

Mesure de dissimilarité pour les patches utilisant la corrélation

Paul RIOT¹, Andrés ALMANSA², Yann GOUSSEAU¹, Florence TUPIN¹

¹LTCI, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France

²MAP5, CNRS, Université Paris Descartes, 75006 Paris, France

paul.riot@telecom-paristech.fr

Résumé – Dans ce travail, nous définissons une mesure de dissimilarité entre patches pour une image corrompue par un bruit blanc, de manière à proposer une alternative à la norme L^2 , très largement utilisée pour évaluer les distances et similarités entre images et patches. Nous montrons comment obtenir une mesure de dissimilarité structurelle au moyen de l'autocorrelation.

Abstract – In this work, we address the problem of defining a robust patch dissimilarity measure for an image corrupted by an additive white noise. In particular, the L^2 -norm is very widely used to evaluate distances and similarities between images or patches. However, we claim that a better and more structural dissimilarity measure can be defined using the autocorrelation.

1 Introduction

Dans cet article, nous définissons une nouvelle mesure de dissimilarité entre patches issus d'images naturelles corrompues par un bruit blanc. L'hypothèse courante de blancheur du bruit, bien qu'imprécise sur des images déjà traitées, est réaliste sur des images RAW.

Pourtant, la plupart des algorithmes n'utilisent pas complètement le potentiel de cette hypothèse de blancheur. En effet, elle n'est presque jamais contrainte ou vérifiée sur les résultats obtenus. En particulier, les méthodes de débruitage non locales utilisent souvent la norme L^2 pour comparer les patches, et le SNR ou MSE sont communément utilisés pour évaluer les performances de débruitage. Ici, nous proposons d'évaluer la blancheur des patches en utilisant leur autocorrelation comme une alternative à la norme L^2 classique, obtenant ainsi une mesure plus robuste au bruit et plus discriminante.

Nous montrons le potentiel de la méthode proposée pour trois applications : (a) la détection de patches similaires (b) le débruitage non local, et (c) l'évaluation de qualité de débruitage sans oracle (contrairement au SNR ou SSIM). La recherche de meilleures mesures de dissimilarité entre patches n'est pas nouvelle et a déjà été abordée pour ces trois applications. Notre approche utilise l'autocorrelation comme mesure pour discriminer le bruit blanc non structuré des caractéristiques structurées des images naturelles. A notre connaissance, peu de travaux proposent de tirer partie des statistiques du bruit de cette façon [1, 5, 6, 7, 9, 12]. Tous ces travaux se concentrent sur une formulation variationnelle où un terme de blancheur est utilisé en complément des termes d'attache aux données et de régularisation, afin de contraindre la solution à respecter les statistiques du premier et second ordre du bruit que l'on cherche à estimer et retirer.

Notre approche est différente et, d'une certaine façon, plus générale. Nous proposons une nouvelle mesure de dissimilarité entre patches dont l'intérêt dans le cadre du débruitage dépasse

l'application variationnelle proposée dans [9]. Nous utilisons cette mesure en substitution de la norme L^2 pour identifier de manière plus robuste les patches similaires dans les méthodes de débruitage non locales.

Finalement, nous nous intéressons au problème de l'évaluation de qualité dans le cadre du débruitage. Il est connu que le SNR, très fréquemment utilisé, fournit peu d'informations structurelles. Le SSIM, présenté dans [13], a été construit pour offrir une mesure de dissimilarité plus pertinente visuellement. Pourtant, des doutes peuvent être émis quant à sa pertinence [4] dans le contexte du débruitage. Nous proposons une alternative (sans oracle) pour évaluer la qualité d'une image débruitée qui repose sur l'autocorrelation de la différence entre les patches.

Dans la première partie de cet article, nous expliquons comment l'autocorrelation peut être utilisée pour définir un nouveau critère de similarité entre patches, ainsi qu'évaluer la qualité du débruitage. Dans la seconde partie de ce travail, nous présentons quelques expériences démontrant le potentiel de cet outil pour les tâches définies précédemment. Une version plus détaillée de ce travail, en anglais, se trouve ici [10].

