

Etude de différentes techniques de Super-Résolution de séquences vidéos

C. Riedinger¹, M. N. Khemakhem¹ et G. Chollet¹

¹Département Traitement du Signal et des Images UMR CNRS 5141

Institut Télécom-ParisTech

46, rue Barrault 75634 PARIS Cedex 13 - FRANCE

Résumé—Cet article propose une étude comparative de différentes techniques de super-résolution (SR) : la méthode fréquentielle, la reconstruction par interpolation, le filtrage de Wiener, le Maximum A Posteriori (MAP), la reconstruction par visages propres (eigenfaces) et la projection sur des ensembles convexes (Projection Onto Convex Sets (POCS)). Plusieurs de ces algorithmes ont été implémentés et expérimentés sur des images en niveaux de gris : l'algorithme de reconstruction par estimation MAP [10], l'algorithme des eigenfaces [11], la méthode de Wiener [1] et la méthode POCS [18]. Les performances de ces algorithmes sont comparés sur des protocoles de test standard.

Les méthodes POCS et MAP donnent des résultats similaires. La SR en aveugle [9] ainsi que l'amélioration des algorithmes pour qu'ils fonctionnent en temps réel constituent des thèmes de recherche prometteurs ainsi que les méthodes hybrides POCS-ML.

Mots clés—Super résolution, Interpolation, Méthode fréquentielle, Wiener, MAP, Eigenfaces, POCS.

I. INTRODUCTION

LA super-résolution (SR) consiste en l'obtention d'une image de haute-résolution (HR) à partir de plusieurs images de basse-résolution (BR). L'image HR est de qualité supérieure aux images d'origine. L'utilité de la reconstruction par la SR n'est de nos jours plus à démontrer. Il existe, en effet, de nombreuses applications civiles ou militaires dans lesquelles il est fait usage de techniques de SR. En effet, l'image SR offre des images de plus grande finesse que les images d'entrée originales. La SR peut, donc, être utilisée quand on dispose d'images BR ou en qualité dégradée afin d'obtenir une image sur laquelle apparaîtront des détails qu'on n'obtenait pas dans l'image originale. Les techniques de SR dans un contexte lié à la vidéo de surveillance ont été utilisées pour le projet ANR CSOSG2007 KIVAOU¹. Celles-ci nous permettront d'obtenir des images de visage de meilleure qualité que celles fournies par les caméras de surveillance. Dans tous les cas, des séquences vidéo sont utilisées en entrée. Les images successives de la séquence vidéo constituent la base d'images utilisées pour effectuer la reconstruction par SR.

1. Ce projet vise à développer un démonstrateur comprenant des outils innovants d'analyse vidéo dédiés à deux problématiques : 1) Outil mobile (valise) d'identification et d'indexation biométrique faciale portable par analyse temps réel vidéo et 2) Plate-forme d'analyse de vidéos multiples enregistrées lors d'un événement, utilisant la synchronisation de vidéos, l'extraction de signatures pour les personnes, et la constitution de trajectoires. Le but est de permettre ou de faciliter une analyse *a posteriori* des données enregistrées en un lieu et pendant une même période à des fins d'investigations.

L'idée la plus intuitive pour effectuer une reconstruction par SR est de superposer les différentes images BR zoomées afin d'obtenir une image somme sur laquelle les détails sont rehaussés du fait du supplément d'information qu'apporte l'utilisation de plusieurs images. Il faut pour cela que l'objet que l'on cherche à reconstruire n'ait subi que de faibles variations de pose et d'illumination d'une image à l'autre et qu'il n'y ait pas d'occlusion dans les différentes images de la séquence. Les problèmes d'occlusion en SR constituent un problème encore mal résolu.

Pour pouvoir effectuer une superposition des images (qui représente l'idée principale d'une reconstruction par SR), il faut, d'abord, passer par une étape de recalage des images BR nécessaire pour que les différentes images de l'objet à reconstruire soient superposables. Cette étape fondamentale à toute SR est, d'abord, abordée avant d'exposer les différentes techniques de reconstruction.

II. LE RECALAGE DES IMAGES

Les principales techniques existantes sont la corrélation de phase [12], la méthode d'évaluation du déplacement par développement en série de Taylor [1] et la méthode de calcul du flot optique (algorithme de Lucas Kanade) [3]. Il a été choisi de développer une méthode de recalage par correspondance de blocs (**block matching**), plus lente, mais dont la robustesse est avérée quand on effectue un recalage d'images peu bruitées.

Le principe du recalage par block matching [2] d'une image BR par rapport à une autre est de calculer la corrélation entre des régions environnant un pixel sur chacune des images et de choisir le pixel sur l'image à recaler pour lequel la corrélation est maximale avec le pixel sur l'image de référence. Ces pixels représentent le même point de l'objet. En calculant la différence de position entre ces deux pixels et en décalant le pixel à recaler de cette différence, on effectue un recalage de l'image pour que les deux soient superposables. Supposant que la résolution de l'image BR soit augmentée d'un facteur m quand on passe en SR, il faut que la précision du recalage soit de $\frac{1}{m}$ des pixels BR pour que l'on puisse recaler correctement les images ([5], [6], [4]). En effet, c'est à la résolution de l'image HR que l'on va superposer les images. Pour ce faire on zoome (interpolation à un plus grand facteur d'échelle) les images BR et on effectue les recherches par corrélation des régions sur les images zoomées (figure 1).

