

# Mesures de similarité entre images dans le cadre des problèmes inverses

Sandrine ANTHOINE, Michel BARLAUD

Laboratoire I3S - (UNSA - CNRS), 2000 route des lucioles, BP.121, 06903 Sophia Antipolis Cedex, France  
anthoine@i3s.unice.fr, barlaud@i3s.unice.fr

**Résumé** – Nous proposons dans ce papier d’étudier les métriques de similarité entre images dans le cadre des problèmes inverses tels que la déconvolution et la séparation de sources. La métrique de similarité que nous proposons est basée sur les notions de théorie de l’information (distance de Kullback-Leibler) et est combinée avec une transformée en ondelettes. L’application visée est un problème de déconvolution et séparation de sources en astrophysique.

**Abstract** – We study in this paper similarity measures for images in the case of inverse problems such as deconvolution and source separation. The measure we propose is based on information theory (Kullback-Leibler distance) and is combined with wavelet transforms. The targeted application is a deconvolution and source separation problem in astrophysics.

## 1 Introduction :

Un problème inverse est la recherche d’un objet  $X$  à partir d’une observation  $Y=f(X)$  (par exemple recherche de l’original à partir d’une photo floue). Le problème est souvent mal-posé (pas d’unicité ou de continuité de la solution) et l’on doit donc le “régulariser” pour trouver une solution raisonnable. Il s’agit alors de trouver une estimation  $\tilde{X}$  qui générerait des observations proches des données  $f(\tilde{X})\approx Y$  (c’est l’adéquation aux données) et qui en même temps possède de bonnes propriétés - par exemple la netteté dans le cas de la photo - (on parle de régularité de la solution).

Les techniques actuelles les plus performantes exploitent l’idée de contraindre  $\tilde{X}$  à avoir de bonnes propriétés. Pour cela, on passe généralement de l’espace des pixels à un autre espace via une transformation (en ondelettes par exemple) qui permet de mieux concentrer l’information utile pour l’objet et donc de la séparer du reste. “Avoir de bonnes propriétés” se résume alors à avoir une représentation parcimonieuse dans l’espace transformé, un critère que l’on sait maintenant optimiser ([4, 6]).

La question de l’adéquation aux données, c’est-à-dire “Quand peut-on dire  $f(X)\approx Y$ ?” a reçu moins d’attention. La plupart des méthodes évaluent l’adéquation aux données par un terme du type  $\phi(Y-f(X))$  où  $\phi$  est en général une norme quadratique. Ceci correspond à modéliser les incertitudes sur l’observation par un bruit additif :  $Y=f(X)+N$ , où  $N$  désigne le bruit. Le cas gaussien i.i.d. conduit naturellement à  $\phi(\cdot)=\|\cdot\|_2^2$  qui est le plus couramment utilisée. Des travaux ponctuels considèrent un bruit de Poisson. Mais de manière générale, peu d’attention est portée à ce sujet.

L’utilisation en traitement d’images de termes d’adéquation aux données du type

$$\|Y-f(X)\|_2^2 = \sum_{i \in \{pixel\}} |Y(i)-f(X)(i)|^2$$

présente le gros désavantage qu’ils ne permettent pas de détecter la présence de structures spatiales dans le résiduel si celles-ci sont de faible énergie. Or même de faible énergie ces structures peuvent être cruciales.

Prenons l’exemple de l’estimation de cartes en astrophysique à partir de données multispectrales. Il s’agit d’estimer différents phénomènes physiques à partir d’observations floues et bruitées de mélanges de ces derniers, acquises à différentes longueurs d’onde [1]. Le modèle mathématique pour ce problème est le suivant : on recherche les  $M$  composantes  $x_1, \dots, x_M$  à partir de  $L$  observations  $y_1, \dots, y_L$  où

$$\forall l=1\dots L, \quad y_l = b_l * \left[ \sum_{m=1}^M a_{m,l} x_m \right] + n_l \quad (1)$$

