



HAL
open science

SuperPatchMatch : Un algorithme de correspondances robustes de patches de superpixels

Rémi Giraud, Vinh-Thong Ta, Aurélie Bugeau, Pierrick Coupé, Nicolas Papadakis

► **To cite this version:**

Rémi Giraud, Vinh-Thong Ta, Aurélie Bugeau, Pierrick Coupé, Nicolas Papadakis. SuperPatchMatch : Un algorithme de correspondances robustes de patches de superpixels. Congrès Reconnaissance des Formes, Image, Apprentissage et Perception (RFIAP'18), May 2018, Marne-la-Vallée, France. hal-01702277

HAL Id: hal-01702277

<https://hal.science/hal-01702277>

Submitted on 26 May 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Public Domain

SuperPatchMatch : Un algorithme de correspondances robustes de patchs de superpixels

Rémi Giraud^{1,2} Vinh-Thong Ta^{1,2} Aurélie Bugeau¹ Pierrick Coupé¹ Nicolas Papadakis³

¹ Univ. Bordeaux, LaBRI, CNRS, UMR 5800, F-33400 Talence, France

² Bordeaux INP, LaBRI, UMR 5800, F-33405 Talence, France

³ Univ. Bordeaux, IMB, CNRS, UMR 5251, F-33400 Talence, France

remi.giraud@u-bordeaux.fr

Résumé

Les superpixels sont devenus très populaires dans de nombreuses applications de vision par ordinateur. Néanmoins, ils restent sous-exploités du fait de l'irrégularité des décompositions qui diffèrent selon les images. Dans ce travail, nous introduisons d'abord une nouvelle structure, un patch de superpixels, appelée SuperPatch. La structure proposée, basée sur le voisinage du superpixel, définit un descripteur robuste incluant les relations spatiales entre superpixels voisins. La généralisation de la méthode de recherche de correspondance PatchMatch aux SuperPatches, nommée SuperPatchMatch, est alors introduite. Enfin, nous proposons une adaptation de la méthode à l'étiquetage automatique depuis une bibliothèque d'images d'exemples. Nous démontrons alors le potentiel de notre approche en obtenant des résultats supérieurs à ceux d'approches basées apprentissages, sur des expériences d'étiquetage de visages.

Mots Clef

Méthode basée patchs, Superpixels, Patchs de superpixels

Abstract

Superpixels have become very popular in many computer vision applications. Nevertheless, they remain underexploited because of the irregularity of the decomposition that differ according to the images. In this work, we first introduce a novel structure, a superpixel-based patch, called SuperPatch. The proposed structure, based on superpixel neighborhood, leads to a robust descriptor including spatial relations between neighboring superpixels. The generalization of the search matching method PatchMatch to these SuperPatches, named SuperPatchMatch, is introduced. Finally, we propose a framework to perform automatic labeling from a library of example images. We demonstrate the potential of our approach by outperforming learning-based approaches on face labeling experiments.

Keywords

Patch-based method, Superpixels, Patches of superpixels

1 Introduction

Depuis leur popularisation pour la synthèse de texture [7] ou le débruitage [4], les méthodes basées patchs ont obtenu des résultats état de l'art sur de nombreuses applications de traitement et d'analyse d'images. Dans la plupart de ces méthodes, un algorithme de recherche de plus proches voisins approchés (ANN, pour Approximate Nearest Neighbor) est utilisé pour trouver des correspondances. Parmi ces méthodes, PatchMatch (PM) [2] a été conçu pour calculer rapidement des correspondances de patchs de pixels.

Lorsque la recherche est effectuée avec PM pour de grandes images, ou dans une bibliothèque d'images, la recherche d'ANN peut devenir coûteuse, et nécessiter de nombreuses itérations du processus afin de converger. Un PM multirésolution peut alors être considéré pour initialiser grossièrement la carte de correspondance. Cependant, comme on l'observe généralement avec de telles approches hiérarchiques régulières ne respectant pas les contours des objets, les détails sont perdus et des ANN peu précis sont estimés pour les petites structures.

Dans ce contexte, l'utilisation des superpixels [1] peut être intéressante. Les superpixels groupent en effet les pixels de l'image en zones homogènes tout en respectant les contours des objets (voir Figure 1). Des modèles locaux de correspondance de superpixels ont été proposés, par exemple pour le suivi d'objets dans des vidéos [18]. Cependant, les algorithmes de recherche d'ANN basés superpixels ont peu été étudiés, ce qui limite l'utilisation de ces décompositions. Par exemple, dans [17] des correspondances de superpixels sont calculées entre deux images, en utilisant un modèle de transport optimal complexe au temps de calcul conséquent.

De plus, l'irrégularité des décompositions en superpixels empêche l'utilisation directe des outils basés patchs considérant le voisinage des pixels. La prise en compte de l'information contenue dans le voisinage autour d'un pixel est pourtant cruciale dans le processus de recherche de correspondances, et les descripteurs standards de vision par ordinateur intègrent cette information [14, 21].

Par conséquent, afin d'à la fois réduire le nombre d'éléments à traiter, de conserver l'information géométrique, et de trouver des correspondances précises, il apparaît nécessaire de considérer des descripteurs spatiaux intégrant le voisinage, afin de pouvoir définir des méthodes robustes de correspondances de superpixels. Des tentatives d'utilisation de l'information de voisinage des superpixels ont été proposées [16, 19]. Cependant, ces méthodes ne sont pas adaptées au contexte de la recherche d'ANN, puisqu'elles effectuent une régularisation sur un graphe construit depuis les superpixels voisins, mais n'incluent pas directement l'information spatiale relative entre superpixels dans une structure dédiée.

Contributions

Dans ce travail, nous proposons une nouvelle structure de voisinage de superpixels appelée *superpatch* [8]. Puisque les voisinages des superpixels sont irréguliers en termes de formes et de nombre d'éléments, une mesure pour comparer des superpatches est également introduite.