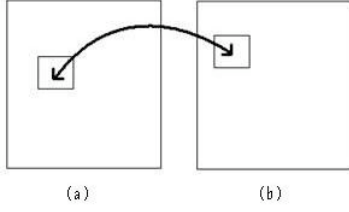


FIGURE 1. On calcule la corrélation entre la région sur l'image (a) (petit carré) et la région sur l'image (b) ($2^{\text{ème}}$ petit carré). La région sur l'image (b) bouge sur toute l'image (b). Lorsque l'on a trouvé un maximum de corrélation entre ces deux régions, on a trouvé de combien s'est déplacé le point central de la région entre l'image (a) et l'image (b).

L'algorithme d'interpolation à un plus grand facteur d'échelle sert à zoomer les images pour pouvoir effectuer le recalage. Son principe est le suivant : soit une série 1D de points ayant des valeurs non nulles qui constituent les points connus de la fonction à interpoler (ceux d'une ligne de l'image BR). A l'initialisation de l'algorithme, ces points sont espacés entre eux par d'autres ayant une valeur nulle et où l'on cherche la valeur de la fonction à interpoler. La valeur de chaque point de valeur initiale nulle est mise à jour en calculant la somme de celle-ci et de ses adjacents divisée par 3. Les points de valeurs non nulles sont maintenus inchangés. Cette étape est réitérée en maintenant inchangés les points qui, originellement, étaient de valeurs non nulles. Les valeurs des autres (originellement nulles) sont réévaluées en calculant la somme de la valeur de ces points et des deux adjacents divisée par 3. L'algorithme est réitéré une vingtaine de fois. La suite de points de valeurs initialement nulle tend vers une suite dont les valeurs sont celles de l'interpolation linéaire des points de valeurs originelle non nulles. Il est possible, ainsi, d'interpoler linéairement une ligne de points espacés de façon non uniforme. Une fois que sont interpolées toutes les lignes de l'image pour lesquels il existe des valeurs non nulles à l'origine des itérations, il est fait de même avec les colonnes. A la fin, une interpolation d'une grille de points espacés de façon non uniforme est obtenue (voir figures 2, 3, 4).

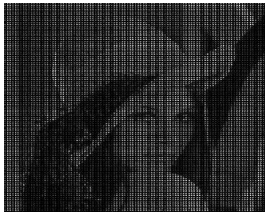


FIGURE 2. Image reconstruite à partir de points uniformément espacés.

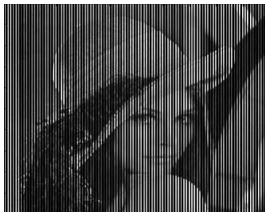


FIGURE 3. Interpolation sur les lignes de l'image 2.



FIGURE 4. Interpolation sur les colonnes de l'image 3.

Pour calculer la corrélation entre deux régions situées autour de deux pixels sur deux images BR zoomées (voir figure 1), plusieurs critères ont été proposés. Citons le calcul du coefficient de corrélation et le calcul de la somme des différences absolues [2].

$$corr = \frac{\sum_{m,n} x_{m,n} y_{m,n}}{\sqrt{\sum_{m,n} x_{m,n}^2 \sum_{m,n} y_{m,n}^2}} \quad (1)$$

$$sda = \sum_i |x_{m,n} - y_{m,n}| \quad (2)$$

$x_{m,n}$ et $y_{m,n}$ sont les valeurs des pixels sur l'image 1 et l'image 2 en un point (m,n) du masque de corrélation.

Dans leur article [2], Borman et al. montrent que les deux critères ne produisent pas des résultats très différents. Le critère de la somme des différences absolues a été choisi. Il a été décrit précédemment comment on pouvait apparier deux pixels sur des images différentes en calculant la corrélation des régions situées autour de ces pixels. Il faut, en effet, trouver un pixel correspondant à chacun de la première image BR zoomée ce qui ralentit considérablement l'algorithme. Des régions (patch) de corrélation de taille 11×11 entourant le pixel central ont été choisies.

III. LA MÉTHODE PAR FUSION

La méthode la plus intuitive pour effectuer une reconstruction par SR est de faire simplement la fusion (somme) des images zoomées recalées et de diviser la valeur des pixels par le nombre d'images utilisées. Une image floue est généralement obtenue. Il est possible d'améliorer la qualité de cette image en utilisant la méthode du filtrage de Wiener [1].

IV. LE FILTRAGE DE WIENER

Soit $X(m,n)$ une matrice qui représente l'image floue obtenue par fusion. On note $\hat{X}(w_m, w_n)$ sa transformée de Fourier. La formule de Wiener (3) définit le filtre optimum qui permet de retrouver une image la plus nette possible.

$$\hat{X}_{optimum} = \frac{H^*}{|H|^2 + \frac{S_n}{S_I}} \hat{X} \quad (3)$$

$H(w_m, w_n)$ est la fonction de transfert du filtre qui permet de passer de l'image nette à l'image floue obtenue par fusion (celui-ci est inconnu), S_n et S_I sont la densité spectrale de puissance (DSP) du bruit ajouté et de l'image originale respectivement. Lors de nos expérimentations, H a été estimé

comme étant un filtre passe bas de transmission uniforme dans la bande passante (voir figure 2). Toutes les possibilités de filtres telle que la largeur de la bande passante soit encore comprise entre $[-N, N]$, où $[-N, N]$ est l'intervalle de points de la transformée de Fourier discrète (**TFD**) de l'image (donc le nombre de points de l'image), ont été testées. On définit le rapport $\frac{S_n}{S_f}$ comme deux demi gaussiennes dont la largeur dépend du nombre de points N et dont les centres se situent en 0 et en N (Alam et al., dans leur article, utilisent un ratio « densité spectrale du bruit sur densité spectrale du signal constant » [1]). La meilleure image de la série d'images obtenue à partir des différents filtres H a été gardée comme définitive.