avec  $a_{m,l}$  scalaire,  $b_l$  la psf à la longueur d’onde  $l$  et  $n_l$  du bruit (supposé gaussien). Fig. 1 donne un exemple de données multispectrales d’observation du fond cosmique. Dans ce cas particulier, une mauvaise estimation de la composante  $x_4$ , qui représente les points sources, a très peu d’incidence dans les termes d’adéquation quadratique du type  $\|y_l - b_l * [\sum_{m=1}^M a_{m,l} x_m]\|^2$  du fait que les points sources sont très compacts. En revanche, elle conduit aussi à une mauvaise estimation de  $x_2$ , la carte des amas de galaxies car ceux-ci sont aussi très compacts mais de moindre contribution dans les observations (voir [1] pour plus de détails).

Ainsi, de telles mesures de similarité sont inadaptées. On souhaite ici définir des mesures de similarité entre images qui permettent de s’assurer que toutes les structures spatiales présentes dans les observations sont expliquées. On ne peut évidemment pas prévoir quel type de telles structures apparaissent dans le résiduel  $Y-f(X)$  : dans le cas précédent, on pourrait voir des résidus de chaque composante  $x_m$  qui ont des structures spatiales différentes. On souhaite donc plutôt comparer directement les structures spatiales de  $Y$  et  $f(X)$ . Pour cela on se pro-

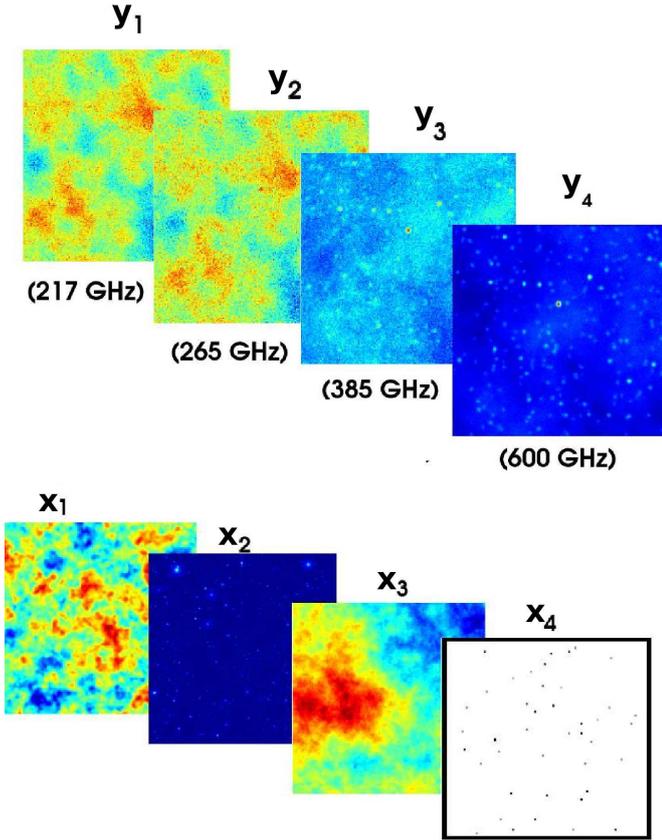


FIG. 1 – Un exemple de données multifréquence. Haut : observations, bas : composantes physiques à estimer.

pose d'une part d'utiliser la transformée en ondelettes - leurs propriétés de localisation permettant de détecter les structures spatiales - et d'autre part de mesurer la ressemblance par une distance de Kullback-Leibler (mesure en théorie de l'information) qui permet de comparer directement les propriétés statistiques des éléments à comparer sans passer par le résiduel.

La section 2 traite de la définition d'une telle mesure de similarité. La section suivante concerne sa mise en oeuvre pratique.

## 2 Mesure de similarité :

Dans cette section, nous définissons un critère de similarité entre deux images  $I_1$  et  $I_2$  dont le but est de détecter si des structures spatiales présentes sont similaires dans les deux images.

