

Un Algorithme d’Inpainting par Motif Utilisant des Heuristiques Locales et Globales pour une Meilleure Cohérence Géométrique

Maxime DAISY, Pierre BUYSENS, 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 – Dans cet article nous proposons deux améliorations du célèbre inpainting par patch originalement formulé par Criminisi et al. [5]. Tout d’abord nous introduisons un terme de donnée basé sur les tenseurs de structures pour améliorer l’ordre de reconstruction. Ensuite, nous proposons une heuristique améliorant la recherche d’un bon patch candidat pour la reconstruction. Ces deux contributions permettent à l’algorithme d’inpainting de produire des images géométriquement plus cohérentes, et d’être plus rapide. Nous illustrons la performance de ces contributions en comparant notre méthode à celles de l’état de l’art.

Abstract – In this paper, we propose two major improvements to the exemplar-based image inpainting algorithm, initially formulated by Criminisi et al. [5]. First, we introduce a structure-tensor-based data-term for a better selection of pixel candidates to fill in based on priority. Then, we propose a new lookup heuristic in order to locate the best source patches to copy/paste to these targeted points. These two contributions clearly make the inpainting algorithm reconstruct geometrically more coherent images, as well as speed up the process drastically. We illustrate the great performances of our approach compared to existing state-of-the-art methods.

1 Introduction

L’inpainting vise à compléter des zones manquantes d’une image d’une façon que le résultat soit le plus plausible possible. Ce type d’algorithme suscite un large engouement dans la communauté du traitement d’image grâce aux applications possibles comme la suppression d’objets ou l’interpolation. La littérature de l’inpainting revue dans [9] montre deux catégories principales : les méthodes basées géométrie et basées motif.

- Les méthodes basées *géométrie* [14, 4, 18] complètent les images en utilisant des interpolations géométriques *semi-locales*. Ces méthodes produisent des résultats intéressants concernant la géométrie mais ne reproduisent souvent aucune texture.
- Les méthodes orientées *motif* [5, 15, 16, 19, 10] sont la suite de travaux sur la synthèse de textures [8]. Elles utilisent l’auto-similarité, et leur but est de trouver des bouts de patchs collant le mieux aux données connues au bord du masque. Ces méthodes fonctionnent très bien pour de larges zones à remplir car elles produisent des textures cohérentes avec les données connues de l’image. En revanche, ces méthodes n’assurent pas une géométrie globale assez cohérente, conduisant à des reconstructions peu réalistes et présentant des effets de bloc. Des méthodes hybrides ont été proposées dans la littérature [11, 1] pour essayer de combiner les méthodes basées géométrie et basées motif, mais des artéfacts restent. Dans cet article nous nous focalisons sur les méthodes basées motifs initiées par [5], qui sont des références dans le domaine de l’inpainting.

Cet article (comme [7]) propose deux ajustements importants de la méthode originale et montre qu’ils améliorent clairement la cohérence géométrique des résultats. Nous reformulons d’abord le terme de donnée afin de mieux sélectionner le patch à reconstruire en priorité. Ensuite, nous introduisons une heuristique permettant une recherche de motif efficace tant en qualité qu’en temps de calcul. Ces deux contributions améliorent les résultats d’inpainting de [5], ce que nous illustrons avec différents cas de reconstructions difficiles sur des images réelles.

2 Contexte et Contributions

Inpainting basé motif : En 2004, Criminisi et al. ont proposé un algorithme d’inpainting glouton [5] qui reconstruit la région non connue Ω d’une image \mathcal{I} morceau par morceau en utilisant un principe de priorisation. L’ordre de reconstruction a été défini par une fonction de priorité P_p estimée pour $p \in \delta\Omega$ (où $\delta\Omega$ est le bord de Ω). Les priorités $P_p = C_p \times D_p$ sont définies dans $[0, 1]$. Alors que le terme de *confiance* C_p ne dépend uniquement que de la topologie locale de Ω , le terme de *donnée* D_p dépend des structures locales dans l’image.

$$D_p = \frac{|\nabla \vec{I}_p^\perp \cdot \vec{n}_p|}{\alpha} \text{ avec } \nabla \vec{I}_p^\perp = \{ \nabla \vec{I}_q^\perp \mid \arg \max_{q \in ((\mathcal{I} - \Omega) \cap \Psi_p)} \|\nabla \vec{I}_q\| \} \quad (1)$$

où Ψ_p est un patch de taille fixe centré en p , \vec{n}_p est la normale au masque en p et α un facteur de normalisation constant.

