JOINT BLUR IDENTIFICATION AND HIGH-RESOLUTION IMAGE ESTIMATION BASED ON WEIGHTED MIXED-NORM WITH OUTLIER REJECTION

Osama A. Omer and Toshihisa Tanaka

Department of Electrical and Electronic Engineering,
Tokyo University of Agriculture and Technology, Tokyo, Japan.
E-mails: osama@sip.tuat.ac.jp, tanakat@cc.tuat.ac.jp.

ABSTRACT

We address problems of conventional super-resolution (SR) methods having the following limitations. First, most of the existing SR algorithms can not cope with local motions and hence not suitable for video sequences. Second, the blurring operator is assumed to be known in advance and constant for all the low-resolution (LR) images. Finally, SR noise is assumed to be either Gaussian or Laplacian. To solve these problems, we propose a general cost function that consists of weighted L_1- and L_2-norms considering the SR noise model where the weights are generated from the error of registration and penalize parts that are inaccurately registered. Both the super-resolved images and blurring operators are jointly estimated. The objective and subjective results are shown to demonstrate the effectiveness of the proposed algorithm.

Index Terms— Super-resolution, image fusion, image registration, mixed-norm, affine motion, outlier rejection

1. INTRODUCTION

Super-resolution is an approach to obtain high-resolution (HR) image(s) from a set of low-resolution images. The most important steps of SR algorithms are image registration, data fusion, and restoration. Image registration, which has been paid much attention for the last two decades, is the process to align images (frames) to the reference image (frame) [1]. However, image registration of images containing locally moving parts is still a challenging task. Data fusion is the process which fuses the registered images onto HR grid. Restoration is the process to estimate the HR image from the fused data on the HR grid.

In most of the recently proposed SR algorithms [2–4] the SR results depend on fusion step. As a cost function, the L_2- or L_1-norm is used to fuse LR images [2–6]. It can be seen [3] that the choice of norm depends on the distortion assumption of fusion error: whether the error is Gaussian (L_2-norm [2, 4]) or Laplacian (L_1-norm [3]). It should be noted that the use of L_2-norm implicitly assumes that the extra resolution content is equally distributed over all LR images [2]. Then the resulting HR image is an average of the contributions from all LR images. This implies that the averaging process leads to propagation of the outlier pixels from any of the LR images into the HR image. To avoid this averaging effect, the use of global weight for each image has been proposed [4, 5]. Although these methods can be suitable with global motion, they indeed fail with sequences that include moving objects.

On the other hand, it is known that L_1-norm is robust against outliers [3]. It can not however cope with errors resulting from occlusion that happens in video sequences that contain local motions. The failure in the case of occlusion is due to two reasons. First, in video sequences that have local and fast motions, Laplacian distribution assumption is not proper. Second, the model converges to the median over the measured data without pre-weighting the LR images, which may lead to failure in the case of occlusion.

On the context of image restoration, a weighted L_2-norm has been also proposed [7], in which weighted and regularized L_2-norm is used to restore one frame from a set of K blurred frames. Actually, this is a special case of the SR where decimation factor is one [7]. Another suggestive idea in image restoration is the use of a mixed-norm [8], where second and fourth norms are combined for image restoration. The relative importance of LMS and least mean fourth (LMF) is used as a function of the kurtosis of the noise. This algorithm also represents a special case of SR algorithms since only one frame is used to restore an enhanced frame with the same dimensions.

In summary, the main problems of the previous SR algorithms are as follows. Most of these algorithms [2–6] assume global motion so they are not suitable for video sequences including locally moving parts. Also, they assume that blurring operator is known. Moreover, even if L_2-norm is suitable for Gaussian noise and L_1-norm is suitable for Laplacian noise, non of them can cope with mixed noise distributions. Motivated by these problems and previous image restoration methods [7, 8], we propose a robust method that simultaneously estimates a blurring matrix and a high resolution image in the presence of the local registration errors by global motion estimation technique. To do that, we introduce a globally weighted average of locally weighted L_2- and L_1-norms as a cost function, where the global weights penalize the SR noise distribution model and the local weights penalizes the local registration error. The global weight is determined in the basis of the SR noise distribution function and the local weighting function is chosen so as to be a function of the local registration error between the reference LR frame and the other LR frames.

