MULTI-FRAME SUPER-RESOLUTION FROM OBSERVATIONS WITH ZOOMING MOTION

*Yushuang Tian and Kim-Hui Yap**

School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
Email: Tian0042@e.ntu.edu.sg, ekhyap@ntu.edu.sg

ABSTRACT

This paper proposes a new multi-frame super-resolution (SR) approach to reconstruct a high-resolution (HR) image from low-resolution (LR) images with zooming motion. Existing SR methods often assume that the relative motion between acquired LR images consists of only translation and possibly rotation. This restricts the application of these methods in cases when there is zooming motion among the LR images. There are currently only a few methods focusing on zooming SR. Their formulation usually ignores the registration errors or assumes them to be negligible during the HR reconstruction. In view of this, this paper presents a new SR reconstruction approach to handle a more flexible motion model including translation, rotation and zooming. An iterative framework is developed to estimate the motion parameters and HR image progressively. Both simulated and real-life experiments show that the proposed method is effective in performing image SR.

Index Terms — image super-resolution, zooming motion

1. INTRODUCTION

Image super-resolution (SR) is a process that fuses a sequence of low-resolution (LR) observations to reconstruct a high-resolution (HR) image. Various SR algorithms and techniques have been proposed in recent years [1]-[5]. Most existing SR algorithms typically assume the resolution of all LR observations to be the same. This assumption, however, may not be applicable when there is relative zooming motion between the LR images, as illustrated in Fig. 1. Hence, this motivates the study of SR algorithms for LR observations with relative zooming motion.

In [6], Li proposed a SR approach for synthetic zooming. In his work, all the LR images are related to each other by employing a line-geometry model. Joshi et al. [7] presented a zooming SR algorithm using Markov random field (MRF) model under the assumption that the SR field is homogenous. In [8], Shen et al. employed a linear motion model to represent the relationship between the desired HR image and the captured LR images with arbitrary zooming factors. Recently, a zooming SR approach with total variation (TV) prior was presented in [9]. Generally, these methods assume the motion parameters estimated by existing registration methods to be error-free. This is

Fig. 1 Illustration of HR reconstruction from LR images with zooming motion.

inconsistent with many practical applications as current registration algorithms still experience various degrees of errors. Besides, the motion models in these algorithms consist of mainly zooming motion alone, with a few work including the translation as well.

In view of this, this paper proposes an iterative method for joint image registration and HR reconstruction from LR observations with different resolutions. In contrast with conventional SR methods that focus mainly on in-plane motion (e.g. translation and rotation), the proposed method takes into account simultaneous translation, rotation and zooming between the acquired LR images. The more flexible motion model enables the proposed method to handle more real-life applications. To improve the SR reconstruction, the proposed method integrates image registration and HR reconstruction into an iterative joint estimation process. This is promising as more accurate motion parameters can be incorporated into the subsequent HR reconstruction.

2. PROBLEM FORMULATION

Let us model the kth ($1 \le k \le N$) acquired LR image $g_k(m, n)$ by rotating the HR image $f(x, y)$ by θ_k, zooming it by a zooming factor a_k, shifting it by a translational vector (s_{xk}, s_{yk}), blurring the result by a point-spread function (PSF) h_k, and down-sampling it to the resolution of the observed image by a decimation factor ρ. Hence, the degradation process can be expressed as:

$$\begin{aligned} g_k(m,n) &= \big(f(a_k x \cos\theta_k - a_k y \sin\theta_k + s_{xk},\, a_k x \sin\theta_k \\ &\quad + a_k y \cos\theta_k + s_{yk}) \otimes h_k \otimes h_c \big) \downarrow_\rho + n_k(m,n) \quad (1) \\ &= \big(f(x_k, y_k) \otimes h_k \otimes h_c \big) \downarrow_\rho + n_k(m,n) \end{aligned}$$

where $\otimes$ is the 2D convolution operator and n_k denotes the additive white Gaussian noise (AWGN). $\downarrow_\rho$ denotes the down-sampling operator with decimation factor ρ. h_k represents the camera lens blur in each LR image. h_c represents the effect of spatial integration of light intensity over a square surface region to simulate image acquisition by the sensors. Hence, h_c takes the form of a uniform PSF with support $(\rho\times\rho)$. In this work, it is assumed that the lens condition (h_k) is known. Based on the motion model, the coordinate vector can be written as,

$$\begin{bmatrix} x_k \\ y_k \end{bmatrix} = \boldsymbol{R}\boldsymbol{a}_k, \text{ where } \boldsymbol{R} = \begin{bmatrix} x & -y & 1 & 0 \\ y & x & 0 & 1 \end{bmatrix} \quad (2)$$
$$\text{and } \boldsymbol{a}_k = [a_k\cos\theta_k,\ a_k\sin\theta_k,\ s_{xk},\ s_{yk}]^T$$