Une nouvelle méthode, appelée *SuperPatchMatch* (SPM), qui étend l'algorithme PM, est alors proposée pour effectuer des recherches rapides et précises d'ANN de superpatches dans une bibliothèque d'images. À notre connaissance, la combinaison de PM avec les superpixels a seulement été proposée dans deux travaux. Dans [10], des correspondances sont calculées entre simples superpixels par des mouvements similaires à ceux de PM. Dans [15], PM est considéré de façon plus réduite pour l'estimation de flot optique. L'algorithme est utilisé pour raffiner les résultats à l'intérieur du rectangle entourant le superpixel traité. Le but de notre travail est donc différent puisque nous souhaitons comparer des voisinages de superpixels définis sur des domaines irréguliers de l'image.

Pour démontrer quantitativement l'intérêt de notre méthode, nous proposons d'appliquer SPM à la segmentation et l'étiquetage depuis une bibliothèque d'images d'exemples. SPM est particulièrement adapté pour gérer de grandes bases d'images en constante évolution puisqu'aucune phase d'entraînement n'est nécessaire, contrairement aux approches basées sur des apprentissages.

2 Superpatch

2.1 Définition du superpatch

De façon similaire au patch de pixels, un *superpatch* est un ensemble de superpixels voisins spatialement. Considérons une image A , décomposée par une quelconque méthode de superpixels, en $|A|$ éléments. Un superpatch \mathbf{A}_i est centré sur un superpixel A_i et est composé de ses superpixels voisins $A_{i'}$. On pose ainsi $\mathbf{A}_i = \{A_{i'}, \text{ tel que } \|c_i - c_{i'}\|_2 \leq R\}$, avec c_i le barycentre spatial des pixels contenus dans A_i . En d'autres termes, le superpatch centré sur un superpixel est défini en considérant tous les superpixels dont le barycentre est inclus dans un rayon de recherche fixé $R \geq 0$. Notons que chaque superpatch \mathbf{A}_i contient au

moins le superpixel A_i . La Figure 1 illustre la définition du superpatch.

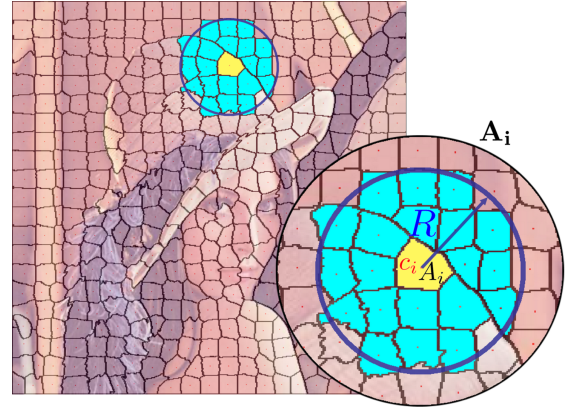


FIGURE 1 – Illustration de la définition du superpatch. En bleu : cercle de recherche de rayon R centré sur c_i , le barycentre de A_i . Le superpatch \mathbf{A}_i est composé par tous les superpixels ayant leur barycentre à l'intérieur du cercle.

2.2 Comparaison de superpatches

La comparaison entre deux patchs réguliers est généralement effectuée en calculant une somme des différences au carré (SSD). Dans le cadre usuel, les deux patchs réguliers ayant la même forme, cette distance est calculée entre les intensités des pixels correspondants spatialement dans chaque patch. Cependant, lorsque l'on considère deux superpatches, leur nombre d'éléments et leur géométrie sont généralement différents. Dans la suite, pour illustrer la méthode, nous considérons deux superpatches \mathbf{A}_i et \mathbf{B}_j , dans deux images potentiellement différentes A et B , décomposées en superpixels de tailles similaires.

Pour comparer deux superpatches, nous recalons tout d'abord les positions relatives de tous les superpixels appartenant aux superpatches. Pour superposer deux superpatches, toutes les positions $c_{j'}$ des superpixels $B_{j'} \in \mathbf{B}_j$ sont recalées par le vecteur $v_{ij} = c_i - c_j$, où c_i et c_j sont les barycentres spatiaux respectifs de A_i et B_j . Contrairement au schéma classique des patchs réguliers, un superpixel recalé $B_{j'}$ peut se superposer avec plusieurs superpixels $A_{i'}$, et cette information doit être considérée.

Pour calculer une distance entre des structures irrégulières, telles que les superpixels, les auteurs de [19] proposent d'utiliser la distance d'édition. Cependant, une telle distance calcule des correspondances uniques entre deux éléments de la structure et ne peut pas traiter efficacement la superposition de superpixels qui nécessite le calcul de correspondances entre un et plusieurs éléments. Une autre limite est le fait que cette distance mélange deux types d'information : la similarité de deux éléments et le coût d'enlever ou d'ajouter des superpixels. Ainsi, la distance d'édition doit être paramétrée avec précision selon l'application.

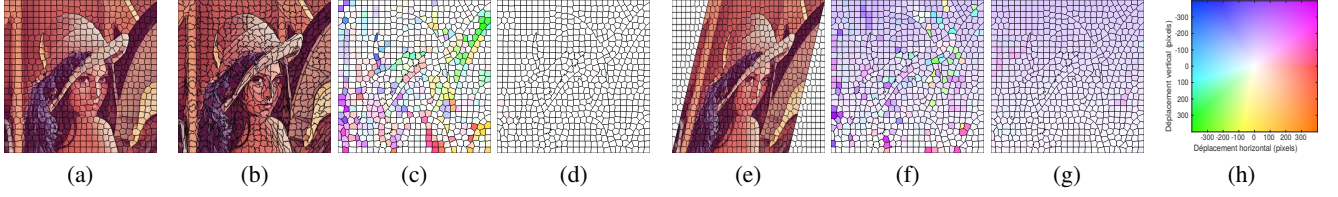


FIGURE 2 – Superpixels vs superpatches pour la correspondance de superpixels. (a) et (b) : deux décompositions par [9] et [5]. (c) et (d) : résultat de correspondances de superpixels [10] et notre résultat de correspondances de superpatch. La même expérience est effectuée entre (a) et l'image cisailée (e), pour laquelle nous présentons les résultats de correspondances de superpixels [10] (f) et de superpatches (g). Le déplacement est illustré avec la représentation standard du flot optique (h).

