

Apprentissage de dictionnaires parcimonieux adaptés pour la séparation d'images

Gabriel PEYRÉ¹, Jalal FADILI², Jean-Luc STARCK³

¹CEREMADE / Université Paris Dauphine

²GREYC, ENSICAEN

³Service d'Astrophysique CEA/Saclay

gabriel.peyre@ceremade.dauphine.fr, jalal.fadili@greyc.ensicaen.fr, jstarck@cea.fr

Résumé – Cet article propose une nouvelle méthode pour séparer une image en une superposition linéaire de composantes morphologiques ayant des structures caractéristiques. Pour chaque composante, un dictionnaire adapté est appris depuis des images d'exemples. Chaque composante est caractérisée par une décomposition parcimonieuse dans le dictionnaire associé. L'algorithme d'analyse en composantes morphologiques permet de résoudre de façon itérative le problème d'optimisation correspondant à la recherche des composantes. L'utilisation de dictionnaires adaptés permet de résoudre certains problèmes inhérents à l'utilisation de dictionnaires fixés à l'avance. Les résultats numériques montrent la nécessité de cette adaptivité pour capturer des motifs texturés complexes.

Abstract – This article proposes a new method for image separation into a linear combination of morphological components. Each component contains specific structures such as a cartoon sketch or various textural patterns. For each of these components, a dictionary is learned from a set of exemplar images. Each component is characterized by a sparse expansion in the corresponding dictionary. The morphological component analysis algorithm allows to solve iteratively the optimization problem of this separation under sparsity constraints. Using adapted dictionary learned from data allows to circumvent some difficulties faced by fixed dictionaries. Numerical results clearly demonstrate that this adaptivity is indeed crucial to capture complex texture patterns.

1 Décomposition variationnelle et parcimonieuse

1.1 Décomposition d'images

Le problème de la séparation d'image consiste à décomposer une image $f \in \mathbb{R}^N$ de N pixels en différentes composantes $f = \sum_s u_s$ où chaque u_s capture un certain type de structures. Cette séparation peut se formaliser dans un cadre variationnel comme l'optimisation de $(u_s)_{s=1}^S$ pour résoudre

$$\min_{u_1, \dots, u_S} \frac{1}{2} \left\| f - \sum_{s=1}^S u_s \right\|_{\ell^2}^2 + \lambda \sum_{s=1}^S E_s(u_s), \quad (1)$$

où chaque énergie $E_s : \mathbb{R}^N \mapsto \mathbb{R}^+$ favorise les images ayant un type spécifique de structures.

Débruitage. Le problème le plus simple correspond au débruitage où l'on cherche une seule composante $f \approx u_1$ qui capture les structures telles que les contours de l'image. Le résiduel $f - u_1$ est seulement supposé d'énergie bornée et le paramètre λ doit être fixé en fonction de ce niveau de bruit. Le modèle de Rudin-Osher-Fatemi [ROF92] correspond ainsi à l'utilisation de

$$E_1(u_1) \stackrel{\text{def.}}{=} \|u_1\|_{\text{TV}} \stackrel{\text{def.}}{=} \int |\nabla_x u_1| dx,$$

où la norme de variation totale $\|\cdot\|_{\text{TV}}$ impose aux contours de u_1 d'avoir un faible périmètre et retire de u_1 les oscillations dues au bruit et à la texture.

Décomposition structure/texture. Yves Meyer étend ce modèle en incluant une composante additionnelle $f = u_1 + u_2$ où u_2 capture les structures oscillantes des textures [Mey01]. Ceci correspond à l'utilisation des énergies

$$E_1(u_1) \stackrel{\text{def.}}{=} \|u_1\|_{\text{TV}} \quad (2)$$

$$\text{et } E_2(u_2) \stackrel{\text{def.}}{=} \|u_2\|_{\text{G}} = \min_{u_1 = \text{div}(g)} \|g\|_{\ell^\infty}. \quad (3)$$

La norme $\|\cdot\|_{\text{G}}$ est proche de la norme duale de la variation totale et favorise donc les fonctions très oscillantes. Ce modèle est implémenté en pratique à l'aide d'algorithmes d'optimisation itératifs comme par exemple celui développé par Aujol et al. [AABFC05].

