To Appear in: 2008 Iranian Conference on Machine Vision and Image Processing (MVIP08), Tabriz University, Tabriz, Iran, Nov. 4-7, 2008.

Reconstruction+Synthesis: A Hybrid Method for Multi-Frame Super-Resolution

Mahmood Amintoosi[1,2] Mahmood Fathy[1]
Nasser Mozayani[1]
[1]Computer Engineering Department,Iran University of Science and Technology
Narmak, Tehran, Iran
[2]Mathematics Department, Tarbiat Moallem University of Sabzevar, Sabzevar, Iran
{mAmintoosi,mahFathy,Mozayani}@iust.ac.ir

Abstract

In this paper a new method for multi-frame Super-Resolution is proposed. Having a low resolution(LR) video and a few high resolution(HR) images from a specific scene is frequently occurred. In addition it is usual that between images and video frames there is some differences, because of different of exposure time, moving objects, camera movement and so on. In the proposed method a Super-Resolution reconstruction method, applies on low-resolution frames for producing a reconstructed HR image. After that a synthesized image with mapping the high resolution training image to the reconstruction result, is produced using a proper transformation model. Fusion of the synthesized image with the reconstruction result makes the final desired HR image of the given LR frames. Experimental results show that our approach is competitive both for quality and quantity with some famous super-resolution methods.

Keywords: Super-Resolution, Synthesis, Reconstruction, Fusion, Homography

1 Introduction

Nowadays using image capturing devices such as mobile sets, which capture low resolution videos and high resolution images have been popular. Enhancing such low resolution images or videos is the subject of Super-Resolution(SR) algorithms. The multi-frame super-resolution problem was first addressed in [1].

The Super-Resolution (SR) techniques fuse a sequence of low-resolution images to produce a higher resolution image. The low resolution images may be noisy, blurred and have some displacement with each other. A common matrix notation which is used to formulate the super-resolution problem [2, 3] is as follows:

$$\underline{Y}_k = DHF_k\underline{X} + \underline{V}_k, \quad k = 1, \ldots, N \tag{1}$$

where $[r^2M^2 \times r^2M^2]$ matrix F_k is the geometric motion operator between the high-resolution frame $\underline{X}$ (of size $[r^2M^2 \times 1]$) and the k^{th} low-resolution frame $\underline{Y}_k$ (of size $[M^2 \times 1]$) which are rearranged in lexicographic order and r is the resolution enhancement factor. The camera's point spread function (PSF) is modeled by the $[r^2M^2 \times r^2M^2]$ blur matrix H, and $[M^2 \times r^2M^2]$ matrix D represents the decimation operator. $[M^2 \times 1]$ vector V is the system noise and N is the number of available low-resolution frames. We assumed that decimation operator D and blur matrix H is same for all images. As in [3] we consider $\underline{Z} = H\underline{X}$, so $\underline{Z}$ is the blurred version of the ideal high-resolution image $\underline{X}$ and the SR problem is broken in two separate steps:

1) Finding a blurred high-resolution image from the low-resolution measurements ($\hat{\underline{Z}}$).
2) Estimating the de-blurred image $\hat{\underline{X}}$ from $\hat{\underline{Z}}$.

Hence the SR problem can be formulated as follows:

$$\hat{\underline{Z}} = \operatorname*{Arg\,Min}_{\underline{Z}} \left[\sum_{k=1}^{N} ||DF_k\underline{Z} - \underline{Y}_k||_p^p \right] \tag{2}$$

In addition of many academic researches about to image and video super resolution[4], recently a commercially product named MotionDSP [1] has been released, which enhances the resolution of movies. The major works in SR domain back to those which try to produce a high resolution(HR) still image or video from a set of low resolution(LR) images. The analysis performed by Lin and Shum[5] indicates that to achieve super resolution at large magnification factors, reconstruction based algorithms are not favorable and one should try other kinds of super resolution algorithms, such as recognition-based algorithms. Hence, recent advances in Super-Resolution techniques show trends towards methods which consider some prior knowledge or models, in addition of LR images as the input of the

[1]http://www.motiondsp.com/

SR algorithm [6, 7]. It can be considered as a special class of SR methods, named learning-based methods[8]. These model-based approaches differ from the reconstruction-based approach in the final step where high-frequency details are recovered from the reconstructed (but possibly blurry) HR image after fusion. Instead of deconvolution, the model-based approach imports plausible high- frequency textures from an image database into the HR image. These methods has gained significant interests in recent years because it promises to overcome the limit of reconstruction-based SR [7]. In [8], Freeman *et al.*used a set of HR images as training data set. For each patch of LR image they searched the training set for finding a match. The corresponding high frequencies patch of the best match has been selected for enhancing the resolution of the LR patch. The output of [9] is not significant different with median filtering.