V. LA MÉTHODE FRÉQUENTIELLE

La méthode fréquentielle [14] repose sur :

- 1) l'utilisation de la formule de Poisson (Théorème de l'aliasing) :

$$F^e(m, n) = \sum_{p=-L}^{L-1} \sum_{q=-L}^{L-1} F^c \left(\frac{2\pi m}{MT_x} + pw_x, \frac{2\pi n}{MT_y} + qw_y \right) \quad (4)$$

(Où M est la longueur de la TFD, T_x est la période d'échantillonnage suivant x , F^e est la transformée de Fourier (**TF**) d'une image BR, F^c est la transformée de Fourier continue (**TFC**) de l'image originale HR, m et n sont entiers variant de $-\frac{M}{2}$ à $\frac{M}{2}$).

- 2) la relation entre la TF d'une image et celle de la même image décalée :

$$F^d(u, v) = F^c(u, v) e^{2i\pi(\delta_x u + \delta_y v)} \quad (5)$$

Du fait de 2), la méthode fréquentielle ne fonctionne que sur une série d'images qui ont subi des translations l'une par rapport à l'autre alors que les autres méthodes fonctionnent également pour des transformations plus complexes entre les images si l'algorithme de recalage permet de retrouver le mouvement entre deux images dû à cette transformation complexe. Les théorèmes 1 et 2 permettent de relier la valeur en certains points de la TFC de l'image à toutes les TF numériques des images BR dont on dispose au départ. Le nombre de points de TFC est supérieur à celui de la TFD. Le système d'équation obtenu reliant les TF est mal conditionné. Il faut, donc, le régulariser. La forme de la régularisation est la suivante : Soit I la TF de l'image HR recherchée. Les points de I doivent vérifier le système d'équation et en même temps vérifier que la norme de I moins A (une approximation de la solution) est minimale. L'approximation de la solution est obtenue en faisant la TF d'une des images BR zoomée et correctement recalée. Le système peut être résolu par une méthode de descente du gradient ou par le calcul du pseudo inverse.

VI. LA MÉTHODE MAP

Dans la méthode MAP ([15], [13], [7]), on construit un estimateur MAP de l'image solution à partir des images BR

et d'une fonction de régularisation. La fonction de coût à minimiser est la suivante :

$$J(X) = \sum_i \|M_i X - Y_i\|_l^l + \lambda F(X) \quad (6)$$

l représente la norme l de l'expression. On note $\{S\}$ le système $Y_i = M_i X$. Dans l'équation 6, le premier terme de fidélité aux données contraint la solution à vérifier le système au sens du maximum de vraisemblance. Le deuxième terme, $F(X)$, de régularisation permet de régulariser le système S généralement mal conditionné. En effet, si N est le nombre de pixels de l'image BR, on dispose de N équations et $N \times \text{fact} \times \text{fact}$ inconnues (où fact est le facteur de SR). Plusieurs normes ont été proposées dans la littérature pour l'attache aux données et la régularisation. Nous avons implémenté la norme L^1 pour la régularisation et la norme L^2 pour l'attache aux données. Plusieurs formes de régularisation ont été proposées dans la littérature : la régularisation de Tikhonov et la minimisation du laplacien ou du gradient de l'image. La minimisation du gradient de l'image a été retenue. Le système $\{Y = M_i X\}_i$ représente l'effet du système optique qui fait passer une image HR X à une image BR Y (les images sont vectorisées, i est l'indice de sommation sur le nombre d'images BR). Les paramètres de ce système et donc de la matrice M_i sont, en général, inconnus. Il faut, donc, les estimer empiriquement. La matrice M_i se compose de 3 sous matrices, $M = DW_i B$ [15]. B est en général une matrice de floutage de l'image originale par un noyau gaussien dont la largeur dépend de la caméra utilisée. W_i est la matrice qui associe un point de l'image HR à un autre de l'image HR recalée (Warping). Son effet est le contraire de l'opération de recalage détaillée plus haut. D est la matrice de décimation. Elle transforme un point de l'image HR en un autre de l'image BR ou un noyau de points de l'image HR en un point de l'image BR. La première solution a été choisie car, dans le calcul du gradient que l'on va voir par la suite, l'opération de transposition de D dans la deuxième solution transforme un point BR en un noyau HR dont tous les points ont la même valeur (on a, donc, un effet de pixellisation très important). Les paramètres de W sont estimés par recalage. B est choisi empiriquement. D dépend du facteur de la SR que l'on souhaite avoir.

Note 1: Les équations s'écrivent formellement avec des images vectorisées mais, dans l'implémentation, nous avons préféré interpréter les opérations matricielles comme des opérations sur les images non vectorisées car la notation sparse est très pratique, ce qui évite pour D , B , W d'avoir à écrire des matrices gigantesques dont une grande partie est remplie par des zéros.