The degradation process in (1) can be expressed in a matrix-vector form as,

$$\boldsymbol{g} = \boldsymbol{DHS}(\boldsymbol{\alpha})\boldsymbol{f} + \boldsymbol{n} \quad (3)$$

where $\boldsymbol{f}$ represents the lexicographically ordered original image f, $\boldsymbol{g} = [\boldsymbol{g}_1^T, \cdots, \boldsymbol{g}_N^T]^T$ and $\boldsymbol{n} = [\boldsymbol{n}_1^T, \cdots, \boldsymbol{n}_N^T]^T$ are the vectors representing the discrete, concatenated and lexicographically ordered g_k and n_k, respectively. $\boldsymbol{D}$ represents the down-sampling operator. $\boldsymbol{H} = [\boldsymbol{H}_1^T, \cdots, \boldsymbol{H}_N^T]^T$ and $\boldsymbol{H}_k$ ($1 \le k \le N$) denotes the corresponding matrix representing the blurring operator $h_k \otimes h_c$. The matrix $\boldsymbol{S}(\boldsymbol{\alpha})$ is the warping operator formed by nonlinear, differentiable functions of unknown motion parametric vector $\boldsymbol{a}$, where $\boldsymbol{a} = [\boldsymbol{a}_1^T, \boldsymbol{a}_2^T, \cdots, \boldsymbol{a}_N^T]^T$. $\boldsymbol{a}_k$ represents the motion parametric vector of the kth LR image. Hence, $\boldsymbol{S}(\boldsymbol{a}) = [\boldsymbol{S}_1(\boldsymbol{a}_1)^T, \boldsymbol{S}_2(\boldsymbol{a}_2)^T, \cdots, \boldsymbol{S}_N(\boldsymbol{a}_N)^T]^T$ and $\boldsymbol{S}_k(\boldsymbol{\alpha}_k)$ ($1 \le k \le N$) denotes the shifting operator for the kth LR image. The objective of image SR is to reconstruct the HR image $\boldsymbol{f}$ from N LR observations with the unknown motion parametric vector $\boldsymbol{a}$.

3. ITERATIVE SCHEME FOR SIMULTANEOUS IMAGE REGISTRATION AND SR

Given the image formation model in (3), the estimated HR image $\boldsymbol{f}$ and the unknown motion parametric vector $\boldsymbol{a}$ can be obtained by minimizing the following cost function:

$$E(\boldsymbol{\alpha}, \boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{DHS}(\boldsymbol{\alpha})\boldsymbol{f}\|^2 + \lambda\|\boldsymbol{Lf}\|^2 = \left\| \begin{matrix} \boldsymbol{r}(\boldsymbol{\alpha}, \boldsymbol{f}) \\ \sqrt{\lambda}\boldsymbol{Lf} \end{matrix} \right\|^2 \quad (4)$$

where $\|\bullet\|$ denotes the L_2-norm, $\boldsymbol{r}(\boldsymbol{a}, \boldsymbol{f}) = \boldsymbol{g} - \boldsymbol{DHS}(\boldsymbol{a})\boldsymbol{f}$ is the fidelity residual vector, $\boldsymbol{L}$ denotes the total variation (TV) operator on the desired HR image and λ is the regularization parameter.

