Reconstruction+Synthesis: a Hybrid method for Video Super-Resolution

Mahmood Amintoosi Mahmood Fathy
Nasser Mozayani
Computer Engineering Department,Iran University of Science and Technology
Narmak, Tehran, Iran
{mAmintoosi,mahFathy,Mozayani}@iust.ac.ir

Abstract

In this paper a new method for video 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 the proposed method a Super-Resolution reconstruction method, applies on each sliding window of low-resolution frames for producing a reconstructed image. Then the HR training image is mapped to the reconstructed image coordinates using a proper transformation model. Fusion of the this image with the reconstruction result makes the final desired HR frame of the given LR frames. Repeating this process for other subsets of LR frames produces the final HR video. Experimental results show the superiour performance of the proposed approach.

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

1 Introduction

Nowadays there are some capturing devices such as mobile sets, which capture low resolution videos and high resolution images. 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 Tsai and Huang [1984].

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 [Elad and Feuer, 1997, Farsiu et al., 2003] 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 Farsiu et al. [2003] 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:

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

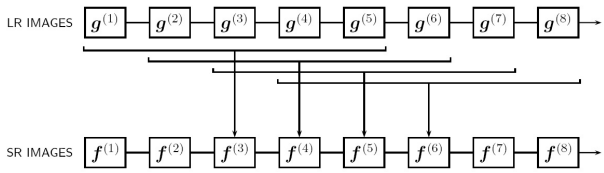


Figure 1: "sliding window" technique for video super resolution [Borman, 2004].

In addition to many academic researches about image and video super resolution, recently a commercially product named as 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 [2004] 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 to LR images as the input of the SR algorithm [Baker and Kanade, 2000, Pham, 2006]. This can be considered as a special class of SR methods, named as learning-based methods [Freeman et al., 2002]. 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 following applying 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 [Pham, 2006]. Freeman et al. [2002]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 Chang et al. [2004] is not significantly different from applying median filtering.

All of the multi frame super resolution techniques may be applied to video restoration by using a shifting window of processed frames as illustrated in figure 1. For a given superresolution frame, a "sliding window" determines the subset of LR frames to be processed to produce a given super-resolution output frame. The window is moved forward in time to produce successive super resolved frames in the output sequence [Borman, 2004].

In this paper a different approach for video Super-Resolution is proposed which is based on this additional assumption that some high resolution images of the same scene is available. Here instead of taking small training patches from HR images, we considered the whole part of HR image for increasing the resolution of the input low resolution video. It is supposed that the HR images may be different from the LR frames from the following aspects:

View Point , due to camera movement,

Illumination , due to different of exposure time or taking photos in distinctive times.

Resolution , due to unequal zooming or changing the resolution setting of camera, or using different devices for image capturing.

In many situations we encountered with the above conditions. Some cases, which some one may have LR and HR images from a specified scene are as follows:

- The owner of digital camera, takes photos with different resolution by manually setting the capturing resolution, because of storage capacity limitation.
- Reducing the camera resolution mistakenly by the owner or some other one.
- Having photos which are captured in different times or by different devices from a scene, that have some differences about view point, resolution and lighting.
- Having LR video frames and HR images, due to camera limitations.

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

In the above situations he/she likes to enhance his/her LR video using HR images. Sometimes our HR images could not cover the entire scene of LR image. This leads to resulting images with spatially varying resolution. Only the resolution of those parts of the LR image which have correspondence in HR images will be increased. One application of this case is enhancing the resolution of a desired region of LR frame, for instance the licence plate of a car in a LR frame of an accident scene, or the face of homicide in a terror scene, using another HR image of the desired region. In contrast to Szeliski [2006], which noted that *"variable resolution image representations and viewers are **not** common"*, based on the aforementioned situations, we believe that *"variable resolution image representations and viewers are **now** common"*.

In Amintoosi et al. [2009] we described a method for increasing the resolution of a single LR image using some HR training images. In this paper we discuss how to use the mentioned method for video super-resolution. The main idea of the proposed method is reconstructing a HR image from LR frames using a usual reconstructing approachs and then synthesizing the result using the method described in Amintoosi et al. [2009].

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

Since the proposed method is based on Amintoosi et al. [2009], at first we will have a quick look to this approach; and then we will discuss how the mentioned method is used for video resolution enhancement.