2 Critère de Similarité Proposé

2.1 Évaluer la blancheur avec l'autocorrelation

Notre but est d'enrichir la norme L^2 sur un patch bruité pour inclure des statistiques qui contiennent des informations structurelles. Parmi les différents outils possibles pour évaluer la blancheur [1], nous avons choisi d'utiliser l'autocorrelation qui nous a semblé simple et efficace pour détecter les structures.

Nous définissons l'estimateur classique de l'autocorrelation r_P sur un patch bruité P de taille $s \times s$ pour toutes les translations possibles $(l, m) \in \{-\frac{s}{2}, \dots, \frac{s}{2}\}^2$:

$$r_P(l, m) = \sum_{(x,y)} P(x, y)P(x + l, y + m) \quad (1)$$

où les indices sont circulaires.

Choisir une convolution circulaire permet d'accélérer les calculs par l'utilisation de la FFT, de simplifier les modèles ainsi que d'améliorer l'estimation des basses fréquences par l'augmentation du nombre d'échantillons utilisés. On peut noter que grâce à l'hypothèse iid, la convolution circulaire ne présente pas de désavantage particulier.

2.2 L'autocorrelation pour définir une nouvelle mesure de dissimilarité entre patches

Dans les algorithmes de débruitage non locaux, la norme L^2 de la différence entre les patches $P_1 - P_2$ est un moyen naturel d'évaluer leur dissimilarité. En effet, en supposant que les patches sans bruit sous-jacents sont égaux, on sait que $P_1 - P_2$ est un bruit blanc Gaussien de variance $2\sigma^2$. Ainsi, la norme L^2 peut être justifiée mathématiquement [3] comme un rapport de vraisemblance généralisé. Pourtant, elle n'est pas toujours robuste pour des bruits de forte variance.

Quand le rapport signal sur bruit est faible, il a été montré dans [8] que la norme L^2 va souvent apparier les réalisations de bruit plutôt que le contenu de l'image. Ce phénomène montre que la distance L^2 est relativement sensible au niveau de bruit. En effet, elle n'évalue que l'énergie de la différence entre les patches $P_1 - P_2$ et ne prend en compte aucune information structurelle.

Dans ce travail, nous montrons que calculer la matrice d'autocorrelation $r_{P_1 - P_2}$ est plus fiable que la norme L^2 , contenant davantage d'informations de structure. Ainsi, on construit une nouvelle mesure de dissimilarité en utilisant l'estimateur $r_{P_1 - P_2}$ défini précédemment.

On peut remarquer que $r_{P_1 - P_2}(0, 0) = \|P_1 - P_2\|^2$ correspond à l'énergie de $P_1 - P_2$, c'est à dire le carré de la norme L^2 . Les autres translations donnent des informations sur la structure et les dépendances de $P_1 - P_2$. Plus les coefficients $|r_{P_1 - P_2}(l, m)|$ pour $(l, m) \neq (0, 0)$ sont importants, plus $P_1 - P_2$ est structuré, signifiant que les patches non bruités sous-jacents ne sont pas similaires.

Grâce à cette observation, une nouvelle mesure de dissimilarité, que nous nommons Whiteness Dissimilarity Measure (WDM), peut être construite :

$$d_{w}^2(P_1, P_2) = \|r_{P_1 - P_2}\|_2^2. \quad (2)$$

Cette mesure est expérimentalement plus robuste que L^2 à des niveaux de bruit élevés. Elle est aussi plus fiable sur des patches de taille importante car l'estimation de l'autocorrelation devient plus robuste quand le nombre d'échantillons augmente. Nous pouvons néanmoins montrer expérimentalement que les résultats sont améliorés même pour les petites tailles de patches.