In order to solve the minimization problem in (4), we develop a regularized nonlinear least square (RNLS) method to perform registration and SR simultaneously. It is noted that the residual vector $\boldsymbol{r}(\boldsymbol{\alpha}, \boldsymbol{f})$ is nonlinear with respect to $\boldsymbol{\alpha}$, due to the inclusion of rotation and zooming. Here, we extend the principle of nonlinear parametric estimation to derive a linear approximation for $\boldsymbol{r}(\boldsymbol{\alpha}, \boldsymbol{f})$. Let $\Delta\boldsymbol{f}$ denote a small change in the HR image $\boldsymbol{f}$ and $\Delta\boldsymbol{\alpha}$ denote a small change in the motion vector $\boldsymbol{\alpha}$. The residual vector $\boldsymbol{r}(\boldsymbol{\alpha}, \boldsymbol{f})$ can be linearized with respect to $\Delta\boldsymbol{f}$ and $\Delta\boldsymbol{\alpha}$ as follows:

$$\boldsymbol{r}(\boldsymbol{\alpha} + \Delta\boldsymbol{\alpha}, \boldsymbol{f} + \Delta\boldsymbol{f}) = \boldsymbol{r}(\boldsymbol{\alpha}, \boldsymbol{f}) - \boldsymbol{DHG}\Delta\boldsymbol{\alpha} - \boldsymbol{DHS}(\boldsymbol{\alpha})\Delta\boldsymbol{f} \quad (5)$$

where $\boldsymbol{G} = \partial(\boldsymbol{S}(\boldsymbol{\alpha})\boldsymbol{f})/\partial\boldsymbol{\alpha}$.

Therefore, given the current estimate of unknown vector $\boldsymbol{\alpha}^i$ and HR image $\boldsymbol{f}^i$, the minimization problem (4) can be rewritten as

$$\min_{\Delta\boldsymbol{\alpha}^i, \Delta\boldsymbol{f}^i} \left\| \begin{pmatrix} \boldsymbol{DHG}^i & \boldsymbol{DHS}(\boldsymbol{\alpha}^i) \\ 0 & \sqrt{\lambda}\boldsymbol{L} \\ \boldsymbol{Q} & 0 \end{pmatrix} \begin{pmatrix} \Delta\boldsymbol{\alpha}^i \\ \Delta\boldsymbol{f}^i \end{pmatrix} + \begin{pmatrix} -\boldsymbol{r}(\boldsymbol{\alpha}^i, \boldsymbol{f}^i) \\ \sqrt{\lambda}\boldsymbol{Lf}^i \\ 0 \end{pmatrix} \right\|^2 \quad (6)$$

where $\boldsymbol{Q}$ is a diagonal matrix of positive weights, similar to the idea in [10].

The proposed iterative algorithm for joint image registration and SR using RNLS technique can be summarized as follows,

1) Initialize $\boldsymbol{\alpha}^0$ by using an existing image registration method and estimating $\boldsymbol{f}^0$ by solving (4).
2) Construct $\boldsymbol{S}(\boldsymbol{\alpha}^0)$, $\boldsymbol{G}^0$ and $\boldsymbol{r}(\boldsymbol{\alpha}^0, \boldsymbol{f}^0)$.
3) Solve (6) for $\Delta\boldsymbol{\alpha}^i$ and $\Delta\boldsymbol{f}^i$ using preconditioned conjugate gradient (PCG) method.
4) Update $\boldsymbol{\alpha}^{i+1} = \boldsymbol{\alpha}^i + \Delta\boldsymbol{\alpha}^i$ and $\boldsymbol{f}^{i+1} = \boldsymbol{f}^i + \Delta\boldsymbol{f}^i$.
5) Construct $\boldsymbol{S}(\boldsymbol{\alpha}^{i+1})$, $\boldsymbol{G}^{i+1}$ and $\boldsymbol{r}(\boldsymbol{\alpha}^{i+1}, \boldsymbol{f}^{i+1})$.
6) Go to 3) and update $i=i+1$ until convergence or a maximum number of iterations is reached.

The order of complexity of the proposed algorithm is $O(IJM)$, where M is the pixel number of the HR image, I and J are the numbers of iterations for the algorithm and PCG optimization, respectively.

4. DERIVATION OF LINEAR APPROXIMATION FOR $\partial(\boldsymbol{S}(\boldsymbol{\alpha})\boldsymbol{f})/\partial\boldsymbol{\alpha}$

In this section, we will derive a linear approximation for $\partial(\boldsymbol{S}(\boldsymbol{\alpha})\boldsymbol{f})/\partial\boldsymbol{\alpha}$. $\partial(\boldsymbol{S}(\boldsymbol{\alpha})\boldsymbol{f})/\partial\boldsymbol{\alpha}$ can be expressed as,

$$\frac{\partial\left(\boldsymbol{S}(\boldsymbol{\alpha})\boldsymbol{f}\right)}{\partial\boldsymbol{\alpha}}=diag\{\frac{\partial\left(\boldsymbol{S}(\boldsymbol{\alpha}_1)\boldsymbol{f}\right)}{\partial\boldsymbol{\alpha}_1},\cdots,\frac{\partial\left(\boldsymbol{S}(\boldsymbol{\alpha}_N)\boldsymbol{f}\right)}{\partial\boldsymbol{\alpha}_N}\} \quad (7)$$

