

Robust Video Super-Resolution with Registration Efficiency Adaptation

Xinfeng Zhang^a, Ruiqin Xiong^b, Siwei Ma^b, Li Zhang^b, Wen Gao^b

^aInstitute of Computing Technology, Chinese Academy of Sciences, Beijing, China

^bSchool of Electronics Engineering and Computer Science, Peking University, Beijing, China
{xfzhang, rqxiong, swma, zhanglili, wgao}@jdl.ac.cn

ABSTRACT

Super-Resolution (SR) is a technique to construct a high-resolution (HR) frame by fusing a group of low-resolution (LR) frames describing the same scene. The effectiveness of the conventional super-resolution techniques, when applied on video sequences, strongly relies on the efficiency of motion alignment achieved by image registration. Unfortunately, such efficiency is limited by the motion complexity in the video and the capability of adopted motion model. In image regions with severe registration errors, annoying artifacts usually appear in the produced super-resolution video. This paper proposes a robust video super-resolution technique that adapts itself to the spatially-varying registration efficiency. The reliability of each reference pixel is measured by the corresponding registration error and incorporated into the optimization objective function of SR reconstruction. This makes the SR reconstruction highly immune to the registration errors, as outliers with higher registration errors are assigned lower weights in the objective function. In particular, we carefully design a mechanism to assign weights according to registration errors. The proposed super-resolution scheme has been tested with various video sequences and experimental results clearly demonstrate the effectiveness of the proposed method.

Keywords: super-resolution, image fusion, image registration, registration efficiency

1. INTRODUCTION

Super-resolution (SR) is a technique to fuse several low-quality and downsampled images into an image of higher quality and resolution. The basic idea is to estimate the missing high-resolution information of an image from a group of low-resolution images describing the same scene, by exploiting the spatio-temporal correspondence among them. In general, the optimal SR reconstruction is reached by minimizing a cost function, e.g, L_2 -norm of the difference between the observed LR image and the one formed by observation model. However this is an ill-posed problem due to insufficient information in LR images and ill-conditioned blur operators. The conventional SR algorithms utilize regularization to stabilize the inversion of the ill-posed problem [1]. The SR process usually consists of three steps. The LR images are firstly aligned through motion estimation, which is referred to as image registration. Then the registered LR images are fused to reconstruct a HR image. A deblurring process is finally used to sharpen the edges and textures in HR image. When applied on video sequences, the effectiveness of the conventional super-resolution technique depends heavily on the efficiency of image registration. In the last two decades, image registration technique has a noticeable improvement, but most of them [2] are based on the assumption of global motion. However, real video sequences may frequently contain complex local motions, which cannot be captured by the motion model in registration and ultimately cause large registration errors. Moreover the registration errors can propagate to multiple pixels during the data fusion process. Therefore, registration errors must be considered in the process of SR reconstruction, especially when there are local and fast motions.

In conventional SR algorithms, all the reference frames are regarded as equally reliable and the different magnitude of registration errors are not taken into consideration. Therefore, it does not work well on video sequences, especially for regions with complex local motion where a perfect image registration is impossible. In order to suppress the negative effect of registration errors in data fusion, He and Kondi [3] proposes to add a weighting term in the cost function, according to the registration errors of each frame. However, such frame-based weights can not reflect the spatial variation of registration errors, and thus cannot reflect the reliability of reference pixels in difference regions. Therefore,

a pixel-based weighted L_2 -norm approach is proposed in [4]. They select weight for each pixel using an exponential function, and the absolute registration error is used as the exponent part. Obviously the weights in [4] are more adaptive than that in [3], but one of the shortcomings is that the weights, each determined by the registration error on a single pixel, are very sensitive to noise and can not reflect the real registration reliability. In the following work [5] and [6], they proposed the region-based weighted L_2 -norm approaches to improve the robustness of the weights, in which a weight is computed using all the registration errors in a region and is assigned to each pixel in the local region. The main problem of [5] and [6] is that it can not reflect the motion difference in the same region, especially when parts of regions are occluded. S. Farsiu et al. proposed a robust super-resolution method, which uses L_1 -norm minimization and robust regularization based on a bilateral prior to deal with different data and noise models in [9] instead of the weighted L_2 -norm approaches.

