

Une approche multi-échelle de super-résolution à partir de séquences d'images

Fabien DEKEYSER, Patrick BOUTHEMY et Patrick PÉREZ

IRISA/INRIA
Campus de Beaulieu
35042 Rennes Cedex
prenom.nom@irisa.fr

Résumé – Nous considérons le problème de la reconstruction d'une image de haute résolution à partir d'une séquence d'images de basse résolution et de la mesure du mouvement entre images. Il est posé comme un problème inverse et est résolu à l'aide d'une méthode itérative appelée algorithme de Kaczmarz généralisé. Nous décrivons une implantation à faible coût de calcul basée sur l'estimation d'un modèle paramétrique 2D du mouvement global. Nous proposons également deux améliorations de cet algorithme. La première est l'utilisation d'une carte de détection du mouvement permettant de gérer les régions où le modèle de mouvement global n'est pas valide. La seconde est une approche multi-échelle permettant d'améliorer la qualité de l'estimation de mouvement au fur et à mesure du traitement de super-résolution.

Abstract – The reconstruction of a high resolution image from a low resolution image sequence using motion information is considered. This issue is formulated as an inverse problem and is solved using an iterative method called generalized Kaczmarz algorithm. We describe a low computational cost implementation of the method based on the estimation of a 2D parametric motion model. We also propose two improvements to this algorithm. The first one is the introduction of a motion detection map to deal with areas where the global motion model is not valid. The second one is a temporal multiscale approach allowing us to improve motion estimation within the super-resolution process.

1 Introduction

Lors de l'acquisition d'une image, le système optique se comporte plus ou moins comme un filtre passe-bas. Certains détails dans les hautes fréquences de l'image sont perdus ou atténués. Malgré cela, les capteurs numériques échantillonnent généralement les images en ne respectant pas le critère de Nyquist. En effet, de hautes fréquences spatiales (correspondant aux contours) restent présentes dans l'image. Les images sont alors entachées d'erreurs de recouvrement du spectre (aliasage).

La super-résolution consiste en l'obtention d'images de résolution plus grande que celle fournie par le capteur utilisé. L'objectif visé est de sur-échantillonner l'image tout en réduisant les erreurs d'aliasage (et éventuellement le flou). Nous utilisons pour cela des informations temporelles extraites des séquences d'images. Nous exploitons plus précisément la présence de mouvements «sous-pixeliques».

On peut répartir les méthodes de résolution de ce problème en trois grandes catégories. La première regroupe les techniques qui résolvent le problème dans le domaine de Fourier. Leur principal inconvénient est de n'être efficaces que lorsque le mouvement entre les images de la séquence est un mouvement de translation globale. La deuxième classe de méthodes compense le mouvement des images de la séquence pour construire une seule image, mais dont

les échantillons ne sont pas uniformément répartis. Cette image est ensuite ré-échantillonnée sur une grille uniforme de plus haute résolution. Cependant, ceci implique que des valeurs erronées (à cause de l'aliasage) sont interpolées dans la reconstruction de l'image haute résolution. Il est donc impossible de retrouver les détails de la scène originale de cette façon. La troisième classe de méthodes consiste à modéliser le processus de dégradation et à poser le problème de la super-résolution comme un problème inverse. On trouvera une description plus complète de toutes ces méthodes dans [1].

Notre méthode appartient à la troisième catégorie. Nous modélisons les dégradations subies par l'image originale «théorique». Nous obtenons ainsi un système linéaire que nous cherchons à résoudre en utilisant un algorithme itératif. La matrice correspondant à ce système a une structure par blocs et notre algorithme itératif exploite cette structure particulière. Ceci permet d'augmenter la vitesse de convergence et permet une implantation efficace. La méthode proposée s'apparente à la méthode de rétro-projection itérative de Irani et Peleg [5] ainsi qu'à la méthode de projection sur des ensembles convexes de Patti, Sezan et Tekalp [8]. Notons que nous n'introduisons pas de régularisation explicite du problème, mais celle-ci est obtenue par l'arrêt des itérations avant convergence.

La modélisation du problème nécessite la connaissance du mouvement apparent. Pour l'estimation des paramètres de mouvement, nous utilisons une méthode robuste multirésolution présentée dans [7]. L'utilisation d'un seul mo-

1. Ce travail a été effectué en collaboration avec Thales-Optronique et partiellement financé par la DGA.

dèle de mouvement global ne permet d'estimer que le mouvement dominant dans la séquence, mais son estimation est beaucoup plus rapide que celle d'un champ de déplacements denses. L'utilisation d'une carte de détection du mouvement permet de gérer les régions non conformes au mouvement dominant. De plus, nous avons introduit une version multiéchelle (au sens temporel) de l'algorithme de super-résolution qui permet de mener conjointement cette dernière et l'estimation du mouvement. Chacune des étapes bénéficie ainsi de la qualité accrue de l'autre.

