

Super-résolution rapide sur flot optique

AURELIEN PLYER, FREDERIC CHAMPAGNAT, GUY LE BESNERAIS

ONERA/DTIM

Chemin de la Hunière et des Joncherettes, BP 80100, FR-91123

PALAISEAU CEDEX, France

aurelien.plyer@onera.fr, frederic.champagnat@onera.fr, guy.le_besnerais@onera.fr

Résumé - Nous proposons un cadre complet de super-résolution (SR) utilisant un recalage des images par flot optique rapide. En se basant sur un modèle de formation d'image approximatif conduisant à des simplifications algorithmiques, nous présentons d'une part une solution directe non itérative très rapide, permettant d'atteindre la cadence vidéo sur des exemples de dimensions réalistes ; d'autre part, en se fondant sur une formulation par pénalisation robuste, nous obtenons des versions itératives à coût limité conduisant à des améliorations significatives de qualité image y compris dans des séquences présentant des mouvements complexes (déformations, occultations).

Abstract – We propose a complete framework for super-resolution (SR) using a fast optical flow for low resolution image registration. Using an approximate model for image formation which allows a major simplification of the algorithms, we present, on one hand, a very fast non iterative SR technique which enables video rate processing on realistic contexts. On the other hand, we derive iterative versions of limited computational cost based on regularization with robust penalization functions. With such solutions, we demonstrate image enhancement even on video with complex motion fields (local deformation, occlusions).

1 Introduction

1.1 Principe de la super-résolution

La super-résolution (SR) désigne un ensemble de traitements spatio-temporels de vidéo qui visent à améliorer la qualité image en termes de bruit, contraste et résolution. Le principe général de ces techniques est d'utiliser le mouvement inter-image pour échantillonner le plan image à une échelle plus fine que le pixel. Le traitement comprend deux étapes principales : le recalage, qui consiste à estimer les mouvements inter-images avec une précision qui doit être subpixelique ; la reconstruction d'image SR qui peut se formuler comme l'inversion d'un système linéaire mal-conditionné de très grandes dimensions. Le cadre le plus simple de la SR concerne la production d'une image super-résolue (image SR) par l'utilisation de plusieurs images "basse résolution" (BR). Dans ce papier, on considère la production d'une séquence SR à partir d'une séquence BR, le traitement étant effectué sur une fenêtre temporelle glissante.

Un ensemble de contributions récentes sur la SR est présentée dans la référence [5]. La plupart des techniques publiées supposent un mouvement global paramétrique entre les images, par exemple un modèle de translation ou un modèle affine. Un petit nombre de références s'intéressent à l'utilisation d'un champ de mouvements non paramétrique produit par un estimateur de flot optique [6][10]. Ces travaux présentent des résultats probants sur des séquences comprenant des objets mobiles, mais reposent sur des choix de méthodes performantes mais gourmandes en

coût de calcul, ce qui conduit à des traitements relativement long malgré une implémentation sur GPU.

1.2 Contribution

Nous utilisons aussi le GPU pour accélérer les calculs, mais le point de vue est très différent de [6][10] en ce que nous recherchons un compromis coût/qualité permettant de produire des résultats à cadence élevée : l'objectif est au mieux la cadence vidéo (25Hz) et, au pire, une cadence de quelques Hz. Nous parlons de traitement en temps interactif, au sens où il permet à un opérateur d'observer les mouvements sur la séquence SR – par opposition à un simple arrêt sur image SR – tout en réglant éventuellement les paramètres de traitement qui sont remis à jour en ligne.

Pour ce faire nous avons choisi de recalibrer les images par l'algorithme de flot optique rapide FOLKI développé depuis 2005 à l'ONERA/DTIM (cf. [3] et [8], Chapitre 3) et de faire reposer la reconstruction SR sur un modèle de formation d'image approximatif. Ces deux choix visent à diminuer le coût de calcul et sont à l'opposé des orientations de [6][10] par exemple. Ils permettent de définir un algorithme de SR direct très rapide, capable d'atteindre la cadence vidéo pour le traitement SRx6 de 21 images BR 256x312. Sur cette base nous avons développé ensuite des versions itératives introduisant des normes robustes dans le critère d'estimation qui permettent de prendre en compte les défauts des données (et les erreurs de recalage) et de produire des reconstructions à préservation de contour. On peut ainsi traiter des séquences avec des objets mobiles pour une cadence de traitement de l'ordre de quelques Hz (toujours SRx6 de 21 images BR 256x312).