1.2 Modèles parcimonieux dans des dictionnaires globaux.

Une alternative à ces énergies exploitant des espaces fonctionnels consiste à exploiter la décomposition de chaque composante $u_s = D_s x_s$ dans un dictionnaire $D_s \in \mathbb{R}^{N \times m_s}$, $m_s \geq N$. Le degré de parcimonie d'une telle décomposition peut se mesurer à l'aide de la norme ℓ^1

$$\|x_s\|_{\ell^1} \stackrel{\text{def.}}{=} \sum_j |x_s[j]|.$$

Pour des dictionnaires redondants $m_s > N$, un tel jeu de coefficients $x_s \in \mathbb{R}^{m_s}$ n'est pas unique, mais on peut définir une énergie de décomposition comme

$$E_s(g) \stackrel{\text{def.}}{=} \min_{x \in \mathbb{R}^{m_s}} \frac{\mu}{2} \|g - D_s x\|_{\ell^2}^2 + \|x\|_{\ell^1}, \quad (4)$$

où le produit $\lambda\mu$ pondère l'erreur de reconstruction $D_s x_s \approx u_s$ avec l'erreur de décomposition $f \approx \sum u_s$.

Une telle énergie parcimonieuse nécessite l'utilisation d'un dictionnaire D_s adéquat pour capturer efficacement des structures intéressantes de la composante u_s . De nombreux dictionnaires ont été proposés pour capturer différents types de caractéristiques des images naturelles.

- Les ondelettes [Mal98] constituent un outil important en compression d'images et permet de capturer de façon optimale les images à variations bornées. La norme ℓ^1 des coefficients d'ondelettes $\|D_W^T v\|_{\ell^1}$, où D_W est une trame ajustée (tight frame) d'ondelettes invariante par translation, s'apparente à la norme de variation totale $\|v\|_{TV}$.
- Pour les images de type dessin animé, qui sont régulières à l'extérieur de contours réguliers, les ondelettes ne constituent pas une représentation optimale. La trame ajustée de curvelets [CD04], introduite par Candès et Donoho permet de capturer efficacement les contours des images.
- Les textures localement oscillantes et stationnaires peuvent être capturés par un dictionnaire redondant D_C de cosinus locaux [Mal98], qui constitue aussi une trame ajustée. La norme $\|D_C^T v\|_{\ell^1}$ d'une image dans cette trame s'apparente à la norme oscillante de Meyer $\|v\|_G$.

Dans cet article, nous allons utiliser les dictionnaires D_W et D_C correspondant à des trames ajustées pour capturer respectivement la composante à variations bornées et la partie oscillante des images. Ces dictionnaires ont été utilisés entre autres par Starck et al. [SED04], ce qui donne une décomposition semblable à celle de Meyer (2). La suite de cet article explique comment adjoindre à ces dictionnaires fixes des dictionnaires appris pour capturer des structures complexes.

1.3 Modèles parcimonieux dans des dictionnaires locaux.

Cet article utilise non seulement des dictionnaires $D_s \in \mathbb{R}^{N \times m_s}$ globaux sur l'ensemble des N pixels de l'image, mais également des dictionnaires locaux pour capturer des structures fines des textures. Un tel dictionnaire $D_s \in \mathbb{R}^{n \times m_s}$ est utilisable pour représenter les imagerie (patches) $R_k(g) \in \mathbb{R}^n$ de $n \stackrel{\text{def}}{=} \tau \times \tau$ pixels extraites d'une image g ,

$$\forall 0 \leq k < n, \quad R_k(g)[i] = g(k_1 + i_1, k_2 + i_2),$$

où $k = k_1 + \tau k_2$ avec $0 \leq k_i < \sqrt{N}$ est l'index du pixel (k_1, k_2) .