Il est important de noter que WDM n'est pas une distance. En effet, il est possible de montrer qu'il vérifie les propriétés de non-négativité, d'identité des indiscernables et de symétrie, mais qu'il ne satisfait pas l'inégalité triangulaire.

2.3 Évaluation du débruitage au delà du SNR

Le SNR et ses variantes comme le MSE ou le PSNR sont les métriques utilisées le plus fréquemment pour évaluer et comparer les performances de débruitage. Fondamentalement, il s'agit de mesurer l'erreur en calculant son énergie $\|f - u\|^2$. C'est parfaitement pertinent d'un point de vue mathématique. Pourtant, cela n'est pas toujours informatif visuellement. En effet, il est reconnu que le SNR n'est pas directement corrélé avec une évaluation de qualité visuelle. C'est la raison pour laquelle de nouvelles métriques ont été développées comme le SSIM [13]. Malgré tout, de nombreuses critiques ont été émises concernant ses propriétés structurelles [4]. De plus, le SSIM et le SNR nécessitent un oracle pour estimer la qualité du débruitage.

Dans ce travail, nous montrons que l'autocorrelation du résidu peut être utilisée comme un complément du SNR pour évaluer la quantité d'informations structurelles qui a été retirée durant le débruitage. Nous montrons également des expériences suggérant que cette quantité est liée à une information de qualité visuelle. En particulier, elle est sensible aux textures endommagées et aux artefacts, et de ce fait pertinente pour évaluer des résultats de débruitage. De plus, notre métrique est construite pour fonctionner sans oracle en évaluant la blancheur du résidu.

L'autocorrelation du résidu, normalisée par son espérance, peut être calculée sur chacun des patches R_k du résidu :

$$A(R_k) = \frac{\|r_{R_k}\|_2^2 - r_{R_k}(0, 0)^2}{\sigma_k^4 s^2 (s^2 - 1)} \quad (3)$$

où $\sigma_k^2 = \sum_{(i,j)} R_k(i, j)^2$ est l'énergie du résidu estimée sur le patch R_k et s^2 est la taille de patch.

Ensuite, ces coefficients calculés sur chacun des patches peuvent être agrégés pour obtenir une carte de qualité en chaque pixel que l'on nomme W-map. Cette carte contient une estimation locale de la quantité de contenu structuré qui a été retirée lors du processus de débruitage. Un score global W-score peut être obtenu avec une simple somme normalisée.

Ce score donne une estimation globale de la qualité du débruitage. Quand il est très proche de 1, l'algorithme n'a pas abîmé l'image sous-jacente. Quand il augmente, cela signifie que le résidu est structuré, et ainsi que du contenu de l'image sans bruit a été retiré. Nous prétendons que cette propriété est pertinente visuellement et permet de préférer un algorithme qui conserve le contenu de l'image en ne débruitant que légèrement les zones texturées, à une méthode qui va débruiter trop fortement et délayer certaines textures. Par ailleurs, l'approche peut être utile pour d'autres étapes de traitement de l'image comme la détection ou la reconnaissance d'objet car elle permet de mieux garantir la conservation de l'information.

Il est important de souligner que le W-score ne devrait pas être évalué seul car il ne contient aucune information sur l'énergie du résidu (nous avons retiré le terme $r_{R_k}(0, 0)^2$). En effet, ajouter un bruit blanc à une image déjà bruitée donnerait un W-score parfait de 1, puisque le résidu serait blanc. Ainsi, le W-score doit être couplé au SNR pour être bien interprété.

Nous avons choisi de ne pas considérer $r_{R_k}(0,0)$ dans la définition de notre mesure, car ce terme contient l'énergie du résidu et que la meilleure stratégie à adopter concernant cette énergie n'est pas toujours claire. Si le modèle est localement pertinent un débruitage complet peut être réalisé et donc la cible pour $r_{R_k}(0,0)$ devrait être σ^2 , mais si la fonction de régularisation n'est localement pas adaptée au contenu, alors il est plus prudent de ne pas débruiter l'image et donc d'avoir $r_{R_k}(0,0)$ proche de zéro. En d'autres termes, il n'est pas simple d'évaluer en même temps des informations de nature très différentes, les autres coefficients contenant avant tout une information fréquentielle.