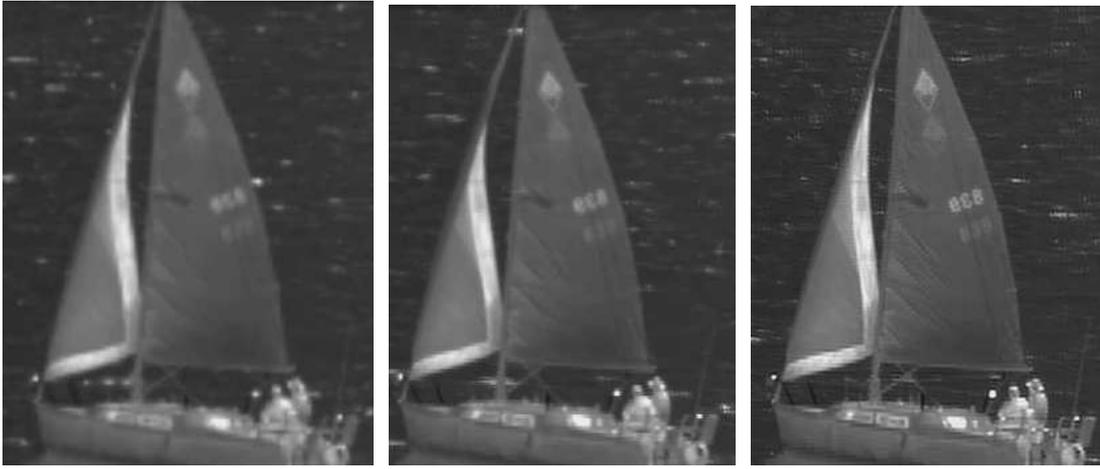


Figure 1 : Résultat de super-résolution rapide. En haut à gauche image BR zoomée (zoom spatial), à droite, image issue de l'opération de fusion. En bas, image issue de la fusion + déconvolution par Small Kernell (SK).

2 Méthode et résultats

2.1 Formulation générale

Nous utilisons un critère classique de régularisation pénalisée :

$$E_{\phi\phi}(\mathbf{I}_{\text{HR}}) = \sum_{i=1}^M \phi^{\alpha} (\|\mathbf{I}_{\text{BR}}^i - \mathbf{D}\mathbf{W}_i \mathbf{B}\mathbf{I}_{\text{HR}}\|^2) + \lambda\phi^{\gamma} (\|\nabla\mathbf{I}_{\text{HR}}\|^2), \quad (1)$$

où \mathbf{I}_{BR}^i désigne une des M images BR en entrée, \mathbf{I}_{HR} désigne l'image SR recherchée, dont les dimensions sont multipliées par le facteur f_{SR} par rapport à celle des images BR. Le modèle de formation d'image comprend une matrice de déformation \mathbf{W}_i (en référence à l'anglais *warp*) modélisant le mouvement entre l'image BR i et l'image BR 1 (qui définit la géométrie de référence). En pratique, cette matrice est constituée à l'aide du flot optique estimé entre les deux images par l'algorithme rapide FOLKI [3]. Lorsque les mouvements locaux sont importants, il est intéressant d'ajouter une étape de flot aller/retour qui permet de rejeter des zones qui ne satisfont pas à l'hypothèse du flot optique (e.g. des occultations partielles), cf. [8]. Le capteur est modélisé par un opérateur \mathbf{B} de convolution 2D en général fondé sur une tache image gaussienne, et \mathbf{D} , un opérateur de décimation de facteur f_{SR} sur les deux axes. Les fonctions de pénalisation ϕ sont soit quadratiques, soit des fonctions robustes de Huber, et le paramètre de régularisation est noté λ .