Lorsque les paramètres des matrices ont été choisis, on peut minimiser le critère de l'équation (6), en général, par une méthode de descente du gradient ce qui a été l'approche dans cet article. L'expression du gradient de la fonction coût est :

$$\frac{\delta J}{\delta X} = \sum_i 2B^T W_i^T D^T (DW_i B X - Y_i) + \lambda \frac{\delta \sum_{m,n} \|grad X\|_1}{\delta X} \quad (7)$$

Lorsque l'on prend une norme L^2 pour la fidélité aux données et que l'on régularise en imposant le gradient de l'image soit le plus petit possible. j est l'indice de sommation sur les pixels de l'image.

Lorsque l'on prend une norme L^1 pour la fidélité aux données, l'expression du gradient est :

$$\frac{\delta J}{\delta X} = \sum_i B^T W_i^T D^T \text{signe}(DW_i B X - Y_i) + \lambda \frac{\delta \sum_{m,n} \|\text{grad} X\|_1}{\delta X} \quad (8)$$

Un algorithme de type itératif a ensuite été utilisé pour estimer l'image HR.

$$X^{n+1} = X^n - \rho \frac{\delta J}{\delta X} \quad (9)$$

Suivant la norme choisie, le terme de fidélité aux données effectuée, à chaque itération de l'algorithme, soit la moyenne (norme L^2), soit la médiane des différentes images BR recalées et décimées (norme L^1). Le terme de régularisation choisi (la norme du gradient de l'image doit être minimale) se traduit par le calcul de la différence de valeurs entre les pixels avoisinant le pixel calculé (éq. 10).

La fonction *signe* retourne -1 ou 1 si le signe de la différence de deux valeurs est négatif ou positif respectivement.

$$\left(\frac{\delta \|\text{grad} X\|_1}{\delta X} \right)_{m,n} = \text{signe}(x_{m,n} - x_{m-1,n}) + \text{signe}(x_{m,n} - x_{m+1,n}) + \text{signe}(x_{m,n} - x_{m,n-1}) + \text{signe}(x_{m,n} - x_{m,n+1}) \quad (10)$$

x est la valeur du pixel (m, n) . On lui ajoute 1 s'il a une valeur inférieure à ses voisins. Dans le cas contraire, on lui retranche 1. Pour une image plate, on ajoute et on retranche 0 au pixel calculé ce qui correspond bien au critère recherché. Les opérateurs de différenciation classique ont, initialement, été employés pour le calcul du gradient de l'image (ce qui aboutit à l'équation 10). Ceux-ci sont inopérants pour les lignes à 45° dans l'image. Il a, donc, été choisi d'utiliser un masque 3×3 qui est le suivant :

$$\begin{pmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{pmatrix}$$

Pour les 8 pixels du bord, on prend le signe de la valeur du pixel central moins celle du pixel du bord et on ajoute les 8 signes obtenus pour calculer le gradient par rapport à X du gradient de l'image. Dans l'expression précédente, on ne sommait que 4 signes.

Farsiu et al., dans leur article, utilisent une version un peu différente de cet opérateur (Bilateral Total Variation Regularisation). L'expression qu'ils utilisent est donnée dans [10].

VII. LA MÉTHODE POCS

Le principal intérêt de la méthode POCS est la simplicité avec laquelle il est possible d'introduire des informations *a priori* hétérogènes sur la solution désirée. Il suffit de définir un ensemble convexe qui les incorpore. Si ces ensembles sont bien choisis, la détermination de l'opérateur de projection peut être très simple et conduire à des algorithmes faciles à implanter.

On se place dans l'espace vectoriel $\mathbb{R}^{M \times N}$ qui représente l'espace de toutes les solutions possibles. Chaque propriété connue de l'image inconnue $f \in \mathbb{R}^{M \times N}$ restreint celle-ci à un sous-ensemble $\{C_i\}_{i=1,2,\dots,m}$ de cet espace. Ces ensembles représentent les caractéristiques souhaitées de la solution. L'ensemble des solutions possibles est donc l'intersection des ensembles C_i . La méthode de projection sur des ensembles consiste à projeter un point de l'espace $\mathbb{R}^{M \times N}$ sur chacun des ensembles C_i , et à répéter l'opération jusqu'à l'obtention d'un point appartenant à l'intersection de tous les ensembles. La convergence est assurée si les ensembles C_i sont fermés et convexes, et bien sûr, si l'intersection est non vide.

Soit C un ensemble de l'espace vectoriel. Pour tout f , la projection Pf de f sur C est l'élément de C le plus proche de f . Lorsque C est fermé et convexe, Pf existe et est unique. L'opérateur de projection ainsi défini est, en général, non linéaire.

Cette méthode a été utilisée par Stark et Oskoui [19] pour la reconstruction d'images HR à partir d'images obtenues par translations et rotations d'un détecteur. Le système optique projette une image $X(u, v)$ sur un ensemble de $L_1 \times L_2$ capteurs rectangulaires de réponse impulsionnelle $\sigma_i(u, v)$. La sortie du $i^{\text{ème}}$ capteur est donnée par :

$$y_i = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X(u, v) \cdot \sigma_i(u, v) \cdot du \cdot dv, \quad i \in \{1, 2, \dots, L_1 L_2\} \quad (11)$$

L'ensemble des y_i nous donne naturellement une image de $L_1 \times L_2$ pixels. Pour obtenir une image de meilleure résolution, on commence par discrétiser cette intégration sur une grille de reconstruction fine de taille de $M \times N$. Nous obtenons :

$$y_i = \sum_{m=1}^M \sum_{n=1}^N X(m, n) \cdot \sigma_i(m, n) \quad (12)$$

On cherche à déterminer $X(m, n)$, la luminance du pixel HR. Si on suppose que le capteur a une réponse unitaire et uniforme sur toute sa surface, on a :

$$\sigma_i(m, n) =$$

$$\begin{cases} 0 & \text{si le pixel HR centré sur } (m, n) \text{ est totalement en dehors de la surface du capteur } i \\ 1 & \text{si le pixel HR centré sur } (m, n) \text{ est totalement compris dans la surface du capteur } i \\ q_i & (0 < q_i < 1) \text{ si le pixel HR centré sur } (m, n) \text{ est partiellement compris dans la surface du capteur } i \end{cases}$$

q_i est la fonction de la surface du pixel en (m, n) contenue dans la surface du capteur i .

