http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 13 (14): 2277-2283, 2014
ISSN 1812-5638 / DOI: 10.3923/itj.2014.2277.2283

Multi-frame Image Super Resolution using ICM Enhancement Method

[1,2]Zhai Haitian, [1]Li Hui and [1]Gao Weiting
[1]School of Electronics and Information, Northwestern Polytechnical University, Shaanxi, 710129, China
[2]Swanson School of Engineering, University of Pittsburgh, PA 15213, USA

Abstract: Super resolution reconstruction estimates the high resolution image from a set of low-resolution observations of the same scene. Most of the reconstruction methods are very sensitive to their assumed model of data and noise, which limits the utility. This study reviewed some of these methods and proposed an alternative approach using iterated conditional modes to deal with different data and noise models. This computationally inexpensive method is robust to errors in motion and blur estimation and results in images with sharp edges. Simulation results confirmed the effectiveness of our method and demonstrated its superiority to other super resolution methods.

Key words: Image reconstruction, super resolution, ICM, sub-pixel registration

INTRODUCTION

High-resolution (HR) images are essential for enhanced diagnosis and analysis in medical imaging, biometrics identification, satellite imaging and so on. To increase the images resolution, we can either reduce the pixel size by sensor manufacturing techniques or increase the chip size of sensors, which are severely constrained by the physical limitation of imaging systems. Therefore, one should turn to algorithmic techniques to achieve resolution enhancement. Super-resolution (SR) algorithms attempt to generate a single HR image from one or more Low Resolution (LR) images of the same scene (Ren *et al*., 2013). The main challenge is to recover the high-frequency information that was lost in the process of generating the low-resolution inputs. If the low resolution images were directly captured by a camera, for example, this information was eliminated by the band-limiting filter of the photographic process due to imperfections in the optics and integration over the pixels of the sensor. If the low resolution images were the result of software down-sampling, this information was lost through the filtering process of anti-aliasing. The goal of the SR algorithms is to recover this missing information in a way that approximates the original high-resolution image as closely as possible. This is an important problem in several communities and has applications which include object recognition, video transmission, image compression, etc.

For the super resolution algorithms, there are frequency domain and spatial domain algorithms. In the frequency domain, a learning-based SR method to synthesis an HR video sequence was introduced and DCT coefficients for feature vector components and design an example selection procedure to construct a compact database were adopted. An extension of the combined Fourier wavelet deconvolution and denoising algorithm for the multi-frame SR application has been presented (Le Meur *et al*., 2013). They use a fast Fourier based multi-frame image restoration method to produce a sharp, but noisy estimation of the HR image and then apply a space-variant non-linear wavelet threshold technique that addresses the non-stationarity inherent in resolution-enhanced fused images. In spatial domain, one based on a Bayesian formulation employing the expectation maximization algorithm (Babacan *et al*., 2011) and another based on a MAP formulation solved iteratively by cyclic coordinate descent (Shen *et al*., 2007). In their approach, noise variance, regularization and registration are all treated as unknown parameters and are estimated jointly using all the available data. Shen *et al*. (2007) also proposed a joint formulation based on the MAP framework, which judiciously combines motion estimation, segmentation and SR together. Robinson *et al*. (2010) proposed a model that can represent the hyper spectral observations from different wavelengths as weighted linear combinations of a small number of basis image planes. Their method fused information from multiple observations and spectral bands to improve spatial resolution and reconstruct the spectrum of the observed scene.

Corresponding Author: Zhai Haitian, School of Electronics and Information, Northwestern Polytechnical University, Shaanxi, 710129, China

A major drawback of most super resolution methods is that they employ a number of unknown parameters that need to be tuned. This tuning process can be cumbersome and time-consuming since the parameter values have to be chosen differently for each image and degradation condition. Moreover, the algorithmic performance depends significantly on the appropriate choice of parameters, such that generally a long supervised process is needed to obtain useful results.

In this study, we propose a novel iterated conditional modes super resolution methods which address both of the above mentioned issues. Parameters are estimated optimally in a stochastic sense, which provides high reconstruction performance. We show that the proposed methods are very robust to errors in initial motion estimates due to adaptive parameter and motion estimation. We demonstrate with experimental results that the proposed methods provide HR images with high quality and accurate motion information and compare favorably to existing SR methods.