De façon similaire à l'énergie (4) associée à un dictionnaire global, on peut définir une énergie $E_s(g)$ associé à un dictionnaire local D_s . Cette énergie permet de contrôler la parcimonie des décompositions de toutes les imagerie $R_k(g)$ dans D_s . L'énergie $E_s(g)$ est définie comme

$$\min_{(x_k)_k \in \mathbb{R}^{m_s \times n}} \frac{\mu}{2N} \sum_{k=0}^{N-1} \|R_k(g) - D_s x_k\|_{\ell^2}^2 + \|x_k\|_{\ell^1}. \quad (5)$$

Dans cette énergie, chaque x_k correspond aux coefficients de la décomposition de l'imagerie $R_k(g)$ à l'aide du dictionnaire D_s .

2 Analyse en Composante Morphologiques

L'algorithme d'analyse en composantes morphologiques (MCA) [SED04] permet de résoudre de façon itérative le problème de séparation variationnelle (1) dans le cas d'énergies parcimonieuses E_s comme défini à l'équation (4). La caractéristique principal de cet algorithme est que le paramètre λ de régularisation décroît de façon linéaire à chaque itération pour atteindre une valeur λ_{\min} finale qui correspond au niveau de régularisation souhaité, et vaut $\lambda_{\min} = 0$ dans le cas d'une séparation sans bruit. Cette décroissance de λ permet d'accélérer la convergence de l'algorithme et permet de calculer le minimiseur à chaque étape à l'aide d'un seuillage non-linéaire.

Pour la décomposition d'une image en sa partie géométrique et sa partie texturée, l'approche originale [SED04] utilise des dictionnaires fixé d'ondelettes D_W (voire de curvelets pour améliorer l'extraction des contours) ainsi que de cosinus locaux D_C . Cet article étend l'algorithme MCA au cas d'énergies E_s issue de dictionnaires locaux D_s comme défini à l'équation (5).

L'algorithme MCA opère en optimisant successivement sur chaque composante u_s en conservant les autres composantes $(u_\ell)_{\ell \neq s}$ constantes. Ceci demande, à chaque étape, la minimisation de

$$\min_{v \in \mathbb{R}^n} \frac{1}{2} \|r_s - v\|_{\ell^2}^2 + \lambda E_s(v) \quad \text{où} \quad r_s = f - \sum_{\ell \neq s} u_s.$$

Cette minimisation s'effectue de façon approchée en deux étapes, qui diffèrent selon que l'on considère un dictionnaire global ou local

- **(Calcul des coefficients) Dictionnaire global** : Le calcul des coefficients $x \in \mathbb{R}^{m_s}$ s'effectue par seuillage des coefficients du résidu r_s

$$x^* = S_\lambda(D_s^T r_s) \quad \text{où} \quad S_t(x) = \{s_t(x_j)\}_j \quad (6)$$

$$\text{avec} \quad s_t(a) \stackrel{\text{def}}{=} \begin{cases} |a| + \text{sign}(a)t & \text{si } |a| > t, \\ 0 & \text{si } |a| \leq t. \end{cases} \quad (7)$$

Dictionnaire local : Les coefficients $x^* = (x_k^*)_k$ de la décomposition de chaque sous-image $R_k(r_s)$ se calculent en minimisant

$$x_k^* = \underset{x}{\text{argmin}} \frac{1}{2} \|R_k(r_s) - D_s x\|_{\ell^2}^2 + \lambda \|x\|_{\ell^1}.$$

- **(Reconstruction)** La composante u_s est mise à jour à partir des coefficients parcimonieux calculés.

Dictionnaire global : on pose $u_s \stackrel{\text{def}}{=} D_s x^*$.

Dictionnaire local : à cause du chevauchement des sous-image $R_k(D_s x_k)$, la reconstruction nécessite un moyen-nage

$$u_s(a) = \frac{1}{\tau^2} \sum_{|k-a| \leq \tau} R_k(D_s x_k)(a - k). \quad (8)$$

L'algorithme MCA, résumé dans le pseudo-code 1, effectue la mise à jour successive de chaque u_s . Le paramètre de régularisation λ décroît à chaque itération.