Pour définir une métrique pertinente entre des superpatches, il est donc nécessaire de considérer leur géométrie. Nous proposons de définir la distance symétrique D entre deux superpatches \mathbf{A}_i et \mathbf{B}_j comme :

$$D(\mathbf{A}_i, \mathbf{B}_j) = \frac{\sum_{A_{i'} \in \mathbf{A}_i} \sum_{B_{j'} \in \mathbf{B}_j} w(A_{i'}, B_{j'}) d(F_{A_{i'}}, F_{B_{j'}})}{\sum_{A_{i'} \in \mathbf{A}_i} \sum_{B_{j'} \in \mathbf{B}_j} w(A_{i'}, B_{j'})}, \quad (1)$$

où d est la distance Euclidienne entre les descripteurs des superpixels $F_{A_{i'}}$ et $F_{B_{j'}}$, par exemple, les coordonnées de c_i , la couleur moyenne, ou tout autre descripteur de superpixel, w est un poids qui dépend de la position relative des superpixels $A_{i'}$ et $B_{j'}$.

Distance rapide entre deux superpixels. Pour calculer le poids entre deux superpixels $A_{i'}$ et $B_{j'}$, nous voudrions idéalement mesurer l'aire de leur superposition relative, c'est-à-dire, calculer $w(A_{i'}, B_{j'}) = |A_{i'} \cap B_{j'}| / |A_{i'} \cup B_{j'}|$. Cependant, ce poids nécessite le calcul coûteux de tous les pixels qui se superposent, ce qui annulerait l'avantage de la représentation en superpixels. Une méthode rapide serait de comparer un superpixel $A_{i'}$ uniquement au superpixel le plus proche $B_{j'}$. Néanmoins, pour mieux tenir compte de la géométrie des deux superpatches, nous proposons une méthode plus robuste qui considère un poids symétrique spatial entre deux superpixels $A_{i'}$ et $B_{j'}$:

$$w(A_{i'}, B_{j'}) = \exp\left(-\frac{\|x_{i'j'}\|_2^2}{\sigma_1^2}\right) w_s(A_{i'}) w_s(B_{j'}), \quad (2)$$

où $x_{i'j'} = c_{j'} - c_{i'} + v_{ij}$ est la distance relative entre $c_{i'}$ et $c_{j'}$, tandis que $w_s(A_{i'})$ pondère l'influence de $A_{i'}$ par rapport à sa distance à A_i tel que $w_s(A_{i'}) = \exp(-\|c_i - c_{i'}\|_2^2 / \sigma_2^2)$, avec σ_1 et σ_2 des paramètres d'échelle.

Le réglage de σ_1 dépend de l'échelle de la décomposition. Puisque les superpatches ont été recalés, et que le but est de comparer un superpixel $A_{i'} \in \mathbf{A}_i$ aux plus proches $B_{j'} \in \mathbf{B}_j$, σ_1 peut être fixé à la moitié de la taille moyenne d'un superpixel, c'est-à-dire, la moitié de la distance moyenne entre les barycentres de deux superpixels, tel que $\sigma_1 = \frac{1}{2} \sqrt{|I|/K}$, pour une image $|I|$ décomposée en K superpixels. Enfin, σ_2 dépend de la taille du superpatch et peut être réglé à $\sqrt{2}R$ pour favoriser la contribution des plus proches superpixels.

2.3 Généralisation des patches réguliers

Dans le cas limite où chaque superpixel ne contient qu'un pixel, c'est-à-dire $A_{i'} = c_{i'}$ et $B_{j'} = c_{j'}$, alors \mathbf{A}_i et \mathbf{B}_j ont la même structure régulière et le même nombre d'éléments. Avec $\sigma_1 \rightarrow 0$ et $\sigma_2 \rightarrow \infty$ (2), $w(A_{i'}, B_{j'}) = w(c_{i'}, c_{j'}) = 1$ si $c_{i'} - c_i = c_{j'} - c_j$ et 0 sinon, et on constate que la distance proposée D (1) est une généralisation de la distance entre patches réguliers, puisqu'elle est réduite à une SSD normalisée standard entre deux patches de pixels :

$$D(\mathbf{A}_i, \mathbf{B}_j) = \frac{1}{|\mathbf{A}_i|} \sum_{A_{i'} \in \mathbf{A}_i} \sum_{B_{j'} \in \mathbf{B}_j} d(F_{A_{i'}}, F_{B_{j'}}) \delta_{x_{i'j'}=0}, \quad (3)$$

$$= \text{SSD}(\mathbf{A}_i, \mathbf{B}_j),$$

où $\delta_a = 1$ quand a est vrai et 0 sinon.

2.4 Robustesse du superpatch

La Figure 2 présente deux expériences illustrant la capacité des superpatches à trouver des correspondances de superpixels plus pertinentes que celles obtenues en utilisant uniquement les superpixels [10]. Deux décompositions sont calculées sur la même image en utilisant [9] et [5] (Figures 2(a) et (b)). Le but est de trouver la meilleure correspondance de superpixels entre (a) et (b) en termes de descripteurs de superpixels (ici norme ℓ_2 sur les histogrammes couleurs normalisés dans l'espace RGB). Nous représentons l'amplitude du déplacement entre les correspondances avec la représentation de flot optique standard (Figure 2(h)). Lorsque l'on utilise uniquement les superpixels, comme dans [10], plusieurs erreurs de correspondances sont obtenues (Figure 2(c)), tandis que la correspondance de superpatches fournit des correspondances très précises (Figure 2(d)).

La même expérience sur une image cisailée décomposée avec [1] (Figure 2(e)), fournit un déplacement uniforme (Figure 2(g)) qui indique des correspondances pertinentes entre les superpatches et démontre la robustesse de la structure proposée aux déformations géométriques par rapport aux méthodes comparant uniquement des superpixels.

Ces exemples illustrent clairement l'intérêt de considérer les relations spatiales entre superpixels voisins avec les superpatches. Ils démontrent également que notre méthode permet de considérer n'importe quelles structures irrégulières définies sur le domaine de l'image.