Dans la première section, nous décrivons le modèle utilisé. La seconde donne les détails de l'algorithme itératif de super-résolution. Les troisième et quatrième sections décrivent deux améliorations de l'algorithme : l'utilisation d'une carte de détection de mouvement et la version multi-échelle. Enfin, des résultats sont présentés dans la dernière partie.

2 Modélisation du problème

Notre modèle est similaire à celui proposé par Elad et Feuer [4]. La séquence d'images observées est constituée de M images $\{Y_k\}_{k=1}^M$. Cette séquence est supposée provenir d'une image haute résolution X . Chaque image observée peut être considérée comme le résultat d'une déformation géométrique W_k (correspondant aux mouvements dans la séquence) de l'image X , qui est ensuite convoluée par la réponse impulsionnelle optique du système d'acquisition dont on note B_k l'opérateur correspondant. Enfin, les images sont sous-échantillonnées. On note D_k l'opérateur de sous-échantillonnage. Les images observées sont de plus supposées dégradées par un bruit additif gaussien η_k . On peut finalement écrire :

$$Y_k = D_k B_k W_k X + \eta_k \quad (1)$$

pour $k \in \{1, 2, \dots, M\}$

On supposera dans la suite que le sous-échantillonnage et le flou sont identiques pour chaque image. On posera donc $B_k = B$ et $D_k = D$

La matrice représentant le sous-échantillonnage par un facteur 2 peut être écrite :

$$D = d \otimes d$$

où \otimes est le produit de Kronecker et

$$d = \frac{1}{2} \left\{ \begin{array}{cccccc} \overbrace{\left[\begin{array}{cccccc} 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 1 \end{array} \right]}^l \\ \end{array} \right\} \frac{l}{2}$$

pour des images carrées de taille $l \times l$.

La matrice W_n est un opérateur de compensation de mouvement. Elle est définie par $\hat{X}_k^c(p) = W_k X = \hat{X}(p + V_{\theta_1^k}(p))$ où $V_{\theta_1^k}(p)$ est le vecteur de déplacement calculé grâce aux paramètres de mouvement estimés entre l'image 1 et l'image k au point p . Comme $p + V_{\theta_1^k}(p)$ est généralement non entier, $\hat{X}_k^c(p)$ est construit par interpolation bi-linéaire avec les pixels voisins.

On peut donc écrire :

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_M \end{bmatrix} = \begin{bmatrix} DBW_1 \\ \vdots \\ DBW_M \end{bmatrix} X + \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_M \end{bmatrix} = \begin{bmatrix} H_1 \\ \vdots \\ H_M \end{bmatrix} X + \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_M \end{bmatrix} \quad (2)$$

où nous avons défini $H_k = DBW_k$.

Le problème se ramène donc à la résolution d'un système linéaire :

$$Y = HX + \eta \quad (3)$$

Ce système linéaire a une structure de blocs que nous allons exploiter pour le résoudre, grâce à une méthode appelée *algorithme de Kaczmarz généralisé*.

3 Description de l'Algorithme

On cherche à résoudre un système linéaire $Y = AX$ où A est une matrice réelle de taille $LM \times N$ et X et Y sont des vecteurs réels de taille $N \times 1$ et $LM \times 1$. Ici, L correspond au nombre de pixels de chaque image et $N = 4L$.

Censor, Eggermont et Gordon [2] proposent de partitionner la matrice A de la façon suivante :

$$A = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_M \end{pmatrix} \quad (4)$$

où A_i est une matrice réelle de taille $L \times N$. Le vecteur Y a été partitionné comme suit :

$$Y^T = (Y_1^T Y_2^T \dots Y_M^T) \quad (5)$$

où Y_i est un vecteur de dimension L .

Pour obtenir une solution du système, on part d'un premier estimé X^0 et on applique l'algorithme itératif :

$$X^{n+1} = X^n + A_{i_n}^T \Sigma^n (Y_{i_n} - A_{i_n} X^n) \quad (6)$$

où n est le numéro de l'itération, i_n est le numéro du bloc de la matrice ou du vecteur, Σ^n est une matrice de relaxation, avec $i_n = n \bmod M + 1$.

Dans [2] sont données les conditions pour la convergence vers la solution aux moindres carrés de norme minimale.