2. PROBLEM DESCRIPTION AND NOTATION

Assume that K LR frames of the same scene in lexicographical order denoted by $\underline{Y}^k (1 \leq k \leq K)$, each containing M^2 pixels, are observed, and they are generated from the HR frame denoted by $\underline{X}$, containing L^2 pixels, where $L \geq M$. We use the underscore notation to indicate a vector. The observation of K LR frames are modeled by the

This work is supported in part by the Telecommunications Advancement Fundation.

following degradation process:

$$\underline{Y}^k = DB^kF^k\underline{X} + \underline{V}^k, \tag{1}$$

where F^k, B^k and D^k are the motion operator, the blurring operator (due to camera), and the down-sampling operator respectively, $\underline{X}$ is the unknown HR frame, $\underline{Y}^k$ is the k^{th} observed LR frame, and $\underline{V}^k$ is an additive random noise for the k^{th} frame. Throughout the paper, we assume that D is known and the additive noise is Gaussian with zero mean. Therefore the problem is to reconstruct HR image ($\underline{X}$) while estimating blurring operators (B^k) and motion operators (F^k). Due to the fact that motion operator is not known, it will be estimated from the successive frames. Motion estimation is not always accurate then the estimated motion operator can be described as $F^k + \Delta F^k$, where ΔF^k is the registration error matrix, so the degradation model in (1) can be described as

$$\underline{Y}^k = DB^kF^k\underline{X} + DB^k\Delta F^k\underline{X} + \underline{V}^k = DB^kF^k\underline{X} + \widetilde{\underline{V}}^k, \tag{2}$$

where $\widetilde{\underline{V}}^k$ is the combination of additive noise and motion error. In this paper, we assume the blurring operation as moving averaging as $B^k\underline{X} = \sum_{l=-\rho}^{\rho}\sum_{m=-\rho}^{\rho} b_{l,m,k}S_x^lS_y^m\underline{X}$, where $b_{l,m,k}(-\rho \le l, m \le \rho)$ are the coefficients of the blurring operator and S_x^l and S_y^m are shifting operators by l and m pixels in x and y direction respectively. Throughout the rest of the paper we will use B^k and $\sum_{l=-\rho}^{\rho}\sum_{m=-\rho}^{\rho} b_{l,m,k}S_x^lS_y^m$ interchangeably.

3. SR BASED ON WEIGHTED MIXED-NORM

In [6], we showed that $\widetilde{\underline{V}}^k$ is suitable to be described as Laplacian in case of slow (global) motions where accurate motion estimation is possible while in case of fast (local) motion it is suitable to be modeled by Gaussian. Based on these results we propose to use a weighted average of L_2- and L_1-norms where the global weights are adaptive with respect to error distribution. Two types of weighting functions are introduced, the first is locally adaptive with respect to the registration error and the second is globally adaptive with respect to error distribution. Hence by using a general regularization term, the proposed cost function is described as follows:

$$\begin{aligned} J(\underline{X}, B) = \sum_{k=1}^{K} \Big[\zeta_2^k(\underline{X}) \left\| DB^kF^k\underline{X} - \underline{Y}^k \right\|_{W_2^k}^2 + \\ \zeta_1^k(\underline{X}) \left\| DB^kF^k\underline{X} - \underline{Y}^k \right\|_{W_1^k}^1 \Big] + \lambda(\underline{X}, B) \left\| C\underline{X} \right\|_p^p \end{aligned} \tag{3}$$

where $\zeta_1^k(\underline{X})$ and $\zeta_2^k(\underline{X})$ are the global weights, W_1^k and W_2^k are the local weights, C is a general high pass operator, and $\lambda(\underline{X}, B)$ is the regularization factor.