Le terme D_p favorise la reconstruction des structures linéaires locales orthogonales à Ω en p . Quand les priorités P_p sont calculées, le pixel p de plus haute priorité est choisi comme prochaine zone de reconstruction. Pour ce point p , le patch $\psi_{\hat{p}} = \{\psi_q \mid \arg \min_{q \mid \psi_q \in (\mathcal{I} - \Omega)} \|\Psi_p - \psi_q\|^2\}$ est recherché dans toute l'image (généralement dans une fenêtre de recherche carrée de taille fixe centrée à p) et est ensuite dessiné en $\Psi_p \cap \Omega$. Les confiances sont ensuite mises à jour avec $\forall q \in \Psi_p \cap \Omega, C_q = C_p$, ainsi que $\delta\Omega$. Le processus se répète jusqu'à $\delta\Omega = \emptyset$.

Cet algorithme est devenu une référence pour la qualité globale de ses reconstructions ainsi que son temps raisonnable de calcul. Le terme de priorité P_p est assez difficile à améliorer (des approches dans cette direction ont été proposées dans [9, 13]). L'algorithme possède quelques désavantages que nous nous proposons de réduire. Il faut savoir que comme le processus d'inpainting est glouton, le résultat à une étape dépend fortement du résultat de l'étape précédente. De fait, en améliorant plusieurs aspects de l'algorithme original permet une augmentation certaine de la qualité des résultats.

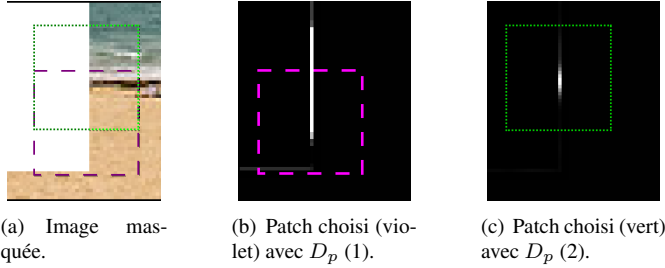


FIGURE 1 – Illustration de l'impact des termes de donnée (1) (original) et (2) (proposé) pour sélectionner le patch de plus haute priorité sur un cas local (gauche).

Des priorités plus robustes : Le terme de donnée D_p indique si des structures linéaires sont présentes à un point p et avec quel angle elles croisent Ω . Malheureusement, le gradient en $p \in \delta\Omega$ ne peut pas être estimé correctement comme des données manquent dans le voisinage de p . Dans [5] il est suggéré d'approximer $\vec{\nabla} \hat{I}_p$ comme le gradient de valeur maximum dans le voisinage connu $\Psi_p \cap \mathcal{I}$ de p (Eq. (1)). Cette approximation ne permet pas d'obtenir une analyse robuste de la géométrie locale : quand la taille de patch est grande, D_p ne sera élevé que là où un contour entre dans Ω , mais aussi où les candidats sont à une distance inférieure à N de ce contour. Cet effet de dilatation donne trop d'importance aux points du voisinage de p , ce qui leur permet d'être sélectionné en priorité au lieu du point de contour, en particulier si la confiance y est meilleure. La Fig.1(a)-(b) illustre ce phénomène : ici, D_p est élevé pour tous les points cibles autour de la frontière plage/mer, si bien que le patch choisi pour la reconstruction (en violet) ne sera pas centré sur le contour. Un patch contenant majoritairement du sable sera probablement choisi pour la reconstruction et brisera les structures locales. Afin de résoudre ce problème, nous proposons une approche plus robuste utilisant des tenseurs de

structure [20] pour modéliser les variations locales de l'image à l'intérieur d'un patch Ψ_p à reconstruire. Tout d'abord, les tenseurs ont l'intérêt de mieux définir les structures locales d'une image multi-canal. Ensuite, nous utilisons les propriétés algébriques des tenseurs pour permettre à des patches contenant des structures avec de multiples orientations (les textures typiquement) d'être choisis quelque soit la normale \vec{n}_p à $\delta\Omega$. Nous proposons de définir \tilde{D}_p comme :

$$\tilde{D}_p = \|\mathbf{G}_p \vec{n}_p\| \quad \text{avec} \quad \mathbf{G}_p = \sum_{q \in \Psi_p \cap (\mathcal{I} - \Omega)} w_q \vec{\nabla} \hat{I}_q \vec{\nabla} \hat{I}_q^T \quad (2)$$

\mathbf{G} est une moyenne pondérée des tenseurs de structure estimés dans la partie connue $(\mathcal{I} - \Omega)$ du patch Ψ_p à reconstruire. w_p est une fonction gaussienne 2d centrée à p d'écart type donné.