IMAGE OBSERVATION MODEL

To comprehensively analyze the multi-frame reconstruction problem, first of all it is necessary to formulate the image formation model that relates the high-resolution image to the low resolution images. Several observation models have been proposed in the literature (Pelletier and Cooperstock, 2012; Kner *et al*., 2009) and they can be broadly divided into the models for still images and for video sequence. To present a basic concept of SR reconstruction techniques, we employ the observation model for still images in this article, since it is rather straightforward to extend the still image model to the video sequence model. Consider F(x, y) is the desired high resolution image. As is showed in Fig. 1, the observed low resolution images usually sampled below the Nyquist rate from the ideal high resolution image.

Let $f = (f_1, f_2, f_3, \ldots, f_{L1M1} \times f_{L2M2})$ denotes the HR image, where $L_1M_1 \times L_2M_2$ represents the HR image size. Let $I_k = (I_{k,1}, I_{k,2}, \ldots, I_{k,M1\times M1})$ be the kth LR image, where k = 1, 2,..., N with N is the number of the low resolution image. So, the SR image observation model as follows Eq. 1:

$$I_k = DH_k W_k f + n_k \tag{1}$$

where, D is the down sample matrix with the size of $M_1M_2 \times L_1M_1L_2M_2$, W_k is the motion matrix with the size of $L_1M_1L_2M_2 \times L_1M_1L_2M_2$, H_k represents the image blurring matrix of size $L_1M_1L_2M_2 \times L_1M_1L_2M_2$, n_k is the noise vector of size M_1M_2.

Blurring may be caused by an optical system (e.g., out of focus, diffraction limit, aberration, etc.), relative motion between the imaging system and the original scene and the Point Spread Function (PSF) of the LR sensor. In single image restoration applications, the optical or motion blur is usually considered. In the SR image reconstruction, however, the finiteness of a physical dimension in LR sensors is an important factor of blur. In the use of SR reconstruction methods, the characteristics of the blur are assumed to be known. However, if it is difficult to obtain this information, blur identification should be incorporated into the reconstruction procedure.

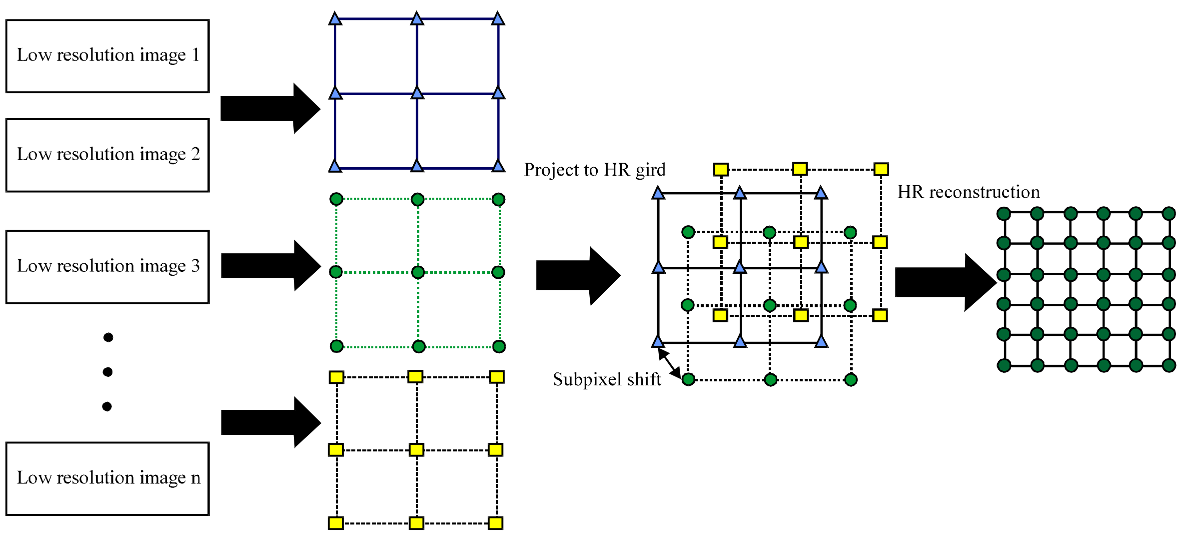

Fig. 1: Schematic diagram of the super resolution reconstruction

The subsampling matrix D generates aliased LR images from the warped and blurred HR image. Although, the size of LR images is the same here, in more general cases, we can address the different size of LR images by using a different subsampling matrix. Although the blurring acts more or less as an anti-aliasing filter, in SR image reconstruction, it is assumed that aliasing is always present in LR images.