In [10] we described a method for increasing the resolution of a single LR image using a HR training image. In this paper we discuss how to use the mentioned method for multi frame super-resolution.

The rest of this paper is organized as follows: Section 2 explains the proposed method. Section 3 provides experimental results and section 4 describes the concluding remarks and future works.

2 The Proposed Method

In [10] we proposed a method for single image super-resolution with mapping a HR training image to a LR image, which summarized as follows:

1. Resizing the LR image, for producing an LR image with desired number of pixels,
2. Finding interest points of this resized LR image and the HR image,
3. Removing outliers and estimating the transformation model,
4. Mapping HR image to LR image,
5. Producing a synthesized HR image with fusion of mapped HR image and resized LR Image.

We name the result of the above method as synthesized image. The main idea of this paper is reconstructing a HR image from LR frames using a usual reconstructing approach and then synthesizing the result using method described in [10] by us.

Suppose $Syn(I_1, I_2)$ denotes the synthesized image of LR image I_1 using HR image I_2 based on [10]. Let I_H represent our HR training image. Hence we can formulate the proposed method as follows[2]:

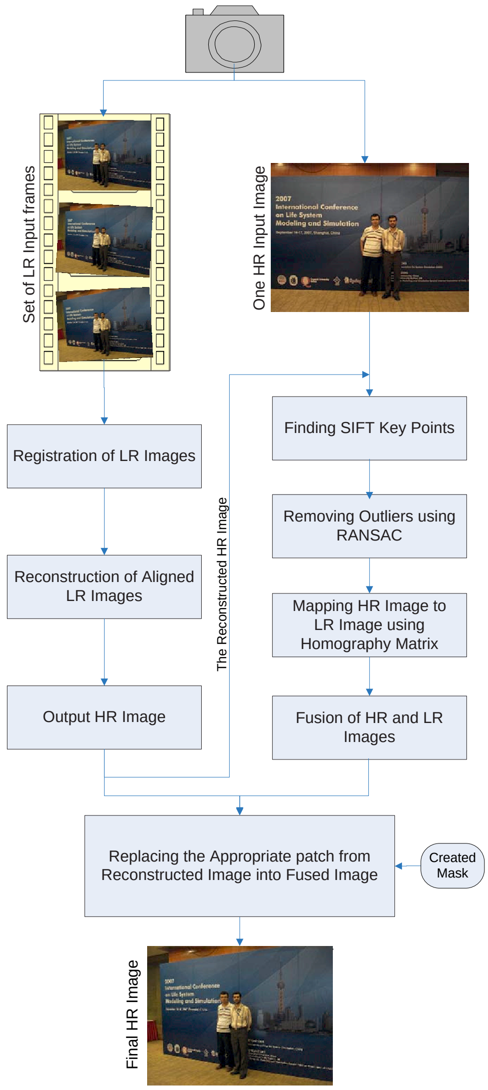


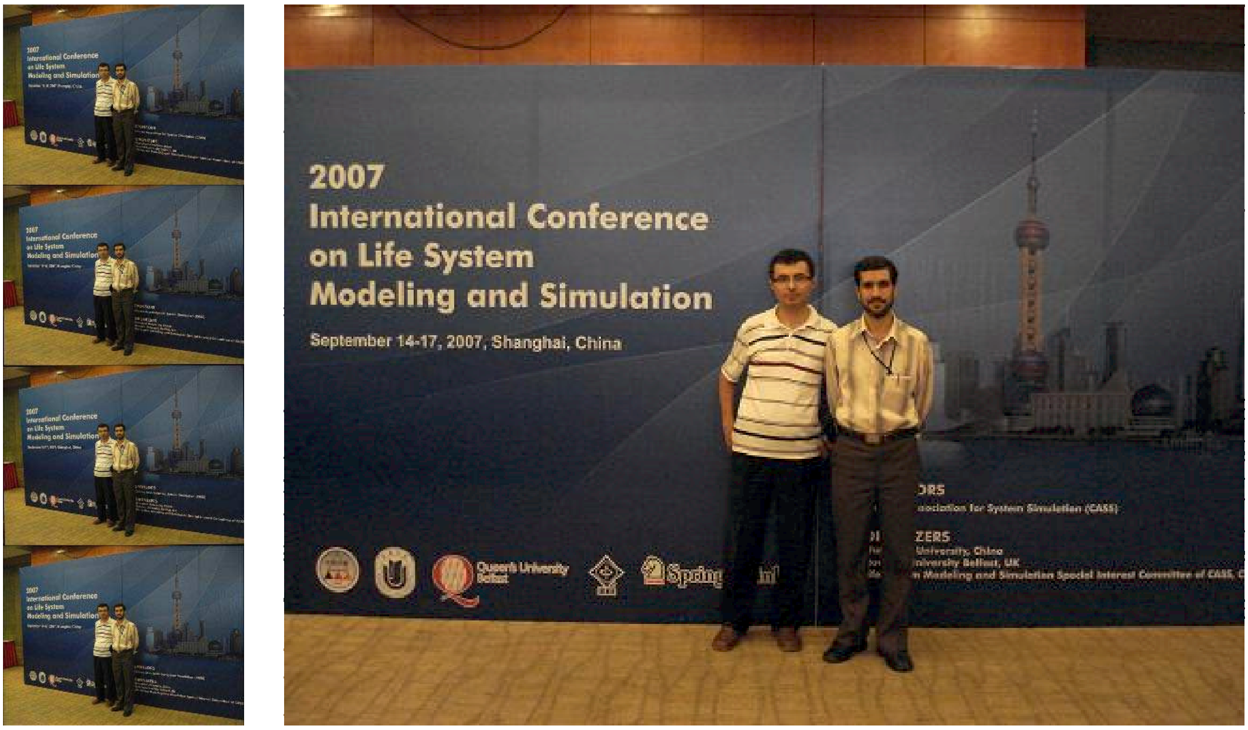
Figure 1: The Overall Framework.