- Lorsque \mathbf{G} est anisotrope ($\mathbf{G} \approx \vec{u} \vec{u}^T$), un contour principal orienté selon \vec{u}^\perp existe dans Ψ_p . (2) tend vers $|\langle \vec{u}, \vec{n}_p \rangle|$ et sera élevé lorsque le contour croisera Ω orthogonalement. Il est à noter que D_p sera plus élevé pour des points p localisés précisément sur un contour (grâce aux poids plus élevés au centre de Ψ_p). L'effet de dilatation n'apparaît plus et le point localisé sur le contour aura le plus haut terme de donnée (Fig.1c).

- Quand \mathbf{G} est isotrope (c.-à-d. $\mathbf{G} \approx \lambda \mathbf{Id}$), (2) est quasiment égal à λ , et très bas dans les zones homogènes. En revanche, il reste élevé lorsque Ψ_p contient de nombreuses variations avec des directions différentes (λ élevé) quelque soit l'orientation de \vec{n}_p . Au contraire, avec la formulation originale, (1) peut être bas si le gradient de valeur maximum et la normale au masque \vec{n}_p sont orthogonaux.

En utilisant (2), nous améliorons la recherche du point de plus haute priorité pour la reconstruction. Cela a une grande incidence sur la qualité de l'image reconstruite comme la nature gloutonne du processus propage les erreurs de reconstruction (Fig.3(b)-(c) montre un exemple de différence).

Une recherche de patches améliorée : Notre seconde contribution est l'amélioration de la technique de recherche de patch pour l'algorithme [5]. Nous nous inspirons principalement d'idées provenant de *Patchmatch* [2] et de ses extensions récentes [3, 10]. Ces algorithmes proposent de calculer efficacement une carte de voisins approchant (ANN) qui permet pour chaque point p d'une image, de retrouver rapidement un patch très similaire à celui centré en p . Ce concept a été utilisé avec des approches d'inpainting pixel par pixel utilisant la multi-résolution. En revanche, lorsqu'il s'agit de synthèse patch par patch, cela devient moins efficace du fait qu'il faut une étape de mise à jour à chaque morceau de patch recollé.

L'algorithme de recherche proposé se base de même sur l'utilisation de décalages afin de réduire l'espace de recherche. Soient p_N le centre du patch Ψ_{p_N} à reconstruire, $\mathcal{A} = \{p_0, \dots, p_{N-1}\}$ l'ensemble des centres des patches $\{\Psi_{p_0}, \dots, \Psi_{p_{N-1}}\}$ déjà reconstruits aux itérations précédentes, $NN(p_i) = \hat{p}_i$ le centre du meilleur patch $\psi_{\hat{p}_i}$ trouvé pour Ψ_{p_i} , et $NNF = \{\hat{p}_0, \dots, \hat{p}_{N-1}\}$ l'ensemble des centres des meilleurs patches.

Nous définissons tout d'abord le vecteur de décalage entre p_n et les points de \mathcal{A} par $\Delta_{p_N}^{p_i} = p_N - p_i, \forall i \in \{0, \dots, N-1\}$. Soit S l'ensemble des points p_i de \mathcal{A} pour lesquels $\|\Delta_{p_N}^{p_i}\| < T$.