Hence, in order to further improve robustness of the above SR schemes, we propose a registration efficiency model to reflect the reliability of the registration results for each pixel. In this model, instead of considering the “lack of fit” of a single pixel, we utilize the weighted registration errors in a small window to compute the reliability of a reference pixel. This method is much less sensitive to the random noises in the LR images [7] [8]. In light of block-based registration method in our paper, we add the block registration errors to reflect the registration reliability of blocks. We also use the time distance from the current frame to reflect the reliability of the whole reference frame. We apply our model to the weighted L_2 -norm SR reconstruction method, and experimental results demonstrate that it outperforms the previous methods both in objective and subjective qualities.

The rest of this paper is organized as follows. Section 2 introduces the formulation of observation model for SR problem. Section 3 first introduces the weighted L_2 -norm SR algorithm formulation. Then, it introduces our proposed registration efficiency model in detail. Section 4 gives the performance comparison of our proposed method with other methods. A brief conclusion is given in Section 5.

2. OBSERVATION MODEL AND NOTATION

Consider the desired HR image of size $L_1N_1 \times L_2N_2$ written in lexicographical notation as the vector $\mathbf{X}=[x_1, x_2, \dots, x_N]^T$, where $N=L_1N_1 \times L_2N_2$. The observed LR images are of size $N_1 \times N_2$. The parameters L_1 and L_2 represent the scaling factors in the horizontal and vertical directions, respectively. Let the k th LR image be denoted in lexicographic notation as $\mathbf{Y}_k=[y_{k,1}, y_{k,2}, \dots, y_{k,M}]^T$, for $k=1, 2, \dots, K$ and $M=N_1 \times N_2$. We assume that the LR images are acquired from the same scene with HR image by motion, blurring and downsampling. Meanwhile, assuming that the procedure of LR images acquisition is corrupted by additive noise, we can formulate the k th observed LR image acquisition procedure as follows,

$$\mathbf{Y}_k=\mathbf{D}_k\mathbf{H}_k\mathbf{W}_k\mathbf{X}+\boldsymbol{\eta}_k \quad (1)$$

where \mathbf{W}_k is a warp matrix of size $L_1N_1L_2N_2 \times L_1N_1L_2N_2$, \mathbf{H}_k represents a blur matrix, \mathbf{D}_k is a $(N_1N_2) \times L_1N_1L_2N_2$ downsampling matrix, and $\boldsymbol{\eta}_k$ represents an additive noise for the k th frame in lexicographically order. If we assume the downsampling and blurring are constant for all the LR images, (1) can be simplified as

$$\mathbf{Y}_k=\mathbf{D}\mathbf{H}\mathbf{W}_k\mathbf{X}+\boldsymbol{\eta}_k \quad (2)$$

In this paper, we assume that D and H are already known.

3. SUPER-RESOLUTION WITH REGISTRATION RELIABILITY MODEL

3.1 Weighted L_2 -norm based video super-resolution algorithm

Taking into account the previous works [3]-[6], the proposed cost function of weighted L_2 -norm based video super-resolution algorithm can be formulated as follows,

$$J(\mathbf{X})=\sum_{k=1}^K(\mathbf{Y}_k-\mathbf{D}\mathbf{H}\mathbf{W}_k\mathbf{X})^T P_k(\mathbf{Y}_k-\mathbf{D}\mathbf{H}\mathbf{W}_k\mathbf{X})+\lambda\|\mathbf{C}\mathbf{X}\|^2 \quad (3)$$

where P_k is the weights matrix for the k th LR frame, which is a diagonal matrix. And the operation C is generally a high-pass filter, $\|\bullet\|$ represents a L_2 -norm and λ is the regularization parameter. The first item represents the fidelity to real data and the second item is a priori knowledge concerning that a desirable solution should be smooth. The second item will stabilize the solution, especially when the number of LR images is not sufficient. The cost function in (3) is convex and differentiable with the use of a quadratic regularization item. Therefore, we can find a unique estimate HR frame $\hat{\mathbf{X}}$ through the following iteration process,

$$\mathbf{X}^{i+1} = \mathbf{X}^i + \alpha^i \left\{ \sum_{k=1}^K (\mathbf{DHW}_k)^T P_k [\mathbf{Y}_k - (\mathbf{DHW}_k)\mathbf{X}] - \lambda \mathbf{C}^T \mathbf{C} \mathbf{X} \right\} \quad (4)$$

where α are the step size of convergence.