La relation (6) appliquée à notre modèle nous donne :

$$X^{n+1} = X^n + W_{i_n}^T B^T D^T \Sigma^n (Y_{i_n} - DBW_{i_n} X^n) \quad (7)$$

En pratique, les matrices étant très creuses, nous n'avons pas intérêt à les construire pour limiter le coût de calcul. Nous allons maintenant étudier une implantation efficace de l'algorithme.

Par construction de la matrice D , la matrice D^T est, à un facteur 4 près, une matrice d'interpolation d'ordre 0 (interpolation par duplication). Si la réponse impulsionnelle optique est symétrique, $B^T = B$. De plus, si W est un opérateur compensant un mouvement de translation uniforme, il est facile de vérifier que W^T compense le mouvement de translation opposé. Dans un premier temps, nous choisissons de donner le même poids à tous les pixels lors de la reconstruction. On pose donc $\Sigma = \lambda I$ où λ est un paramètre de relaxation (ou de gain) et I la matrice identité.

L'algorithme est donc équivalent au schéma-bloc de la figure 2. Chacun des blocs peut être calculé efficacement, sans passer par une représentation matricielle, en utilisant des algorithmes traitant le problème pixel par pixel.

L'opérateur W transposé ne correspond à l'opérateur de mouvement inverse que dans le cas d'un mouvement de translation globale. Puisque nous ne souhaitons pas construire la matrice W , nous ne pouvons pas calculer sa transposée. Nous remplaçons donc l'opérateur transposé par la transformation géométrique correspondant au mouvement inverse. Notre schéma-bloc ne correspond alors plus exactement à la relation (6). La convergence de notre schéma-bloc n'est donc théoriquement pas démontrée pour des mouvements plus complexes que la translation globale. Nous l'avons néanmoins utilisé avec succès avec des modèles paramétriques de mouvement affines et quadratiques, ainsi qu'avec des champs de déplacements denses. Le mouvement inverse doit être estimé, ce qui rend particulièrement coûteux l'utilisation d'un champ dense de vitesse. Cependant, dans le cas de modèles de mouvement de 6 paramètres ou moins, il est inutile de réestimer le mouvement car le modèle inverse peut être exprimé de façon analytique en fonction du modèle direct.

4 Carte de détection du mouvement

Nous utilisons pour l'estimation du mouvement un modèle paramétrique du mouvement global. Ceci nous permet de nettement diminuer le coût de calcul. Il se peut cependant que ce modèle de mouvement ne soit pas valide pour tous les pixels.

Nous allons chercher à déterminer une carte binaire qui nous permettra de discriminer les pixels conformes au modèle de mouvement dominant et les pixels non conformes. Ces pixels non conformes sont des pixels dont le mouvement est mal approximé par le modèle paramétrique, ou bien correspondant à des objets mobiles dans l'image et dont le mouvement n'est pas pris en compte par le modèle global.

Nous utilisons le critère du maximum *a posteriori* et une modélisation markovienne pour déterminer la carte de détection de mouvement [6]. La carte de détection \hat{d} est donnée par le minimum d'une fonction d'énergie :

$$\hat{d} = \arg \min_d U_1(d, o) + U_2(d) \quad (8)$$

où $U_1(d, o)$ est le terme d'énergie liant le champ des observations o et le champ des étiquettes binaires d , et $U_2(d)$ le terme d'énergie rendant compte de notre *a priori* sur les étiquettes. Les observations utilisées sont données par la moyenne des vitesses normales pondérées par le carré du module du gradient spatial de l'intensité. Le terme U_2 (à travers la fonction de potentiel qui le définit) cherche à favoriser des étiquettes voisines similaires [6].

Par ailleurs, pour le calcul du mouvement global, un estimateur robuste est introduit conduisant à une méthode de moindres carrés pondérés itérés. A la fin de la procédure, nous avons donc un poids qui est affecté à chaque pixel. Plus ce poids est fort, plus le déplacement du pixel correspond au modèle paramétrique estimé. Nous pouvons donc utiliser cette carte de poids pour pondérer la carte

binaire de détection. Cette carte binaire pondérée est utilisée pour construire la matrice de relaxation Σ . Plus précisément, Σ est une matrice diagonale dont les éléments sont les coefficients issus de la carte binaire pondérée. Lors de la mise à jour de l'estimée courante dans la procédure itérative, un pixel a ainsi une importance proportionnel à la confiance que nous avons en la qualité de l'estimation de son mouvement.

