 3
- [2] M. Ashikhmin. Synthesizing natural textures. In *Proceedings of the 2001 symposium on Interactive 3D graphics*, pages 217–226. ACM, 2001. 1
- [3] Raphaël Bornard, Emmanuelle Lecan, Louis Laborelli, and Jean-Hugues Chenot. Missing data correction in still images and image sequences. In *Proceedings of the tenth ACM international conference on Multimedia*, pages 355–361, 2002. 1
- [4] F. Cao, Y. Gousseau, S. Masnou, and P. Pérez. Geometrically guided exemplar-based inpainting. *SIAM Journal on Imaging Sciences*, 4(4) :1143–1179, 2011. 1
- [5] A. Criminisi, P. Pérez, and K. Toyama. Object removal by exemplar-based inpainting. In *Computer Vision and Pattern Recognition*, volume 2, pages II–721. IEEE, 2003. 1
- [6] A. Criminisi, P. Pérez, and K. Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on Image Processing*, 13(9) :1200–1212, sep 2004. 1
- [7] M. Daisy, P. Buysens, D. Tschumperlé, and O. Lézoray. A smarter exemplar-based inpainting algorithm using local and global heuristics for more geometry coherence. In *International Conference on Image Processing*, pages 4622–4626, Paris, France, 2014. 2, 3
- [8] M. Daisy, D. Tschumperlé, and O. Lézoray. A fast spatial patch blending algorithm for artefact reduction in pattern-based image inpainting. In *SIGGRAPH Asia 2013 Technical Briefs*, 2013. 2, 3, 4
- [9] A.A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *International Conference on Computer Vision*, volume 2, pages 1033–1038. IEEE, 1999. 1
- [10] C. Guillemot and O. Le Meur. Image inpainting : Overview and recent advances. *IEEE Signal Processing Magazine*, 31(1) :127–144, 2014. 1
- [11] Kaiming He and Jian Sun. Statistics of patch offsets for image completion. In *Computer Vision–ECCV 2012*, pages 16–29. Springer, 2012. 1
- [12] J. Jia and C-K. Tang. Inference of segmented color and texture description by tensor voting. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 26(6) :771–786, 2004. 1
- [13] O. Le Meur, M. Ebdelli, and C. Guillemot. Hierarchical super-resolution-based inpainting. *IEEE Transactions on Image Processing*, 22(10) :3779–3790, 2013. 1
- [14] A. Newson, A. Almansa, M. Fradet, Y. Gousseau, P. Pérez, et al. Video inpainting of complex scenes. In *SIAM Journal on Imaging Sciences*, volume 7 :4, pages 1993–2019, 2014. 3
- [15] K. A Patwardhan, G. Sapiro, and M. Bertalmio. Video inpainting of occluding and occluded objects. In *IEEE International Conference on Image Processing*, volume 2, pages II–69. IEEE, 2005. 1
- [16] P. Pérez, M. Gangnet, and A. Blake. Poisson image editing. *ACM Transactions on Graphics*, 22(3) :313–318, jul 2003. 1
- [17] D. Tschumperlé and R. Deriche. Vector-valued image regularization with pdes : A common framework for different applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4) :506–517, 2005. 2
- [18] Y. Wexler, E. Shechtman, and M. Irani. Space-time completion of video. *IEEE Transaction Pattern Analysis and Machine Intelligence*, 29(3) :463–476, mar 2007. 1, 3
- [19] Zongben Xu and Jian Sun. Image inpainting by patch propagation using patch sparsity. *Image Processing, IEEE Transactions on*, 19(5) :1153–1165, 2010. 1