Region-Based Super Resolution for Video Sequences Considering Registration Error

Osama A. Omer* and Toshihisa Tanaka

Department of Electrical and Electronic Engineering, Tokyo University of Agriculture and Technology, 2-24-16, Nakacho, Koganei-shi, Tokyo 184-8588, Japan
osama@sip.tuat.ac.jp, tanakat@cc.tuat.ac.jp

Abstract. Super-resolution (SR) for video sequences is a technique to obtain a higher resolution image by fusing multiple low-resolution (LR) frames of the same scene. In a typical super-resolution algorithm, image registration is one of the most affective steps. The difficulty of this step results in the fact that most of the existing SR algorithms can not cope with local motions because they assume global motion. In this paper, we propose a SR algorithm that takes into account inaccurate estimates of the registration parameters and the point spread function. When frames obey the assumed global motion model, these inaccurate estimates, along with the additive Gaussian noise in the low-resolution image sequence, result in different noise level for each frame. However, in case of existence of local motion and/or occlusion, regions that have local motion and/or occlusion have different noise level. To cope with this problem, we propose to adaptively weight each segment according to its reliability. The segments are generated by segmenting the reference frame using watershed segmentation. The experimental results using real video sequences show the effectiveness of the proposed algorithm compared to three state-of-the-art SR algorithms.

Keywords: Super-resolution, affine model, image registration, resolution enhancement, region-based global weight.

1 Introduction

In many applications such as remote sensing, video surveillance, and medical diagnostics, the demand for high-resolution images is gradually increasing since high resolution images offer more details that provide to the viewer. One way to obtain high-resolution images is to physically reduce the pixel size and therefore increase the number of pixels per unit area. However, since a reduction of pixel size causes a decrease in the amount of light, shot noise is generated that severely degrades the image quality. Instead of altering the sensor manufacturing technology, digital image processing methods to obtain a high resolution image from low-resolution observations have been investigated by many researchers [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19].

Super-resolution (SR) is an approach to obtain HR image(s) from a set of low-resolution (LR) images. The most important steps of SR algorithms are image registration and data fusion. Image registration process has been paid more attention for

* Part of this work has been done while the first author with Nokia R&D Tokyo-Japan center. This work is supported in part by Egyptian Ministry of High Education.

T. Wada, F. Huang, and S. Lin (Eds.): PSIVT 2009, LNCS 5414, pp. 944–954, 2009.

the last two decades [1, 2]. However, image registration of images containing locally moving parts is still a challenging task. Data fusion is the process which fuses the registered images to construct the HR image. In most recently proposed SR algorithms [3, 4, 5, 6, 7, 8, 9, 10] the SR results depend on fusion step. As a cost function, the L_2- or L_1-norm is used to fuse LR images [3, 4, 5]. Also, the weighted L_2-norm is used measure function in [6, 7, 8, 9, 10]. The main problems of the previous SR algorithms are as follows. Even if L_2-norm is suitable for Gaussian noise, it is implicitly assumed that the extra resolution content is equally distributed among all LR images [3]. Therefore the result is obtained by averaging the contributions from all LR images. The averaging process leads to propagation of the outlier pixels from any of the LR images into the HR image which means that it fails with local motion. In spite of the fact that L_1-norm is suitable for Laplacian noise and robust against outliers[4, 5], it can not cope with errors resulting from occlusion that happens in video sequences containing local motions. The failure of this algorithm in case of occlusion is due to that, the model converges to the median over the measured data without pre-weighting the LR images, which may lead to failure in case of existence of locally moving parts in the scene.

To overcome the problems of registration error in locally moving parts, three techniques are appeared in the literature. The first is to use different global (or local) weight for different registration error level [6, 7, 8, 9, 10, 11]. The main idea behind this technique is to weight the frames (or pixels) that have high registration error with small weight or even discard them. The second is to use local motion (or multi-motion) estimation to improve the accuracy of registration in the locally moving parts [12, 13, 14]. The main idea behind this technique is to incorporate information from different frames as much as possible. The third is a combination of the previous categories.

