REGION-BASED WEIGHTED-NORM APPROACH TO VIDEO SUPER-RESOLUTION WITH ADAPTIVE REGULARIZATION

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 propose a super-resolution (SR) algorithm that takes into account inaccurate estimates of the registration parameters. 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 region according to its reliability and the regularization parameter is simultaneously estimated for each region. The regions 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.

***Index Terms*—** Super-resolution, affine model, image registration, region-based global weight, adaptive regularization

1. INTRODUCTION

In many applications such as remote sensing and video surveillance the demand for extracting a high-resolution image from low resolution video sequence 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 [2–10].

Super-resolution (SR) is an approach to obtain high-resolution (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 the last two decades [1]. Data fusion is the process which fuses the registered images to construct the HR image. However, registration of images containing locally moving objects is still a challenging task. To overcome the problems of registration error in locally moving parts, three techniques appeared in the literature. The first is to use different global (or local) weight for different registration error level [4–7]. 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 [8]. 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 [7]. Among them, the first technique is widely used in super-resolution.

As stated before, the idea of using different weight for different registration is based on rejecting pixels or even whole frames that have high registration error. We can categorize existing techniques into two classes. The first is the global weighting [4], where each frame is weighted with certain weight based on the error in the whole frame. The second is the local weighting [5, 6], where each pixel has its individual weight. In [4], it has been 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 has a problem in case of existence of occlusion and/or local motion. In [5], local weights have been proposed, however, the weights are not adapted to registration error, thus the algorithm in [5] can not cope with the error resulting from inaccurate image registration. In [6], 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 have been used before to enhance resolution [9, 10]. In [9], 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 [10], the image is segmented into background and different objects and each of these are super-resolved separately using traditional technique and then the super-resolved regions are merged to construct the HR image. This algorithm is very complex since it requires detection of moving objects and registration of each object separately.

Hence, the motivation of this paper is to develop a robust algorithm that takes into account inaccurate registration in case of existence of moving objects. To do that, we propose to segment the reference frame into arbitrary shaped regions and to use a global weight for each region and the regularization parameter is simultaneously estimated 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. Also, the regularization parameter increases as the registration error increases and decreases as the smoothness of the region increases. This technique can achieve bet-

This work is supported in part by Egyptian Ministry of High Education.

ter 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

2.1. Observation Model

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

$$\underline{Y}_k = DHF_k\underline{X} + \underline{V}_k. \tag{1}$$

where F_k, H and D are the motion operator of the k^{th} frame, 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 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 matrices inversion, super-resolution problem is usually solved iteratively. The method of iterative gradient descent is commonly employed [5] in order to minimize the following error function:

$$J(\underline{X}) = \sum_{k=1}^{K} \rho\left(DHF_k\underline{X} - \underline{Y}_k\right) + \lambda_k Z(\underline{X}) \tag{2}$$

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

$$\underline{X}^{n+1} = \underline{X}^n + \beta \sum_{k=1}^{K} \underline{R}_k^n \tag{3}$$

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

$$\underline{R}_k^n = F_k^T H^T D^T \psi\left(DHF_k\underline{X}^n - \underline{Y}_k\right) + \lambda_k^n \Phi(\underline{X}^n) \tag{4}$$

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* [5] 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

3.1. Proposed Cost Function

Two cost functions are proposed, namely, the adaptively regularized weighted L_1-norm and the adaptively regularized weighted L_2-norm. Then the data fidelity terms and the regularization terms are described as follows:

$$\rho_1\left(DHF_k\underline{X} - \underline{Y}_k\right) = \left\|DHF_k\underline{X} - \underline{Y}_k\right\|_{1,W^k}^1,$$
$$\rho_2\left(DHF_k\underline{X} - \underline{Y}_k\right) = \left\|DHF_k\underline{X} - \underline{Y}_k\right\|_{2,W^k}^2,$$