5 Version multi-échelle temporelle

La méthode de super-résolution décrite plus haut prend comme référence une image dans la séquence d'images observées et estime le mouvement entre cette image et les autres images de la séquence. Nous allons introduire une version multi-échelle. Nous traitons maintenant la séquence originale par groupes de trois images et nous calculons une estimée super-résolue de l'image centrale de chaque groupe. Le résultat est une nouvelle séquence d'images. Chaque image a une meilleure résolution, mais la séquence est sous-échantillonnée temporellement. Le procédé peut être itéré sur cette nouvelle séquence d'images. Nous appelons ce schéma multi-échelle temporel, *schéma pyramidal*, car cela revient à construire une pyramide ayant plusieurs niveaux de résolution temporelle (Fig. 1).

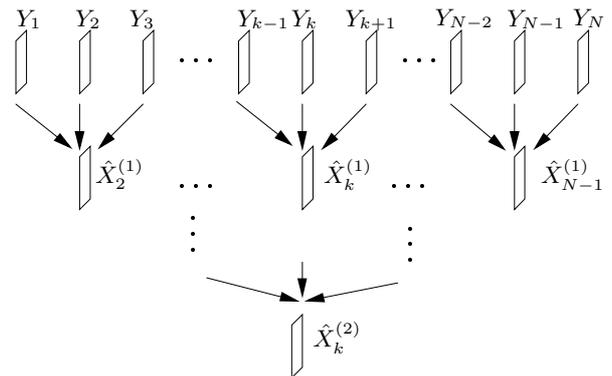


FIG. 1: Le schéma de super-résolution pyramidal

On réalise ainsi à chaque étape un compromis entre résolution spatiale et résolution temporelle. Ce schéma pyramidal s'avère plus robuste que le premier schéma mais il converge moins rapidement. Un avantage majeur est de permettre la ré-estimation des paramètres du mouvement à chaque niveau de la pyramide. Le mouvement estimé sur des images de meilleure qualité est plus précis et peut permettre une reconstruction de meilleure qualité dans les niveaux suivants de la pyramide.

6 Résultats

Nous présentons à la figure 3 les résultats de l'algorithme multi-échelle pour une séquence de 15 images.¹ En comparant les images 3e et 3f, on remarque une nette amélioration de la résolution de l'image, notamment sur les rochers à l'arrière plan. La non prise en compte du mouvement propre des automobiles entraîne cependant d'importants artefacts. L'utilisation de la carte de détection de mouvement élimine ces artefacts. Les régions non conformes au mouvement dominant gardent alors leur résolution d'origine.

1. La séquence a été fournie par l'INA, département innovation.

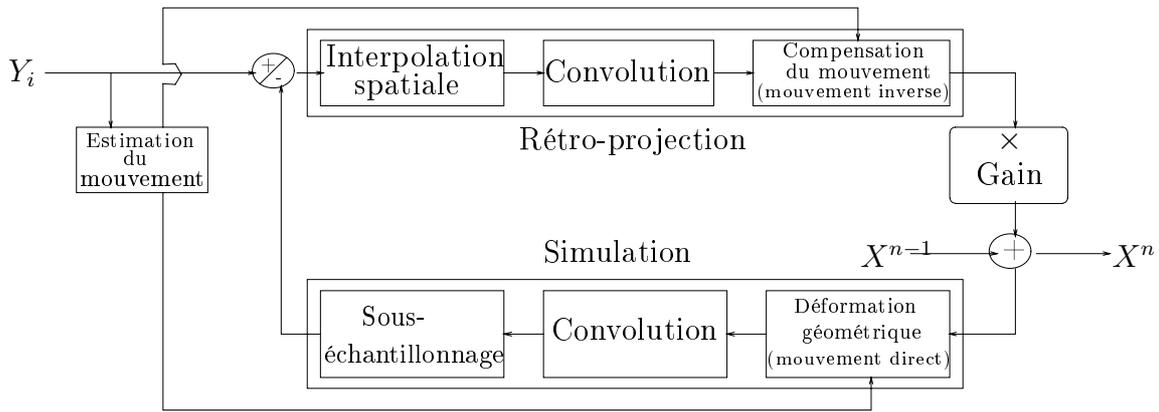


FIG. 2: schéma-bloc de l'algorithme de super-résolution.