### 2.1 Détection de structures spatiales via les ondelettes

La transformée en ondelettes d'une image  $I$  est l'analyse de cette image par une famille de fonctions  $\{\psi_{j,k}\}_{j,k}$  constituée de translatées et dilatées d'une fonction  $\psi$  appelée l'ondelette mère. Du fait des propriétés de localisation en espace et en fréquence de l'ondelette mère, le coefficient d'ondelette  $w(I)_{j,k} = \langle \psi_{j,k}, I \rangle$  donne des informations sur le contenu l'image  $I$  aux alentours du point

$k$  et dans une bande de fréquence proche de l'échelle  $j$ . Si l'image considérée est relativement régulière (régions homogènes séparées par des discontinuités) alors la transformée en ondelette concentre la plupart de l'information spatio-fréquentielle de l'image (i.e. ses structures spatiales) en quelques coefficients de large amplitude (représentation parcimonieuse).

En première approximation, ces coefficients sont décorrélés ce qui donne lieu au traitement par seuillage des coefficients en ondelettes, très efficace en débruitage et compression d'images. Mais en réalité les coefficients d'ondelettes à différentes échelles sont corrélés. Par exemple la présence d'une discontinuité le long d'une courbe se traduit en un point  $k_0$  de cette courbe par des coefficients grands à toutes les échelles ( $\forall j, w(I)_{j,k_0}$  grand). Différents modèles de dépendance entre coefficients ont été proposés pour améliorer la description de structures spatiales fines [9, 10]. En particulier on note une dépendance entre un coefficient d'ondelette  $w(I)_{j,k}$  et ses plus proches voisins en échelle ( $w(I)_{j-1,k}$ ) ou en espace ( $w(I)_{j,k\pm 1_x}, w(I)_{j,k\pm 1_y}$ ).

Nous avons montré dans [1], que les statistiques de vecteurs de coefficients en ondelettes de la forme

$$\mathbf{w}(I)_{j,k} = (w(I)_{j,k}, w(I)_{j-1,k}, w(I)_{j,k\pm 1_x}, w(I)_{j,k\pm 1_y}) \quad (2)$$

permettent de caractériser les structures spatiales de nature très différentes (celles présentes dans chacun des phénomènes astrophysiques recherchés dans Eq.(1)). Pour cela, il suffit d'ajuster pour chaque phénomène un modèle de mélange de gaussiennes pour décrire la probabilité jointe de ces vecteurs. Dans le cas présent, on ne sait pas quels types de structures seront présentes dans les observations, on ne peut donc pas fixer de modèle. Néanmoins on espère que la distribution de ces vecteurs sera représentative des structures spatiales présentes dans l'image. On souhaite donc définir une mesure prenant en compte la probabilité jointe des vecteurs voisinages d'ondelettes  $\mathbf{w}(I)_{j,k}$  de façon non-paramétrique.

### 2.2 Critères entropiques

Comme vu précédemment, des critères quadratiques pénalisent les structures de faible énergie et correspondent à un a priori gaussien sur le résiduel. De même, utiliser d'autres normes  $L_p$  ou des combinaisons de celles-ci sur le résiduel reviendrait à paramétrer la distribution de ce dernier. Compte-tenu de la variabilité des structures spatiales que l'on peut rencontrer dans le résiduel, le choix d'une paramétrisation serait difficile à justifier. Nous proposons d'introduire des métriques de similarité valables sans paramétrisation de la distribution des voisinages : les métriques issues de la théorie de l'information telles que l'entropie des voisinages du résiduel, l'information mutuelle ou la distance de Kullback-Leibler entre les distributions des voisinages de coefficients d'ondelettes des deux images.

Supposons qu'un voisinage  $\mathbf{w}(I)_{j,k}$  contienne  $d$  coefficients. La distribution (ou densité de probabilité) de l'ensemble des voisinages de l'image  $I$  est notée  $p_{\mathbf{w}(I)}$  et vérifie :  $p_{\mathbf{w}(I)} : \mathbb{R}^d \rightarrow \mathbb{R}^+$  et  $\int_{\mathbb{R}^d} p_{\mathbf{w}(I)}(x) dx = 1$ .