and

$$Z_1(\underline{X}) = \sum_{l=-P}^{P}\sum_{m=-P}^{P} \alpha^{|l|+|m|} \left\|\underline{X} - S_x^l S_y^m \underline{X}\right\|_1^1, \qquad Z_2(\underline{X}) = \left\|C\underline{X}\right\|_2^2,$$

where weighted L_1- and L_2-norms for vector $\underline{X}$ of size M^2 are defined as

$$\|\underline{X}\|_{1,W^k}^1 = \sum_{i=1}^{M^2} W^k(i,i)|\underline{X}(i)|,$$
$$\|\underline{X}\|_{2,W^k}^2 = \left(\underline{X}^T W^k \underline{X}\right)^{1/2},$$

respectively, S_x^l and S_y^m are the shifting operators in x and y by l and m respectively, λ_k is a matrix that contains a regularization parameter corresponding to each region, W^k is the weighting matrix containing different weight for different region, β is a scalar representing the step size in the direction of the gradient, C is a high pass operator and α is a parameter that adjust the bilateral filter ($0 < \alpha < 1$).

3.2. Choice of Weights

In this paper we propose to use an exponential function to globally weight each region rather than the whole frame. Also, the regularization parameter is adaptively estimated for each region. It is assumed that each region have the same motion and 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 matrix at each frame k be

$$\underline{E}^k = DHF^k\underline{X} - \underline{Y}^k, \tag{5}$$

we define the error value at each region as

$$E_{\Re_i}^k = \frac{1}{N_\Re} \sum_{j\in\Re_i} |\underline{E}^k(j)| \tag{6}$$

and $N_{\Re_i}$ is the number of pixels in region $\Re_i$. Then the weight for all pixels in region $\Re_i$ at frame k is used as

$$W_{\Re_i}^k = \exp\left(-E_{\Re_i}^k\right), \tag{7}$$

The weights are normalized so that the summation of the weights for the same region in all frames equals the number of frames.

3.3. Choice of Regularization Parameter

The regularization parameter is adaptively estimated for each region as follows:

1. in case of adaptively regularized weighted L_1-norm

$$\lambda_{\Re_i}^k = \frac{\sum_{j\in\Re_i} |\underline{E}^k(j)|}{\gamma_{\Re_i}^k - \sum_{j\in\Re_i} \mathcal{Z}_1(j)} \tag{8}$$

where $\mathcal{Z}_1(j) = \sum_{l=-P}^{P}\sum_{m=-P}^{P} \alpha^{|l|+|m|} |\underline{X}(j) - \underline{X}^{l,m}(j)|$, $\underline{X}^{l,m} = S_x^l S_y^m \underline{X}$, $\gamma_{\Re_i}^k$ is chosen so that $\lambda_{\Re_i}^k$ is non-negative; therefore it can be chosen as $\gamma_{\Re_i}^k > \sum_{j\in\Re_i} \mathcal{Z}_1(j)$. We used $\gamma_{\Re_i}^k$ as the summation of L_1-norm of these LR region $\Re_i$ as $\gamma_{\Re_i}^k = \sum_{k=1}^{K} \sum_{j\in\Re_i} |Y^k(j)|$.

Algorithm 1 : the proposed algorithm.

Pre-compute:

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

Iterate until convergence:

1. determine weights for each region using Eqs. 5 to 7,
2. determine the regularization parameter for each region using Eq. 8 or 9,
3. update HR image using steepest decent.

2. in case of adaptively regularized weighted L_2-norm

$$\lambda^k_{\Re_i} = \frac{\sum_{j\in\Re_i}(\underline{E}^k(j))^2}{\gamma^k_{\Re_i} - \sum_{j\in\Re_i}\mathcal{Z}_2^2(j)} \tag{9}$$

where $\mathcal{Z}_2 = C\underline{X}$, $\gamma^k_{\Re_i}$ is chosen so that $\lambda^k_{\Re_i}$ is non-negative; therefore it can be chosen as [11] $\gamma^k_{\Re_i} > \sum_{j\in\Re_i}\mathcal{Z}_2^2(j)$. We used $\gamma^k_{\Re_i}$ as the summation of L_2-norm of these LR region $\Re_i$ as $\gamma^k_{\Re_i} = \sum_{k=1}^{K}\sum_{j\in\Re_i}(Y^k(j))^2$.