3.1. Choice of Global Weights

The global weights are introduced in Eq. (3) to adjust the relative importance of the contributions of L_1- and L_2- norms. It is desired from these weights that in cases of Gaussian noise the contribution of L_2-norm increases and that of L_1-norm decreases, while in cases of Laplacian noise the contribution L_2-norm decreases and the contribution of L_1-norm increases. On the other hand, for cases of mixed noise these weights should adjust the relative importance on Gaussian and Laplacian. To achieve that, we propose to use global weights that are function of the generalized likelihood ratio test (GLRT) value of the SR noise [3] since it has a large value (> 0.7602) for Gaussian while it has smaller value (< 0.7602) for Laplacian. Also, to avoid the different scale of the L_1- and L_2-norms we use a normalization with respect to the regularization function [4]. $\zeta_1^k(\underline{X})$ and $\zeta_2^k(\underline{X})$ are defined as

$$\zeta_1^k(\underline{X}) = \xi_1(\underline{X}) \ln\left(T_1^k(\underline{X}) \cdot \frac{\|C\underline{X}\|_2^2}{\|DB^kF^k\underline{X} - \underline{Y}^k\|_1^1} + 1 \right),$$

$$\zeta_2^k(\underline{X}) = \xi_2(\underline{X}) \ln\left(T_2^k(\underline{X}) \cdot \frac{\|C\underline{X}\|_2^2}{\|DB^kF^k\underline{X} - \underline{Y}^k\|_2^2} + 1 \right),$$

where

$$T_1^k(\underline{X}) = \frac{\sum_{i=1}^{K}\|DB^iF^i\underline{X} - \underline{Y}^i\|_1}{\|DB^kF^k\underline{X} - \underline{Y}^k\|_1}, T_2^k(\underline{X}) = \frac{\sum_{i=1}^{K}\|DB^iF^i\underline{X} - \underline{Y}^i\|_2^2}{\|DB^iF^k\underline{X} - \underline{Y}^k\|_2^2},$$

with $\xi_1(\underline{X})$ and $\xi_2(\underline{X})$ being the average of $\lambda_1^k(\underline{X})$ ($1 \le k \le K$) and $\lambda_2^k(\underline{X})$ ($1 \le k \le K$) over the K frames respectively, where $\lambda_1^k(\underline{X})$ and $\lambda_2^k(\underline{X})$ are defined as follows:

$$\lambda_1^k(\underline{X}) = \begin{cases} 0.5e^{-\left|\mathrm{GLRT}^k(\underline{X}) - 0.7602\right|}, & \text{if } \mathrm{GLRT}^k(\underline{X}) \ge 0.7602 \\ 1 - 0.5e^{-\left|\mathrm{GLRT}^k(\underline{X}) - 0.7602\right|}, & \text{if } \mathrm{GLRT}^k(\underline{X}) < 0.7602 \end{cases}, \tag{4}$$

and $\lambda_2^k(\underline{X}) = 1 - \lambda_1^k(\underline{X})$, where $\mathrm{GLRT}^k(\underline{X})$ is the $\mathrm{GLRT}(\underline{X})$ of the errors at k^{th} frame.

3.2. Choice of Local Weights

To overcome the errors resulting from inaccurate registration, we propose to use local weights for both L_2- and L_1-norms. Weighted L_1- and L_2-norms for vector $\underline{X}$ of size N are defined as

$$\|\underline{X}\|_{W_2} = \left(\underline{X}^T W_2 \underline{X}\right)^{1/2}, \quad \|\underline{X}\|_{W_1} = \sum_{i=1}^{N} W_1(i,i)|X(i)|,$$

respectively, where W_1 and W_2 are diagonal matrices with non-negative components. To use these norms in (3), the local weights can be adaptive with respect to the registration error as follows

$$W_2^k(i,i) = e^{-\left|\underline{Y}^k(i) - (F_\downarrow^k \underline{Y}^{ref})(i)\right|}, \tag{5}$$