The chief idea of the proposed method in Amintoosi et al. [2009] is mapping each HR image to a suitable region of the LR image and fusing the result. The resulting output takes its low frquency details from input LR image and the high frequncy details from HR training images, in the common areas. As the main steps are identical for every HR image, the method is described for one HR image and the process is repeated for other HR images. The main steps of the proposed method are summarized as follows:

1. Resizing the LR image, for producing an LR image with the 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. Tuning the model by an area-based registration method,
5. Mapping HR image to LR image,
6. Producing a synthesized HR image with fusion of mapped HR image and resized LR Image.

Depending on the existance of a propoer HR image to each region of LR image, the resolution of that region may be enhanced according to the HR resolution. We name the result of the above method as synthesized image. In the following we will see how to use this method for video enhancement.

2.1 Video Super Resolution Through Image Synthesizing

The aim of this section is to enhance a low resolution video frames, based on the usual reconstruction methods and using a high resolution training image. It is supposed that the LR video frames have some displacement to eachother, otherwise the usual super resolution reconstruction will fail to produce an output better than input frames. This assumption is held in situations such as recording video by a hand held camera. An instance HR training image in addition to four synthesized LR video frames from the same scene of HR image is shown in figure 2. The main idea of the proposed method is first reconstructing a HR image from LR frames using a usual reconstructing approach and then synthesizing the result using method described in Amintoosi et al. [2009].

Let N is the window size of the 'sliding window' technique and $\hat{Z}^{(i)}$ is the reconstruction method of the i^{th} window. Also suppose that running the proposed algortihm in Amintoosi et al. [2009] on $\hat{Z}^{(i)}$ and $g^{(i)}$ using HR training image T yields output images $f^{(i)}$ and $h^{(i)}$, respectively. If the reconstruction result produces good result, then $f(i)$ has better quality than both $\hat{Z}^{(i)}$ and $h^{(i)}$. The reason is that $f^{(i)}$ benefits the information of both the reconstructed image $\hat{Z}^{(i)}$ and the synthesized image $h^{(i)}$.

(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 the clear differences in illumination and place of persons between LR frames and HR image. Photos was taken by a Sony DSC-W30 digital camera.

The method proposed in Amintoosi et al. [2009] has been used for still images, hance the problem of moving objects was not addressed there. If the common region of LR frame and HR image contains some different patches like moving objects, we need a method for dealing it. The usual methods for background and foreground detection such as Amintoosi et al. [2007], which are based on subtraction technique, are not suitable here. In addition of illumination changing, there are some displacements between LR frames, which makes the background modeling a difficult task. Instead of using such methods, here we create a proper mask for each frame, indicating the common regions of two images which will be synthesized.The two appraoches are as follows

Manual, If the moving object regions are obvious in the first frame, mask M is created manually for the first frame. Then the i^{th} mask will be constructed by warping the previuos mask according to the motion parameters between frames i and $i-1$.

Automatic, Each mask is created by a simple subtraction and thresholding method between reconstructed frame,$\hat{Z}^{(i)}$ and registered HR training image, $T(\mathbf{W}(\mathbf{x};\mathbf{p}))$.

Also, the fusion stage of Amintoosi et al. [2009] was not a seamless blending approach. Here we used a version of the multi-band blending approach Burt and Adelson [1983] as a powerfull image fusion technique. With this fusion method one can determine which regions of each image contributed in the final composite image by a mask. We produce the final HR frame $f^{(i)}$ by compositing the static part of the scene from the registered HR image and the regions related to the dynamic parts of the reconstructed image $\hat{Z}^{(i)}$ using the aformentioned mask M. The multi-band blending approach guaranties the smoothness of the transition between this parts, so we have a seam-less result.

Algorithm 1 shows the overall framework of the proposed method. In the mentioned algorithm, $\mathbf{W}(\mathbf{x};\mathbf{p})$ denotes the parameterized set of allowed warps, $\mathbf{p} = (p_1, \ldots, p_n)^T$ is a vector of parameters; $T(\mathbf{W}(\mathbf{x};\mathbf{p}))$ is the training image T warped back onto the coordinate frame of the reconstruction result $\hat{Z}$ and $\mathbf{x} = (x, y)^T$ is a column vector containing the pixel coordinates. The warp $\mathbf{W}(\mathbf{x};\mathbf{p})$ takes the pixel $\mathbf{x}$ in the coordinate frame of the image T and maps it to the sub-pixel location $\mathbf{W}(\mathbf{x};\mathbf{p})$ in the coordinate frame of the image $\hat{Z}$ [Baker et al., 2004]. The warp model may be any transformation model such as affine, homography or optical flow. But in this paper we concentrated on homography model.