The idea of using different weight for different registration error levels is used in the super resolution literature [6, 7, 8, 9, 10]. This idea is based on rejecting pixels or even whole frames that have high registration error. Two categories are used in the literature, the first is the global weighting [6, 7, 8], where each frame is weighted with certain weight based on the error in the whole frame. The second is the local weighting [9, 10], where each pixel has its individual weight. In [6], for each frame the weight is chosen so that it decreases as the error increases and increases as the smoothness of the HR image increases. Three weighting functions are presented in [6], namely, linear, square root and logarithmic functions. The main problem of this method is that it globally weights each frame, then in case of existence of occlusion and/or local motion the whole frame will be weighted with small weight even if only the parts that have occlusion and/or local motion parts are inaccurately registered then they should be penalized by small weight. Moreover, assuming affine motion model can result in different error level for different region if the actual motion is projective motion. A similar algorithm is proposed in [7]. In this algorithm, the authors proposed to use a global weight for each frame and also have different regularization parameter so that as the error increases the weight decreases and the regularization parameter increases. This algorithm suggests different regularization for different frame, however it is still have the same problem as that in [6] in case of existence of occlusion and/or local motion. In [8], the global weighting function is used as exponential function of the registration error. This

algorithm is a modified version of [6], where the convexity of weighting function is taken into consideration.

On the other hand, the use of local weights has been proposed [9, 10, 11, 12, 13]. However, in [9], the weights are not adaptive to the registration error and then it can not cope with the error resulting from inaccurate image registration. In [13], weights are determined based on the information about the corrupting noise which is not known in practical. In [11], a pixel-level selection strategy for outlier rejection is proposed. In this algorithm, a similarity measure is used to determine the reliability of each LR pixel in the SR estimation process. Pixel is discarded from the estimation process if the measure evaluated at this pixel location is greater than certain threshold. The performance of this algorithm highly depends on the selection of the threshold. In [10], local weights have been selected for each pixel using an exponential function. The registration error (the absolute difference between the reference frame and the warped frames) has been used as exponent. The main problem of this algorithms is that the local weights are sensitive to noise because they are determined pixel by pixel.

In addition, region segmentation has been used before to enhance resolution [15, 16, 17, 18]. In [15, 16], a region-based super-resolution algorithm is proposed in which different filters are used according to the type of region. But in this method the segmentation information is not fully used where it is used only to classify regions into homogeneous and inhomogeneous regions. In [17], the image is segmented into different types of regions according to the local higher order statistics (HOS). The weight function of the regularization term is determined by the segmentation label. This method achieves anisotropic diffusion for edge pixel and isotropic diffusion for pixel in smooth region. In [18], the image is segmented into background and different objects and each of these are super-resolved separately using traditional technique [19] and then the super-resolved regions are merged to construct the HR image.

The motivation of this paper is therefore to develop a robust algorithm that can cope with the local registration errors even by using global motion estimation technique. To do that, we propose to segment the reference frame into arbitrary shaped regions and to use a global weight for each region. For each region, the weight is adjusted so that region with high registration error (due to local motion in this region) will be considered with small weight or even discarded depending on the amount of error. This technique can achieve better results than both global and local weighting techniques because it combines the advantages of both techniques, where weights are less sensitive to noise and also regions which don't suffer from registration error will not be affected by weights.

2 Problem Description and Error Modeling

2.1 Observation Model

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 following degradation process:

$$\underline{Y}_k = D_k H_k F_k \underline{X} + \underline{V}_k, \tag{1}$$

where F_k, H_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. We assume that H_k are constant for all the K frames ($H_k = H$ for all $1 \leq k \leq K$), and D_k are constant for all the K frames ($D_k = D$ for all $1 \leq k \leq K$). Then the degradation model is simplified as

$$\underline{Y}_k = DHF_k \underline{X} + \underline{V}_k. \tag{2}$$

Throughout the paper, we assume that D and H are known and the additive noise is Gaussian with zero mean. Therefore the problem here in this paper is to find the original image $\underline{X}$.

2.2 Iterative Super-Resolution

To avoid matrixes inversion, super-resolution problem is usually solved iteratively. In order to minimize the error function in (3), the method of iterative gradient descent is commonly employed [9].