where W_2^k is the weighting matrix of size $M^2 \times M^2$ for L_2-norm at k^{th} frame, $\underline{Y}^{ref}$ is the reference LR frame in lexicographical order, and $F_\downarrow^k$ is the motion operator in the LR domain corresponding F^k in HR domain. The weighting matrix for L_1-norm at k^{th} frame (W_1^k) is chosen as a logic version of W_2^k, that is,

$$W_1^k(i,i) = \begin{cases} 1 & \text{if } W_2^k(i,i) < \theta(i) \\ 0 & \text{otherwise} \end{cases}, \tag{6}$$

where $\theta(i)$ is chosen as $\theta(i) = \frac{1}{K}\sum_{k=1}^{K} W_2^k(i,i)$.

3.3. Alternative Minimization

The optimization of $J(\underline{X}, B)$ is performed with respect to $\underline{X}$ and $b_{l,m,k}$ alternatively. The gradients of $J(\underline{X}, B)$ with respect to $\underline{X}$ may be approximated by

$$\begin{aligned} \nabla_{\underline{X}} J(\underline{X}, B) \simeq \sum_{k=1}^{K} F^{k^T} B^{k^T} D^T \Big(2 \cdot \zeta_2^k(\underline{X}) W_2^k (DB^kF^k\underline{X} - \underline{Y}^k) + \\ \zeta_1^k(\underline{X}) W_1^k \mathrm{sign}(DB^kF^k\underline{X} - \underline{Y}^k) \Big) + \lambda(\underline{X}, B) C^T C\underline{X}, \end{aligned}$$

where we assume that the gradient of the weighs $\zeta_1^k(\underline{X})$ and $\zeta_2^k(\underline{X})$ with respect to $\underline{X}$ is negligible with respect gradients of the L_1- and L_2-norms. The gradient of $J(\underline{X}, B)$ with respect to $b_{l,m,k}$ is given by

$$\nabla_{b_{l,m,k}} J(\underline{X}, B) = \underline{X}^T F^{k^T} S_y^{-m} S_x^{-l} D^T \Bigg(2 \cdot \zeta_2^k(\underline{X}) W_2^k (\sum_{l=-\rho}^{\rho} \sum_{m=-\rho}^{\rho} b_{l,m} D S_x^l S_y^m F \underline{X} - \underline{Y}^k) + \zeta_1^k(\underline{X}) W_1^k \mathrm{sign}(\sum_{l=-\rho}^{\rho} \sum_{m=-\rho}^{\rho} b_{l,m} D S_x^l S_y^m F \underline{X} - \underline{Y}^k) \Bigg).$$

The steepest decent updating equations for HR image ($\underline{X}$) and blurring coefficients ($b_{l,m,k}$) can be given as

$$\underline{X}^{(n+1)} = \underline{X}^{(n)} - \beta_X^{(n)} \nabla_{\underline{X}^{(n)}} J(\underline{X}, B), \qquad b_{l,m,k}^{(n+1)} = b_{l,m,k}^{(n)} - \beta_B^{(n)} \nabla_{b_{l,m,k}^{(n)}} J(\underline{X}, B),$$

where $\underline{X}^{(n)}$ and $b_{l,m,k}^{(n)}$ are $\underline{X}$ and $b_{l,m,k}$ at the n^{th} iteration respectively. At each iteration, the blurring coefficients are constrained to be bi-symmetric, non-negative, and unit-sum.