The relative position between a zoomed, rotated and shifted HR images $\boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f}$ and the reference HR image $\boldsymbol{f}$ is shown in Fig. 2. We denote $[d_k, e_k]^T$ as the distance vector between the pixel $f(x_k, y_k)$ and the pixel f_{tl} at the top-left position in the reference HR grid. As illustrated in Fig. 2, d_k and e_k satisfy the following relationship:

$$d_k = x_k - \text{floor}(x_k),\quad e_k = y_k - \text{floor}(y_k) \quad (8)$$

where floor(•) denotes the operator rounding the number to the nearest integer less than or equal to itself. Using bilinear interpolation, $\boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f}$ can be obtained by

$$\begin{aligned}\boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f} = &(1-\boldsymbol{d}_k)\odot(1-\boldsymbol{e}_k)\odot\boldsymbol{f}_{tl}+(1-\boldsymbol{d}_k)\odot\boldsymbol{e}_k\odot\boldsymbol{f}_{tr}\\ &+\boldsymbol{d}_k\odot(1-\boldsymbol{e}_k)\odot\boldsymbol{f}_{bl}+\boldsymbol{d}_k\odot\boldsymbol{e}_k\odot\boldsymbol{f}_{br}\end{aligned} \quad (9)$$

where $\odot$ is an entry-by-entry multiplication operator. $\boldsymbol{f}_{tl}$, $\boldsymbol{f}_{tr}$, $\boldsymbol{f}_{bl}$, $\boldsymbol{f}_{br}$, $\boldsymbol{d}_k$ and $\boldsymbol{e}_k$ are vectors representing the lexicographically ordered f_{tl}, f_{tr}, f_{bl}, f_{br}, d_k and e_k, respectively.

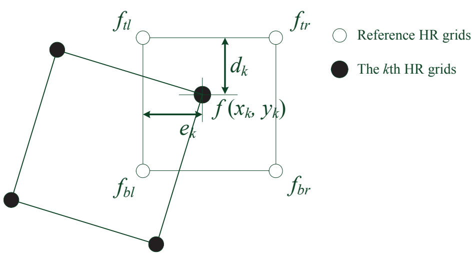

Fig. 2 The relative position between the kth HR grid and the reference HR grid.

Therefore, $\partial\left(\boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f}\right)/\partial\left([\boldsymbol{x}_k^T,\ \boldsymbol{y}_k^T]^T\right)$ can be written as

$$\begin{aligned}\boldsymbol{E}_k = [&diag\{(1-\boldsymbol{e}_k)\odot(\boldsymbol{f}_{bl}-\boldsymbol{f}_{tl})+\boldsymbol{e}_k\odot(\boldsymbol{f}_{br}-\boldsymbol{f}_{tr})\}\\ &diag\{(1-\boldsymbol{d}_k)\odot(\boldsymbol{f}_{tr}-\boldsymbol{f}_{tl})+\boldsymbol{d}_k\odot(\boldsymbol{f}_{br}-\boldsymbol{f}_{bl})\}]\end{aligned} \quad (10)$$

Combining (2) and (10), $\partial\left(\boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f}\right)/\partial\boldsymbol{\alpha}_k$ can be expressed as follows,

$$\frac{\partial\left(\boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f}\right)}{\partial\boldsymbol{\alpha}_k}=\frac{\partial\left(\boldsymbol{S}_k(\boldsymbol{\alpha}_k)\boldsymbol{f}\right)}{\partial\left([\boldsymbol{x}_k^T,\ \boldsymbol{y}_k^T]^T\right)}\frac{\partial\left([\boldsymbol{x}_k^T,\ \boldsymbol{y}_k^T]^T\right)}{\partial\boldsymbol{\alpha}_k}=\boldsymbol{E}_k\boldsymbol{R} \quad (11)$$

5. EXPERIMENTAL RESULTS

In this section, we will demonstrate the effectiveness of the proposed method. Three 256×256 images in Fig. 3 are selected as the test images. We conduct various experiments and compare the results of our method with other two methods, namely, the Shen' method [8] and the Ng's method [9], respectively. Peak signal-to-noise ratio (PSNR) and normalized mean square error (NMSE) are chosen to evaluate the estimated HR image and motion parameters, respectively.