Le critère (1) repose sur un modèle de formation d'image approximatif, dit "warpBR" dans [4] qui applique le mouvement après la convolution – le modèle généralement admis étant appelé "warpHR" au sens où le mouvement est appliqué en géométrie HR avant de convoluer puis décimer. Les deux modèles, comparés dans [9] ne sont rigoureusement identiques que dans le cas d'un mouvement de translation. En pratique cependant, le modèle approximatif warpBR donne des

résultats exploitables pour des mouvements plus généraux [4]. Son avantage essentiel est de conduire à définir les matrices de déformation en géométrie BR et non HR, ce qui limite le coût de calcul et de stockage de la solution et ce qui permet surtout d'utiliser directement le flot optique estimé entre les images BR sans avoir à l'interpoler comme dans les approches warpHR. Notons que nous utilisons une seconde approximation pour accélérer la mise en œuvre qui consiste à tronquer les mouvements inter-images issus du flot optique à des valeurs entières du pas HR.

2.2 Cadre quadratique

Lorsque l'on se restreint à des potentiels quadratiques, on peut montrer que l'optimisation de (1) peut être approximativement conduite par un procédé non itératif en deux étapes [1]: l'une de fusion, dans laquelle on produit une image en géométrie HR en distribuant les intensités relevées dans les images BR et en interpolant les trous ; l'autre de déconvolution, ie. d'inversion régularisée de \mathbf{B} . Les travaux de [8] montrent que ces deux étapes peuvent être codées par des opérations très rapides sur GPU. Grâce à la troncature des déplacements mentionnée en section précédente, la fusion se fait par vote à un coût quasi-indépendant du facteur SR, tant que la mémoire n'est pas saturée. La seconde étape est implémentée par une convolution avec un noyau de déconvolution non séparable déduit de \mathbf{B} par une approche de filtrage de Wiener à horizon limité (small kernel, SK) [7]. Les filtres de Wiener SK sont paramétrés par un exposant de décroissance fréquentielle (fixé à 2), par une largeur de PSF gaussienne (fournie avec la séquence ou réglée par l'utilisateur) et par un paramètre de régularisation, réglé par l'utilisateur.

Les résultats de ces deux étapes de traitement sont présentés sur un exemple en Figure 1.