In the next section we will mention the experimental results of the proposed algorithm for video enhancement.

Algorithm 1 Regional Varying Video Enhancement.

Input: LR video frames $g^{(1)}, \ldots, g^{(n)}$, HR training image T, magnification factor r, window size N.
Output: HR video frames $f^{(1)}, \ldots, f^{(n)}$.

1: Find the SIFT key-points of HR training image.
2: **for** $i = 1$ to n-N+1 **do**
3: Select the next LR frames $\{g^{(i)}, \ldots, g^{(i+N-1)}\}$, as $\{Y_1, \ldots, Y_N\}$,
4: Run a multiframe reconstruction method on $\{Y_1, \ldots, Y_N\}$ for achieving $\hat{Z}$ in eq (2).
5: Find SIFT key-points of $\hat{Z}$,
6: Remove outliers and estimate the transformation model, $\mathbf{W}(\mathbf{x}; \mathbf{p})$ for mapping T onto coordinate frame of $\hat{Z}$,
7: Tune the warp model by Lucas-Kanade Algorithm [Lucas and Kanade, 1981],
8: Warp T based on $\mathbf{W}(\mathbf{x}; \mathbf{p})$ onto coordinate frame of $\hat{Z}$, $(T(\mathbf{W}(\mathbf{x}; \mathbf{p})))$,
9: Create mask M by thresholding of subtraction of $\hat{Z}$ and $T(\mathbf{W}(\mathbf{x}; \mathbf{p}))$ for detecting moving objects.
10: Produce $f^{(i)}$ by fusion of $\hat{Z}$ and $T(\mathbf{W}(\mathbf{x}; \mathbf{p}))$ according to mask M with multi-band blending approach [Burt and Adelson, 1983].
11: **end for**

3 Experimental Results

We categorised the experiments into two cases. In the first case we used a synthesized short sequence for visual comparison of the proposed method against some others. In the second part we will use some real video sequences, which corrupted with noise.

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 the 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 the clear differences in illumination and place of persons between two pictures. Here our aim is to increase the resolution of the first LR frame using other LR frames and the HR one. Figure 3 shows some output results of various methods and a magnified region of each image for comparison. Note the better quality of text in the proposed method (3(h)) with respect to 3(b) and 3(f) due to using the synthesised method of Amintoosi et al. [2009] and better quality of persons in 3(h) with respect to 3(b) and 3(d) because of the reconstruction method.

Better investigation of figures 3(h) and 3(d) clears that the proposed method (3(h)) has better mapping result with respect to 3(d) (compare the small text '14-17, 2007, Shanghai, China'). The mapping method of both techniques is the approach described in Amintoosi et al. [2009]. The reason is that the proposed method maps the training image onto coordinate frame of $\hat{Z}$, the result of reconstruction method, which has higher frequency details with respect to the LR frame used in 3(d). Hence the SIFT key points are localized more precisely and thus the mapping process is done more precisely.

The mask M has been created manually and is shown in figure 4.

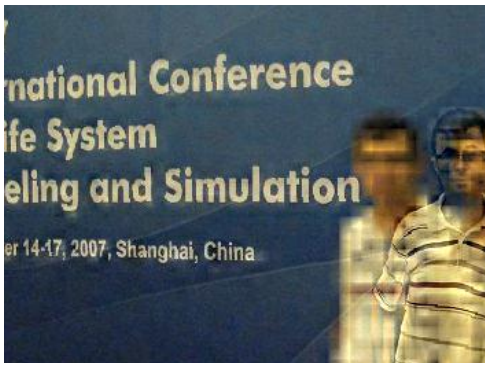
In the figures, Rep,Syn,POCS,IN,RS are denoted for 'Replication', 'Synthesizing'[Amintoosi et al., 2009] (without tuning stage), 'Projection Onto Convex Sets', 'Interpolation', 'Robust'[Zomet et al., 2001] reconstruction methods, respectively. and RecSyn as the 'Reconstruction' following by 'Synthesis' denotes the proposed approach in algorithm 1.