Une seule acquisition d'image nous donne un système sous-déterminé de $L_1 * L_2$ équations. On obtient des équations supplémentaires en réalisant de nouvelles acquisitions après rotation et translation des capteurs. On sait que pour un capteur rectangulaire de dimension $a * b$, translaté par un vecteur (u_j, v_j) et soumis à une rotation d'angle θ_j :

$$\sigma_i(m, n) = \text{rect} \left[\frac{(u-u_j) \cos \theta_j + (v-v_j) \sin \theta_j}{a} \right] * \text{rect} \left[\frac{(v-v_j) \cos \theta_j + (u-u_j) \sin \theta_j}{b} \right] \quad (13)$$

En ordonnant de manière lexicographique $X(m, n)$ et $\sigma_i(m, n)$ pour obtenir les vecteurs colonnes X et σ_i^T , l'équation (12) s'écrit : $y_i = \sigma_i^T X$.

Pour chacune des M images, on ordonne de manière lexicographique les pixels y_i pour obtenir le vecteur y_k . Pour la $k^{\text{ème}}$ image, le modèle est de la forme : $y_k = H_k X$. Cependant, considérer les rotations et les translations du détecteur est moins général que de considérer les transformations géométriques de l'image. En effet, on ne peut prendre en compte ni les effets de perspective, ni les objets mobiles.

On considère maintenant les ensembles C_i tels que :

$$C_i = \{X; \sigma_i^T X = y_i\}, 1 \leq i \leq L_1 L_2 \quad (14)$$

L'ensemble C_i est l'ensemble de toutes les images SR pour lesquelles la réponse du $i^{\text{ème}}$ détecteur est la valeur observée y_i . On définit autant d'ensembles C_i qu'il y a de pixels dans la séquence d'images observée. On peut montrer que les ensembles C_i sont fermés et convexes, et que le projeté g d'un vecteur h sur C_i est donné par :

$$g = \begin{cases} h & \text{si } \sigma_i^T X = y_i \\ h + \frac{y_i - \sigma_i^T X}{\sigma_i^T \sigma_i} \sigma_i & \text{sinon} \end{cases} \quad (15)$$

Pour obtenir une image SR, la méthode de projection sur des ensembles convexes consiste à projeter itérativement sur chacun des ensembles C_i une image estimée initiale jusqu'à convergence. La solution n'est, généralement, pas unique et dépend de l'initialisation. Notons que si l'on utilise que les ensembles C_i , qui sont des contraintes d'attache aux données, on a équivalence avec la technique de reconstruction algébrique utilisée en tomographie. Il est toutefois possible d'utiliser également des contraintes qui représentent une connaissance *a priori* sur la solution.

L'une des plus simples est la contrainte d'amplitude qui donne un ensemble C_A défini par :

$$C_A = \{X(u, v); \alpha \leq X(u, v) \leq \beta\} \quad (16)$$

Dans le cas d'images codées sur 8 bits, on peut imposer par exemple $\alpha = 0$ et $\beta = 255$.

Une autre contrainte possible est la contrainte d'énergie :

$$C_E = \{X; \|X\|^2 \leq E\} \quad (17)$$

où E est l'énergie maximale permise pour l'image reconstruite.

Cette méthode a été reprise et modifiée par Tekalp, Özkan et Sezan pour prendre en compte le bruit blanc additif dans le cas de séquences d'images soumises à un mouvement de translation uniforme [20]. Les ensembles C_i deviennent alors :

$$C_i = \{X; |r_i| < \delta_i\} \quad (18)$$

où r_i est le $i^{\text{ème}}$ élément du résiduel $r = Y - HX$ qui correspond au bruit d'observation. δ_i représente donc la confiance que nous avons dans les observations. Sa valeur dépend des statistiques du bruit. Dans ce cas, le projeté g d'un vecteur h sur C_i est donné par :

$$g = \begin{cases} h + \frac{r_i + \delta_i}{H_i^T H_i} H_i & \text{si } r_i \notin [-\delta_i, \delta_i] \\ h & \text{si } r_i \in [-\delta_i, \delta_i] \end{cases} \quad (19)$$

Patti, Sezan et Tekalp ont ensuite étendu cette méthode pour prendre en compte dans le modèle le flou de bouger entre les images de la séquence [17], puis une grille d'échantillonnage arbitraire [18].

Jusqu'ici, le modèle de discrétisation de l'image continue a toujours été considéré comme étant d'ordre 0, c'est-à-dire que l'image continue est supposée constante sur la surface du pixel HR. Patti et Altunbasak ont introduit explicitement une interpolation d'ordre plus élevé dans le modèle [16]. Ils ajoutent également une méthode de régularisation adaptative afin d'éviter les phénomènes de rebond souvent observés près des contours. Pour cela, ils estiment la direction des contours et pondèrent la fonction de flou le long du contour par une fonction proportionnelle au gradient d'intensité.