To obtain different looks at the same scene, some relative scene motions must exist from frame to frame via multiple scenes or video sequences. Multiple scenes can be obtained from one camera with several captures or from multiple cameras located in different positions. These scene motions can occur due to the controlled motions in imaging systems, e.g., images acquired from orbiting satellites. The same is true of uncontrolled motions, e.g., movement of local objects or vibrating imaging systems. If these scene motions are known or can be estimated within sub-pixel accuracy and if we combine these LR images, SR image reconstruction is possible. Shown in Fig. 1 is the schematic of the super resolution reconstruction.

The goal of the image restoration is to recover a degraded (e.g., blurred, noisy) image. In fact, restoration and SR reconstruction are closely related theoretically and SR reconstruction can be considered as a second-generation problem of image restoration. The differences among the several proposed works are subject to what type of reconstruction method is employed, which observation model is assumed, in which particular domain (spatial or frequency) the algorithm is applied, what kind of methods is used to capture LR images and so on.

MATERIALS AND METHODS

In this section, we explain the proposed method.

Sub-pixel motion estimation: For the problem of sub-pixel estimation (Dempsey *et al*., 2011; Guizar-Sicairos *et al*., 2008), the input image f(x, y), can be reconstructed numerically from measurements of the magnitude of the Fourier transform of g(x, y). In this context, a reconstruction g(x, y) is considered successful even if it has a global coordinate translation (x_0, y_0) or is multiplied by an arbitrary constant ε. The quality of the reconstruction must then be assessed through an error metric that is invariant to these operations. Normalized Root-Mean-Square Error (NRMSE) between $I_R(x, y)$ and g(x, y) is used to obtain the sub-pixel movement, defined by Eq. 2:

$$\min_{\alpha, x_0, y_0} \frac{\sum |\alpha g(x-x_0, y-y_0) - f(x, y)|^2}{\sum_{x,y} |f(x, y)|^2} = 1 - \frac{\max_{x_0, y_0} |r(x_0, y_0)|^2}{\sum_{x,y} |g(x, y)|^2 \sum_{x,y} |f(x, y)|^2} \tag{2}$$

where, summations are taken over all image points (x, y):

$$\begin{aligned} r(x_0, y_0) &= \sum_{x,y} f(x,y) g^*(x-x_0, y-y_0) \\ &= \sum_{u,v} F(u,v) G^*(u,v) \exp\left[i2\pi\left(\frac{ux_0}{M} + \frac{vy_0}{N}\right)\right] \end{aligned} \tag{3}$$

Is the cross correlation of f(x, y) and g(x, y); N and M are the image dimensions; g* represents complex conjugation of g; F denotes the DFT of f, as given by the Eq. 4:

$$F(u, v) = \sum_{x,y} \frac{f(x, y)}{\sqrt{MN}} \exp\left[-i2\pi\left(\frac{ux}{M} + \frac{vy}{N}\right)\right] \tag{4}$$

Thus, evaluation of the NRMSE by Eq. 2 requires solving the more general problem of sub-pixel image registration by locating the peak of the cross correlation $r(x_0, y_0)$. We incorporate the algorithm that refines the initial estimate using a nonlinear optimization conjugate gradient routine to maximize $|r(x_0, y_0)|^2$ that significantly improve performance without sacrificing accuracy. Its partial derivative with respect to x_0 is (Eq. 5):

$$\begin{aligned} \frac{\partial |r(x_0, y_0)|^2}{\partial x_0} = 2\,\mathrm{Im}\Bigg\{ & r(x_0, y_0) \sum_{u,v} \frac{2\pi u}{M} F^*(u, v) \\ & \times G(u, v) \exp\left[-i2\pi\left(\frac{ux_0}{M} + \frac{vy_0}{N}\right)\right]\Bigg\} \end{aligned} \tag{5}$$

With a similar expression for the partial derivative with respect to y_0. This algorithm iteratively searches for the image displacement (x_0, y_0) that maximizes r (x_0, y_0) and can achieve registration precision to within an arbitrary fraction of a pixel.

Bayesian based super resolution: Super-resolution reconstruction is considered as an ill-posed problem, thus it usually requires some kind of regularization. A set of low resolution observations must be considered into the estimation of the HR reconstruction. Considering the Bayesian formulation which can provide an effective way of imposing a priori constraints to the estimation. So, that we can estimate the HR image by maximizing the conditional probability with Bayesian formulation (Eq. 6):

$$\hat{F} = \arg\max_F \{p(F \mid I)\} \tag{6}$$

where, $\hat{F}$ is the high resolution estimation.