3 Résultats Expérimentaux

3.1 Détection de patches similaires avec de forts niveaux de bruit

Pour déterminer la pertinence de WDM appliqué à la détection de patches similaires, nous suivons le protocole qui a été récemment proposé dans [3]. Les performances respectives de WDM, L^1 , L^2 et L^4 sont comparées sur un dictionnaire composé de 196 patches sans bruit de taille $s = 8$. Les patches bruités sont obtenus avec un bruit blanc Gaussien d'écart type $\sigma = 100$. Les mesures de dissimilarité sont évaluées pour toutes les paires de patches bruités. Le processus est répété 20 fois avec des réalisations de bruit indépendantes. Numériquement, les performances des critères sont données au moyen d'une courbe ROC (Receiver Operating Characteristic), c'est à dire la courbe de probabilité de détection comme une fonction de la probabilité de fausse alarme. Les résultats sont affichés sur la Figure 1.

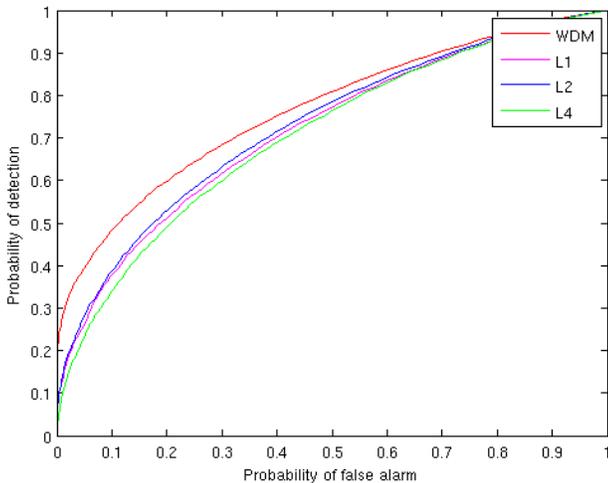


FIGURE 1 – Courbe ROC obtenue lors de la détection de patches similaires soumis à un fort bruit Gaussien ($\sigma = 100$).

Il apparaît clairement sur cette courbe que WDM obtient de meilleures performances que L^1 , L^2 et L^4 . Comme il a été dit précédemment, cela est particulièrement vrai pour de faibles

rapports signal sur bruit.

3.2 Amélioration de l'algorithme Non Local Means (NLM)

Le nouveau critère de similarité défini précédemment peut être appliqué à n'importe quel algorithme fondé sur les patches. Mais, pour des raisons d'évaluation, nous allons ici nous concentrer sur l'algorithme original des moyennes non locales (NLM) [2] qui permet le débruitage par l'agrégation de patches similaires. Bien que beaucoup de méthodes plus récentes et efficaces existent, nous pensons, comme cela est expliqué dans [3], que cet algorithme est un bon moyen de comparer des critères de similarité. On peut aussi noter qu'il existe d'autres possibilités pour rendre cet algorithme robuste au bruit, comme l'approche par ACP proposée par [11].

Dans cette partie, les performances de cet algorithme seront comparées en utilisant respectivement WDM et L^2 . Pour limiter le champ de notre étude, une taille de fenêtre de recherche $W = 21$ est fixée ainsi que la taille des patches $s = 5$. Afin de comparer les résultats de façon juste, le paramètre de bande passante h (l'écart type de la gaussienne permettant de calculer les poids) est optimisé pour chacun des cas pour obtenir le meilleur SNR possible. Le SNR et le SSIM des résultats pour différentes images sont disponibles dans le tableau 1. On observe que changer la distance L^2 pour WDM améliore toujours les performances d'au moins 0.5 dB, souvent plus. Le SSIM est aussi systématiquement amélioré.