La méthode de recherche proposée est d'étendre l'espace de recherche en considérant plusieurs *sous-fenêtres de recherche* centrées en chaque point $NN(q) - \Delta_{p_N}^q, \forall q \in S$. Ce procédé permet de rechercher de bons patches candidats dans des fenêtres de recherche potentiellement plus intéressantes en relation avec les patches déjà collés. Il est à noter que la fenêtre centrée en p_N est aussi utilisée. Afin de réduire le nombre total de candidats, nous utilisons une taille $sz_N = sz/\sqrt{\text{Card}(S)}$ pour les sous-fenêtres, où sz paramètre la taille de la fenêtre de recherche initiale. Fixer la taille de cette manière permet de garder une complexité similaire pour chaque étape. Une vue

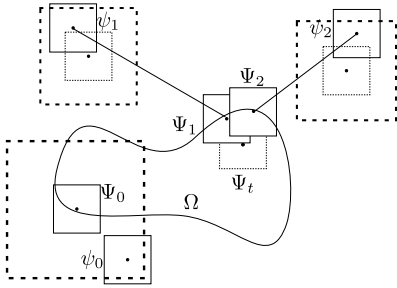


FIGURE 2 – Vue résumée de l'algorithme de recherche proposé.

schématique du processus est illustré Fig. 2, où le patch cible Ψ_{p_3} est le carré pointillé à l'intérieur de Ω , et les deux fenêtres de recherche (carrés en tirets) centrées sur les candidats initiaux obtenus par les décalages de $\Delta_{p_1}^{p_3}$ et $\Delta_{p_2}^{p_3}$ à partir de \hat{p}_1 et \hat{p}_2 . Pour la clarté, la fenêtre de p_3 n'est pas dessinée. Il est à noter que la fenêtre pour Ψ_0 est plus grande car elle n'a pas de voisin déjà reconstruit (c.-à-d. $S = \emptyset$). Les avantages du procédé proposé sont :

- Comme les sous-fenêtres se chevauchent en pratique, le nombre de candidats est moins grand que dans la recherche de base [5], ce qui permet d'accélérer le processus (voir Fig.3(a)-(d)).
- La taille de la zone de recherche étant réduit, le meilleur candidat approchant n'est pas forcément optimal (au sens de la SSD), mais en pratique, cela donne visuellement de meilleurs résultats. Les fenêtres de recherche sont centrés aux endroits géométriquement intéressants par rapport aux patches déjà reconstruits. Par conséquent, notre processus d'inpainting modifié devient moins sensible à la taille de patch utilisée, permettant de recouvrir de large portions de structure avec des patches de taille inférieure qu'avec le processus original [5].

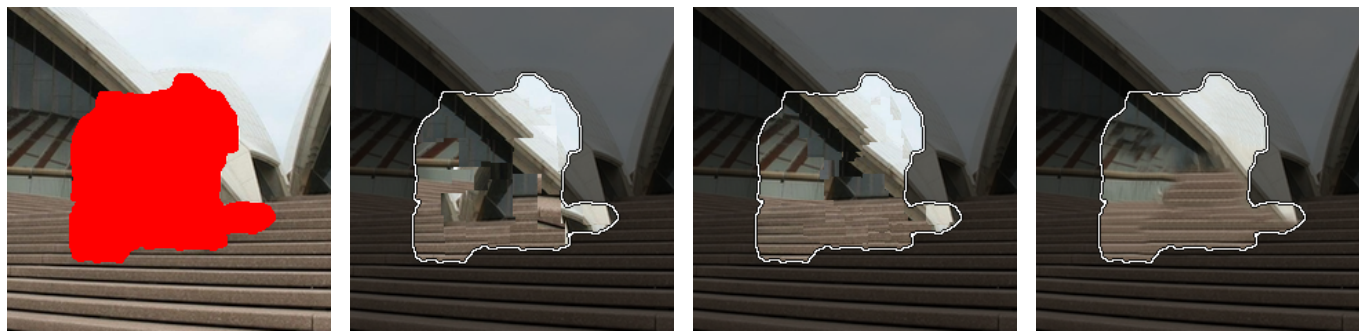
3 Résultats et Conclusion

Les résultats de la Fig. 3(a)-(d) montrent les améliorations de l'algorithme original [5] apportées par nos contributions. La grande taille du masque ainsi que les structures complexes présentes dans l'image en font un cas difficile. Il apparaît clairement que nos améliorations permettent d'obtenir des résultats plus en accord avec le contenu initial de l'image. La Fig. 3 montre que notre algorithme rivalise clairement avec ceux de l'état de l'art [12] et l'algorithme de Photoshop combinant [19] et [3]. Pour chaque résultat, une étape de *fondus spatial de patches*

(un de nos travaux publié dans [6]) a été utilisée afin d'améliorer la cohérence visuelle. Pour faciliter la reproductibilité de nos travaux, nous proposons des versions code source C++ et binaires de nos travaux dans le framework *G'MIC* [17].