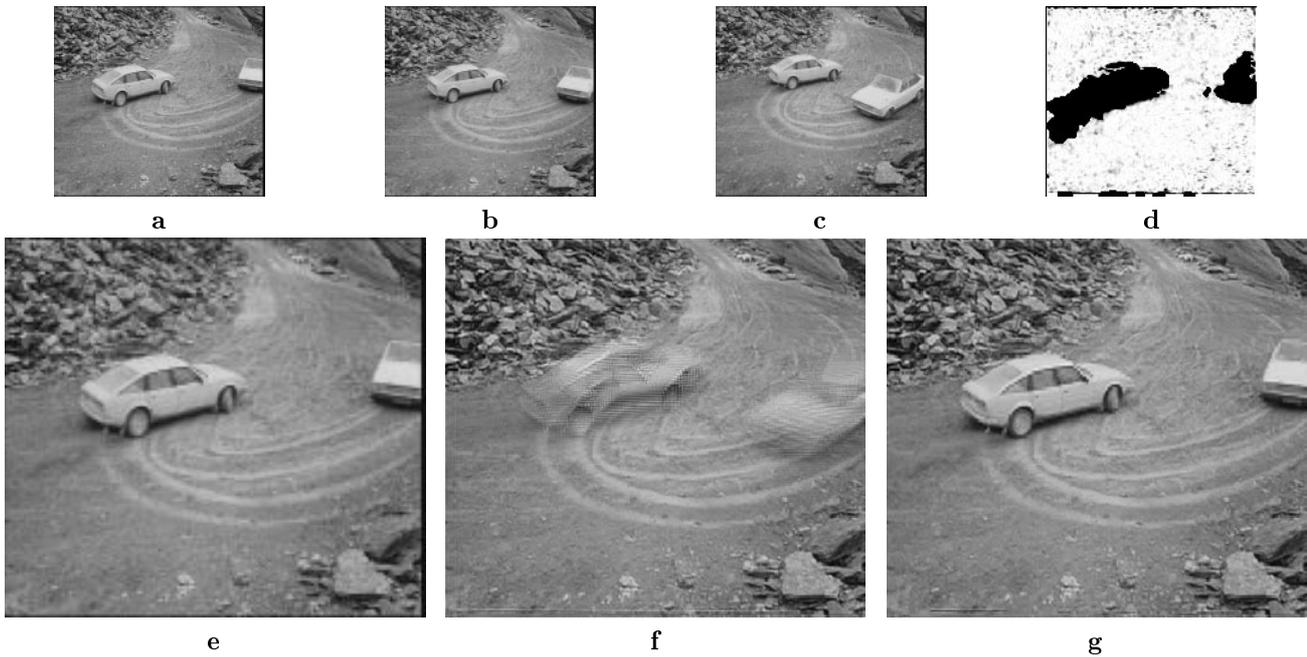


FIG. 3: (a, b, c) 3 images d'une séquence de 15 images de basse résolution, (d) carte des facteurs de relaxation pour chaque pixel déduite de la détection du mouvement et de l'estimation du mouvement dominant (poids), les valeurs sont représentées en niveaux de gris : noir pour 0, blanc pour 1, (e) image 3.a interpolée par simple interpolation spatiale, (f) image 3.a super-résolue par la méthode multi-échelle mais sans le module de détection de mouvement, (g) image 3.a super-résolue par la méthode multi-échelle et avec le module de détection du mouvement.

Références

- [1] S. Borman and R. Stevenson. Spatial resolution enhancement of low-resolution image sequences. Technical report, Laboratory for Image and Signal Analysis, University of Notre Dame, Notre Dame, IN, 1998.
- [2] Y. Censor, P. P. B. Eggermont, and D. Gordon. Strong underrelaxation in Kaczmarz's method for inconsistent systems. *Numerische Mathematik*, 41:83–92, 1983.
- [3] F. Dekeyser, P. Bouthemy, P. Pérez, and E. Payot. Super-resolution from noisy image sequences exploiting a 2D parametric motion model. In *International Conference on Pattern Recognition, ICPR'2000*, Barcelone, septembre 2000.
- [4] M. Elad and A. Feuer. Restoration of a single super-resolution image from several blurred, noisy and undersampled measured images. *IEEE Trans. Image Processing*, 6(12):1646–1658, décembre 1997.
- [5] M. Irani and S. Peleg. Motion analysis for image enhancement: Resolution, occlusion and transparency. *JVCIR*, 4(4):324–335, décembre 1993.
- [6] J. M. Odobez and P. Bouthemy. Separation of moving regions from background in an image sequence acquired with a mobile camera. In *Video Data Compression for Multimedia Computing*, chapter 8, pages 295–311. H. H. Li, S. Sun, and H. Derin, eds, Kluwer, 1997.
- [7] J.M. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *Jour. of Vis. Com. and Im. Rep.*, 6(4):348–365, décembre 1995.
- [8] A. J. Patti, M. I. Sezan, and A. M. Tekalp. Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time. *IEEE Trans. on Image Processing*, 6(8):1064–1076, août 1997.