L'entropie différentielle de Shannon de  $p_{\mathbf{w}(I)}$  est définie

par

$$H(p_{\mathbf{w}(I)}) = - \int_{\mathbb{R}^d} p_{\mathbf{w}(I)}(x) \log p_{\mathbf{w}(I)}(x) dx. \quad (3)$$

Elle mesure la quantité d'information contenue dans cette distribution. La distance de Kullback-Leibler :

$$D_{KL}(p_{\mathbf{w}(I_1)} || p_{\mathbf{w}(I_2)}) = \int_{\mathbb{R}^d} p_{\mathbf{w}(I_1)}(x) \log \frac{p_{\mathbf{w}(I_1)}(x)}{p_{\mathbf{w}(I_2)}(x)} dx \quad (4)$$

est une mesure de ressemblance entre les distributions  $p_{\mathbf{w}(I_1)}$  et  $p_{\mathbf{w}(I_2)}$  basée sur ce concept d'information. Notons que si on définit l'entropie croisée par :

$$H_x(p_{\mathbf{w}(I_1)}, p_{\mathbf{w}(I_2)}) = - \int_{\mathbb{R}^d} p_{\mathbf{w}(I_1)}(x) \log p_{\mathbf{w}(I_2)}(x) dx. \quad (5)$$

alors la distance de Kullback-Leibler s'exprime comme une différence d'entropies :

$$D_{KL}(p_{\mathbf{w}(I_1)} || p_{\mathbf{w}(I_2)}) = H_x(p_{\mathbf{w}(I_1)}, p_{\mathbf{w}(I_2)}) - H(p_{\mathbf{w}(I_1)}) \quad (6)$$

L'utilisation de ces mesures sur les distributions de l'intensité des pixels d'une image donne de bons résultats dans le domaine de la segmentation, ainsi que du réalignement d'images [8, 2, 11]. Une distance de Kullback dans l'espace des ondelettes a aussi été proposée pour le problème d'indexation dans [5, 12]. Plus précisément, dans ces deux articles, les auteurs paramètrent la distribution des coefficients d'ondelettes pour chaque échelle  $j$  par une gaussienne généralisée et somment les distances de Kullback obtenues à chaque échelle pour obtenir la similarité entre deux images.

Nous proposons d'étudier des mesures similaires pour déterminer la ressemblance entre deux images mais avec deux différences majeures. D'une part, comme vu à la section 1, les coefficients d'ondelettes à différentes échelles ne sont pas indépendants. Or sommer les distances de Kullback à chaque échelle correspond à les supposer cette indépendance. Nous considérons donc les entropies jointes de coefficients, en particulier, celle des voisinages précédemment décrits. D'autre part, nous ne paramétrons pas les distributions en jeu.

Nous proposons la mesure la similarité entre les images  $I_1$  et  $I_2$  de la manière suivante :

$$S(I_1, I_2) = \sum_j \alpha_j D_{KL}(p_{\mathbf{w}_j(I_1)} || p_{\mathbf{w}_j(I_2)}) \quad (7)$$

où

- $p_{\mathbf{w}_j(I_i)}$  est la distribution non-paramétrique des voisinages de coefficients d'ondelette de l'image  $I_i$  à l'échelle  $j$ .
- les  $\alpha_j > 0$  sont des poids de normalisation à fixer selon la redondance du système d'ondelettes utilisé.

### 3 Mise en pratique

#### Calcul du critère : estimation par les plus proches voisins

Nous expliquons dans cette section la méthode que nous utilisons pour calculer le critère proposé à l'Eq.(7).

Dans les cas de [5] et [12], la seule estimation nécessaire est celle des paramètres des gaussiennes généralisées représentant les distributions des coefficients (à chaque échelle et pour chaque image). En effet, la distance de Kullback entre deux gaussiennes généralisées peut être calculée sous forme analytique. Ici, l'estimation des distances de Kullback mises en jeu est plus ardue d'une part du fait que les distributions ne sont pas paramétriques et d'autre part car l'espace des échantillons est de grande dimension.