The output image of POCS was not demonstrates well the high frequency details of the image. The SSIM maps shown in figure 5 clears the mentioned note. As can be seen in SSIM map of POCS method in figure 5, the resulting image of POCS has poor high frequency details such as the borders of the texts or buildings compared to the proposed method (RecSyn). In summery in this experiment, the proposed method outperforms other methods in terms of final perceived quality.

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

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

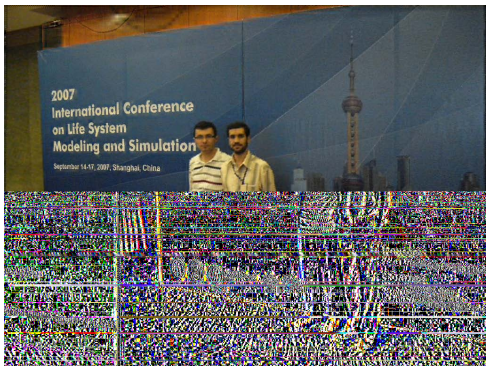
(c) Synthesizing the first LR image using method describe in Amintoosi et al. [2009] with HR image 2(b) and Wavelet transform as fusion stage.



(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)

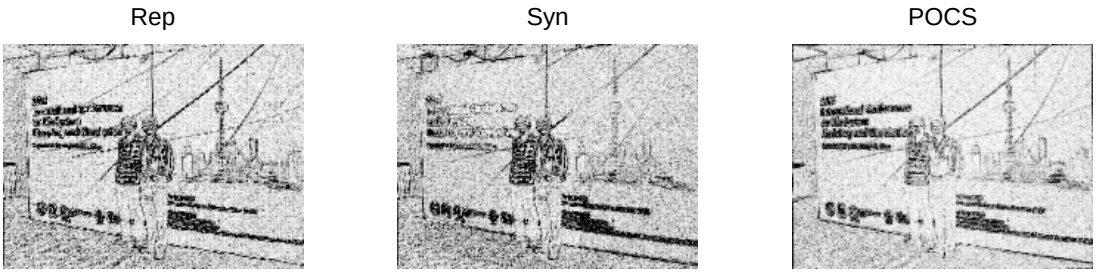


(g) Synthesizing the reconstruction result shown in 3(e) with HR image 2(b), with the proposed algorithm (Algorithm 1)

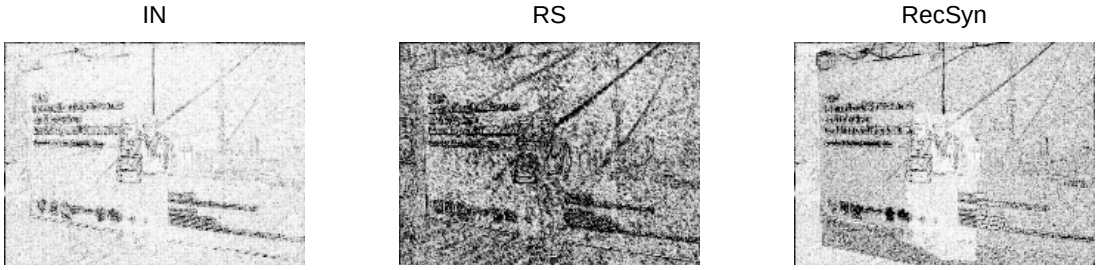
(h) A magnified region of 3(g)

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

Figure 4: The manually created mask used in the results shown in figure 3.



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

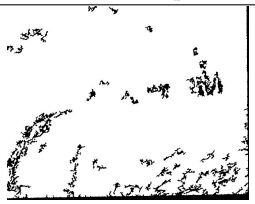
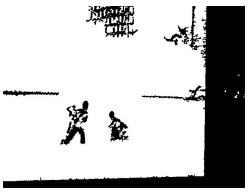
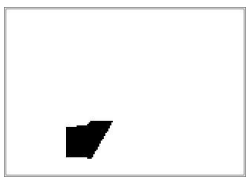


(b) SSIM Maps of Interpolation and Robust reconstruction methods and the proposed method.

Figure 5: SSIM Map of various methods evaluated in this paper for comparison.