$$J(\underline{X}) = \sum_{k=1}^{K} \rho\left(DHF_k \underline{X} - \underline{Y}_k\right) + \lambda Z(\underline{X}) \tag{3}$$

where ρ is a general data fidelity function, Z is the property function and λ is the regularization parameter. This optimization technique seeks to converge towards a local minimum following the trajectory defined by the negative gradient. That is, at iteration n, the high-resolution image according to observation $\underline{Y}^k$, is updated as

$$\underline{X}^{n+1} = \underline{X}^n + \beta \sum_{k=1}^{K} \underline{R}_k^n \tag{4}$$

where $\underline{R}_k^n$ is the residual gradient at for frame k at iteration n. It is computed as

$$\underline{R}_k^n = F_k^T H^T D^T \psi\left(DHF_k \underline{X}^n - \underline{Y}_k\right) + \lambda \Phi(\underline{X}^n) \tag{5}$$

where ψ is the gradient of the data fidelity term, and Φ is the gradient of the regularization term. This equation reveals that the iterative super-resolution method is in fact an iterative fusion of the gradients of the cost function. Using this idea, Mejdi et. al proposed to use LMS-based adaptive weight for gradient at each pixel. Also, the global weighting method can be seen as globally weighting the gradient at each frame.

3 Region-Based Weight for Super-Resolution

In this paper we propose to use exponential function to globally weight each region rather than the whole frame. It is assumed that each region have the same motion and

Algorithm 1. The proposed algorithm.

Pre-compute:

1. register low-resolution frames with respect to the reference frame using Lucas-Kanade affine motion model [2],
2. segment the reference frame into sub-regions using watershed segmentation.

Iterate until convergence:

1. determine weights for each region using Eqs. 6 to 8,
2. update HR image using steepest decent using Eq. 4.

then have the same error level. Therefore, weighting each region with same weight is a reasonable choice. The weight for region $\Re_i$ in frame k is weighted as follows: let the error vector at frame k be

$$\underline{E}^k = DHF^k\underline{X} - \underline{Y}^k, \tag{6}$$

we define the error value at each region as

$$E^k_{\Re_i} = \frac{1}{N_{\Re}} \sum_{j \in \Re_i} |\underline{E}^k(j)| \tag{7}$$

and $N_{\Re_i}$ is the number of pixels in region $\Re_i$. To spread the range of $E^k_{\Re_i}$ between 0 and 1, the values $E^k_{\Re_i}$ for all regions are divided by the maximum value for the same region among the frames $P_{\Re_i}$. Then the weight at each region is used as

$$W^k_{\Re_i} = \exp\left(-\frac{E^k_{\Re_i}}{P_{\Re_i}}\right), \tag{8}$$

The weights are normalized so that the summation of the weights for the same region equals the number of frames. The whole algorithm is described in 1. The data fidelity term in the error function is used as the weighted L_1-norm. While the regularization term is used as bilateral total variation [4, 5]. The updating equation can be described as:

$$\begin{aligned}\underline{X}^{n+1} = \underline{X}^n + \beta \Bigg\{ & \sum_{k=1}^{K} {F^k}^T H^T D^T W^k sign\left(DHF^k\underline{X}^n - \underline{Y}^k\right) \\ & + \lambda \sum_{l=-P}^{P} \sum_{m=-P}^{P} \alpha^{|l|+|m|} \left(\mathcal{I} - S_x^{-l} S_y^{-m}\right) sign\left(\underline{X}^n - S_x^l S_y^m \underline{X}^n\right) \Bigg\}\end{aligned} \tag{9}$$

where S_x^l and S_y^m are the shifting operators in x and y by l and m respectively, S_x^{-l} and S_y^{-m} are the inverse operator for S_x^l and S_y^m respectively, $sign$ is the signum function, and $0 < \alpha < 1$, β is a scalar representing the step size in the direction of the gradient.

4 Simulation Results and Discussion

4.1 Data Set

For test, two different video sequences including Table Tennis and Mobile sequence are tested. Both of the two sequences are in SIF format (240×352). Color images are commonly represented by the RGB channels. However, humans are more sensitive to changes in luminance than to changes in color. Thus, instead of using the RGB color model, we use the YCbCr color space where the Y channel represents luminance and the Cb and Cr channels represent chromaticity. In our method, the chromaticity components from the Cb and Cr channels are simply interpolated using bicubic interpolation from the low-resolution image to the target high-resolution image. Hence, only the luminance values from the Y channel are used in the resolution enhancement process. Moreover, we assumed that the sequence is already demosaicked or captured by three CCD sensors.

4.2 Experiment Setup

To test the efficiency of the proposed region-based weight, we compared the proposed algorithm with three state-of-the-art SR algorithms, namely L_2-norm [3], L_1-norm [4] and frame-based weighted L_2-norm [8]. In the simulation, we used 20 steepest decent iterations for all the algorithms.