(a) (b) (c)

Fig. 3 Test images: (a) "Board"; (b) "Barbara"; (c) "Pepper".

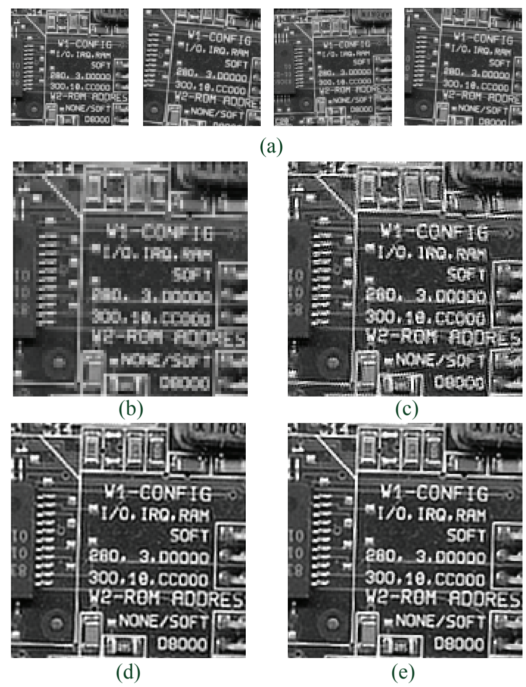

Fig. 4 SR on LR "Board" images with simultaneous zooming, rotation and translation. (a) 4 samples of the LR images. (b) The scaled-up reference image (1st LR image). (c) Reconstructed image based on initial registration estimates. (d) Reconstructed image using the proposed method. (e) Reconstructed image using known exact motion parameters.

In the experiments, the "Board" image in Fig. 3(a) was selected as the test image. To generate 9 LR images, the HR image was rotated by randomly selected angles from a uniform distribution over [−4°, 4°] and shifted by randomly selected translations from a uniform distribution over [−3, 3] pixels. The shifted HR image was then blurred by a 2×2 uniform blur, followed by a down-sampling operator with a randomly selected decimation factor from a uniform distribution over [1.8, 2.2]. The first LR image was selected as the reference and its decimation factor was set to 2. Finally, these LR images were degraded by additive white Gaussian noise (AWGN) to produce a signal-to-noise (SNR) ratio at 35dB. 4 samples of the LR images and a scaled-up reference image are shown in Fig. 4(a) and (b), respectively. We first estimated the motion parameters by using the image registration method in [11]. Then, the registration estimates were used in the subsequent HR reconstruction (similar to [1]) to produce the result in Fig. 4(c). It is noted that it suffers from significant artifacts, especially near the word areas. Next, the proposed method was performed on the LR images and the result is shown in Fig. 4(d). It can be seen that the proposed method has achieved satisfactory HR reconstruction where the overall clarity has been recovered. We also compared our result with the reconstructed HR image using known exact motion parameters, as shown in Fig. 4(e). It can be observed that the HR image reconstructed using the proposed method is almost similar to that reconstructed using the exact motion parameters. Comparison shows the effectiveness of the proposed method in performing SR reconstruction on LR images with simultaneous zooming, rotation and translation.

TABLE I PSNR and NMSE comparison for different motions

	Zooming				Zooming and translation			
	Shen's method [8]		**Proposed method**		Ng's method [9]		**Proposed method**	
	PSNR	NMSE	PSNR	NMSE	PSNR	NMSE	PSNR	NMSE
Barbara	27.91	10.83	**33.58**	**0.11**	27.65	1.34	**34.30**	**0.01**
Pepper	30.33	12.16	**34.83**	**0.07**	30.01	1.43	**34.15**	**0.02**

Next, we will demonstrate the performance of the proposed method in handling various motions. The experiments are divided into two groups based on the motion between the LR images, namely, (i) zooming only, (ii) zooming and translation. To provide a fair comparison, our proposed method is compared with only the respective method that is able to handle the motion model in each case. TABLE I shows the NMSE of the estimated motion parameter and PSNR of the reconstructed HR image obtained by the proposed method and other two methods. Comparison reveals that the proposed method outperforms the other methods. It is also more flexible as it can handle different motion models.