In the second experiment we compare the proposed method with some other methods on some real video frames. we have applied them to a broad variety of low-quality videos, including those corrupted by impulse noise, indoor and outdoor video sequences. Because of our assumptions in the proposed algorithms, we have to use special videos and HR training image such that (i) HR image can be transformed to each frame using planar projective model, (ii) for super-resolution porposes the frames must have some displacements against each other and (iii) the moving objects must not be so large to affect the registration procedure. These restrictions prohibited us from using some common LR videos in SR context, so we used our own collected data. Table 1 shows the description of the used video sequences. The different resolutions between LR video frames and HR training images would be obvious by zooming. Two separate sources of motion were present in each sequence. The first kind of motion was created by moving the camera for each individual frame. The second motion was due to the changing the positions of people or waterfall.

Table 1: Description of test sequences.

Sequence Name:	Tehran Park	Tokyo	Shanghai Garden
Frames	60	60	86
First Original Frame			
Resolution	720×576	640×480	160×112
Device:	Panasonic NV-GS75	Sony HDR-SR12E	Sony DSC-W30
First LR Frame			
Resolution	360×288	320×240	160×112
Noisy	Yes	Yes	Yes
Training Image			
Resolution	720×576	640×480	816×612
HR Training is:	From Seq.	Not in seq.	Not in seq.
An instance mask			
Mask:	Automatic	Automatic	Manual

In the following objective results, when the ground-truth HR image was not available (sequences 'LSMS Opening', 'Shanghai Garden') the bicubic resized of the LR frame (without noise) was used as the reference frame. We used the 'sliding-window' techniques with the Interpolation (IN), Iterated Backprojection(BP)[Irani and Peleg, 1991] and Robust Superresolution (RS)[Zomet et al., 2001] as reconstruction stages along with the area based registration method of Keren et al. [1988] for computing the motion parameters between frames. The magnification factor r and the window size were set to 2 and 4 respectively

Figure 6 shows quantitative comparisons of the mentioned methods based on Mean Absolute Error (MAE), Power Signal to Noise Ratio (PSNR) and SSIM for the entire of the 'Tokyo' sequence, in which:

$$MAE = \frac{\Sigma_{i=1}^{N}\Sigma_{j=1}^{M}\Sigma_{q=1}^{Q}|F^q(i,j) - \hat{F}^q(i,j)|}{N.M.Q} \tag{3}$$

and

$$PSNR = 10 \times log\Big(\frac{255^2}{\frac{1}{N.M.Q}\Sigma_{i=1}^{N}\Sigma_{j=1}^{M}\Sigma_{q=1}^{Q}\big(F^q(i,j) - \hat{F}^q(i,j)\big)^2}\Big) \tag{4}$$

where M, N are the image dimensions, Q is the number of channels of the image ($Q = 3$ for color image), and $F^q(i,j)$ and $\hat{F}^q(i,j)$ denote the qth component of the original image vector and the distorted image,

Table 2: MAE, PSNR and SSIM comparisons of the proposed video enhancement algorithm (Algorithm 1) and some super-resolution reconstruction methods over different sequences. The first and the second best scores are highlighted with **Bold** and *italic* letters in each row, respectively.

Method:	RecSyn(Algorithm 1)	BP	IN	RS
MAE($\times 10^{-3}$)				
Tehran Park	*56.93*	102.46	**54.16**	82.57
Tokyo	**50.90**	133.43	*55.94*	96.70
Shanghai Garden	**43.08**	84.63	*47.46*	67.99
PSNR				
Tehran Park	*21.51*	17.58	**22.44**	19.30
Tokyo	**22.12**	15.44	*22.10*	17.97
Shanghai Garden	*22.69*	19.59	**23.72**	21.25
SSIM				
Tehran Park	*0.49*	0.24	**0.50**	0.31
Tokyo	**0.47**	0.13	*0.39*	0.20
Shanghai Garden	**0.61**	0.31	*0.45*	0.36

at pixel position (i, j), respectively. In these experiments, the mentioned criteria has been computed over gray scale version of images (Q=1). The proposed method has the best result against other methods for all MAE, PSNR and SSIM criteria for the 'Tokyo' sequence.