In the simulation, two scenarios are used to evaluate the efficiency of the proposed algorithm. In the first scenario, we assumed that the available sequence is HR sequence then the LR frames were generated from the original HR video sequences according to the model as in (2), where the frames were blurred by Gaussian operator (5×5 with variance equal 1), down-sampled by a decimation factor of 2 in the horizontal and vertical directions, and distorted by an additive white Gaussian noise with 30 dB signal-to-noise ratio. Then we used different SR algorithm to reverse these operations. This enables us to compare the resulting HR frames with the original HR frame. In the second scenario, we directly applied SR algorithms to enhance the resolution for a given sequence where the original HR frames are unknown. In the following results, we applied the first scenario to the Table Tennis sequence, and we applied the second scenario to the Mobile sequence.

4.3 Results and Discussion

Figure 1 shows the first LR frame and the segmented regions of each of Table Tennis and Mobile sequences. For their importance, the locally moving objects are marked with ellipses or rectangles.

To show the efficiency of the proposed algorithm, zoomed part containing moving objects is shown in Fig. 2. In this figure, zoomed parts of the resulting HR image using different SR algorithms are shown. Obviously shown in this figure that using L2-norm is sensitive to registration error which is obvious at the locally moving parts as ball and train in this example (see Fig. 2a) where the projection of the registration error of each LR frame appears. Also, although it is robust to registration error, L1-norm cannot cope

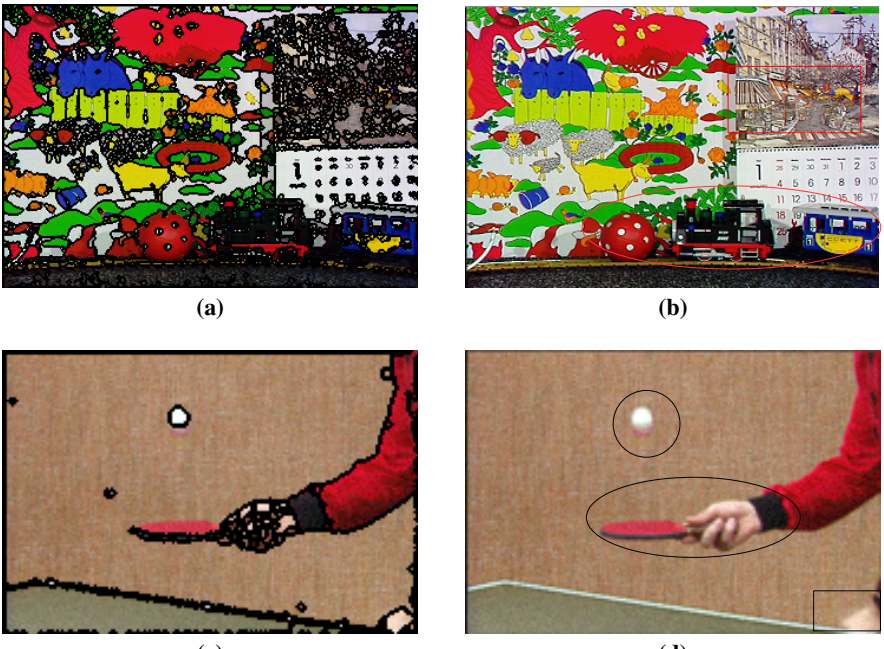

Fig. 1. Mobile sequence: (a) Segmented regions, (b) LR frame, and Table Tennis sequence: (c) Segmented regions, (d) LR frame