3.2 Registration efficiency model

Super-resolution technique reconstructs a HR frame making use of information both in spatial and temporal domain. However, all parts of reference frames may not provide efficient information for HR reconstruction. For example, when there are occlusions between referenced LR frames and current LR frame, the temporal information of these occlusion parts is not helpful in reconstructing HR frame. On the contrary it may destroy image structures. Therefore, we proposed a new registration efficiency model to reflect the reliability of temporal information after image registration. Joint with the block-based registration method in our paper, we also consider all the registration errors in each block integrally. Meanwhile, we use the time distance from the current LR frame to be reconstructed as the reliability measurement for the whole frame of referenced LR images. We describe our registration reliability model with the following pseudocode.

IF $Block(k, m, n).error > T1$ THEN

{ all $p(k, s, t)$ in $Block(k, m, n)$ is set to 0 }

ELSE IF $Pixel(k, s, t).error \leq T2$ THEN

$$p(k, s, t) = \frac{1}{(1 + Block(k, m, n).error)} \cdot e^{-\left(\sum_{j=s-ws}^{s+ws} \sum_{i=t-ws}^{t+ws} w(k, j, i) |Pixel(k, j, i).error|\right)} \cdot \frac{1}{Time_dis(k)}$$

ELSE $p(k, s, t) = 0$;

In the pseudocode, $Block(k, m, n).error$ represents the average of absolute registration errors in $Block(k, m, n)$, where m and n are the block coordination in the k th LR frame, respectively. $Pixel(k, s, t).error$ is the pixel registration error indexed by the coordination (k, s, t) . $w(k, j, i)$ is a weight value against distance between point (j, i) and (s, t) , and point (s, t) is the center of corresponding window in k th frame,. The notation ws is local window size. $Time_dis(k)$ is the distance between the k th referenced LR image and current LR image against time axis. $T1$ and $T2$ are two thresholds. Our model can be divided into three parts. In the first part, we use the $Block(k, m, n).error$ to reflect the registration reliability of blocks. In the second part, we consider each pixel registration reliability by compare pixel registration errors in a small window, which is centered in the current pixel [8]. The pixel registration errors in a window can not only reflect the reliability of current pixel better, but also improve robust to noise. We take the weighted-sum, $w(k, j, i)$ and $Pixel(k, j, i).error$ in the same window as the exponent parts, due to the registration errors approach Gaussian distribution [4]. In the third part, we consider the reliability of whole frame decrease by time distance increasing. For the current LR image, we assign a higher value especially. The model rejects some pixels which may be occlusion points with the help of thresholds $T1$ and $T2$, which are designed against the whole registration results. The adaptive threshold $T1$ is two times of the average of $Block(k, m, n).error$ in all LR frames and $T2$ is equal to the standard deviation of $Pixel(k, j, i).error$ in all LR frames.

4. EXPERIMENT RESULTS

In this section, we show the performance of our proposed technique. The test sequences are *foreman@352×288*, *students@352×288*, *city@704×576* and *mobile@704×480*. We downsample these video sequences into half size along horizontal and vertical directions, respectively, to generate the LR video sequences as the input set.

We apply traditional *bilinear* interpolation method, L_2 -norm based SR without weights, method in [4] and our proposed method to upsample the LR sequences to original resolution. In order to verify our registration reliability model validity, we use the same registration results and regularization. For different SR methods, they all use five LR frames to reconstruct one HR frame and $\lambda=0.12$. For simplicity, we consider that the blurring is only caused by downsampling. Therefore, the matrix H is an identity matrix. In all the experiments, we set the window size ws as 3. The average PSNR of luminance from 2nd to 27th frame is listed in Table 1. PSNR of *foreman* for each frame from the 2nd to the 27th is shown in Figure.1. From objective quality comparison, our proposed method has obvious improvement than other methods. In Figure.2 and 3, the luminance of color images, (a), (c), (e) and (g), are produced by *bilinear*, L_2 -norm SR without weights, method in [4] and our proposed method, respectively. Chroma of (a), (c), (e) and (g) are all interpolated

by *bilinear*. (b), (d), (f) and (h) are the luminance residual between the original frame and reconstructed HR frames (a), (c), (e) and (g), respectively. From Figure.2 and 3, we can see that *bilinear* interpolation blurred the HR image. The L_2 -norm SR without weights has some outlier pixels, e.g. in the red box of Figure 2 (c), because it uses all the information equally from LR frames and some occlusion parts in LR reference images are employed by mistake. Method in [4] and our proposed method, they both get pleasing HR frames. But from the residual frames, our proposed method is more effective on suppressing outliers and persevering sharp edges, e.g in yellow box of Figure 2 (f) and Figure 2 (h). These results can verify that our registration efficiency model is very effective.