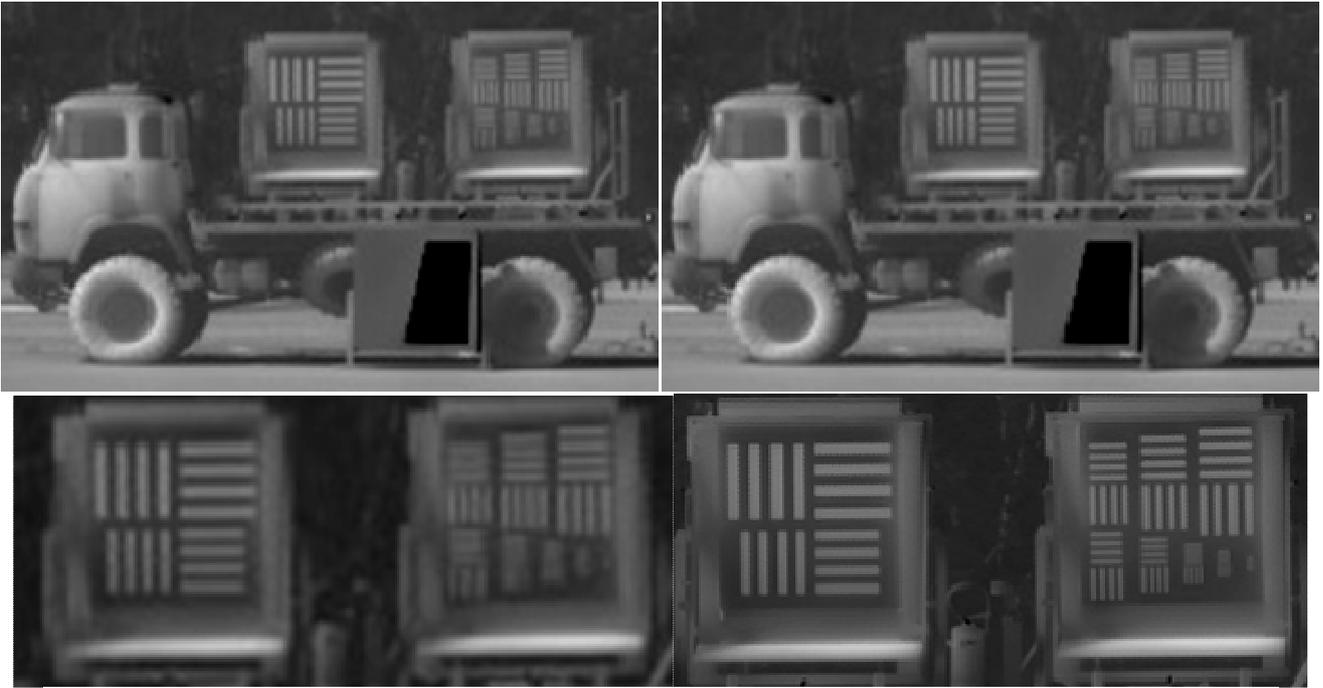


Figure 2 : Résultat de super-résolution robuste (séquence de test issue de [10]). En haut à gauche image BR zoomée (zoom spatial), à droite, image SR. En bas, détails sur les mires permettant d'évaluer le gain en résolution.

Sur cette séquence infra-rouge (IR) qui présente des mouvements locaux (déformation de la voile, mouvement des passagers) le gain de résolution obtenu par la SR rapide est significatif, permettant par exemple la lecture de l'immatriculation. On observe cependant les défauts classiques des techniques quadratiques (rebonds), ainsi que les effets de la troncature des déplacements. Le temps de calcul est de 45ms sur une carte Nvidia GTX480, c'est-à-dire proche de la cadence vidéo pour le traitement de 21 images 256x512 pour un facteur SR $f_{SR} = 5.5$! Cette performance est, à notre connaissance, inédite pour un traitement de SR fondé sur un recalage par flot optique.

2.3 Pénalisations robustes

Dans ce cadre, il faut utiliser une optimisation itérative, ici simplement un gradient à pas fixe, qui donne de bons résultats dans ce problème. Grâce au modèle warpBR on montre que le calcul du gradient de la partie adéquation aux données du critère (1) fait toujours intervenir des opérations de votes, comme dans la méthode rapide, mais affectées d'une pondération variable au cours des itérations. Ceci entraîne un surcoût limité par rapport au cadre quadratique : avec les dimensions des images de la Figure 1, le coût de calcul est multiplié par un facteur 2 à 5 par rapport à la solution rapide. Ce facteur reste néanmoins compatible avec notre objectif de départ d'une cadence de quelques Hz.

L'utilisation d'une optimisation plus précise et de potentiels non quadratiques conduit à des gains en qualité image par rapport à la solution rapide et à la possibilité d'obtenir un meilleur respect des contours, au prix d'un nombre plus important de paramètres à régler.

Un exemple test est présenté en Figure 2 (données acquises par la DGA dans le cadre du groupe OTAN SET 140, cf. [11]). L'acquisition utilise une monture tournante pour produire un mouvement de translation entre les images et l'objet visé est un camion surmonté de mires de résolution, ce qui permet de quantifier approximativement le gain en résolution. Sur cet exemple on obtient un gain proche de 2, valeur limite dans de nombreux cas pratiques [1].

La pénalisation robuste du terme de données, ou, de manière équivalente, la repondération des données au cours des itérations, combinée à un recalage par flot optique permet d'étendre le domaine d'emploi de la SR à des champs de mouvement complexes. En d'autres termes, on peut sortir largement du contexte de mouvement paramétrique (et le plus souvent translationnel) des techniques de SR classiques [5]. La Figure 3 présente à cet égard un exemple particulièrement difficile sur une séquence infra-rouge de piétons issue de [11]. Les piétons sont déformables, éventuellement tournent sur eux-mêmes et se croisent, générant des occultations. La SR est limitée à 5 images : le gain en résolution n'est pas évident à quantifier, mais on observe néanmoins un apport significatif en termes de netteté et de contraste. A notre connaissance, ce type de résultat est peu commun dans la littérature de la SR et les travaux comparables en termes de complexité des mouvements gérés reposent sur des techniques beaucoup plus coûteuses [6][10].