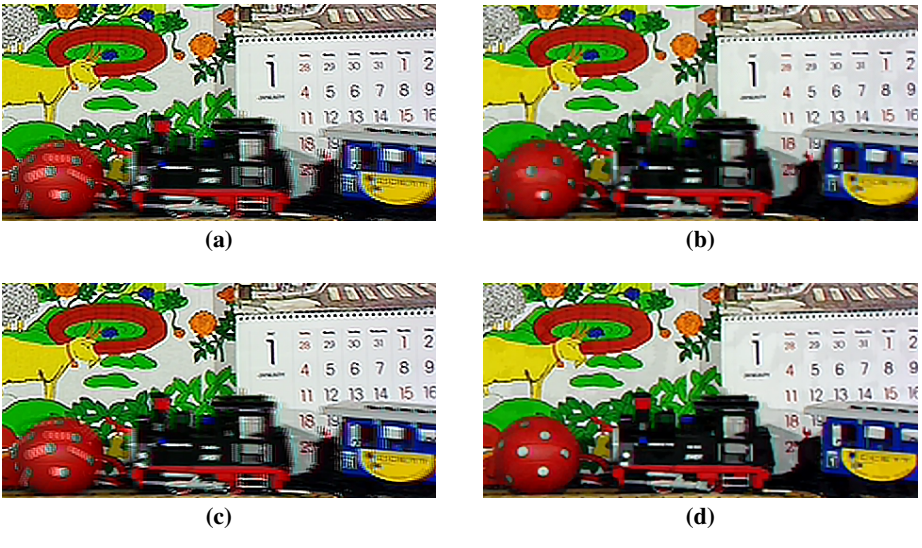

Fig. 2. Mobile sequence: HR frame using; (a) L2-norm [3], (b) L1-norm [4], (c) Global weighted L2-norm [8], and (d) proposed algorithm

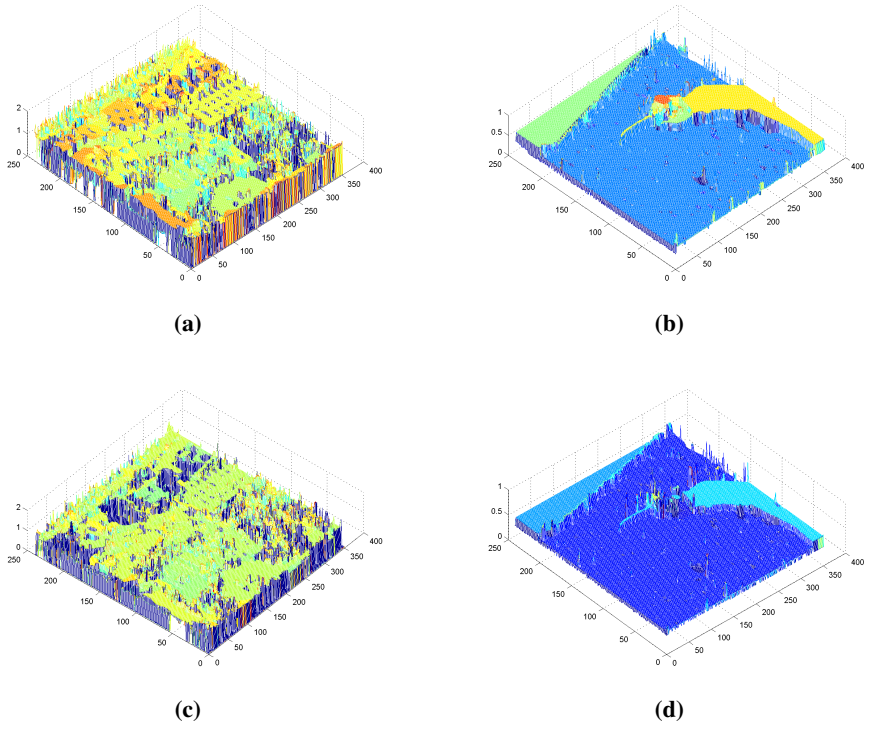

Fig. 3. Local weights for the second frame using proposed method for: (a) mobile sequence, and (b) Tennis sequence, Local weights for the fourth frame for: (c) mobile sequence, and (d) Tennis sequence

with the local registration error. Instead of projecting the registration error of all LR frames, L1-norm select the median over the LR frames which is not suitable for local registration error as shown in Fig. 2b. In addition, using frame-based weight is suitable in case of global registration error and when the area of local errors is big so that global weight can be dominated by these local errors and then frames containing this error can be discarded. However, in case of small moving objects the global weight (frame-based weight) is not suitable as shown in Fig. 2c. On the other hand, using region-based weighting function can overcome the problem of local motion and/or occlusion as shown in Fig. 2d.

Another example to demonstrate the effectiveness of the proposed algorithm is shown in Fig. 4. In this example, the LR frames are generated from known HR frames using the model in Eq. 2. The original HR frame is shown in Fig. 4a. This sequence contains three locally moving objects as marked in Fig. 1d. Due to averaging over the LR frames, L2-norm deforms the moving objects as shown in Fig. 4b where the upper hand and ball are repeated and the lower hand partially disappeared. Also, using L1-norm and frame-based weight still have the same problems with locally moving objects as lower hand and ball. On the other hand, using region-based weight overcame the problem of local motion in most of regions as ball, upper and lower hand. However, a small