To avoid high values of the regularization parameters in case of small regions (which is sensitive to noise), we truncated the regularization parameters in these regions so that they don't increase more than 0.2. The whole algorithm is described in Algorithm 1.

4. SIMULATION RESULTS AND DISCUSSION

For test, the Mobile video sequence with CIF format (240×352) is used as a LR sequence. We assumed that the sequence is already demosaicked or captured by three CCD sensors. To test the efficiency of the proposed region-based weight with adaptive regularization, we compared the proposed algorithm with three state-of-the-art SR algorithms, namely L_2-norm [2], L_1-norm [3] and frame-based weighted L_2-norm with adaptive regularization [4]. In the simulation, we used 20 steepest decent iterations for all the algorithms, α is fixed to 0.2.

Figure 1 shows the results of different SR algorithms. In this figure, a zoomed part of the LR frame and the corresponding segments are shown in Figs. 1a and 1b respectively. Figure 1c shows that using L_2-norm is sensitive to registration error which is obvious at the locally moving parts as the ball and train in this example where the projection of the registration error of each LR frame appears. Also, although it is robust to registration error, L_1-norm cannot cope with the registration error for moving objects. Instead of projecting the registration error of all LR frames, L_1-norm select the median over the LR frames which is not suitable for local registration error as shown in Fig. 1d. In addition, using frame-based weight with adaptive regularization 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. 1e. On the other hand, using region-based weights can overcome the problem of local motion and/or occlusion as shown in Figs. 1f and 1g.

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. 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 with adaptive regularization. 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.

6. REFERENCES

[1] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision", in Proc. of the Int. Joint Conf. on Artificial Intelligence, 1981.

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

[3] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multi-frame super-resolution," IEEE Trans. on Image Processing, vol. 13, no. 10, pp. 1327-1344 , October 2004.

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

[5] M. Trimeche, R. C. Bilcu, and J. Yrjanainen, "Adaptive outlier rejection in image super-resolution," EURASIP Journal on Applied Signal Processing, vol. 2006 (2006), Article ID 38052.

[6] O. A. Omer and T. Tanaka, "Multiframe image and video super-resolution algorithm with inaccurate motion registration errors rejection," in Proc. of the SPIE Conf. on Visual Comm. and Image Processing, San Jose, California, January 2008.

[7] R. R. Schultz, R. L. Stevenson, "Extraction of high-resolution frames from video sequences," IEEE Trans. on Image Processing, vol. 5, no. 6, June 1996.