TABLE 1 – Performances en termes de SNR (dB) et SSIM de l'algorithme NLM pour les critères de similarités L^2 et WDM avec $\sigma = 25$ ($W = 21$, $s = 5$).

	L^2	WDM
Lena	22.10 / 0.869	22.65 / 0.878
Building	17.85 / 0.883	18.37 / 0.897
Mandrill	19.06 / 0.821	19.29 / 0.836
House	25.73 / 0.811	26.57 / 0.829
Fingerprint	21.26 / 0.921	21.95 / 0.943
Flintstones	22.52 / 0.940	23.28 / 0.951
Boat	22.59 / 0.852	23.24 / 0.869
Barbara	21.91 / 0.895	22.68 / 0.907

D'autres expériences, qui ne sont pas affichées ici pour des raisons de place, ont montré que WDM obtient des performances significativement meilleures que L^2 quelle que soit la taille du patch s . De plus, l'amélioration est d'autant plus importante que le niveau de bruit augmente. Pourtant, il est intéressant de noter que WDM obtient des performances très proches de L^2 quand les poids sont calculés directement sur l'image oracle. Cela confirme la pertinence de WDM, même en l'absence de bruit. Finalement, nous avons aussi remarqué que les performances de WDM sont plus stables quant au choix du paramètre de bande passante h .

3.3 Évaluation de qualité du résidu

Cette dernière expérience s'intéresse à la pertinence de la mesure de qualité introduite dans la Section 2.3 avec $s = 15$. Rappelons que celle-ci est calculée sur le résidu et ainsi ne nécessite pas d'oracle, contrairement au SNR et au SSIM.

Cette expérience très simple est seulement fournie comme une preuve de concept. Elle consiste à comparer les W-scores et W-map pour des résultats de débruitage ayant le même SNR mais étant visuellement très différents. Un exemple est visible sur la Figure 2, où on compare des versions débruitées de l'image Barbara corrompue par un bruit blanc Gaussien d'écart type $\sigma = 25$. Ces images ont été obtenues avec l'algorithme NLM et des paramètres de bande passante h différents. Elles présentent un SNR identique de 19.50dB.

Le résultat 2.(a) est relativement conservateur et seulement une petite partie du bruit a été retiré. Les textures et le contraste ont bien été préservés. Ainsi, aucune structure n'est détectée dans le résidu et notre critère est proche de la valeur 1. A l'inverse, le résultat 2.(b) a été trop fortement débruité. La plupart des textures ont été délavées et le résultat présente de fortes pertes de contraste. Notre critère d'évaluation semble pouvoir détecter les zones abîmées comme les textures sur les vêtements qui correspondent à des valeurs importantes de W-map. On peut remarquer que les zones homogènes présentent de bonnes valeurs pour W-map car elles ont été bien débruitées. Globalement, le W-score pour cette image est de 1.72.

Bien sûr, cette simple expérience n'est pas suffisante pour affirmer l'intérêt perceptuel de la mesure proposée. Cette tâche mérite d'autres travaux, probablement par une étude utilisateur.