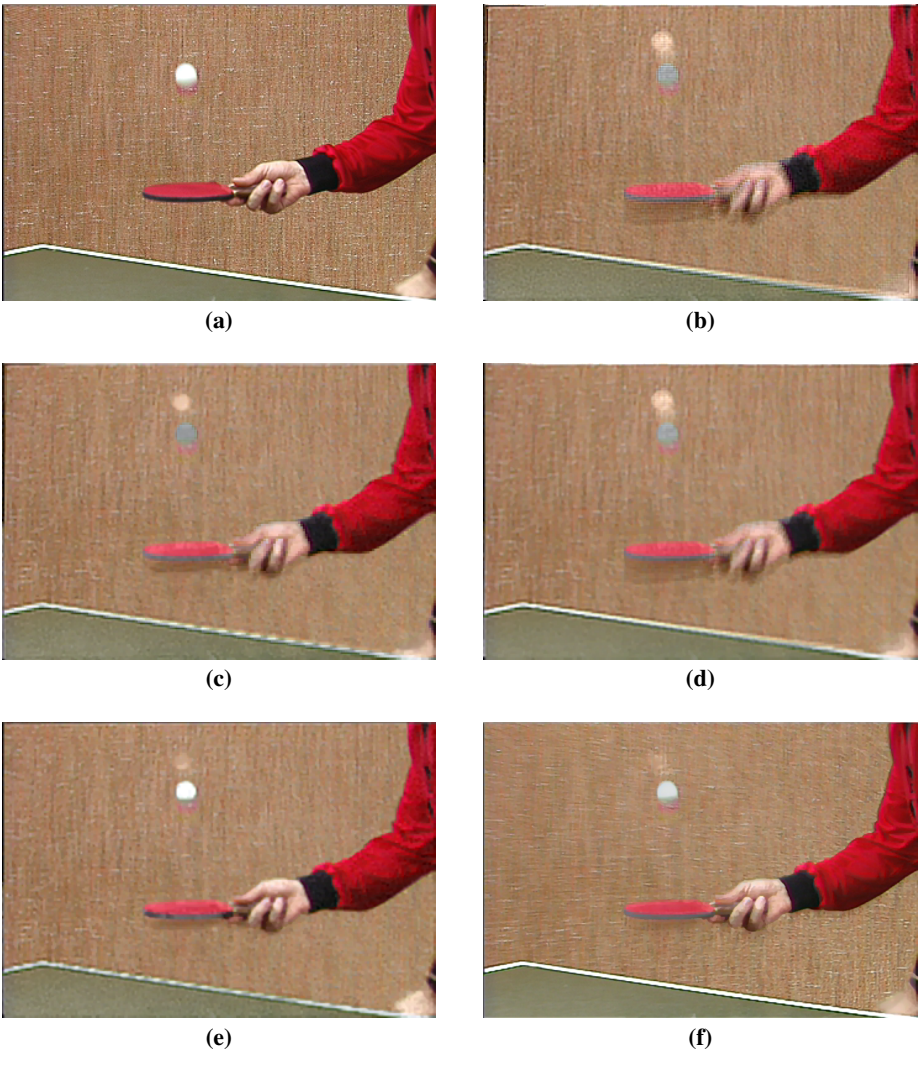

Fig. 4. Table Tennis sequence, (a) Original HR frame, HR frame using; (b) L2-norm [3], (c) L1-norm [4], (d) Global weighted L2-norm [8], (e) proposed algorithm, and the HR frame beyond the available resolution using proposed algorithm

deformation in the background at the occluded part appears. This deformation is due to that the weights of some inaccurately registered regions are very small but not zero so small error still appears in the HR frame as shown in Fig. 4e. Moreover, to show the efficiency of the proposed algorithm, the resolution is increased beyond the available resolution (the resolution of the original frames) as shown in Fig. 4f.

The weights for two different frames (second and fourth frames) for the tested sequences are plotted in three dimensions in Fig. 3. This figure shows how the proposed weighting function penalizes the local registration error for each region.

5 Conclusion

In this paper, we presented an algorithm for image and video resolution enhancement. The proposed algorithm takes into account inaccurate estimates of the registration parameters and the point spread function. These inaccurate estimates, along with the additive Gaussian noise in the low-resolution image sequence, result in different noise level for each frame. However, in case of existence of local motion and/or occlusion, regions that have local motion and/or occlusion have different noise level. The proposed algorithm is based on global weighting for each region. The weights are determined globally for each region. The regions are determined by segmenting the reference frame into sub-regions using watershed segmentation. The proposed algorithm can cope with the local motion and occlusion problems. Affine motion model is assumed. For color video sequences, only the luminance is processed with the super-resolution algorithm while the chrominance is interpolated using bicubic interpolation. Also, the sequences are assumed to be demosaicked or being captured by three CCD sensors.