La méthode standard pour estimer des quantités entropiques dans le cas non-paramétrique consiste à estimer les densités de probabilité intervenant puis à calculer l'intégrale correspondant. Pour estimer les distributions, les méthodes de Parzen qui estiment un histogramme lissé élaboré à partir des données au moyen d'un noyau de lissage sont les plus techniques actuellement couramment utilisées. Elles souffrent néanmoins de certains inconvénients parmi lesquels on peut citer : le problème du choix du noyau, le fait que la taille de fenêtre de lissage soit fixe (ce qui pose un problème d'estimation aux points où l'on n'a pas d'échantillons) et la difficulté d'étendre la méthode aux variables aléatoires de dimension supérieure à un.

Nous utilisons donc méthode différente pour estimer les distances de Kullback entre distributions des voisinages de coefficients qui sont des vecteurs de  $\mathbb{R}^d$ . Il s'agit de la méthode de calcul de termes entropiques dite des  $k$  plus proches voisins ( $k$  nearest neighbors en anglais ou  $k$ NN). Cette méthode est duale des méthodes de Parzen et permet d'estimer directement les termes entropiques sans estimer les distributions sous-jacentes et ceci même en grande dimension.

Supposons que  $\mathcal{V}$  est l'ensemble des  $N_{\mathcal{V}}$  échantillons  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{N_{\mathcal{V}}}$  de loi  $p_{\mathcal{V}}$  dont on cherche l'entropie ( $\mathbf{w}_i \in \mathbb{R}^d, \forall i$ ). Fixons  $k$ , un entier non nul. On note  $v_d$  le volume de la boule unité dans  $\mathbb{R}^d$ ,  $\psi$  la fonction digamma.

On définit  $\rho_{k, \mathcal{V}}(s)$  la distance du point  $s \in \mathbb{R}^d$  à son  $k$ -ième plus proche voisin dans l'ensemble  $\mathcal{V} - \{s\}$ . On note  $\mu_{\mathcal{V}}(g)$  la moyenne de  $g$  sur  $\mathcal{V}$  :  $\mu_{\mathcal{V}}(g) = \frac{1}{N_{\mathcal{V}}} \sum_{n=1}^{N_{\mathcal{V}}} g(\mathbf{w}_n)$ .

La méthode  $k$ NN se base sur la remarque que la densité de probabilité au point  $s$  est inversement proportionnelle à la taille de la boule contenant les  $k$  échantillons les plus proches de  $s$  dans  $\mathcal{V}$  soit :

$$p_{\mathcal{V}}(s) = \frac{k}{v_d \rho_{k, \mathcal{V}}^d(s)}$$

La méthode  $k$ NN est donc duale de celle de Parzen : au lieu de fixer une fenêtre d'intégration puis estimer la fréquence des échantillons dans cette fenêtre pour estimer la densité, on estime la densité, la taille de la fenêtre s'adaptant à chaque point pour capturer ses  $k$  plus proches voisins dans les échantillons.

Un estimateur non biaisé de l'entropie  $\widehat{H}(p_{\mathcal{V}})$  est ([7]) :

$$\widehat{H}(p_{\mathcal{V}}) = \log[(N_{\mathcal{V}} - 1)v_d] - \psi(k) + d\mu_{\mathcal{V}}(\log[\rho_{k, \mathcal{V}}]) \quad (8)$$

et un estimateur de l'entropie croisée est ([3]) :

$$\widehat{H}_x(p_{\mathcal{V}^1}, p_{\mathcal{V}^2}) = \log[(N_{\mathcal{V}^2})v_d] - \psi(k) + d\mu_{\mathcal{V}^2}(\log[\rho_{k, \mathcal{V}^1}]) \quad (9)$$

et enfin on en déduit un estimateur de la distance de

Kullback-Leibler ([3]) :

$$\widehat{D}_{KL}(p_{\mathcal{V}^1} || p_{\mathcal{V}^2}) = \log\left[\frac{N_{\mathcal{V}^2}}{N_{\mathcal{V}^1}-1}\right] + d\mu_{\mathcal{V}^2}(\log[\rho_{k,\mathcal{V}^1}]) - d\mu_{\mathcal{V}^1}(\log[\rho_{k,\mathcal{V}^1}]) \quad (10)$$