ICM based super resolution: The maximization in Eq. (6) usually demands high computational power. Therefore, maximization of the local probabilities is an alternative. The ICM algorithm (Meng *et al*., 2010) uses a greedy strategy to the local maximization using a MRF prior model. The algorithm sequentially updates the labels F_i of each pixel I = 1, 2,..., M_1M_2, by maximizing the posterior probability (Eq. 7):

$$p(F_i \mid I, F_{nb(i)}) \tag{7}$$

where, nb(i) is the set of neighbors of pixel i, according to a neighborhood system.

If considering I_i^l is a set of low resolution pixels, from the Bayes theorem, it follows that (Eq. 8):

$$p(F_i \mid I, F_{nb(i)}) \propto p\left(I_i^l \mid F_i\right) p\left(F_i \mid F_{nb(i)}\right) \tag{8}$$

Considering the image formation model in Eq. (2), in the presence of zero mean independent Gaussian white noise, the likelihood distribution is given by (Eq. 9):

$$p(I \mid F) = \frac{1}{\left(2\pi\sigma^2\right)^{q\frac{N^2}{2}}} \exp\left\{-\sum_{k=1}^{q} \frac{\left\|I_k - DH_k\hat{F}\right\|^2}{2\sigma^2}\right\} \tag{9}$$

where, σ^2 is the noise variance.

High resolution reconstruction: We can estimate the local conditional distribution $p(I_i^l|F_i)$ by Eq. 10:

$$p\left(I_i^l \mid F_i\right) \propto \exp\left\{-\sum_{r_i=1}^{m} \frac{\left\|I^{(r_i)} - D^{(r_i)}H^{(r_i)}\hat{F}\right\|^2}{2\sigma^2}\right\} \tag{10}$$

where, m denotes the number of LR pixels that lay over pixel i. $I^{(ri)}$ is the rth low-resolution pixel that lay over pixel i and $D^{(ri)}H^{(ri)}$ is the corresponding pixel generated by the estimation $\hat{F}$.

Moreover, in the ICM algorithm, we also need to know $p(F_i|F_{nb(i)})$. The solution of this probability is given by Eq. 11:

$$p\left(F_i \mid F_{nb(i)}\right) = \frac{1}{Z} \exp\left\{-U\left(F_i \mid F_{li}\right)\right\} \tag{11}$$

where, the potential function $U(F_i|F_{nb(i)})$ is defined as (Eq. 12):

$$U\left(F_i \mid F_{nb(i)}\right) = \sum_{i' \in nb(i)} \beta\left[1 - 2\exp\left(-\left(F_i - F_{i'}\right)^2\right)\right] \tag{12}$$

In which Z is called partition function and β can be viewed as an interaction coefficient.

As discussed above, we can estimate the HR image by the following steps:

- Step 1: Define a MRF model for the true values of F_i
- Step 2: Choose an initial high resolution estimation
- Step 3: For i from 1 to $M_1 \times M_2$, update F_i by the value that maximizes $p(I_i^l|F_i)\, p(F_i|F_{nb(i)})$
- Step 4: Repeat step (3) N times, where N is the maximum number of iterations

This model is meaningful in texture representation and is easy to process in the image reconstruction procedure.

Refinement step: The refinement step contains two terms, the first term tries to deblur the sequence and make it more robust and the second term is the Bilateral Total Variation (BTV) filter which was introduced by Farsiu. In this study, a generalized form of the 2-D BTV filter is proposed. The 3-D BTV filter introduced in this step can be considered as a Regularizer (Mochizuki *et al*., 2011). Since this problem is an under determined problem, having a regularizer will help to remove artifacts from answers and also having sharper edges. The proposed regularizer term is shown in (Eq. 8):

$$Z = \text{argmin}\left[\left\|GZ - X\right\|_1 + \lambda R_{BTV}(Z)\right] \tag{13}$$

where, Z is the desired high resolution video sequence, X is the current interpolated video sequence and G applies Gaussian blur matrix on the sequence. In the second part of the statement, λ is the regularizer parameter and $R_{BTY}(Z)$ is the regularization function which is the Bilateral Total Variation:

$$R_{BTV}(Z) = \sum_{l=-P}^{P} \sum_{m=-P}^{P} \sum_{k=-P}^{P} \alpha_1^{|m|+|l|} \alpha_2^{|k|} \left\|\left(Z - S_x^l S_y^m S_t^k Z\right)\right\|_1 \tag{14}$$