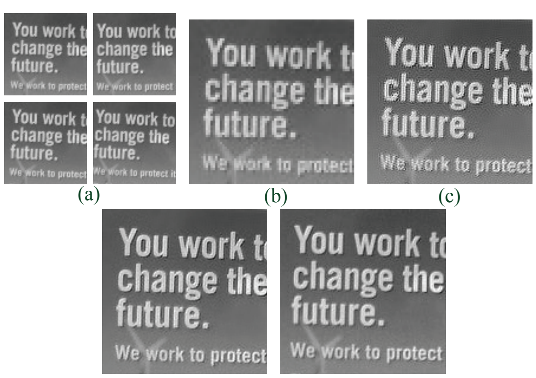

Fig. 5 SR on real-life images. (a) 4 samples of the LR images. (b) The scaled-up reference image (1st LR image). (c) Reconstructed image based on initial registration estimates. (d) Reconstructed image using the proposed method. (e) Ground truth.

We also conducted a real-life experiment by capturing 10 "Magazine" LR images using a web camera with relative zooming, rotation and translation. The LR images are shown in Fig. 5(a). We set the decimation factor to 2. An image captured with the resolution of the HR image is used as the ground truth in Fig. 5(e). Next, the reconstructed images based on initial motion parameter estimates and our proposed method are shown in Fig. 5(c) and (d), respectively. It can be seen that our proposed method produces better image quality. Further, the reconstructed HR image using the proposed method manages to achieve similar visual quality of the ground truth image. Comparison shows that the proposed method can handle real-life SR image reconstruction effectively.

6. CONCLUSIONS

In this paper, we have presented a new SR method that can address the HR reconstruction problem when the motion model consists of zooming, rotation and translation. As opposed to existing zooming SR methods, the proposed method takes the initial registration errors into consideration. An iterative scheme is employed to improve the estimates of motion parameters and HR image simultaneously and progressively. Experimental results show that the proposed method is effective in performing image SR reconstruction.

7. REFERENCE

[1]. S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Advances and challenges in super-resolution," *International Journal of Imaging Systems and Technology,* vol. 14, pp. 47-57, 2004.

[2]. P. Vandewalle, L. Sbaiz, J. Vandewalle, and M. Vetterli, "Super-Resolution From Unregistered and Totally Aliased Signals Using Subspace Methods," *IEEE Transactions on Signal Processing,* vol. 55, pp. 3687-3703, 2007.

[3]. M. Ebrahimi and E. R. Vrscay, "Multi-frame super-resolution with no explicit motion estimation " *International Conference on Image Processing, Computer Vision, and Pattern Recognition, IPCV,* 2008, pp. 455-459.

[4]. K.-H. Yap, Y. He, Y. Tian, and L.-P. Chau, "A nonlinear L1-norm approach for joint image registration and super-resolution," *IEEE Signal Processing Letters,* vol. 16, pp. 981-984, 2009.

[5]. J. Tian and K.-K. Ma, "Stochastic super-resolution image reconstruction," *Journal of Visual Communication and Image Representation,* vol. 21, pp. 232-244, 2010.

[6]. X. Li, "Super-Resolution for Synthetic Zooming," *EURASIP Journal on Applied Signal Processing,* vol. 2006, pp. 1-12, 2006.

[7]. M. V. Joshi, S. Chaudhuri, and R. Panuganti, "Super-resolution imaging: use of zoom as a cue," *Image and Vision Computing,* vol. 22, pp. 1185-1196, 2004.

[8]. M. Shen and P. Xue, "Super-resolution from observations with variable zooming ratios," in *Proceedings of IEEE International Symposium on Circuits and Systems* (ISCAS), 2010, pp. 2622-2625.

[9]. M. K. Ng, H. Shen, S. Chaudhuri, and A. C. Yau, "Zoom-based super-resolution reconstruction approach using prior total variation," *Opical Engineering,* vol. 46, Dec. 2007.

[10]. Y. He, K.-H. Yap, L. Chen, and L.-P. Chau, "A Nonlinear Least Square Technique for Simultaneous Image Registration and Super-Resolution," *IEEE Transactions on Image Processing,* vol. 16, pp. 2830-2841, 2007.

[11]. B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Transactions on Image Processing*, vol. 5, pp. 1266-1271, 1996.