Table 1. PSNR for different sequences with different methods.

sequence	Bilinear	L_2 -norm SR without weights	Method in [4]	Our proposed
foreman	28.9307	30.0645	30.5242	32.1918
students	28.2733	29.1259	29.4137	30.2670
city	27.0728	28.1456	30.2140	30.8164
mobile	20.6430	21.3769	21.1714	22.9716

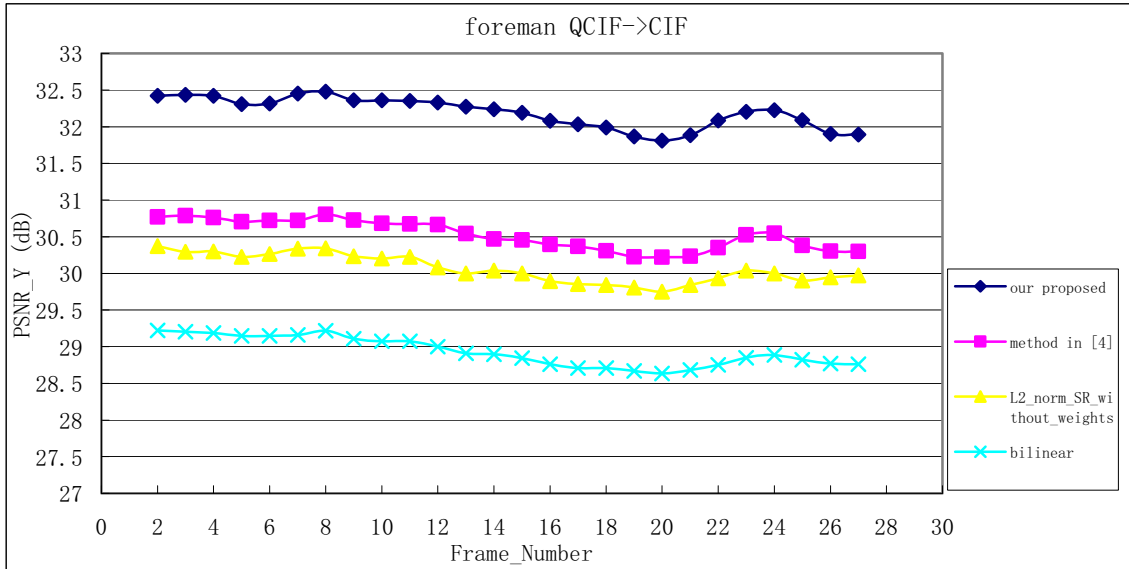


Figure.1 the PSNR results from the 2nd frame to the 27th frame of *foreman*. PSNR results are for luminance component.



(a)



(b)



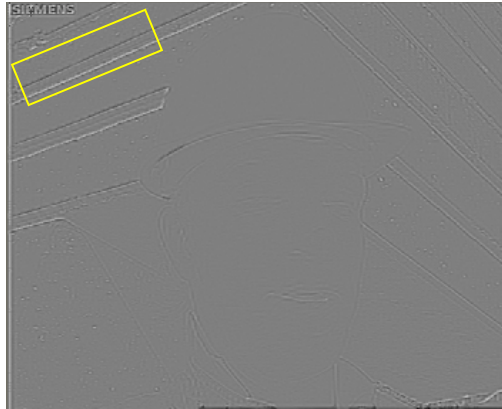
(c)



(d)