$$Syn(\hat{Z}, I_H) \tag{3}$$

For fusion stage of synthesizing I_1 using HR image I_2, we used Wavelet transform as the most common form of transform image fusion. Hence the high frequency details of LR image I_1 are amplified using I_2.

One situation which is mentioned in [10] as a bottleneck of synthesizing method is moving objects. The scene difference due to light changing can be handled well by Wavelet transform, but the problem of changing the location of objects remains unsolved. The usual methods for background and foreground detection such as [11], which are based on subtraction technique, are not efficient for our problem here, because of illumination changing. At least at this stage of our research we use an interactively created mask for dealing with this trouble. Regions corre-

[2] It is supposed that $\hat{Z}(of size[M \times M])$ is the matrix form of $\underline{\hat{Z}}$ (of size $[M^2 \times 1]$).



(a) LR frames.

(b) The input HR image

Figure 2: The input LR frames each of size (256×192) and the input HR image(1024×768). Note to clear differences in illumination and place of persons between LR frames and HR image. Photos was taken by a Sony DSC-W30 digital camera.

spondence to moving objects, which produced false results in synthesized image, will be identified manually. The resulting regions considered as a mask, which we replace the related regions of synthesized image by the correspondence area from the reconstructed image.

Figure 1 shows the overall framework of the proposed method.

3 Experimental Results

For demonstrating the proposed method and having a quantitative comparison, we synthesized 4 LR images from a HR image. 3 frames have some differences about horizontal and vertical shifts and rotation angles relating to first LR image. Figure 2(a) shows the LR frames and 2(b) shows our HR training image which is captured from a different view of the same scene. Note to clear differences in illumination and place of persons between two pictures. Our aim is to increase the resolution of the first LR frame using other LR frames and HR one. Although having some amount of noise in LR frames is usual in SR context, here our LR images are free of noise. Hence it will be excepted that the interpolation method for reconstruction stage will produce the best result among various reconstruction approaches in the case of precise registration.

Figure 3 shows some output results of various methods evaluated here for comparison. Note to better quality of text in proposed method (3(h)) with respect to 3(b) and 3(f) and better quality of persons in 3(h) with respect to 3(b) and 3(d).

The interactively created mask is shown in figure 4. Figures 5, 6 and 7 shows the quantitative comparisons using different criteria. In these figures Rep,Syn,POCS,IN,RS and RecSyn are denoted for 'Replication', 'Synthesizing'[10], 'Projection Onto Convex Sets', 'Interpolation', 'Robust'[12] reconstruction methods and the proposed method, respectively. As can be seen in figures 5 and 6, the proposed method (RecSyn) is competitive with interpolation reconstruction method which is the best in our test images[3]. The comparison presented in figure 6 is based on Wang's paper[14]. The score returned by Wang's method typically has a value between 1 and 10 (10 represents the best quality, 1 the worst). As it can be seen from the mentioned figure, our method and IN produced the best results among other methods.

The bar chart in figure 7 is a quantitative comparison based on Structural SIMilarity (SSIM)[15],

[3]As it was expected, the interpolation method is the best due to noise freely LR images and using Keren method[13] for registration as one of the best registration methods.