Eren, Sezan et Tekalp ont introduit une carte de validité permettant de ne pas prendre en compte les régions avec une mauvaise estimation de mouvement. Pour cela, ils ont utilisé une segmentation manuelle et un module de suivi [8]. Un modèle de mouvement différent peut être calculé pour chaque objet afin d'assurer une meilleure reconstruction.

VIII. LA MÉTHODE DES EIGENFACES

Dans la méthode des eigenfaces [11], on utilise, également, la matrice M (mentionnée dans la section V). Le principe est d'effectuer un apprentissage sur une base d'images HR, de calculer les vecteurs propres de l'ACP appliquée à cette base d'images HR (eigenfaces) et de calculer les vecteurs propres obtenus par ACP sur la version BR de cette même base et de faire correspondre la projection sur l'espace ACP HR recherchée aux projections ACP sur l'espace de toutes les images BR après application de la matrice M à l'image HR (on utilise l'ACP sur les pixels de l'image). Les vecteurs propres de la matrice sont calculés :

$$C = \sum_j I_j I_j^T \quad (20)$$

Où I_j est la $j^{\text{ème}}$ image de la base d'images HR. On fait de même pour les versions I_j BR des images I_j . On garde les vecteurs propres associés au plus grandes valeurs propres

(on garde autant de vecteurs propres qu'il y a d'images dans la base). On a alors :

$$X = \phi a \quad (21)$$

Où X est l'image HR et a est un vecteur de coefficients de projection (features). ϕ est la matrice des vecteurs propres colonnes.

La projection d'une image sur la base de vecteurs propres se traduit, pour les images HR, par la multiplication suivante (projection de X sur la base de vecteurs propres) :

$$a_j = \phi_j^T X \quad (22)$$

avec :

$$a = [a_1 \ a_2 \dots a_L]^T \quad \text{et} \quad \phi = [\phi_1 \ \phi_2 \dots \phi_L]$$

Pour les images BR $I_{j'}$, on note ψ la matrice de vecteurs propres .

$$\hat{a}_j = \psi_j^T Y_i \quad (23)$$

L'image HR vectorisée est reliée à l'image BR vectorisée par :

$$Y_j = M_j X \quad (24)$$

Les projections sont, donc, reliées par :

$$\psi \hat{a}_j = M_j \phi a \quad (25)$$

(ici \hat{a} désigne le vecteur feature pour la $j^{\text{ème}}$ image BR. Il y a un vecteur « feature » pour chaque image BR).

En multipliant la dernière équation par ψ^T et en utilisant le fait que $\psi^T \psi$ est égale à l'identité, on obtient :

$$\hat{a}_j = \psi^T M_j \phi a \quad (26)$$

Cette équation relie la projection sur l'espace BR de l'image Y_i et la projection sur l'espace HR de l'image X . Le vecteur \hat{a}_j est obtenu par simple projection de l'image BR connue. On estime a en résolvant le critère :

$$\operatorname{argmin}_a \left\{ \left\| \psi^T M_i \phi a - \hat{a}_j \right\|^2 \right\} \quad (27)$$

La minimisation de ce critère est résolue par une méthode de descente du gradient comme dans le cas de la méthode MAP. Après plusieurs itérations de l'algorithme, on obtient une estimation de a . On obtient, ensuite, X en faisant la projection inverse à la projection de l'image sur l'espace HR c'est-à-dire :

$$X = \phi a \quad (28)$$

Dans cet article, une base d'images de 90 eigenfaces a été utilisée.

IX. RÉSULTATS

Pour obtenir nos résultats, des films vidéo AVI disponibles sur le site de P. Milanfar ont été utilisés. Des images provenant de la base de données MBGC ont également été utilisées. Le protocole expérimental est le suivant : Pour les figures 5 à 9 on effectue une SR d'un facteur 4 avec 16 images BR en entrée. Deux versions de la méthode MAP ont été implémentées. La première compare l'image HR avec une version recalée sur la grille HR des images BR. La seconde compare l'image HR décimée (donc une image BR) avec les images BR originales. La méthode 2 donne des contours plus nets que la méthode 1. On compare ainsi les images obtenues avec la méthode MAP et la POCS à l'image obtenue par simple interpolation d'une image BR. La POCS et la MAP 2ème version donnent les meilleurs résultats visuellement. Pour la figure (5), les images BR proviennent du site de P. Milanfar. La SR avec la méthode MAP permet d'augmenter la qualité de l'image et ne produit pas d'artefacts comme ceux produits par l'interpolation simple d'une image BR. Pour la figure (6), les images BR proviennent, de même, du site de P. Milanfar. L'image avant application du filtre de Wiener est floue alors que celle après l'application du filtre de Wiener est de qualité supérieure à l'image BR. Le critère de qualité est ici la lisibilité du texte après reconstruction SR. Les images de la figure (7) proviennent de la base de données FRGC : Une image a été exclue de l'ensemble d'apprentissage, décalée aléatoirement, décimée avant d'être reconstruite. Dans cette catégorie d'images, on dispose de la vérité terrain (on possède l'image HR originale que l'on décompose et on essaye de la reconstruire (voir figure 7 - (c))). Les images de la figure (8) proviennent de la base de données MBGC. Les images (9 - (a)) et (9 - (b)) montrent respectivement un fragment d'une mire de télévision BR et la reconstruction SR pour laquelle nous avons utilisé la deuxième méthode MAP. La mire a été filmée par nous même avec une webcam et est disponible à www.tsi.enst.fr/~rieding.