Initialisation : $\forall s, u_s = 0, \lambda = \lambda_{\max}$.
Boucle : Tant que $\lambda > \lambda_{\min}$,
 – Pour chaque composante s , le résidu est calculé :
 $r_s = f - \sum_{\ell \neq s} u_\ell$.
 – (*Décomposition*) Calcul des coefficients x^* de la décomposition parcimonieuse de r_s (dictionnaire global, équation (6)) où de chaque imagerie $(R_k(r_s))_k$ (dictionnaire local, équation (2)).
 – (*Reconstruction*) Calcul de u_s à l'aide $u_s \stackrel{\text{def}}{=} D_s x^*$ (dictionnaire global) où par moyennage (dictionnaire local, équation (8)).
 – (*Mise à jour du seuil*) $\lambda \leftarrow \lambda - \delta$.

Table 1: Pseudo-code pour l'analyse en composantes morphologiques.

3 Apprentissage de dictionnaires

L'énergie (5) définie sur des sous-images de $n = \tau \times \tau$ pixels nécessite l'utilisation d'un dictionnaire D_s adapté aux structures locales des textures. Pour obtenir de tel dictionnaires, Olshausen et Field [OF96] ont proposé d'apprendre la matrice D_s pour représenter de façon optimale un ensemble $Y = (y^k)_{k=1}^p \in \mathbb{R}^{n \times p}$ de $p \gg n$ imagettes. Ils ont montré que lorsque ces imagettes sont extraites d'un ensemble d'images naturelles, le dictionnaire obtenu est similaire au dictionnaire D_W des ondelettes redondantes.

De façon à représenter efficacement des motifs texturés, nous employons une stratégie d'apprentissage similaire, en utilisant uniquement des exemples y_k issus d'une texture homogène. Le dictionnaire D_s ainsi appris est optimisé pour représenter les motifs structurés de cette texture homogène et la composante morphologique u_s de la décomposition (1) contient ces motifs. Pour apprendre ce dictionnaire D_s , nous utilisons l'algorithme K-SVD de Aharon et al. [AEB06]. Cet algorithme cherche à minimiser l'énergie E_s de l'équation (5) en modifiant à la fois les coefficients $(x_k)_k$ et le dictionnaire D_s . Ceci correspond à l'optimisation sur l'ensemble des dictionnaires $D \in \mathbb{R}^{n \times m_s}$ de

$$\min_{(x_k)_k \in \mathbb{R}^{m_s \times p}} \sum_{k=1}^p \frac{\mu}{2} \|y_k - Dx_k\|_{\ell^2}^2 + \|x_k\|_{\ell^1}.$$

sous la contrainte additionnelle de normalisation $\|d_j\|_{\ell^2} = 1$ des colonnes des D considérés.

Cette optimisation est non-linéaire et non-convexe et l'algorithme K-SVD (5) permet de trouver un minima local en alternant sur le calcul des coefficients $(x_k)_k$ optimaux et du dictionnaire D . Le pseudo-code 2 détaille l'enchaînement de ces deux étapes. La figure 1 (d) montre un dictionnaire appris. On observe que les atomes d_k capturent les caractéristiques de la texture utilisée comme exemple montrée figure 1 (a).

Initialisation : Le dictionnaire est initialisé avec une transformée en cosinus $D = D_C$ et $x_k \leftarrow D^T y_k$.
Boucle : Tant que D n'a pas convergé,
 – D fixé : les coefficients (x_k) sont mis à jour pour chaque exemple x_k en minimisant

$$x_k \leftarrow \operatorname{argmin}_x \frac{\mu}{2} \|y_k - Dx\|_{\ell^2}^2 + \|x\|_{\ell^1}.$$

Cette optimisation peut être résolue de façon approchée par matching pursuit, voir [Mal98].

– $(x_k)_k$ fixé : pour chaque atome d_j , on note $I_j = \{k \setminus x_k[j] \neq 0\}$ l'ensemble des exemples qui utilisent l'atome d_k . Cet atome d_j est mis à jour en minimisant

$$d_j \leftarrow \operatorname{argmin}_g \min_x \sum_{k \in I_j} \|\tilde{y}_k - x[k]g\|_{\ell^2}, \quad \tilde{y}_k = y_k - \sum_{\ell \neq j} x_k[\ell]d_\ell.$$

Cette minimisation est équivalente à une approximation de rang 1 de la matrice contenant les signaux \tilde{y}_k pour $k \in I_j$, ce qui peut être résolu par une décomposition en valeurs singulières (SVD).