Cet estimateur de la distance de Kullback-Leibler peut se calculer relativement rapidement et ceci quelque soit la dimension des échantillons. Il est de plus robuste au choix du nombre de plus proches voisins  $k$ .

## 4 Conclusion

Dans ce papier, nous avons proposé une mesure de similarité entre images qui combine les concepts de théorie de l'information et la transformée en ondelettes. Il s'agit de sommer les distances de Kullback à chaque échelle de la distribution de vecteurs appelés voisinages de coefficients en ondelettes. Les voisinages contenant les coefficients à des locations spatiales mais aussi à des échelles proches, ils capturent les dépendances spatiale et inter-échelle des coefficients ce qui permet de détecter des structures spatiales fines. La distance de Kullback sur ces vecteurs est estimée de manière non-paramétrique malgré leur dimension élevée, ceci grâce aux estimateurs entropiques des plus proches voisins.

Dans le cas de convolution simple et avec des voisinages réduits, les résultats préliminaires concluant nous amènent à poursuivre nos travaux d'une part en élargissant les voisinages et d'autre part en étudiant la nécessité d'incorporer des poids modulant la contribution à chaque échelle.

## Références

- [1] S Anthoine. *Different Wavelet-based Approaches for the Separation of Noisy and Blurred Mixtures of Components. Application to Astrophysical Data*. Ph.d. dissertation, Princeton University, 2005.
- [2] S. Boltz, E. Debreuve, and M. Barlaud. A high-dimensional framework for joint color-spatial segmentation. In *ICIP*, San Antonio, USA, September 2007.
- [3] S. Boltz, E. Debreuve, and M. Barlaud. High-dimensional kullback-leibler distance for region-of-interest tracking : Application to combining a soft geometric constraint with radiometry. In *CVPR*, Minneapolis, USA, June 2007.
- [4] I Daubechies, M Defrise, and C De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, 57 :1413–1541, 2004.
- [5] M N. Do and M Vetterli. Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. *IEEE Trans. Image Process.*, 11(2) :146–158, 2002.
- [6] D Donoho and M Elad. Maximal sparsity representation via  $l^1$  minimization. *Proc. Nat. Aca. Sci.*, 100 :2197–2202, 2003.
- [7] M Goria, N Leonenko, V Mergel, and P Novi Inverardi. A new class of random vector entropy estimators and its applications in testing statistical hypotheses. *Journal of Nonparametric Statistics*, 17 :277–298, 2005.
- [8] A. Herbulot, S. Boltz, E. Debreuve, M. Barlaud, and G. Aubert. Space-time segmentation based on a joint entropy with estimation of nonparametric distributions. In *SSVM*, Ischia, Italy, May 2007.
- [9] J Portilla, V Strela, M Wainwright, and E P Simoncelli. Image denoising using a scale mixture of Gaussians in the wavelet domain. *IEEE Trans Image Processing*, 12(11) :1338–1351, 2003.
- [10] J K Romberg, H Choi, and R G Baraniuk. Bayesian tree-structured image modeling using wavelet-domain hidden markov models. *IEEE Transactions on Image Processing*, 10(7) :1056–1068, 2001.
- [11] M R Sabuncu and P J Ramadge. Spatial information in entropy-based image registration. In *WBIR*, pages 132–141, 2003.
- [12] Z Wang, G Wu, H R Sheikh, E P Simoncelli, E-H Yang, and A C Bovik. Quality-aware images. *IEEE Transactions on Image Processing*, 15(6) :1680–1689, 2006.