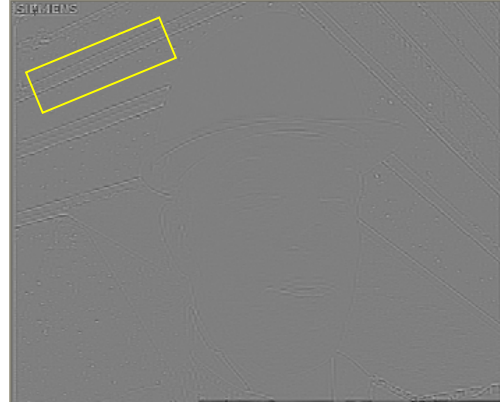
(e)



(f)



(g)

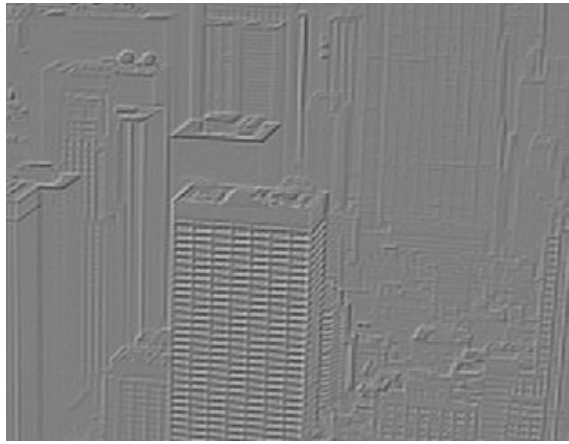


(h)

Figure 2, super-resolution result for *foreman* sequence, (a), (c), (e) and (g) are the HR frames reconstructed by *bilinear*, L_2 -norm SR without weights, method in [4] and our proposed method, respectively. (b), (d), (f) and (h) are the residual images between original and reconstructed HR frames.



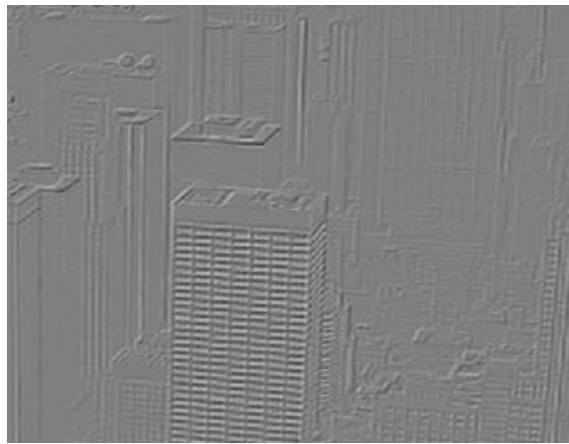
(a)



(b)



(c)



(d)