Table 2: Pseudo-code pour l'algorithme K-SVD.

4 Application à la décomposition d'image

Dans une première application, nous avons utilisé un dictionnaire $D_1 = D_W$ de transformée en ondelettes redondantes pour capturer la partie géométrique de l'image f , et un dictionnaire D_2 appris. Ce dictionnaire est calculé selon la méthode de la section 3 à partir de la partie d'une texture connue f_0 . L'algorithme MCA décrit à la section 2 est appliqué sur une image $f = u_1 + u_2$ où u_2 est une texture visuellement similaire à f_0 . La figure 1 compare la décomposition (e) utilisant un dictionnaire de cosinus local comme proposé par [SED04] et la décomposition (f) obtenue avec le dictionnaire D_2 appris.

Dans une deuxième application, on calcule une séparation $f = \sum_{i=1}^4 u_i$ à l'aide d'un dictionnaire $D_1 = D_W$ d'ondelettes invariantes par translation et de dictionnaires appris $\{D_i\}_{i=2}^4$. Chacun de ces dictionnaire D_2, D_3, D_4 est appris à partir de sous-images y^k extraites autour d'un pixel désigné par l'utilisateur, voir figure 2. La figure 3 montre un autre exemple de décomposition avec une seule composante texturée.

Références

- [AABFC05] J. F. AUJOL, G. AUBERT, L. BLANC-FERAUD et A. CHAMBOLLE : Image decomposition into a bounded variation component and an oscillating component. *Journal of Math. Im. and Vision*, 22(1):71–88, janvier 2005.
- [AEB06] M. AHARON, M. ELAD et A.M. BRUCKSTEIN : The K-SVD : An algorithm for designing of overcomplete dictionaries for sparse representation. *IEEE Trans. On Signal Processing (to appear)*, 2006.