3 SuperPatchMatch

Nous proposons à présent la méthode SuperPatchMatch (SPM), une extension de l’algorithme PatchMatch (PM) [2], dédiée à notre structure de superpatch pour la recherche rapide d’ANN basés sur les superpixels. Le manque de géométrie régulière entre superpatches est le principal problème pour adapter l’algorithme PM. La notion d’adjacence doit notamment être définie pour une décomposition irrégulière de l’image.

Dans cette section, seules les relations d’adjacence directes ont besoin d’être considérées pour définir notre algorithme. Néanmoins, comme pour les pixels décrits par des patches réguliers, nous avons montré dans la Figure 2 que l’information de voisinage avec les superpatches était nécessaire pour que la méthode de recherche soit efficace pour calculer des correspondances précises de superpixels.

Dans la suite, comme dans la section 2.2, nous illustrons la méthode proposée en considérant deux superpatches A_i et B_j , dans deux images différentes A et B . Notre approche sera ensuite appliquée à une bibliothèque d’images. L’ANN d’un superpixel $A_i \in A$, est noté $B_{(i)} = B_{\mathcal{M}(i)}$, avec \mathcal{M} la carte qui stocke pour les superpixels de A , l’indice dans B de leur ANN.

3.1 Étape d’initialisation

Comme dans l’algorithme PM, pour chaque superpixel $A_i \in A$, nous assignons aléatoirement un superpixel $B_{(i)} \in B$. La Figure 3 montre quelques exemples d’initialisation. Après cette étape, une propagation et une recherche aléatoire sont effectuées itérativement pour chaque superpixel afin d’améliorer les correspondances initiales.

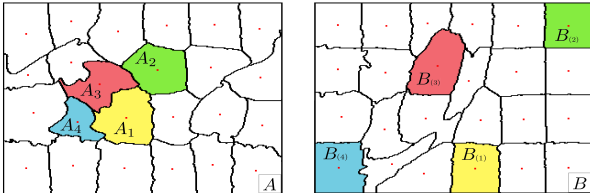


FIGURE 3 – Étape d’initialisation de SPM. Chaque superpixel $A_i \in A$ est associé aléatoirement à un superpixel $B_{(i)} \in B$.

3.2 Étape de propagation

Dans [2], l’étape de propagation tente d’améliorer les ANN actuels en utilisant ceux des voisins adjacents dans A . Les pixels sont traités selon un parcours pré-défini, par exemple depuis le coin haut-gauche au coin bas-droit. La propagation considère uniquement les éléments adjacents traités durant l’étape courante, par exemple les pixels haut-gauche. Leur ANN est décalé afin de respecter les positions relatives des pixels dans A . Le pixel à cette position décalée dans B est alors considéré comme candidat pour améliorer la correspondance du pixel dans A .

Pour les superpixels, nous définissons d’abord l’ordre de traitement des superpixels selon le même sens de parcours à l’échelle pixelique, et proposons de considérer les ANN de tous les superpixels adjacents déjà traités durant l’étape courante. Cette sélection des candidats depuis les superpixels adjacents est illustrée dans la Figure 4. La sélection des voisins adjacents n’est cependant pas directe puisqu’il n’existe pas de géométrie régulière entre les éléments.

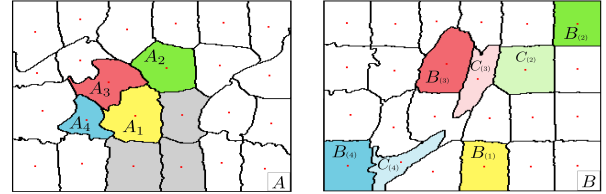


FIGURE 4 – Étape de propagation de SPM. Seuls les superpixels en position haut-gauche par rapport à A_1 sont considérés durant les itérations paires (A_2 , A_3 et A_4). Les voisins restant (en gris) sont testés durant les itérations impaires. Les correspondances actuelles sont notées $B_{(i)}$, et les $C_{(i)}$ représentent les nouveaux candidats à tester comme ANN.

Quand pour un superpixel A_i , son superpixel adjacent $A_{i'}$ et son ANN $B_{(i')}$ sont considérés, un voisin de $B_{(i')}$ est sélectionné comme candidat pour tenter d’améliorer la correspondance de A_i . Cependant, l’ANN $B_{(i')}$ ne peut pas être décalé comme pour les patches réguliers, puisque la géométrie autour des deux superpixels considérés est différente. Afin d’améliorer la correspondance de A_i , il faut donc sélectionner un superpixel candidat pertinent, noté $C_{(i')}$, parmi les superpixels adjacents à $B_{(i')}$. Il est alors important de conserver les positions relatives entre les voisins spatiaux dans les images A et B durant cette étape pour avoir une propagation plus rapide de grandes zones de correspondance [2].

Nous proposons de sélectionner le superpixel candidat $C_{(i')}$ en fonction de sa position relative à $B_{(i')}$, de sorte qu’elle soit la plus similaire à $\theta_{ii'}$, l’angle entre A_i et $A_{i'}$, calculé depuis les barycentres des superpixels :

$$C_{(i')} = \underset{B_k \in V(B_{(i')})}{\operatorname{argmin}} \|(\theta_{ii'} + \pi) - \theta_{i'k}\|, \quad (4)$$

avec $\theta_{i'k}$ l’angle entre $B_{(i')}$ et l’ensemble de ses superpixels adjacents $B_k \in V(B_{(i')})$.

Dans la Figure 5, nous illustrons la sélection du candidat $C_{(3)}$ pour le voisin A_3 , qui est en haut à gauche de A_1 . Afin de conserver la même position relative, le nouveau superpixel à tester comme ANN pour A_1 , est le voisin B_k de $B_{(3)}$ qui est le plus proche de sa position bas-droite, par rapport à l’angle θ_{3k} .