(a) Resized of the first LR image shown in 2(a)

(b) Close-up of a region in 3(a)

(c) Synthesizing the first LR image using method describe in [10] with HR image 2(b) and Wavelet transform as fusion stage.

(d) A magnified region of 3(c)

(e) The result of SR reconstruction with Interpolation method as reconstruction stage on 4 input LR images.

(f) A magnified region of 3(e)

(g) Synthesizing the reconstruction result shown in 3(e) with HR image 2(b)

(h) A magnified region of 3(g)

Figure 3: Visual comparing of the proposed method with some other methods.

Figure 4: The manually created mask for dealing with moving objects.

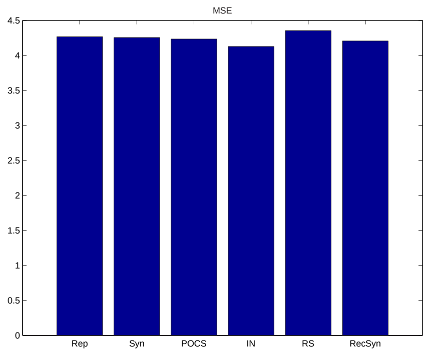


Figure 5: Quantitative comparison of the SR output with different approaches. MSE is between produced HR images and the original HR image.

as a metric for quality assessment of images. Based on this figure, it seems that POCS method is better than the proposed method. But visual comparing the result of POCS (not shown here) and our method showed that the quality of our method is much better than POCS. The output image of POCS was not demonstrates well the high frequency details of the image. The SSIM maps shown in figure 8 clears the mentioned note. As can be seen in SSIM map of POCS method in figure 8, it is far from the original image in high frequency details such as the borders of the texts or buildings with respect to the proposed method (RecSyn). In summery although the quantitative results shows that the proposed method is near the best method here (Interpolation), but as can be seen from subjective comparison in figure 3, the proposed method outperforms other methods in terms of final perceived quality.

4 Conclusion

In this paper a hybrid method for multi-frame Super-Resolution proposed. In summary we combined our idea[10] about using the entire of a training HR image in single image super-resolution, with the usual multi-frame super-resolution reconstruction methods. From a set of LR images, a HR image produced using a reconstruction method such as interpolation or POCS. Then the resulting HR image, synthesized with a training HR image by our recent method [10]. Then the resulting image merged with the reconstruction result in a manner to overcome the moving objects. Our approach is a flexible method,which can be used for super-resolution problems with arbitrary magnification factors up to HR training images. The various comparing approaches showed the good performance of the proposed method. Especially the proposed method outperformed other methods in terms of final perceived quality. As future work we plan to automate dealing with moving objects, which has been done here by a manually created mask.

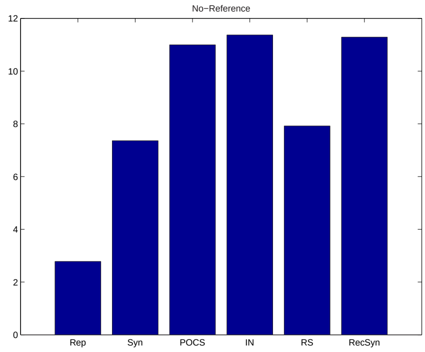


Figure 6: Quantitative comparison using No-Reference Perceptual Quality Assessment method [14].

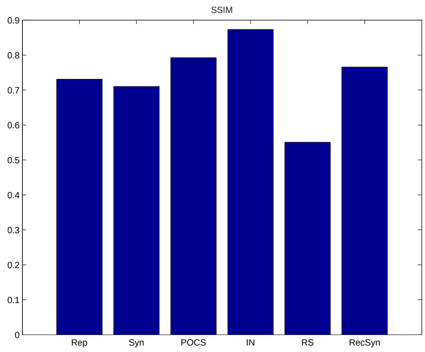
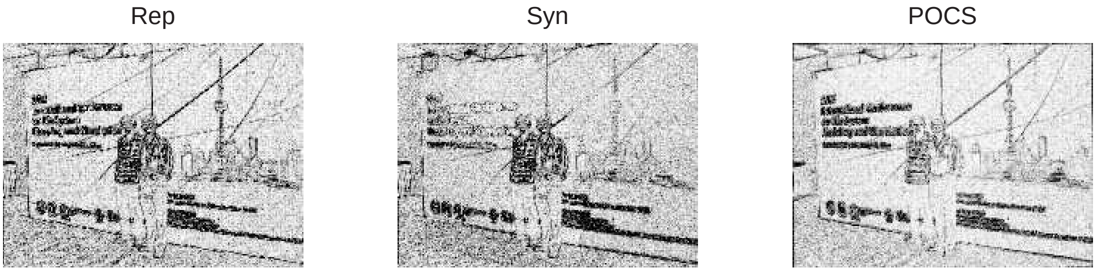


Figure 7: Quantitative comparison using SSIM criteria [15].