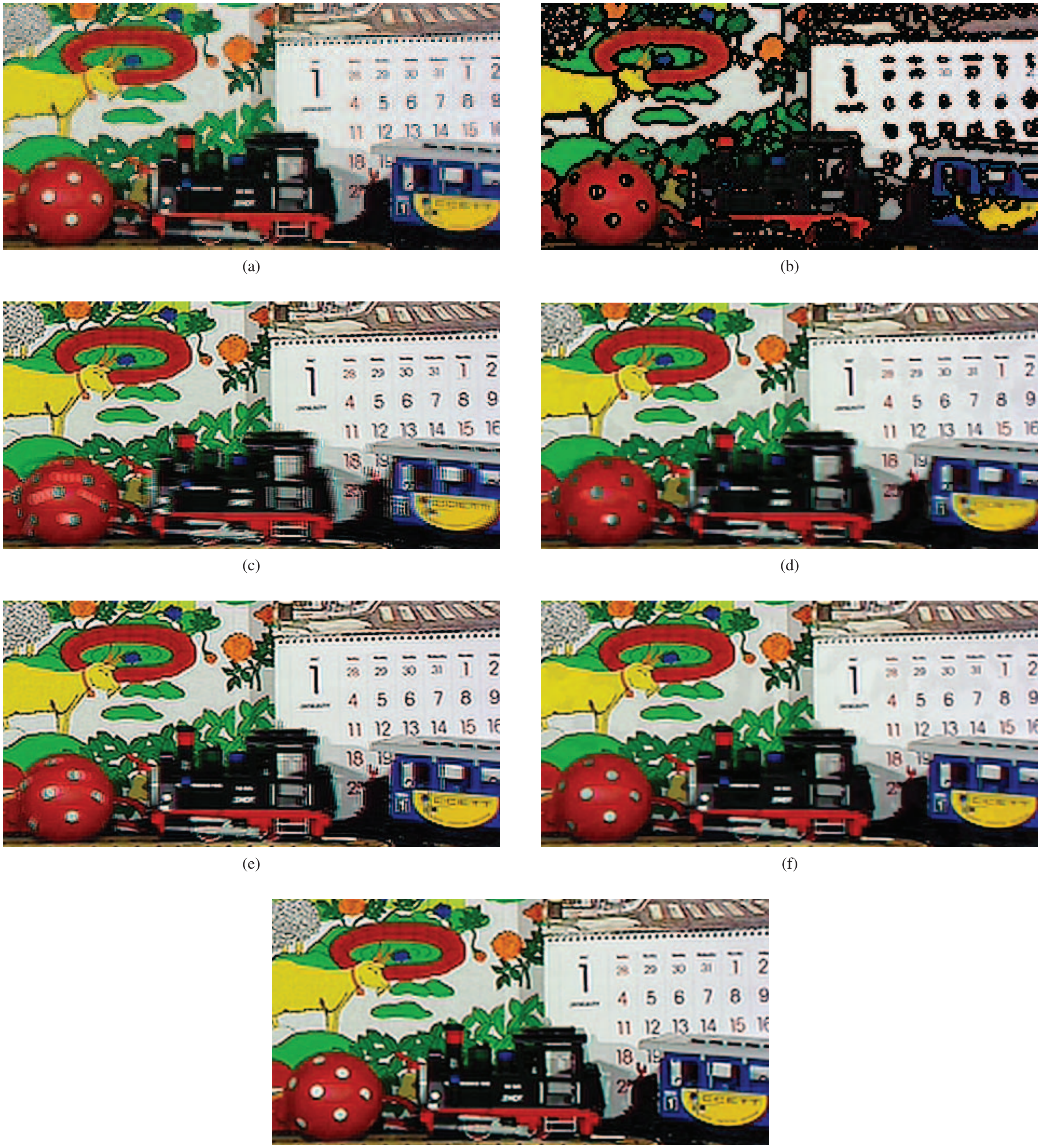
[8] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Robust methods for high-quality stills from interlaced video in the presence of dominant motion," IEEE Trans. on Circuits and Systems for Video Technology, vol. 7, no. 2, April 1997

[9] B. Choi, S.D. Kim and J.B. Ra, "Region-based super-resolution using adaptive diffusion regularization," Optical Engineering 47(2), 027006, February 2008.

[10] A. van Eekeren, K. Schutte, J. Dijk, D.J.J. de Lange and L.J. van Vliet, "Super-resolution on moving objects and background," in Proc. of the Int. Conf. on Image Processing, (ICIP 2006) vol. 2, pp. 2709-2712, 2006.

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

[12] P. De Smet and D. De Vleschauwer, "Performance and scalability of highly optimized rainfalling watershed algorithm," in Proc. of the Int. Conf. on Imaging Science, Systems and Technology, CISST98, pp. 266-273, Las Vegas, NV, USA, July 1998.



(a) (b) (c) (d) (e) (f) (g)

Fig. 1: Mobile sequence: (a) LR frame, (b) Segmented regions, HR frame using; (c) L2-Norm [2], (d) L1-Norm [3], (e) Global weighted L2-Norm [4], (f) the proposed region-based weighted L1-Norm with adaptive regularization, and (g) the proposed region-based weighted L2-Norm with adaptive regularization.