The mean value of MAE, PSNR and SSIM of the proposed video enhancement algorithm (Algorithm 1) and some super-resolution reconstruction methods over test sequences is illustrated in table 2. The best score is highlighted with **Bold** letters for each sequence. As can be seen the proposed method is the best one in the most cases and the second score in the others.

4 Conclusion

In this paper a hybrid method for multi-frame Super-Resolution was proposed. In summary we combined our idea inAmintoosi et al. [2009] 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, was synthesized with a training HR image by our recent method. Hence the proposed approach benefits the information of both the reconstructed image and the synthesized image. Two manual and automatic methods for dealing with dynamic regions of the LR video is prposed and tested in the experiments. 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 objective and subjective comparisons showed the good performance of the proposed method.

Acknowledgment

The authors are indebted to Dr. Vandewalle [Vandewalle et al., 2006] for his Super-Resolution package and Dr. D. Lowe for his SIFT key-points program[2]. They also thank to Dr. Peter Kovesi, Dr. Simon Baker and his co-workers [Baker et al., 2004] for providing many useful MATLAB functions and Dr. Gh. Mohajeri for 'Tokyo' sequence.

References

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*, pages 1–5, Singapore, December 2007. 4

[2]Available online at: http://www.cs.ubc.ca/∼lowe/keypoints/

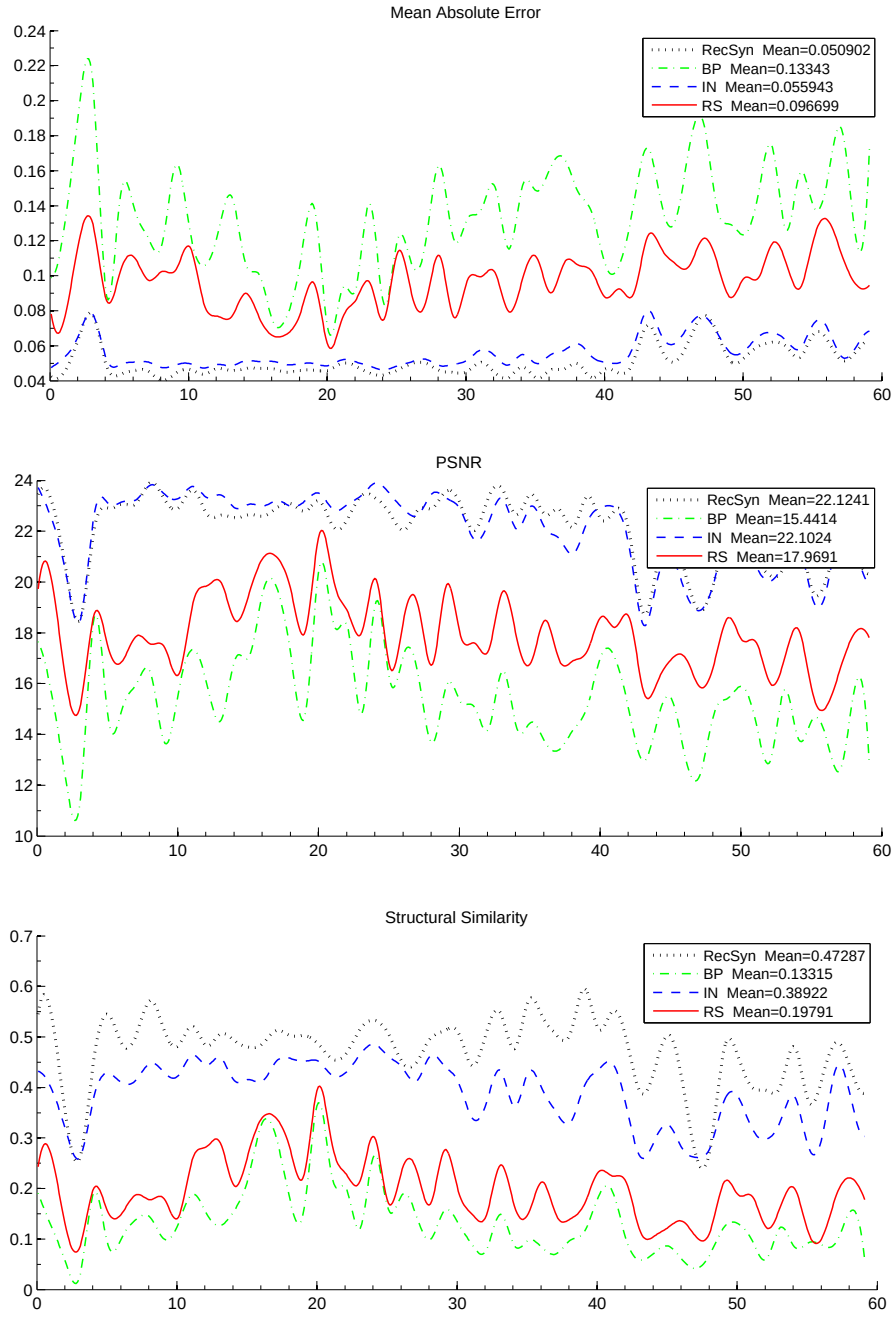


Figure 6: MAE, PSNR and SSIM comparison of the proposed video enhancement algorithm (Algorithm 1) and some super-resolution reconstruction methods for 'Tokyo' sequence .

M. Amintoosi, M. Fathy, and N. Mozayani. Regional varying image super-resolution. In *IEEE International Joint Conference on Computational Sciences and Optimization*, volume 1, pages 913–917, Sanya, China, April 23-26 2009. 3, 4, 5, 6, 9

Simon Baker and Takeo Kanade. Hallucinating faces. In *FG '00: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000*, page 83, Washington, DC, USA, 2000. IEEE Computer Society. ISBN 0-7695-0580-5. 2

Simon Baker, Ralph Gross, and Iain Matthews. Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56:221–255, 2004. 4, 9

Sean Borman. *Topics in Multiframe Superresolution Restoration.* PhD thesis, University of Notre Dame, Notre Dame, IN, May 2004. 2