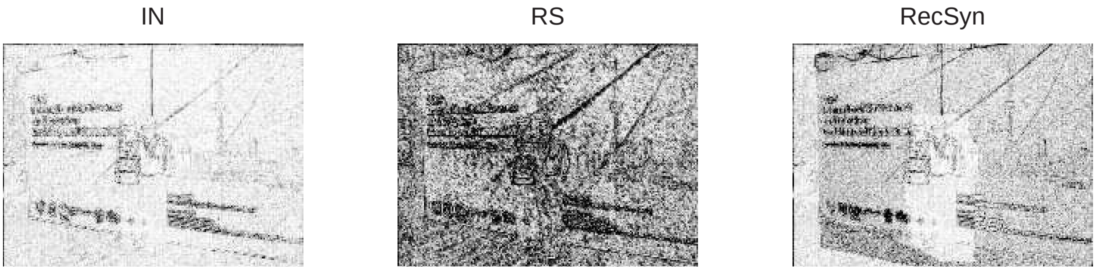
Acknowledgment.
The authors wish to thank to Prof. Milanfar and Dr. Farsiu for providing us with their Super-Resolution package.

References

[1] R. Tsai and T. Huang, "Multiframe image restoration and registration," in *Advances in Computer Vision and Image Processing*, R. Y. Tsai and T. S.



(a) SSIM Maps of Replication, Synthesized[10] and POCS methods.



(b) SSIM Maps of Interpolation and Robust[12] reconstruction methods and the proposed method.

Figure 8: SSIM Map of various methods evaluated in this paper for comparison.

Huang, Eds., vol. 1. JAI Press Inc, 1984, pp. 317–339. 1

[2] M. Elad and A. Feuer, "Restoration of single super-resolution image from several blurred, noisy and downsampled measured images," *IEEE Trans. Image Processing*, vol. 6, pp. 1646–1658, Dec 1997. 1

[3] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Robust shift and add approach to super-resolution," in *Proc. of the 2003 SPIE Conf. on Applications of Digital Signal and Image Processing*, Aug 2003, pp. 121–130. 1

[4] P. Vandewalle, S. Süsstrunk, and M. Vetterli, "A Frequency Domain Approach to Registration of Aliased Images with Application to Super-Resolution," *EURASIP Journal on Applied Signal Processing*, vol. 2006, p. Article ID 71459, 2006. 1

[5] Z. Lin and H. Shum, "Fundamental limits of reconstruction-based superresolution algorithms under local translation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, pp. 83–97, 2004. 1

[6] S. Baker and T. Kanade, "Hallucinating faces," in *FG '00: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000*, USA, 2000, p. 83. 2

[7] T. Q. Pham, "Spatiotonal adaptivity in super-resolution of under-sampled image sequences," Ph.D. dissertation, aan de Technische Universiteit Delft, 2006. 2

[8] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Comput. Graph. Appl.*, vol. 22, no. 2, pp. 56–65, 2002. 2

[9] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 01. Los Alamitos, CA, USA: IEEE Computer Society, 2004, pp. 275–282. 2

[10] M. Amintoosi, N. Mozayani, and M. Fathy, "Spatially varying image super-resolution," in *International Interdisciplinary Conference on Biomedical Mathematics, Promising Directions in Imaging, Therapy Planning and Inverse Problems*, Guizhou, China, 2008, submitted. 2, 3, 4, 5, 6

[11] M. Amintoosi, F. Farbiz, and M. Fathy, "A QR Decomposition based mixture model algorithm for background modeling," in *ICICS2007, Sixth International Conference on Information, Communication and Signal Processing*, Singapore, December 2007, pp. 1–5. 2

[12] A. Zomet, A. Rav-Acha, and S. Peleg, "Robust super resolution," in *Proceedings of the Int. Conf. on Computer Vision and Patern Recognition (CVPR)*, Dec 2001, pp. 645–650. 3, 6

[13] D. Keren, S. Peleg, and R. Brada, "Image sequence enhancement using sub-pixel displacement," in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 1988, pp. 742–746. 3

[14] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of jpeg compressed images," in *IEEE International Conference on Image Processing*, Sept 2002. 3, 5

[15] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004. 3, 5