In (Eq. 14), P is the window size and S_x^l, S_y^m and S_t^k shift the sequence Z, l, m, k pixels in horizontal, vertical and temporal directions, respectively. Also α_i(i = 0, 1,

$0<\alpha_i<1$) gives a decaying effect to the summation of the regularization term. For finding minimum of the argmin, gradient of Eq. 13 is computed. By using steepest descent method the solution to this problem would be:

$$Z_{n+1} = Z_n - \delta\left[G^T \mathrm{sign}(GZ_n - Z_{int}) + \lambda\Gamma(k,m,l)\right] \quad (15)$$

Where:

$$\Gamma(k,m,l) = \sum_{l=-P}^{P}\sum_{m=-P}^{P}\sum_{k=-P}^{P}\alpha_1^{|m|+|l|}\alpha_2^{|k|}(I - S_t^{-k}S_y^{-m}S_x^{-l}Z) \quad (16)$$

δ is a scalar defining the step size in the direction of the gradient.

In generalizing the 2-D BTV (Knoll *et al*., 2013) to 3-D version different decaying parameter have been allocated to the spatial and temporal shifts. This is due to the fact that the behavior of a video sequence in spatial (x, y) and temporal (t) domain are different. A video sequence with fast motion may lose the consistency in time faster than spatial consistency, because frames have motion in respect to each other and thus a single pixel may be more correlated to its adjacent pixels in space than its adjacent pixels in time. Consequently $\alpha_1>\alpha_2$ seems to give better results in these sequences. But in a sequence where motion is very small and frames do not change much with respect to each other, a single pixel is more correlated to the adjacent frames in time and thus in these cases $\alpha_1>\alpha_2$ is chosen. Testing different values for α_1 and α_2 supports the specified criteria for choosing them. Experimental results are driven by choosing the best values for α_i (i = 0, 1) by considering the stated condition.

EXPERIMENT RESULT

For the experimental verification of the proposed method, we provided several examples to compare high resolution image between classical super resolution methods and the proposed method. Table 1 showed the average PSNR values in the experiments we have conducted.

For the first dataset which contains 12 LR images, the results of rotation estimation and translation estimation are shown in Fig. 2. The accuracy of the registration is about 0.02 pixels. As is shown in Fig. 3, which provides the results from different super resolution algorithms, it is clear that the proposed method provide HR image estimation with sharper and fewer ringing artifacts than the other methods. One of the 12 LR images is shown in

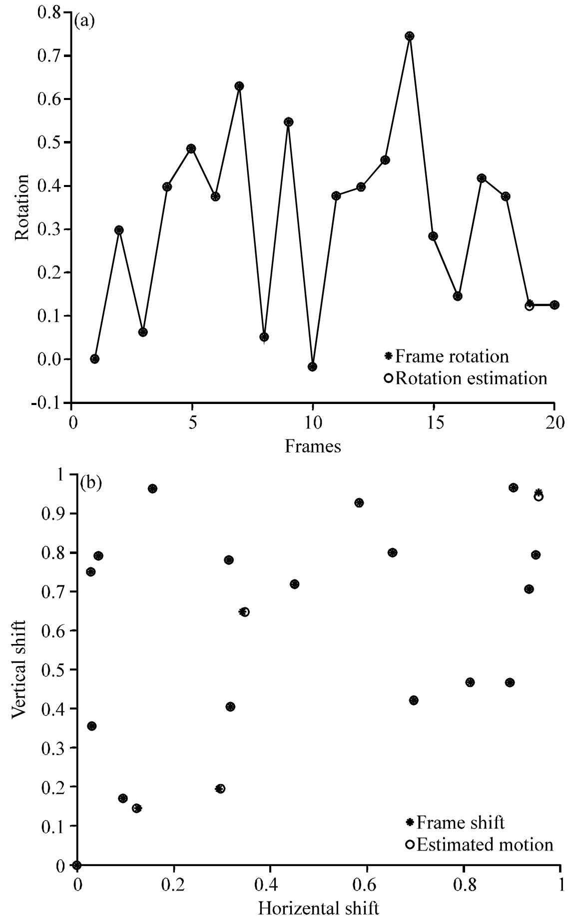

Fig. 2(a-b): Example estimated sub-pixel registration, (a) Rotation estimation and (b) Translation estimation