3 Conclusion

Nous avons présenté un cadre de SR fondé sur un recalage par flot optique qui conduit, grâce à

l'exploitation de l'architecture massivement parallèle des GPU, à une variété de techniques : solution non itérative très rapide et solutions itératives robustes. Ces dernières permettent de traiter des séquences présentant des mouvements locaux à des cadences rapides, compatibles avec l'observation du mouvement par un opérateur sur la séquence SR produite à la volée. Ces développements ont été intégrés dans une plate-forme de traitement de vidéo par SR permettant le réglage de paramètres en ligne, avec la possibilité d'observer directement l'effet

de ces changements de réglage sur la vidéo produite en sortie.

Les perspectives de nos travaux concernent l'extension du domaine de la SR à d'autres cas limites, comme le fait de produire un résultat en géométrie non plane (rendu 3D SR), et le contexte de l'observation à longue portée : SR sur petits objets (détection) et prise en compte des effets de la propagation à travers une atmosphère turbulente.



Figure 3 : Super-résolution robuste sur une séquence présentant des mouvements complexes (données issues de [10]). Les piétons sur cette séquence IR marchent (déformations) et se croisent (occultations). En haut, séquence d'origine zoomée (zoom spatial) en bas résultat de super-résolution.

4 Bibliographie

- [1] Champagnat, F., Le Besnerais, G., & Kulcsár, C. (2009). Statistical performance modeling for superresolution: a discrete data-continuous reconstruction framework. *JOSA A*, 26(7), 1730-1746, 2009.
- [2] M. Elad, Y. Hel-Or, A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur, *IEEE Transactions on Image Processing*, Vol 10, n° 8, Aug 2001.
- [3] Le Besnerais G. et Champagnat, F. (2005). Dense optical flow by iterative local window registration. In *IEEE International Conference on Image Processing (ICIP)*, page 137, 2005.
- [4] A. Létienne, F. Champagnat, C. Kulcsár, G. Le Besnerais, P. Viaris De Lesegno. Fast super resolution on moving objects in video sequences, *EUSIPCO 2008*.
- [5] Milanfar P., *Super-Resolution Imaging*, CRC Press 2012
- [6] Mitzel, D., Pock, T., Schoenemann, T., et Cremers, D. Video super resolution using duality based tv-l1 optical ow. In *Pattern Recognition (Proc. DAGM)*, Jena, Germany, 2009.
- [7] Pereira, S. and Goussard, Y. Unsupervised 3d restoration of tomographic images by constrained wiener filtering. In *Proceedings of 19th Intl Conf. EMBS*, Chicago, 1997.
- [8] Plyer, Aurélien *Architecture massivement parallèles et vision artificielle bas-niveau*, ONERA/DTIM , thèse à l'université Paris 13, février 2013.
- [9] Rochefort, G. Rochefort, G., Champagnat, F., Le Besnerais, G. & Giovannelli, J. F. (2006). An improved observation model for super-resolution under affine motion. *Image Processing, IEEE Transactions on*, 15(11), 3325-3337, 2006.
- [10] Unger, M., Pock, T., Werlberger, M., et Bischof, H. A convex approach for variational super-resolution. In Goesele, M., Roth, S., Kuijper, A., Schiele, B., and Schindler, K., editors, *Pattern Recognition*, volume 6376 of *Lecture Notes in Computer Science*, pages 313-322. Springer Berlin / Heidelberg, 2010.
- [11] Weiss, A., R., Adomeit, U., Chevalier, P., Landeau, S., Bijl, P., Champagnat, F., Dijk, J., Göhler, B., Landini, S., Reynolds, J. P., and Smith, L. N. *A standard data set for performance analysis of advanced IR image processing techniques*. volume 8355, page 835512. SPIE., 2012.