3.3 Étape de recherche aléatoire

L’étape de recherche aléatoire consiste à échantillonner itérativement dans B afin d’échapper à un potentiel minimum local [2]. Les candidats sont sélectionnés à l’intérieur

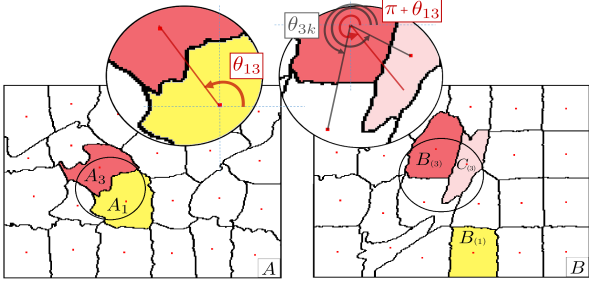


FIGURE 5 – Afin d’améliorer l’ANN du superpixel A_1 , le voisin A_3 est considéré. Son ANN actuel est $B_{(3)}$. Le superpixel sélectionné $C_{(3)}$ à tester comme candidat est le superpixel adjacent à $B_{(3)}$ avec l’orientation relative la plus similaire, c’est-à-dire le superpixel B_k dont l’angle avec $B_{(3)}$, θ_{3k} , est le plus proche de $\pi + \theta_{13}$.

de zones de recherche centrées autour du barycentre de l’ANN actuel, et dont la taille diminue d’un facteur 2 jusqu’à ne contenir qu’un seul pixel. Dans le cas des superpixels, une position dans B est sélectionnée aléatoirement et le superpixel la contenant est le candidat à tester. La Figure 6 illustre la recherche aléatoire, où les zones de recherche sont représentées par des lignes pointillées.

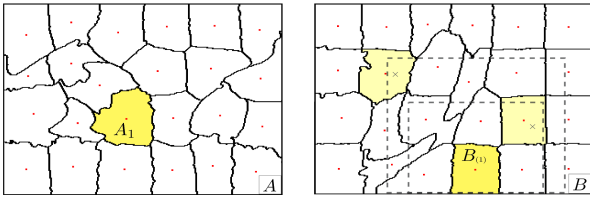


FIGURE 6 – Étape de recherche aléatoire de SPM. L’échantillonnage est effectué à une distance décroissante du barycentre de l’ANN actuel $B_{(1)}$ du superpixel A_1 . Les superpixels contenant les positions sélectionnées (croix) sont testés pour améliorer l’ANN de A_1 .

3.4 Bibliothèque d’images d’exemples

Comme pour PM, un avantage majeur de SPM est sa complexité, qui ne dépend que de la taille de l’image à traiter A et pas de la taille de l’image d’exemple B . Les adaptations de SPM au cas d’une bibliothèque d’images d’exemples T sont directes. L’initialisation est étendue afin de sélectionner aléatoirement des correspondances dans T . L’étape de propagation teste toujours l’ANN décalé des voisins spatiaux dans l’image A . Ces voisins n’ont alors pas nécessairement des ANN se trouvant dans la même image de la bibliothèque, et permettent donc à un superpixel d’accéder potentiellement à de nouvelles images de T .

Enfin, concernant l’étape de recherche aléatoire, le nombre d’éléments de l’image peut être inférieur à celui du nombre d’images d’exemples. Une image dans T qui ne serait pas accédée par une correspondance aléatoire lors de l’initialisation ne pourrait donc jamais être atteinte du-

rant le processus. Nous effectuons donc comme dans [20] la recherche aléatoire dans l’image contenant l’ANN et également dans une image tirée aléatoirement dans T , afin d’augmenter la probabilité d’accéder à toutes les images de la bibliothèque.

Enfin, contrairement à PM, qui ne calcule qu’un ANN, nous proposons de calculer k ANN dans la bibliothèque T avec SPM. On choisit d’effectuer la recherche des k ANN par k SPM indépendants afin de faciliter l’implémentation parallèle.

3.5 Schéma de fusion d’étiquettes

Nous considérons un schéma de fusion permettant l’étiquetage multi-étiquettes de superpixels depuis une bibliothèque d’images d’exemples. On considère une bibliothèque d’exemples T , et on dénote $l(T_j) \in \{1, \dots, M\}$ l’étiquette du superpixel $T_j \in T$. En sortie de l’algorithme SPM, k ANN sont estimés pour chaque superpixel de l’image A . Afin d’obtenir l’étiquetage final d’un superpixel A_i , dont l’ensemble des k ANN T_j , ayant pour étiquettes $l(T_j) = m$, est noté $\mathcal{K}_i^m = \{T_j\}$, la carte de fusion d’étiquettes $L_m(A_i)$ est définie par :

$$L_m(A_i) = \frac{\sum_{T_j \in \mathcal{K}_i^m} \omega(A_i, T_j)}{\sum_{m=1}^M \sum_{T_j \in \mathcal{K}_i^m} \omega(A_i, T_j)}, \quad (5)$$

où $\omega(A_i, T_j)$ est le poids assigné à l’étiquette m , et dépend de la similarité entre le superpatch $\mathbf{A}_i \in A$, et le superpatch ANN $\mathbf{T}_j \in T$. Cette carte d’étiquettes L_m donne la probabilité d’assigner l’étiquette m au superpixel A_i .

Certains types de données, ou certaines bases spécifiques comme les images médicales, considèrent des images recalées où les structures d’intérêt entre les images sont proches spatialement. Sur ces images, de bonnes correspondances de superpatches ne devraient pas être spatialement trop éloignées entre les images comparées, et cette proximité peut également être utilisée afin de renforcer la cohérence spatiale des k ANN sélectionnés. Chaque contribution d’ANN à l’étiquetage peut donc être pondérée par la distance spatiale entre les barycentres $c_i \in A$ et $c_j \in T$, et le poids utilisé dans l’Éq. (5) peut être défini comme :

$$\omega(A_i, T_j) = \exp \left(1 - \left(\frac{D(\mathbf{A}_i, \mathbf{T}_j)}{h(A_i)^2} + \frac{\|c_i - c_j\|_2}{\beta^2} \right) \right), \quad (6)$$

où $h(A_i)^2 = \alpha^2 \min_{T_j \in \cup_m \mathcal{K}_i^m} (D(\mathbf{A}_i, \mathbf{T}_j) + \epsilon)$, où $\epsilon \rightarrow 0$, et α et β

sont des paramètres d’échelle. Notons que si $\beta = \infty$, cela revient à ne pas considérer d’a priori de position. Avec la fonction $h(A_i)$, la distance des contributions à l’étiquetage de A_i est divisée par la distance minimale parmi toutes les contributions des k ANN.