Table 1: Average PSNR comparison for different SR algorithms

Method	Calendar	Lab	Lena
Bicubic	31.25	28.03	25.36
Fast and robust	23.74	27.51	25.21
TV reconstruction	22.11	29.63	26.32
Normalized convolution	21.23	24.81	27.05
Proposed method	20.14	23.51	25.14

Fig. 2a. Figure 2b shows the bicubic interpolation of the low resolution image. The high resolution image reconstructed by the proposed algorithm is shown in Fig. 2d.

The second dataset consists of 12 LR images taken from a low resolution video. The reconstructed HR images by a factor of 4 resolution increase obtained by nearest neighbor, bicubic interpolation and SR algorithms are shown in Fig. 4. Although, there was a slightly mis-registration among the frames, the proposed method still has a good performance.

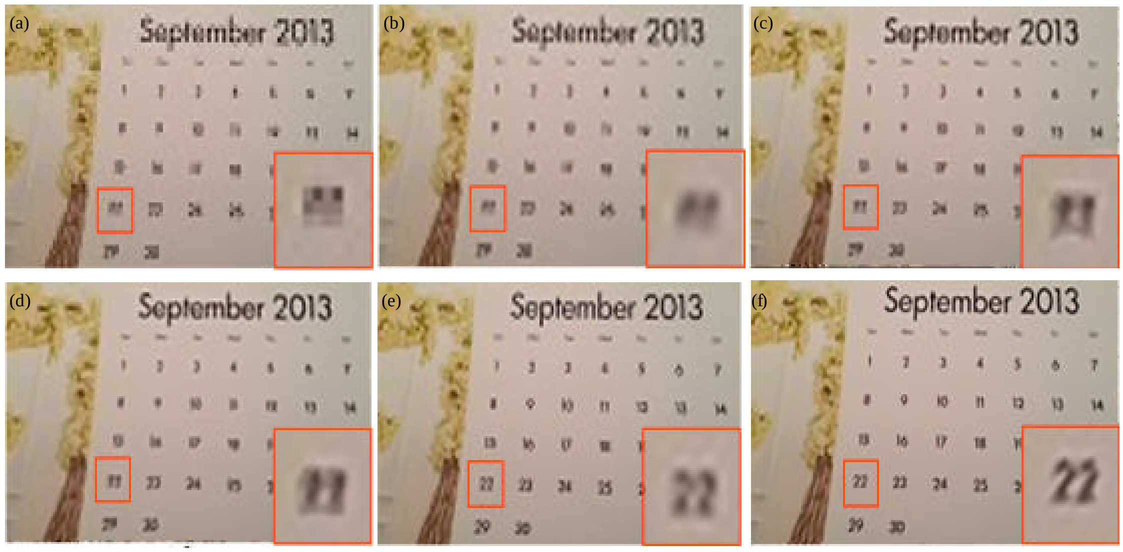

Fig. 3(a-f): Example estimated HR estimation from different super resolution methods. Results (3x resolution increase) by (a) Original low resolution image interpolated by nearest neighbor interpolation, (b) Bicubic interpolation, (c) Fast and robust method, (d) TV reconstruction, (e) Normalized convolution and (f) Proposed method