Peter J. Burt and Edward H. Adelson. A multiresolution spline with application to image mosaics. *ACM Trans. Graph.*, 2(4):217–236, 1983. ISSN 0730-0301. 4, 5

Hong Chang, Dit-Yan Yeung, and Yimin Xiong. Super-resolution through neighbor embedding. In *2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 01, pages 275–282, Los Alamitos, CA, USA, 2004. IEEE Computer Society. doi: http://doi.ieeecomputersociety.org/10.1109/CVPR.2004.243. 2

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

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*, pages 121–130, Aug 2003. 1

William T. Freeman, Thouis R. Jones, and Egon C Pasztor. Example-based super-resolution. *IEEE Comput. Graph. Appl.*, 22(2):56–65, 2002. 2

Michal Irani and Shmuel Peleg. Improving resolution by image registration. *CVGIP: Graph. Models Image Process.*, 53(3):231–239, 1991. ISSN 1049-9652. 8

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)*, pages 742–746, 1988. 8

P. D. Kovesi. MATLAB and Octave functions for computer vision and image processing. School of Computer Science & Software Engineering, The University of Western Australia. Available from: http://www.csse.uwa.edu.au/∼pk/research/matlabfns/. 9

Z. Lin and H.Y. Shum. Fundamental limits of reconstruction-based superresolution algorithms under local translation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26:83–97, 2004. 2

B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI81*, pages 674–679, 1981. 5

Tuan Q. Pham. *Spatiotonal Adaptivity in Super-Resolution of Under-sampled Image Sequences.* PhD thesis, aan de Technische Universiteit Delft, 2006. 2

Richard Szeliski. Image alignment and stitching: a tutorial. *Found. Trends. Comput. Graph. Vis.*, 2(1):1–104, 2006. ISSN 1572-2740. 3

R. Tsai and T. Huang. Multiframe image restoration and registration. In R. Y. Tsai and T. S. Huang, editors, *Advances in Computer Vision and Image Processing*, volume 1, pages 317–339. JAI Press Inc, 1984. 1

Patrick Vandewalle, Sabine Süsstrunk, and Martin Vetterli. A Frequency Domain Approach to Registration of Aliased Images with Application to Super-Resolution. *EURASIP Journal on Applied Signal Processing (special issue on Super-resolution)*, 2006:Article ID 71459, 14 pages, 2006. 9

A. Zomet, A. Rav-Acha, and S. Peleg. Robust super resolution. In *Proceedings of the Int. Conf. on Computer Vision and Patern Recognition (CVPR)*, pages 645–650, Dec 2001. 5, 8