L’étiquetage final $\mathcal{L}(A_i)$ est enfin obtenu avec l’étiquette ayant la plus forte probabilité :

$$\mathcal{L}(A_i) = \operatorname{argmax}_{m \in \{1, \dots, M\}} L_m(A_i). \quad (7)$$

La relation (7) donne une décision à l'échelle des superpixels qui peut s'avérer irrégulière. Comme dans [16], les cartes d'étiquetage L_m (5) peuvent être utilisées comme un terme de données multi-étiquettes et considérées dans le problème de régularisation suivant, où une énergie E est définie sur le graphe construit depuis les superpixels adjacents :

$$E(\mathcal{L}) = \sum_{i=1}^{|\mathcal{A}|} \sum_{A_{i'} \in \mathcal{A}_i} \exp\left(-\frac{d(F_{A_i}, F_{A_{i'}})}{\gamma}\right) \delta_{\mathcal{L}(A_i) \neq \mathcal{L}(A_{i'})} + \sum_{i=1}^{|\mathcal{A}|} (1 - L_{\mathcal{L}(A_i)}(A_i)), \quad (8)$$

où γ est un paramètre de régularisation, le terme de données $1 - L_{\mathcal{L}(A_i)}(A_i)$ est proche de 0 (respectivement 1) quand la probabilité de l'étiquette $\mathcal{L}(A_i)$ est forte (respectivement faible), $\delta_a = 1$ quand a est vrai et 0 sinon. La minimisation de cette énergie incite à donner la même étiquette à deux superpixels voisins spatialement et ayant une faible distance d entre leurs descripteurs F .

4 Application à l'étiquetage de visages

Nous proposons d'évaluer SPM pour l'étiquetage de visages sur la base de données LFW (pour Labeled Faces in the Wild, [11]). Cette base contient des images de taille 250×250 pixels, recalées linéairement, et chaque image est fournie avec une décomposition en environ 225 à 250 superpixels.

LFW est une base de référence, utilisée pour valider les nouvelles méthodes basées sur les superpixels puisqu'elle contient des vérités terrains d'étiquetage (cheveux, visage et fond) associées à des décompositions en superpixels [12] fournies pour chaque image. Ainsi, contrairement aux autres bases standards, les comparaisons avec les méthodes de l'état de l'art ne sont pas biaisées par les décompositions en superpixels, qui peuvent fortement impacter la précision de segmentation.

La segmentation et l'étiquetage de visages sont des applications difficiles, en raison de plusieurs problèmes tels que la diversité de styles de coiffure, de couleurs de peau, les occultations partielles ou encore la présence de plusieurs visages sur une même image.

4.1 Réglage des paramètres

SPM a été implémenté avec MATLAB via du code C-MEX multi-processus. Pour comparer la méthode à celles ayant des résultats de référence sur la base LFW [12, 13], nous utilisons les mêmes 1 500 images d'exemples dans notre bibliothèque T , et les mêmes 927 images de test comme images à étiqueter.

Le nombre d'itérations de SPM est fixé à 5, comme dans [2]. Nous utilisons une norme ℓ_2 entre les histogrammes de gradients d'intensité orientés (HoG) [6] comme distance d dans (1). Dans l'Éq. (2), puisque les images sont de taille

$h \times w = 250 \times 250$ pixels, et décomposées en approximativement $K = 250$ superpixels, σ_1 est fixé à 8 (cf. section 2.2). Dans (6), les paramètres α et β sont fixés empiriquement à 2 et 4. Dans le modèle de régularisation (8), nous fixons γ à 0.5 et utilisons l'algorithme α -expansion [3] pour minimiser la fonctionnelle. Enfin, les temps de calcul dans la Figure 7 incluent la recherche des k ANN, la fusion d'étiquettes, la régularisation et l'étiquetage final.

4.2 Influence des paramètres

Nous analysons tout d'abord l'influence de la taille du superpatch et du nombre d'ANN, c'est-à-dire de k , le nombre de SPM effectués en parallèle. La Figure 7 représente la précision de l'étiquetage à l'échelle superpixelique et le temps de calcul. La précision d'étiquetage est améliorée avec notre structure de superpatch et les meilleurs résultats sont obtenus avec $R=50$ pixels (95.08% avec $k=50$ ANN). Cette taille correspond en moyenne à la capture de trois anneaux de superpixels voisins, puisque les superpixels sont en moyenne de taille 16×16 pixels.

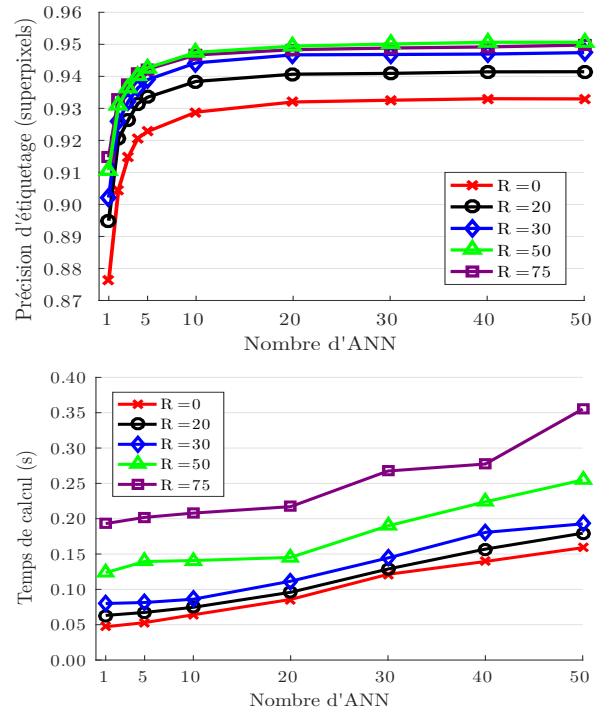


FIGURE 7 – Influence de la taille du superpatch sur la précision d'étiquetage et temps de calcul correspondant.