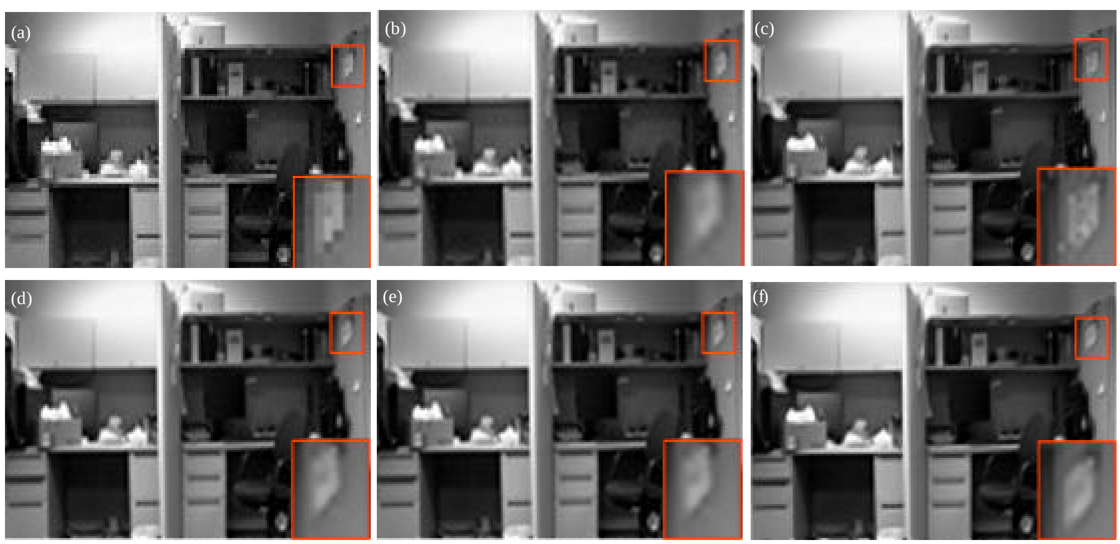

Fig. 4(a-f): Example estimated HR estimation from different super resolution methods. Results (4x resolution increase) by (a) Nearest neighbor interpolation, (b) Bicubic interpolation, (c) Fast and Robust method, (d) TV reconstruction, (e) Normalized convolution and (f) Proposed method

CONCLUSION

We developed an efficient technique for multi-frame images super resolution. First, we incorporated the fourier transform to the sub-pixel registration which is efficiency and accurate. Following, a novel high resolution reconstruction method is derived by maximizing the local conditional probability with iterated conditional modes. Then, the generalized form of the 2-D BTV filter is introduced to deblur the HR image. Finally, we conducted experiments by applying the proposed approach and the experimental results demonstrate that the new super resolution method is computationally inexpensive and robust to errors in motion and blur estimation and results in images with sharp edges.

ACKNOWLEDGMENT

Supported by National Natural Science Foundation of China (Grant No. 61171155), Natural Science Foundation of Shaan Xi Province (Grant No. 2012JM8010) and Doctorate Foundation of Northwestern Polytechnical University (No. CX201316).

REFERENCES

Babacan, S.D., R. Molina and A.K. Katsaggelos, 2011. Variational bayesian super resolution. IEEE Trans. Image Process., 20: 984-999.

Dempsey, G.T., J.C. Vaughan, K.H. Chen, M. Bates and X. Zhuang, 2011. Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. Nat. Methods, 8: 1027-1036.

Guizar-Sicairos, M., S.T. Thurman and J.R. Fienup, 2008. Efficient subpixel image registration algorithms. Opt. Lett., 33: 156-158.

Kner, P., B.B. Chhun, E.R. Griffis, L. Winoto and M.G.L. Gustafsson, 2009. Super-resolution video microscopy of live cells by structured illumination. Nat. Methods, 6: 339-342.

Knoll, F., G. Schultz, K. Bredies, D. Gallichan, M. Zaitsev, J. Hennig and R. Stollberger, 2013. Reconstruction of undersampled radial PatLoc imaging using total generalized variation. Magn. Reson. Med., 70: 40-52.

Le Meur, O., M. Ebdelli and C. Guillemot, 2013. Hierarchical super-resolution-based inpainting. IEEE Trans. Image Process., 22: 3779-3790.

Meng, J., J. Zhang, Y. Chen and Y. Huang, 2010. An iterated conditional modes solution for sparse bayesian factor modeling of transcriptional regulatory networks. Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine, Volume 13, December 18-21, 2010, Hong Kong, pp: 335-340.

Mochizuki, Y., Y. Kameda, A. Imiya, T. Sakai and T. Imaizumi, 2011. Variational method for super-resolution optical flow. Signal Process., 91: 1535-1567.

Pelletier, S. and J.R. Cooperstock, 2012. Preconditioning for edge-preserving image super resolution. IEEE Trans. Image Process., 21: 67-79.

Ren, Z., C. He and Q. Zhang, 2013. Fractional order total variation regularization for image super-resolution. Signal Process., 93: 2408-2421.

Robinson, M.D., C.A. Toth, J.Y. Lo and S. Farsiu, 2010. Efficient fourier-wavelet super-resolution. IEEE Trans. Image Process., 19: 2669-2681.

Shen, H., L. Zhang, B. Huang and P.A. Li, 2007. MAP approach for joint motion estimation, segmentation and super resolution. IEEE Trans. Image Process, 16: 479-490.