3.4. Adaptive Regularization and Step Size

The regularization parameter, $\lambda(\underline{X}, B)$, controls the trade-off between fidelity to the data and smoothness of the solution. The following choice of the regularization parameter [9] can be useful:

$$\lambda(\underline{X}, B) = \tau \Bigg[\sum_{k=1}^{K} \Big\{ \zeta_2^k(\underline{X}) \| D B^k F^k \underline{X} - \underline{Y}^k \|_{W_2}^2 + \zeta_1^k(\underline{X}) \| D B^k F^k \underline{X} - \underline{Y}^k \|_{W_1}^1 \Big\} + \lambda(\underline{X}, B) \| C \underline{X} \|_p^p \Bigg], \tag{7}$$

implying

$$\lambda(\underline{X}, B) = \frac{\sum_{k=1}^{K} \Big\{ \zeta_2^k(\underline{X}) \| D B^k F^k \underline{X} - \underline{Y}^k \|_{W_2}^2 + \zeta_1^k(\underline{X}) \| D B^k F^k \underline{X} - \underline{Y}^k \|_{W_1}^1 \Big\}}{1/\tau - \| C \underline{X} \|_p^p},$$

where τ is chosen so that $\lambda(\underline{X}^n, B^n)$ is nonnegative; therefore it can be chosen as [8] $\frac{1}{\tau} \geq \|C\underline{X}\|_2^2 = \|C\|_2^2 \|\underline{X}\|_2^2 \geq \|\underline{X}\|_2^2$.

The step size $\beta_{\underline{X}}^{(n)}$ and $\beta_B^{(n)}$ are calculated by minimizing the cost function $J(\underline{X}^{(n+1)}, B^{(n)}) = J(\underline{X}^{(n)} - \beta_{\underline{X}}^{(n)} \nabla_{\underline{X}^{(n)}} J, B^{(n)})$ and $J(\underline{X}^{(n)}, B^{(n+1)}) = J(\underline{X}^{(n)}, B^{(n)} - \beta_B^{(n)} \nabla_{B^{(n)}} J)$ with respect to $\beta_{\underline{X}}^{(n)}$ and $\beta_{B^k}^{(n)}$ respectively. Then $\beta_{\underline{X}}^{(n)}$ and $\beta_{B^k}^{(n)}$ are obtained as $\beta_{\underline{X}}^{(n)} = \mathcal{A}^{(n)} / \mathcal{B}^{(n)}$, where

$$\mathcal{A}^{(n)} = \sum_{k=1}^{K} \zeta_2^k(\underline{X}^{(n)}) (\nabla_{\underline{X}^{(n)}} J)^T F^{k^T} B^{k^T} D^T W_2^k (D B^k F^k \underline{X}^{(n)} - \underline{Y}^k) + \lambda(\underline{X}^{(n)}, B) (\nabla_{\underline{X}^{(n)}} J)^T C^T C \underline{X}^{(n)},$$

$$\mathcal{B}^{(n)} = \sum_{k=1}^{K} \zeta_2^k(\underline{X}^{(n)}) (\nabla_{\underline{X}^{(n)}} J)^T F^{k^T} B^{k^T} D^T W_2^k D B^k F^k \nabla_{\underline{X}^{(n)}} J + \lambda(\underline{X}^{(n)}, B) (\nabla_{\underline{X}^{(n)}} J)^T C^T C \nabla_{\underline{X}^{(n)}} J,$$

and

$$\beta_{B^k}^{(n)} = \frac{F^{k^T} B^{k^T} D^T (D B^k F^k \underline{X}^{(n)} - \underline{Y}^k)}{\nabla_{b_{l,m,k}^{(n)}} J F^{k^T} B^{k^T} D^T (D B^k F^k \nabla_{\underline{X}^{(n)}} J)}.$$

4. SIMULATION RESULTS AND DISCUSSION

For the test, we generated the LR frames by following the observation model as in Eq. (1), where frames are blurred by Gaussian operator (5×5) with different variance for each frame, down-sampled by $1:2$ in each direction, and then distorted by an additive white Gaussian noise with signal-to-noise ratio equal to 30 dB. Motion is estimated using six parameters affine model [1]. Moreover, for all the compared models, we used the same regularization technique, where the bilateral total variation is used. L_2-norm ($p = 2$) is used in the regularization weighted L_2-norm [5, 6] and the proposed model, while in case of L_1-norm model we used L_1-norm in the regularization term as proposed by the authors [3]. The bilateral total variation regularization term is defined as follows:

$$\|C\underline{X}\|_p^p = \sum_{l=-\varrho}^{\varrho} \sum_{m=-\varrho}^{\varrho} \alpha^{|l|+|m|} \|\underline{X} - S_x^l S_y^m \underline{X}\|_p^p. \tag{8}$$