Sans la structure de superpatch, c'est-à-dire en calculant seulement la distance de correspondance sur le superpixel central ($R=0$ pixels), comme dans [10], de moins bons ANN sont trouvés, ce qui décroît la précision d'étiquetage (93.29% avec $k=50$ ANN). La taille du superpatch doit être assez grande pour capturer l'information contenue dans le voisinage des superpixels. Cependant, avec des superpatches trop grands, c'est-à-dire $R > 50$, trop de superpixels voisins contribuent, menant à des ANN moins pertinents, à un étiquetage moins précis et à des temps de

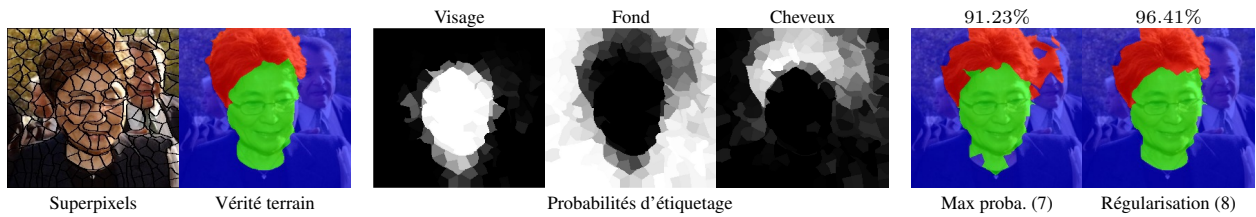


FIGURE 8 – Exemple de probabilités d’étiquetage sur chaque classe et impact de la régularisation sur l’étiquetage final.

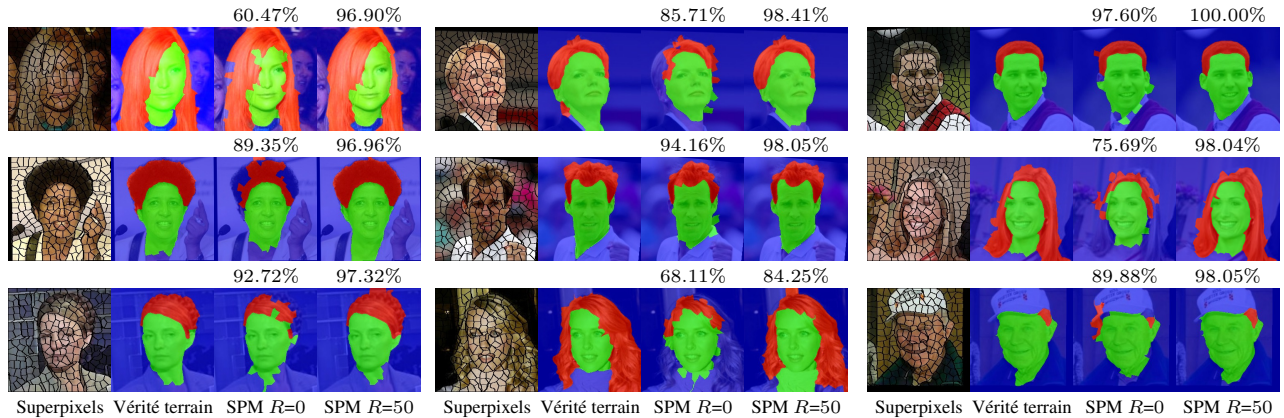


FIGURE 9 – Exemples d’étiquetages avec précision à l’échelle des superpixels obtenus avec SPM pour $R = 0$ et $R = 50$ pixels, c’est-à-dire, en utilisant respectivement les superpixels et les superpatches.

calcul plus élevés. Notons que l’on propose dans (6) une amélioration de l’étape de fusion d’étiquettes pour prendre en considération le recalage de la base, mais nous obtenons des résultats comparables (95.00% au lieu de 95.08%) sans a priori de position, c’est-à-dire $\beta = \infty$ dans (6).

La Figure 8 illustre le processus de régularisation. Les probabilités d’étiquetage (5) obtenues par SPM sont représentées pour chaque étiquette. La régularisation spatiale (8) donne des résultats plus cohérents qu’en assignant directement l’étiquette avec la plus forte probabilité (7).

La Figure 9 montre enfin l’influence du superpatch sur différents exemples. Les erreurs d’étiquetage sont principalement dues à la similarité entre la couleur des cheveux et le fond, et à l’imprécision des segmentations en superpixels fournies par la base.

Afin d’adapter l’étape de propagation de PatchMatch aux superpixels, nous avons proposé dans (4), de prendre comme candidat, le superpixel dans l’image T_j ayant l’orientation la plus similaire à celle des superpixels dans l’image A , en considérant les angles entre les barycentres des superpixels. L’intérêt de cette stratégie peut être validé expérimentalement. En effet, nous obtenons une précision d’étiquetage de 95.08%, alors qu’en considérant dans (4), un voisin aléatoire plutôt que celui avec l’orientation la plus similaire, la précision d’étiquetage descend à 94.41%.

4.3 Comparaison avec l’état de l’art

SPM est comparé aux méthodes de référence appliquées à la base LFW dans le Tableau 1. Dans [12], la méthode GLOC (pour GLObal and LOcal) utilise une machine de

Boltzmann contrainte RBM (pour Restricted Boltzmann Machine) comme complément d’un champ aléatoire conditionnel, noté CRF (pour Conditional Random Field). La combinaison de ces approches permet de réduire les erreurs d’étiquetage obtenues avec des modèles simples qui ne considèrent pas d’a priori de forme globale, aux dépens d’un temps de calcul plus élevé. Dans [13], une architecture basée sur un réseau de neurones convolutionnel profond DCNN (pour Deep Convolutional Neural Network) est proposée. Notons que pour toutes les méthodes comparées, une étape d’apprentissage, qui peut durer plusieurs heures, est nécessaire afin d’entraîner les modèles. De plus, elles considèrent des a priori d’information sémantique dans leur processus d’étiquetage, par exemple, l’étiquette cheveux doit se trouver principalement au dessus de l’étiquette visage. Enfin, nous fournissons également les résultats obtenus avec l’algorithme PatchMatch à l’échelle pixelique, utilisé avec les mêmes paramètres et la même méthode de fusion d’étiquettes que celle présentée dans la section 3.5. Les correspondances sont alors calculées par une somme des différences au carré sur les intensités couleurs dans l’espace RGB de patches de taille 9×9 pixels.