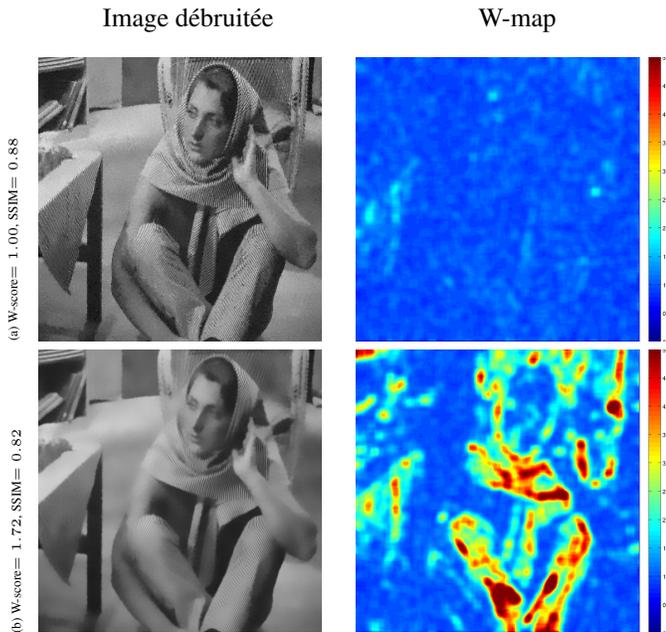


FIGURE 2 – Évaluation de qualité pour 2 méthodes de débruitage avec le même SNR (19.5dB) mais des qualités visuelles différentes.

4 Conclusion

Dans ce travail, nous avons présenté une alternative à la norme L^2 pour comparer des patches bruités. Cette alternative, nommée WDM, semble être plus sensible aux structures de l'image que la norme L^2 . La mesure de dissimilarité WDM, qui repose sur un estimateur de l'autocorrélation, a été appliquée à différentes tâches telles que le débruitage non-local, la détection de patches similaires où l'évaluation de qualité. Elle a montré des propriétés intéressantes pour chacune de ces applications. Ce travail ouvre de nombreuses perspectives comme l'extension à d'autres types de bruits blancs et une validation plus en profondeur de la mesure de qualité par une étude psychophysique. Enfin, beaucoup d'autres applications pourraient bénéficier de la mesure introduite, telles que le débruitage variationnel, la stéréovision ou la détection de mouvement.

Références

- [1] Dominique Brunet, Edward R Vrscay, and Zhou Wang. The use of residuals in image denoising. In *Image Analysis and Recognition*, pages 1–12. Springer, 2009.
- [2] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation*, 4(2):490–530, 2005.
- [3] Charles-Alban Deledalle, Loïc Denis, and Florence Tupin. How to compare noisy patches? patch similarity beyond gaussian noise. *International journal of computer vision*, 99(1):86–102, 2012.
- [4] Richard Dosselmann and Xue Dong Yang. A comprehensive assessment of the structural similarity index. *Signal, Image and Video Processing*, 5(1):81–91, 2011.
- [5] Jérôme Fehrenbach, Mila Nikolova, Gabriele Steidl, and Pierre Weiss. Bilevel image denoising using gaussianity tests. In *Scale Space and Variational Methods in Computer Vision*, pages 117–128. Springer, 2015.
- [6] Alessandro Lanza, Serena Morigi, and Fiorella Sgallari. Variational image restoration with constraints on noise whiteness. *Journal of Mathematical Imaging and Vision*, pages 1–17, 2014.
- [7] Alessandro Lanza, Serena Morigi, Fiorella Sgallari, and Anthony J Yezzi. Variational image denoising based on autocorrelation whiteness. *SIAM Journal on Imaging Sciences*, 6(4):1931–1955, 2013.
- [8] M Lebrun, A Buades, and JM Morel. A nonlocal bayesian image denoising algorithm. *SIAM Journal on Imaging Sciences*, 6(3):1665–1688, 2013.
- [9] Paul Riot, Andrés Almansa, Yann Gousseau, and Florence Tupin. Penalizing local correlations in the residual improves image denoising performance. In *Signal Processing Conference (EUSIPCO), 2016 24th European*, pages 1867–1871. IEEE, 2016.
- [10] Paul Riot, Andrés Almansa, Yann Gousseau, and Florence Tupin. A Correlation-based Dissimilarity Measure for Noisy Patches. working paper or preprint, March 2017.
- [11] Tolga Tasdizen. Principal neighborhood dictionaries for nonlocal means image denoising. *IEEE Transactions on Image Processing*, 18(12):2649–2660, 2009.
- [12] Tanja Teuber, Steffen Remmele, Jürgen Hesser, and Gabriele Steidl. Denoising by second order statistics. *Signal Processing*, 92(12):2837–2847, 2012.
- [13] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.