X. CONCLUSION

Nous avons, dans cet article, présenté plusieurs techniques de SR. Les algorithmes POCS et MAP donnent des résultats similaires pour les images que nous avons testées. La SR en aveugle (blind SR) constitue un thème de recherche prometteur puisqu'il permet d'effectuer une reconstruction par SR sans connaître les caractéristiques du système optique et de sous échantillonnage avec lequel les images ont été acquises. Dans notre cas, nous avons estimé empiriquement les caractéristiques du système optique. Notons que la rapidité de calcul représente un point clé de la reconstruction SR. La méthode par fusion et interpolation suivie d'un filtrage de Wiener constitue, peut être, la solution à une implémentation en temps réel des algorithmes de SR. Enfin, nous devons mentionner le problème de la reconstruction couleur. Celle-ci est plus difficile que la reconstruction en niveaux de gris car les différentes composantes de la couleur de l'image ne suivent pas le critère de Bayer (les décalages entre une image et une autre ne sont pas les mêmes pour les différentes composantes de couleur) (ce problème a été traité dans [9]).



FIGURE 5. (a) : Une des images BR originales, (b) : Image obtenue avec la méthode MAP 1^{ère} version (décimation = 2, nombre d'images = 16), (c) : Image obtenue avec la méthode MAP 2^{ème} version (décimation = 2, nombre d'images = 16), (d) : Image obtenue par interpolation d'une image BR (décimation = 2, nombre d'images = 16), (e) : Image obtenue avec la méthode POCS (décimation = 2, nombre d'images = 16) et (f) : Image obtenue avec la méthode POCS (décimation = 4, nombre d'images = 16) (Source : Site de P. Milanfar).

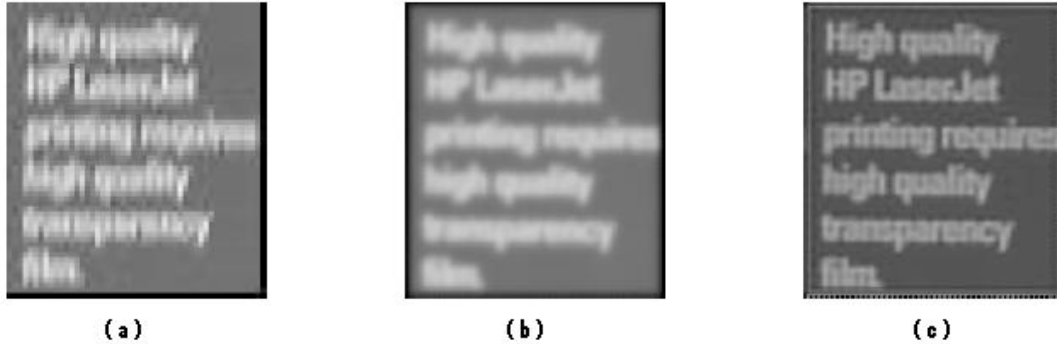


FIGURE 6. (a) : Une des images BR originales, (b) : Image obtenue par fusion + interpolation sans application du filtre de Wiener (décimation = 4, nombre d'images = 16) et (c) : Image après application du filtre de Wiener (décimation = 4, nombre d'images = 16) (Source : Site de P. Milanfar).



FIGURE 7. (a) : Une des images BR originales, (b) : Image HR reconstruite par la méthode des eigenfaces (décimation = 2, nombre d'images = 16) et (c) : L'image HR originale (Source : Base de données FRVT).

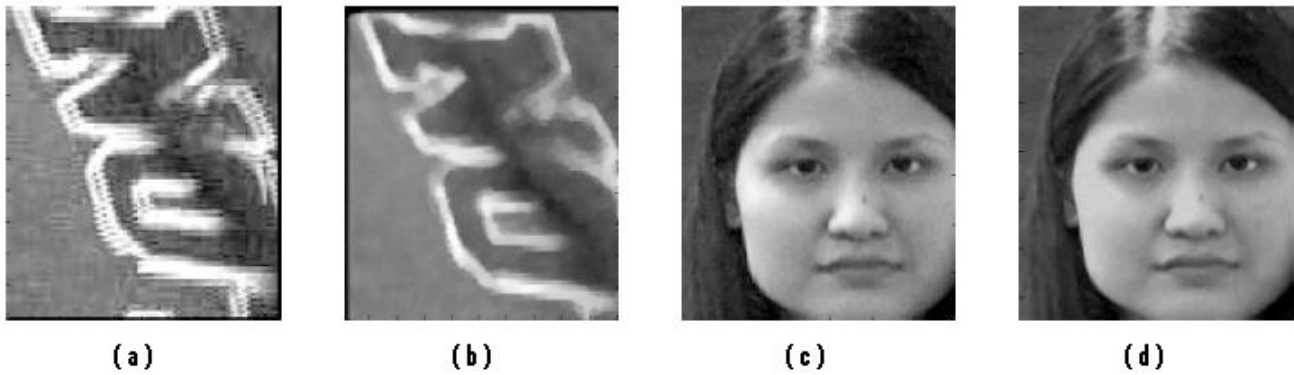


FIGURE 8. (a) et (c) : Deux images BR originales, (b) et (d) : Deux images HR obtenues par la méthode MAP 2^{ème} version (décimation = 2, nombre d'images = 16) (Source : Base de données MBGC).