Références

- [1] P. ARIAS et al. "A Variational Framework for Exemplar-Based Image Inpainting". *IJCV* 93.3 (2011), p. 319–347.
- [2] C. BARNES et al. "PatchMatch : A Randomized Correspondence Algorithm for Structural Image Editing". *ACM Transactions on Graphics (Proc. SIGGRAPH)* 28.3 (2009).
- [3] C. BARNES et al. "The Generalized PatchMatch Correspondence Algorithm". *Proceedings of the European Conference on Computer Vision*. T. LNCS 6313, Part III. 2010, p. 29–43.
- [4] M. BERTALMIO et al. "Image inpainting". *SIGGRAPH*. New York, USA, 2000, p. 417–424.
- [5] A. CRIMINISI, P. PEREZ et K. TOYAMA. "Region filling and object removal by exemplar-based image inpainting". *IEEE Trans. IP* 13.9 (2004), p. 1200–1212.
- [6] M. DAISY, D. TSCHUMPERLÉ et O. LÉZORAY. "A Fast Spatial Patch Blending Algorithm for Artefact Reduction in Pattern-based Image Inpainting". *SIGGRAPH Asia Technical Briefs*. Hong Kong, Hong Kong : ACM, 2013, 8 :1–8 :4.
- [7] Maxime DAISY et al. "A smarter exemplar-based inpainting algorithm using local and global heuristics for more geometric coherence". *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, p. 4622–4626.
- [8] A.A. EFROS et T.K. LEUNG. "Texture synthesis by non-parametric sampling". *ICCV*. T. 2. 1999, p. 1033–1038.
- [9] Christine GUILLEMOT et Olivier LE MEUR. "Image inpainting : Overview and recent advances". *IEEE SPM* 31.1 (2014), p. 127–144.
- [10] K. HE et J. SUN. "Statistics of Patch Offsets for Image Completion". *Proceedings of the 12th European Conference on Computer Vision*. T. LNCS 7573, Part II. Florence, Italy : Springer-Verlag, 2012, p. 16–29.
- [11] N. KAWAI, T. SATO et N. YOKOYA. "Image Inpainting Considering Brightness Change and Spatial Locality of Textures and Its Evaluation". *Proc. of the 3rd PSIVT*. Tokyo, Japan, 2009, p. 271–282.
- [12] O. LE MEUR, J. GAUTIER et C. GUILLEMOT. "Exemplar-based inpainting based on local geometry". *ICIP*. Brussel, Belgium, 2011, p. 3401–3404.
- [13] R. MARTINEZ-NORIEGA, A. ROUMY et G. BLANCHARD. "Exemplar-based image inpainting : Fast priority and coherent nearest neighbor search". *Machine Learning for Signal Processing (MLSP), 2012 IEEE International Workshop on*. 2012, p. 1–6.
- [14] S. MASNOU et J.-M. MOREL. "Level Lines based Disocclusion". *ICIP* (3). 1998, p. 259–263.
- [15] P. PÉREZ, M. GANGNET et A. BLAKE. *PatchWorks : Example-Based Region Tiling for Image Editing*. Rap. tech. Microsoft Research, MSR-TR-2004-04, 2004.
- [16] J. SUN et al. "Image completion with structure propagation". *ACM Trans. Graph.* 24.3 (2005), p. 861–868.
- [17] D. TSCHUMPERLÉ. "G'MIC, open-source framework for image processing". <http://gmic.sourceforge.net> (2014).
- [18] D. TSCHUMPERLÉ et R. DERICHE. "Vector-Valued Image Regularization with PDEs : A Common Framework for Different Applications". *IEEE Trans. PAMI* 27.4 (2005), p. 506–517.
- [19] Y. WEXLER, E. SHECHTMAN et M. IRANI. "Space-Time Completion of Video". *IEEE Trans. PAMI* 29.3 (2007), p. 463–476.
- [20] S.Di ZENZO. "A Note on the Gradient of a Multi-Image". *Computer Vision, Graphics, and Image Processing* 66 (1986), p. 116–125.



(a) Image masquée. (b) Résultat utilisant Criminisi et al. [5] (1m48s). (c) Résultat avec [5] et notre terme de donnée Eq.(2) (1m36s). (d) Résultat avec [5] et toutes les améliorations proposées (0m01s).