The parameters of the bilateral function are chosen as follows: $\alpha = 0.1$ and $\varrho = 1$. Since only LR versions of $\underline{X}$ are available, we used $1/\tau$ as the summation of squared L_2-norm of these LR images as $1/\tau = 2 \sum_{k=1}^{K} \|\underline{Y}^k\|_2^2$.

To test the effectiveness of the proposed model, experiments are conducted by using four frames with different motion and different blurring operator to estimate HR frame and blurring operators for two different video sequences. We compared the proposed model (Mixed) with previously proposed three models, namely

- globally weighted L_2-norm (Global-2) [5],
- locally weighted L_1-norm (Local-1) [1],
- locally weighted L_2-norm (Local-2) [6].

The performance of each of these models is measured by both objective and subjective measures.

Figures 1(a) and 1(b) show the PSNR of the reconstructed HR frames of Football sequence and Table Tennis sequence respectively. From these figures, we can see that mixed-norm model is the most stable among the other models. In addition, a comparison between the proposed algorithm in case of known and estimated blurring operators is shown in Fig. 1(c). From this figure we can see the effectiveness of the proposed algorithm to estimate blurring operators.

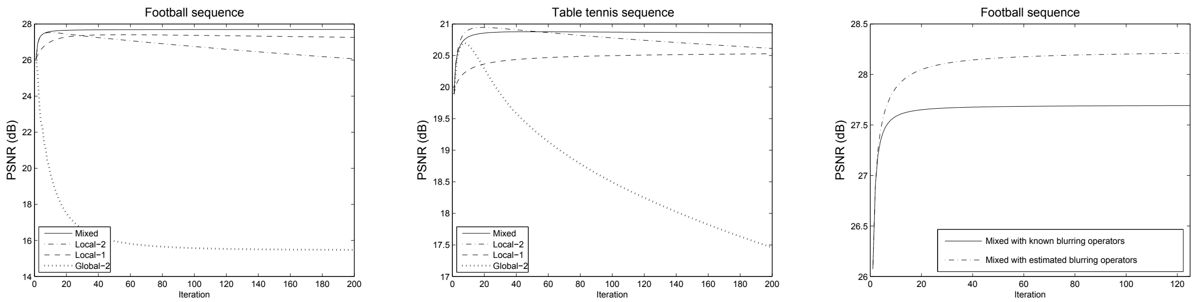
Visual results for Table Tennis and Football sequences are shown in Figs. 2 and 3 respectively. In Figs. 2(c) and 3(c) we can see that even if weighted L_2-norm is suitable for sequences containing locally moving objects, it is affected by the global motion of the whole scene that exists. Also, in Figs. 2 (e) and 3 (e) we can see that because of the existence of locally moving parts with fast motion, weighted L_1-norm fails around these parts because in these parts occlusion may exist therefore median values of the weighted measured frames is not a proper solution. In addition, it is clear in Figs. 2 (f) and 3 (f) that L_2-norm with global weights is not suitable for modeling SR noise especially in case of video sequences that include local motion.

5. REFERENCES

[1] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," Lecture Notes in Comp. Science, Proc. of the Second European Conf. on Comp. Vision, Vol. 588, pp. 237-252, 1992.

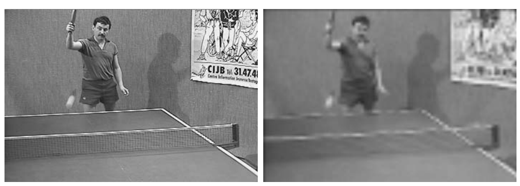
[2] M. Elad and Y. Hel-Or, "A fast super-resolution reconstruction algorithm for pure transnational motion and common space invariant blur," IEEE Trans. Image Processing, Vol.10, no.8, pp.1187-1193, August 2001.