References

1. Bergen, J.R., Anandan, P., Hanna, K.J., Hingorani, R.: Hierarchical model-based motion estimation. In: Sandini, G. (ed.) ECCV 1992. LNCS, vol. 588, pp. 237–252. Springer, Heidelberg (1992)
2. Lucas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proceedings of the International Joint Conference on Artificial Intelligence (1981)
3. Elad, M., Hel-Or, Y.: A fast super-resolution reconstruction algorithm for pure transnational motion and common space invariant blur. IEEE Trans. on Image Processing 10(8), 1187–1193 (2001)
4. Farsiu, S., Robinson, D., Elad, M., Milanfar, P.: Fast and robust multi-frame super-resolution. IEEE Trans. on Image Processing 13(10), 1327–1344 (2004)
5. Farsiu, S., Robinson, D., Elad, M., Milanfar, P.: 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)
6. Lee, E.S., Kang, M.G.: Regularized adaptive high-resolution image reconstruction considering inaccurate subpixel registration. IEEE Trans. on Image Processing 12(7) (July 2003)
7. He, H., Kondi, L.P.: An image super-resolution algorithm for different error levels per frame. IEEE Trans. on Image Processing 15(3), 592–603 (2006)
8. Park, M.K., Kang, M.G., Katsaggelos, A.K.: Regularized Super-Resolution Image Reconstruction Considering Inaccurate Motion Information. SPIE Optical Engineering 46(11), 117004-1–117004-12 (2007)
9. Trimeche, M., Ciprian Bilcu, R., Yrjanainen, J.: Adaptive outlier rejection in image super-resolution. EURASIP Journal on Applied Signal Processing 2006, Article ID 38052 (2006)
10. Omer, O.A., Tanaka, T.: Multiframe image and video super-resolution algorithm with inaccurate motion registration errors rejection. In: Proc. of the 2008 SPIE Conf. on Visual Communication and Image Processing, San Jose, California (January 2008)
11. Ivanovski, Z.A., Panovski, L., Karam, L.J.: Robust super-resolution based on pixel-level selectivity. In: Proceedings of SPIE, vol. 6077 (2006)
12. Schultz, R.R., Stevenson, R.t.L.: Extraction of high-resolution frames from video sequences. IEEE Trans. on Image Processing 5(6) (June 1996)

13. Zhao, W.Y., Sawhney, S.: Is super-resolution with optical flow feasible? In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2350, pp. 599–613. Springer, Heidelberg (2002)
14. Andrew, J., Patti, M.I.: Robust methods for high-quality stills from interlaced video in the presence of dominant motion. IEEE Trans. on Circuits and Systems for Video Technology 7(2) (April 1997)
15. Choi, B., Ra, J.B.: Region-based super-resolution using multiple blurred and noisy undersampled images. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2, pp. 609–612 (2006)
16. Choi, B., Kim, S.D., Ra, J.B.: Region-based super-resolution using adaptive diffusion regularization. Optical Engineering 47(2), 027006 (February 2008)
17. Qiao, J., Liu, J.: HOS-based image super-resolution reconstruction. In: Sebe, N., Liu, Y., Zhuang, Y.-t., Huang, T.S. (eds.) MCAM 2007. LNCS, vol. 4577, pp. 213–222. Springer, Heidelberg (2007)
18. van Eekeren, A., Schutte, K., Dijk, J., de Lange, D.J.J., van Vliet, L.J.: Super-resolution on moving objects and background. In: Proc. Int. Conf. Image Processing (ICIP 2006), vol. 2, pp. 2709–2712 (2006)
19. Hardie, R.C., Barnard, K.J., Bognar, J.G., Armstrong, E., Watson, E.A.: High-resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system. Optical Engineering 37(1), 247–260 (1998)
20. De Smet, P., De Vleschauwer, D.: Performance and scalability of highly optimized rainfalling watershed algorithm. In: Proc. Int. Conf. on Imaging Science, Systems and Technology, CISST 1998, Las Vegas, NV, USA, pp. 266–273 (July 1998)