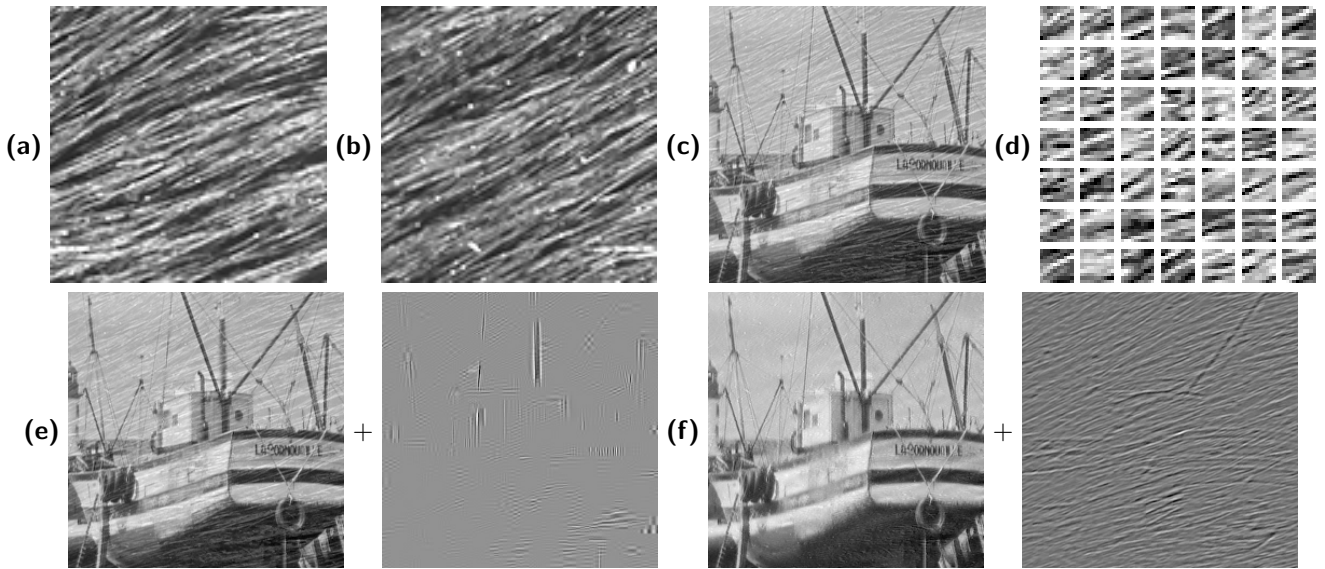


FIG. 1 – (a) Texture connue f_0 , (b) texture à extraire u_2 , (c) image à décomposer $f = u_1 + u_2$, (d) le dictionnaire D_2 appris, (e) décomposition à l'aide d'un dictionnaire d'ondelettes invariantes $D_1 = D_W$ et d'un dictionnaire de cosinus local $D_2 = D_C$, (f) décomposition à l'aide d'un dictionnaire d'ondelettes invariantes $D_1 = D_W$ et du dictionnaire D_2 appris.

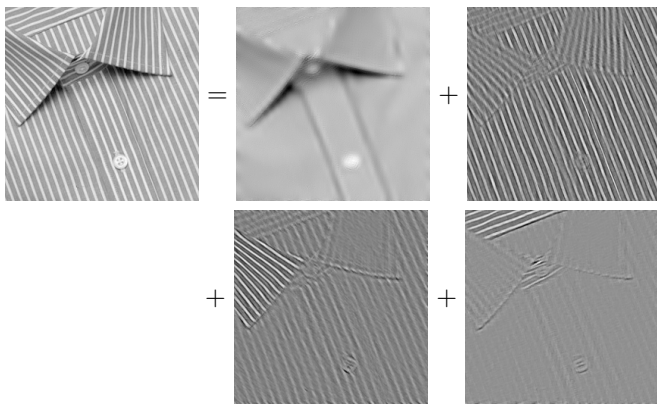


FIG. 2 – Décomposition à l'aide d'un dictionnaire d'ondelettes invariantes et de 3 dictionnaires appris.

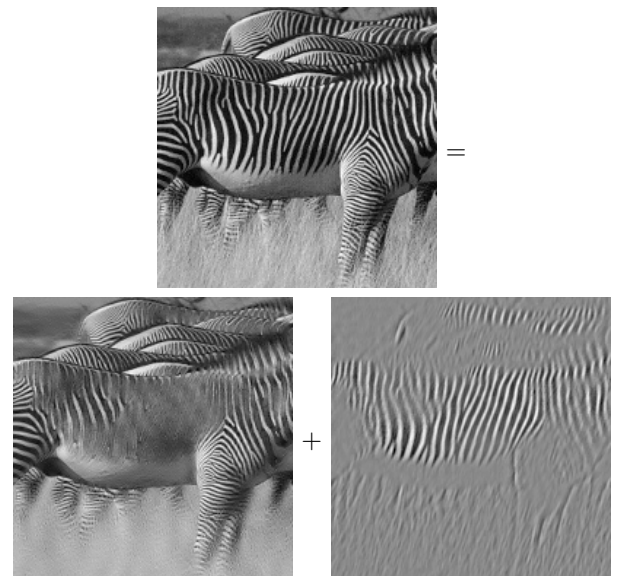


FIG. 3 – Exemple de décomposition à l'aide d'un dictionnaire appris à partir de sous-images extraites de l'image de départ.

- [CD04] E. CANDÈS et D. DONOHO : New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities. *Comm. Pure Appl. Math.*, 57(2):219–266, 2004.
- [Mal98] S. MALLAT : *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, 1998.
- [Mey01] Y. MEYER : *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*. American Mathematical Society, Boston, MA, USA, 2001.
- [OF96] B. A. OLSHAUSEN et D. J. FIELD : Emergence of simple-cell receptive-field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, juin 1996.
- [ROF92] L. I. RUDIN, S. OSHER et E. FATEMI : Non-linear total variation based noise removal algorithms. *Phys. D*, 60(1-4):259–268, 1992.

- [SED04] J.-L. STARCK, M. ELAD et D.L. DONOHO : Redundant multiscale transforms and their application for morphological component analysis. *Advances in Imaging and Electron Physics*, 132, 2004.