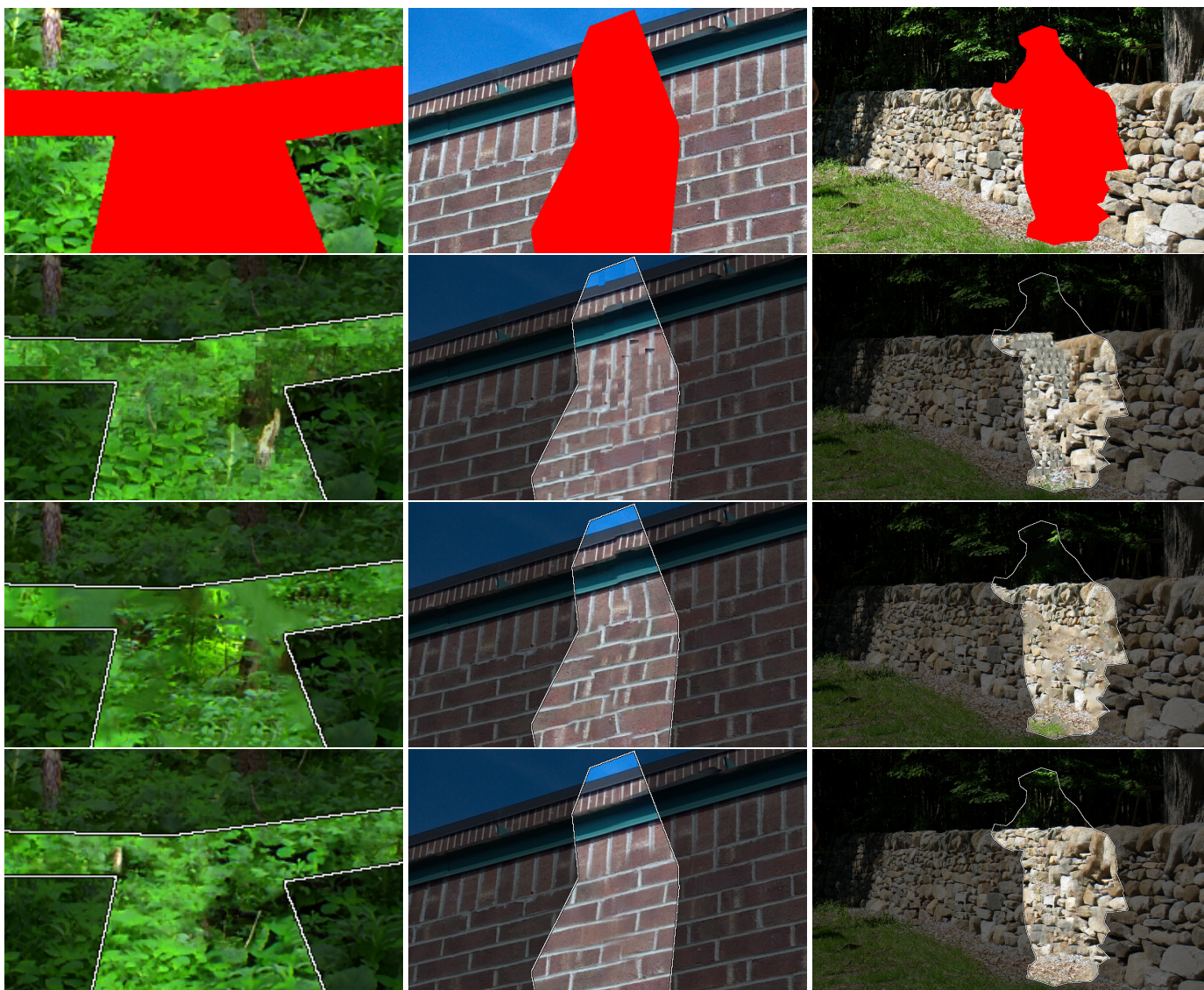


FIGURE 3 – Première ligne : illustration des contributions successives présentées dans cet article. Ensuite de haut en bas, comparaisons avec des méthodes de l'état de l'art : images masquées, résultats avec Lemeur et al. [12], résultats avec Photoshop utilisant Wexler et al. [19] + Patchmatch [3], et enfin les résultats avec notre méthode. Nous avons essayé de trouver les meilleurs paramètres pour chacun de ces résultats.