[1] This algorithm is based on [3] but we used iterative weight by using the proposed weights for comparison.

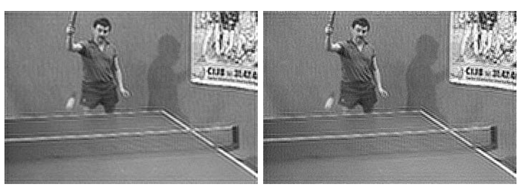


(a) Football sequence (b) Table tennis sequence (c) Blurring estimation evaluation

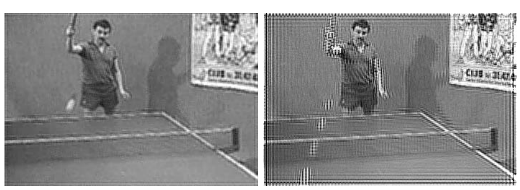
Fig. 1: PSNR for reconstructed SR images.



(a) Original HR frame (b) Bilinear interpolation

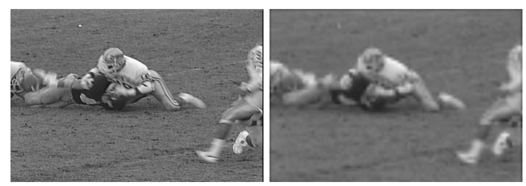


(c) Mixed (d) Local-2

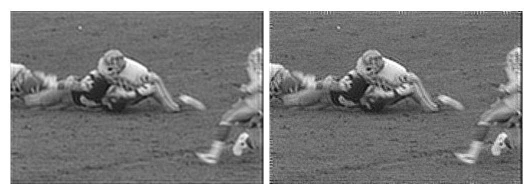


(e) Local-1 (f) Global-2

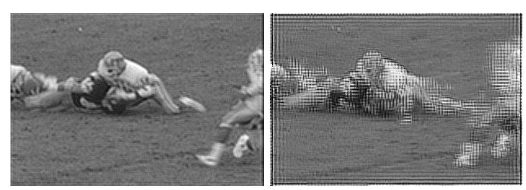
Fig. 2: Estimated HR images using different SR noise models for Table Tennis sequence.



(a) Original HR frame (b) Bilinear interpolation



(c) Mixed (d) Local-2



(e) Local-1 (f) Global-2

Fig. 3: Estimated HR images using different SR noise models for Football sequence.

[3] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Robust shift-and-add approach to super-resolution," in Proc. of the 2003 SPIE Conf. on Applications of Digital Signal and Image Processing, San Diego, California, August 2003.

[4] E. S. Lee and M. G. Kang, "Regularized adaptive high-resolution image reconstruction considering inaccurate sub-pixel registration," IEEE Trans. Image Processing, vol. 12, no. 7, pp. 826-837, July 2003.

[5] H. He and L. P. Kondi, "An image super-resolution algorithm for different error levels per frame," IEEE Trans. Image Processing, vol. 15, no. 3, pp. 592-603, March 2006.

[6] O. A. Omer and T. Tanaka, "Multiframe image and video super-resolution algorithm with inaccurate motion registration errors rejection," in Visual Communication and Image Processing 2008, January 2008.

[7] M. C. Hong, M. G. Kang, and A. Katsaggelos, "An iterative weighted regularized algorithm for improving the resolution of video sequences," in Proc. of Int. Conf. Image Processing, (ICIP 1997) vol. 2, pp. 474-477, October 1997.

[8] M. C. Hong, T. Stathaki, and A. K. Katsaggelos, "Iterative regularized least-mean mixed norm image restoration," Optical Engineering, vol. 41, issue 10, pp. 2515-2524, October 2002.

[9] M. G. Kang and A. K. Katsaggelos, "General choice of the regularization functional in regularized image restoration," IEEE Trans. Image Processing, vol. 4, no. 5, pp. 594-602, May 1995.