Afin de comparer notre méthode aux autres approches de l’état de l’art, nous fournissons dans le Tableau 1 les résultats d’étiquetage à l’échelle des superpixels et des pixels. La précision pixelique peut en effet être différente puisque la précision superpixellique ne considère pas la taille des superpixels. Les résultats sont ceux publiés par les auteurs, par conséquent, les valeurs de certaines métriques ne peuvent être reportées. La précision

TABLE 1 – Précision d’étiquetage sur la base LFW. Les résultats de SPM sont obtenus avec $k=50$, et $R=50$ pixels.

Méthode	Précision à l’échelle des superpixels	Précision à l’échelle des pixels
PatchMatch	87.73%	87.02%
CRF spatial [12]	93.95%	<i>non reporté</i>
CRBM [12]	94.10%	<i>non reporté</i>
GLOC [12]	94.95%	<i>non reporté</i>
DCNN [13]	<i>non reporté</i>	95.24%
SuperPatchMatch	95.08%	95.43%

d’étiquetage superpixellique de SPM (95.08%) surpasse celles des méthodes comparées. La précision pixellique (95.50%) surpasse également le résultat donné pour l’architecture de réseau de neurones DCNN [13], qui a été optimisée pour être appliquée à la base LFW. L’augmentation de la précision pour l’échelle pixellique par rapport à l’échelle superpixellique montre que notre méthode échoue principalement à étiqueter de petits superpixels.

Le temps de calcul est un autre point de comparaison important. Les résultats de SPM donnés dans le Tableau 1 sont obtenus en 0.255s de traitement par image, et des temps similaires sont reportés pour la méthode GLOC [12].

SPM surpasse donc les méthodes de l’état de l’art comparées en termes de précision d’étiquetage sans aucune étape d’entraînement. L’absence d’apprentissage nécessaire pourrait permettre à SPM d’utiliser toutes les images dans une procédure de validation croisée de type *leave-one-out*. Ainsi, SPM atteint une précision superpixellique de 95.26%. Ce résultat souligne l’impact de la diversité des images d’exemples qui peut améliorer la qualité des correspondances. De plus, ce résultat est obtenu sans augmentation du temps de calcul, puisque la complexité de notre algorithme ne dépend que de la taille de l’image traitée.

5 Conclusion

Dans ce travail, nous proposons une nouvelle structure de voisinage de superpixels, permettant de réaliser des comparaisons robustes de motifs visuels sur des domaines irréguliers de l’image. Nous avons aussi introduit SuperPatchMatch, un nouvel algorithme de recherche de correspondance de superpatches. Nous avons démontré l’intérêt de notre méthode en obtenant des résultats état de l’art pour l’étiquetage des visages. SuperPatchMatch ne nécessite aucune phase d’apprentissage, pouvant aller jusqu’à plusieurs heures pour de nombreuses méthodes de la littérature. Ce travail ouvre donc de nouvelles perspectives intéressantes pour les méthodes et les chaînes de traitement utilisant les superpixels.

Remerciements

Ce travail a bénéficié d’un support financier de l’état Français, via l’Agence Nationale de la Recherche (ANR) dans le cadre du projet GOTMI (ANR-16-CE33-0010-01).

Références

- [1] R. Achanta et al., SLIC superpixels compared to state-of-the-art superpixel methods, *PAMI*, vol 34, no 11, pp 2274–2282, 2012.
- [2] C. Barnes et al., PatchMatch : A randomized correspondence algorithm for structural image editing, *ACM ToG*, vol 28, no 3, 2009.
- [3] Y. Boykov et al., Fast approximate energy minimization via graph cuts, *PAMI*, vol 23, no 11, pp 1222–1239, 2001.
- [4] A. Buades et al., A non-local algorithm for image denoising, *CVPR*, 2005.
- [5] P. Buysens et al., Eikonal-based region growing for efficient clustering, *IVC*, vol 32, no 12, pp 1045–1054, 2014.
- [6] N. Dalal et al., Histograms of oriented gradients for human detection, *CVPR*, 2005.
- [7] A. Efros et al., Texture synthesis by non-parametric sampling, *ICCV*, 1999.
- [8] R. Giraud et al., SuperPatchMatch : An algorithm for robust correspondences using superpixel patches, *TIP*, 2017.
- [9] R. Giraud et al., Robust superpixels using color and contour features along linear path, *CVIU*, 2018.
- [10] S. Gould et al., Superpixel graph label transfer with learned distance metric, *ECCV*, 2014.
- [11] G. Huang et al., Labeled faces in the wild : A database for studying face recognition in unconstrained environments, *Technical Report 07-49, Univ. of Massachusetts, Amherst*, 2007.
- [12] A. Kae et al., Augmenting CRFs with Boltzmann machine shape priors for image labeling, *CVPR*, 2013.
- [13] S. Liu et al., Multi-objective convolutional learning for face labeling, *CVPR*, 2015.
- [14] D. Lowe et al., Distinctive image features form scale-invariant keypoints, *IJCV*, vol 60, no 2, pp 91–110, 2004.
- [15] J. Lu et al., PatchMatch filter : Efficient edge-aware filtering meets randomized search for fast correspondence field estimation, *CVPR*, 2013.
- [16] S.-C. Pei et al., Saliency detection using superpixel belief propagation, *ICIP*, 2014.
- [17] J. Rabin et al., Adaptive color transfer with relaxed optimal transport, *ICIP*, 2014.
- [18] M. Reso et al., Temporally consistent superpixels, *ICCV*, 2013.
- [19] R. Sawhney et al., GASP : Geometric association with surface patches, *3DV*, 2014.
- [20] W. Shi et al., Cardiac image super-resolution with global correspondence using multi-atlas PatchMatch, *MICCAI*, 2013.
- [21] E. Trulls et al., Segmentation-aware deformable part models, *CVPR*, 2014.