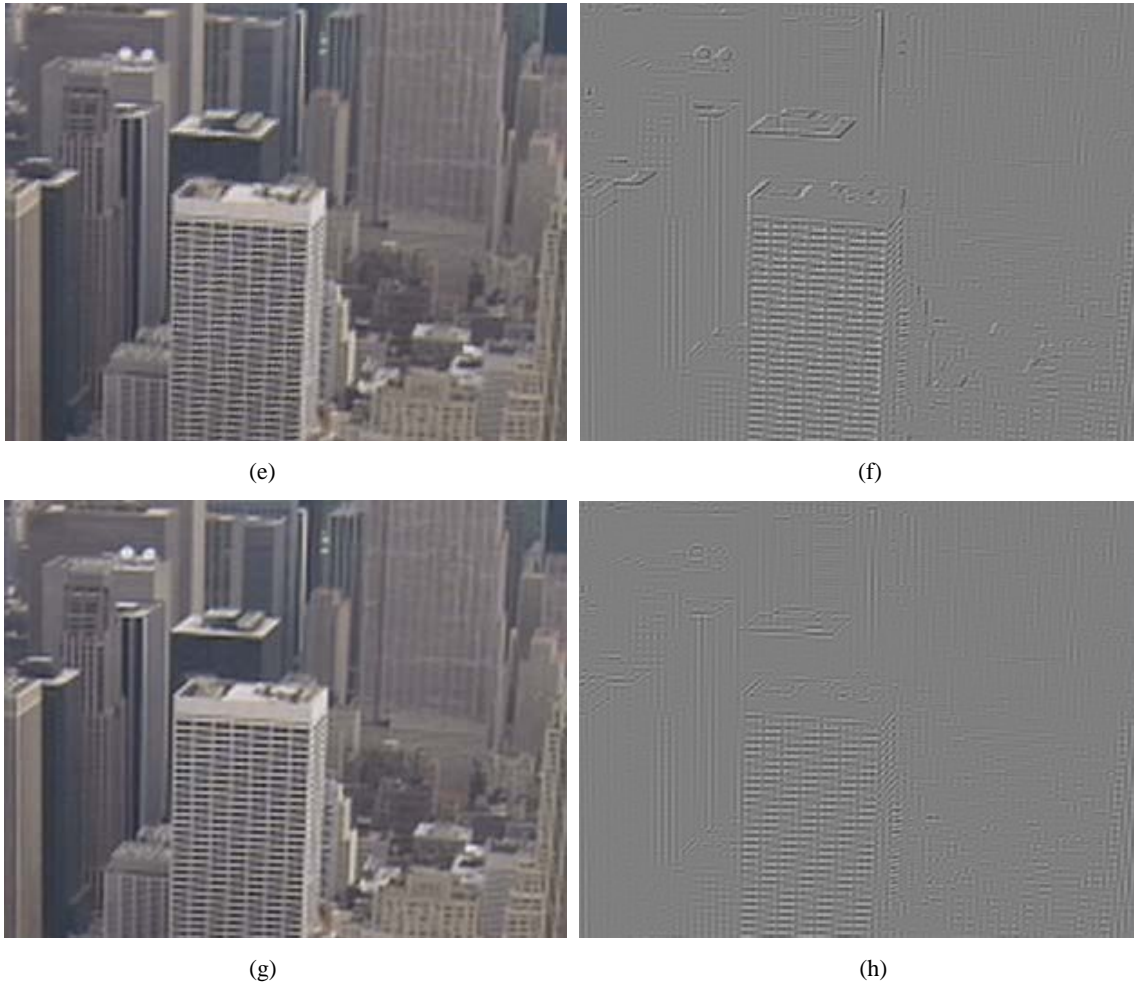


Figure. 3, super-resolution result for *City* sequence, (a), (c), (e) and (g) are the HR frames reconstructed by *bilinear*, L_2 -norm SR without weights, method in [4] and our proposed method, respectively. (b), (d), (f) and (h) are the residual images between original and reconstructed HR frames.

5. CONCLUSION

In this paper, we proposed a new registration efficiency model for weighted L_2 -norm based super-resolution method. Because it can reflect the registration reliability more accurately, it improves the quality of SR reconstruction and outperforms other resolution enhancement methods obviously. Based on the results, it proves that the registration efficiency model is effective.

ACKNOWLEDGEMENT

This work was supported in part by National Science Foundation (60833013, 60803068) and National Basic Research Program of China (973 Program, 2009CB320904)

REFERENCES

1. Park, S., Park, M. and Kang, M.G., "Super-resolution image reconstruction: a technical overview," IEEE signal processing Magazine, vol.20, no.3, pp.21-36, May 2003.

2. Bergen, J. R., Anandan, P., Hanna, K.J., and Hingorani, R., "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.
3. He, H. and Kondi, L. P. "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.
4. Omer, O.A., Tanaka, T., "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, pp. 682222-1-682222-9, Jan. 2008
5. Omer, O.A., Tanaka, T., "Region-Based Super Resolution for Video Sequences Considering Registration Error," in Proc. of the 3rd Pacific-Rim Symposium on Image and Video Technology, Tokyo, Japan, pp.944-954, vol.5414 of Lecture Notes in Computer Science, Jan. 2009
6. Omer, O.A., Tanaka, T., "Region-based weighted-norm approach to video super-resolution with adaptive regularization," in Proc. of 2009 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2009), pp.833-836, Taipei, Taiwan, Apr. 2009.
7. Protter, M., Elad, M., Takeda, H. and Milanfar, P., "Generalizing the non-local-means to super-resolution reconstruction," IEEE Trans. Image Processing, vol. 18, pp.36-51, Jan. 2009.
8. Buades, A. Coll, B. and Morel, J.M. "A non-local algorithm for image denoising", IEEE International Conference on Computer Vision and Pattern Recognition, pp. 60-65, June. 2005.
9. S. Farsiu, D. Robinson, M. Elad, P. Milanfar. "Fast and robust multi-frame super-resolution", IEEE Trans. on Image Processing, vol. 13, no. 10, pp. 1327-1344, Oct. 2004.