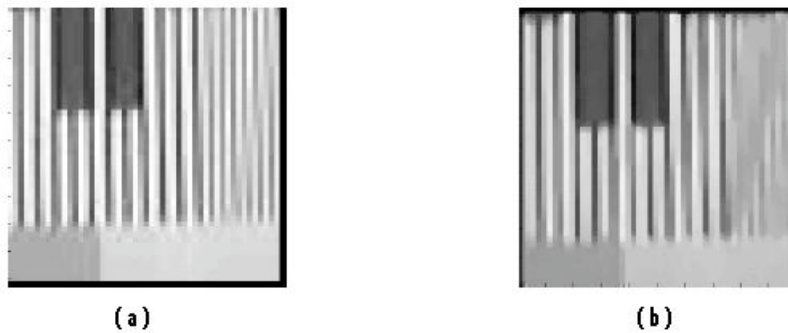


FIGURE 9. Exemple d'une mire télé. (a) Une des images BR originales et (b) : Une image HR obtenue par la méthode MAP 2^{ème} version (décimation = 2, nombre d'images = 16) (Source : La mire a été filmée par nous mêmes et est disponible à www.tsi.enst.fr/~crying).

RÉFÉRENCES

- [1] M.S. Alam, J.G. Bognar, R.C. Hardie, and B.J. Yasuda. Infrared image registration and high-resolution reconstruction using multiple translationally shifted aliased videos. *IEEE transactions on instrumentation and measurement*, 49(5), Oct. 2000.
- [2] S. Borman, R.C. Bognar, R.C. Hardie, and B.J. Yasuda. Block matching subpixel motion estimation from noisy undersampled frames : an empirical performance evaluation. *Visual communications and image processing '99, SPIE proceedings*, 3653, Déc. 1998.
- [3] J.Y. Bouguet. Pyramidal implementation of the lucas kanade feature tracker, description of the algorithm. *Inter corporation*.
- [4] D. Capel. *Image Mosaicing and Super-resolution*. PhD thesis, University of Oxford, 2001.
- [5] D. Capel and A. Zisserman. Automated mosaicing with super-resolution zoom. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Santa Barbara*, pages 885–891, 1998.
- [6] D. Capel and A. Zisserman. Super-resolution enhancement of text image sequences. In *Proc. International Conference on Pattern Recognition*, pages 600–605, 2000.
- [7] D. Capel and A. Zisserman. Super-resolution from multiple views using learnt image models. volume 2, pages II–627–II–634 vol.2, 2001.
- [8] P.E. Eren, M.I. Sezan, and A.M. Tekalp. Robust, object-based high-resolution image reconstruction from low-resolution video. *IEEE Trans. Image Processing*, 6(10) :1446–1451, Oct. 1997.
- [9] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar. *Advances and challenges in super-resolution*. Wiley periodicals, 2004.
- [10] S. Farsiu, M. Dirk Robinson, M. Elad, and P. Milanfar. Fast and robust super resolution. *IEEE transactions on image processing*, 13(10) :1327–1344, Oct. 2004.
- [11] B.K. Gunturk, A.U. Batur, Y. Altunbasak, M.H. Hayes, and R.M. Mersereau. Eigenface-domain super-resolution for face recognition. *IEEE transactions on image processing*, 12(5), Mai 2003.
- [12] F. Humblot. *Détection de petits objets dans une image en utilisant les techniques de super-résolution*. PhD thesis, Université Paris-Sud XI, 2005.
- [13] F. Humblot and A. Mohammad-Djafari. Super-resolution using hidden markov model and bayesian detection estimation framework. *EUR-ASIP Journal on Applied Signal Processing*, Special number on Super-Resolution Imaging : Analysis, Algorithms, and Applications, 2006.
- [14] E.A. Kaltenbacher and R.C. Hardie. High resolution infrared image reconstruction using multiple low-resolution aliased frames. *Proc. SPIE*, 2751(142), 1996.
- [15] S.C. Park, M.K. Park, and M.G. Kang. Super-resolution image reconstruction : a technical overview. *IEEE signal processing magazine*, mai 2003.
- [16] A.J. Patti and Y. Altunbasak. Artifact reduction for set theoretic super resolution image reconstruction with edge adaptive constraints and higher-order interpolants. *IEEE Trans. Image Processing*, 10(1) :179–186, Jan. 2001.
- [17] A.J. Patti, M.I. Sezan, and A.M. Tekalp. High-resolution image reconstruction from a low-resolution image sequence in the presence of time-varying motion blur. In *Proc. of the IEEE Inter. Conf. on Image Processing*, pages 343–347, Novembre 1994.
- [18] A.J. Patti, M.I. Sezan, and A.M. Tekalp. Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time. *IEEE Trans. Image Processing*, 6(8) :1064–1076, Août 1997.
- [19] H. Stark and P. Oskoui. High resolution image recovery from image-plane arrays, using convex projections. *J. Opt. Soc. Am. A*, 6 :1715–1726, 1989.
- [20] A.M. Tekalp, M.K. Ozkan, and M.I. Sezan. High-resolution image reconstruction from lower-resolution image sequences and space varying image restoration. *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), San Francisco, CA*, 3 :169–172, Mars 1992.