Regional Varying Image 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 single image Super-Resolution using some high resolution images is proposed. It is assumed that each high resolution image shows a region of the low resolution image, with some differences about illumination or view point. These differences may be due to taking photos in different times, camera motion or unequal zooming. In the proposed method each high resolution image is mapped to a resized version of the given low resolution image using homography matrix and RANSAC method, which are known in computer vision context. The mapped image is fused with the LR image for producing a synthesized image. The mentioned method is repeated for each of the HR images. The resulting image has higher resolution only on its regions corresponding to HR images. The experimental results show the superior performance of the proposed method against some other methods in term of final perceived quality.

(a) LR Image (b) HR No.1 (c) HR No.2 (d) HR No.3

Figure 1. One LR and three HR images of a portion of bas relief of Darius, which have different resolutions, illuminations and view points.

1. Introduction

In recent years digital cameras have been papular. Many of these devices, have optical zoom and can save photos with different resolutions. Enhancing the images and videos is one of the growing domain of image processing applications. Among the various image enhancement and restoration methods, Super-Resolution(SR) methods are the only ones which produce an output image that has higher resolution than the input image. The special case which we have one input and one output named as Single-Input Single-Output (SISO) category in Super-Resolution context[3]. The origin of the more general from of Super-Resolution known as Multiple Input Single Output(MISO) come back to work of Tsai and Huang [17] in 1984, motivated by the need to improve the resolution of images acquired by the Landsat 4 satellite [3]. The analysis performed by Lin and Shum[12], 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 model-based or example-based algorithms. The model-based approaches import plausible high-frequency textures from an image database into the low resolution image. Figure 1 shows one Low Resolution (LR) and three High Resolution (HR) images –as training data– from bas relief of Darius. The view point, resolution and the illumination of images are slightly different with each other. These methods have gained significant interests in recent years because it promises to overcome the limit of reconstruction-based SR [15].

In previous example-based super-resolution algorithms [2, 6, 10] during the training phase, pairs of LR and the corresponding HR image patches are collected. Then, in the super-resolution phase, each patch of the given LR image is compared to the stored LR patches, and the HR patch corresponding to the nearest LR patch is selected as the output. Freeman *et al*.[6] used a set of HR images as training data set. The super-resolution was performed by the Nearest Neighbor-based estimation of high-frequency patches based

978-0-7695-3605-7/09 $25.00 © 2009 IEEE
DOI 10.1109/CSO.2009.278

on the corresponding patches of input low-frequency image. The corresponding high frequencies patch of the best match has been selected for enhancing the resolution of the LR patch.

Although the mentioned method has already shown an impressive performance, there is still room for improvement if we do not restrict ourself to small patches. In this paper a different approach for single image Super-Resolution is proposed which is based on this extra 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 image. It is supposed that the HR image may be different with the LR image 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. An example is when the owner of digital camera, takes photos with different resolution, because of storage capacity limitation or due to camera restrictions (such as mobile phones). In the these situations he/she likes to enhance his/her LR images using HR images. Sometimes our HR images could not cover the entire scene of LR image. In the proposed method, only the resolution of those parts of the LR image correspond to HR images will be increased. In contrast to [16], 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"*.

The reminder of this paper is organized as follows: the proposed method is explaind in section 2. The experimental results and conclusion are provided in sections 3 and 4, respectively.

2. The Proposed Method

As mentioned in the previous section, the previous model-based methods act on small patches. In the proposed method we use the entire of training HR image for resolution enhancement of the imput LR image.

The chief idea of the proposed method is mapping each HR image to a suitable region of the LR image and fusing the result, so that we have a homogenous output. The overall framework of the proposed method is illustrated in figure 2. Because the main steps are identical for every HR image, the proposed method is described for one HR image. The process repeats for other HR images. The main steps

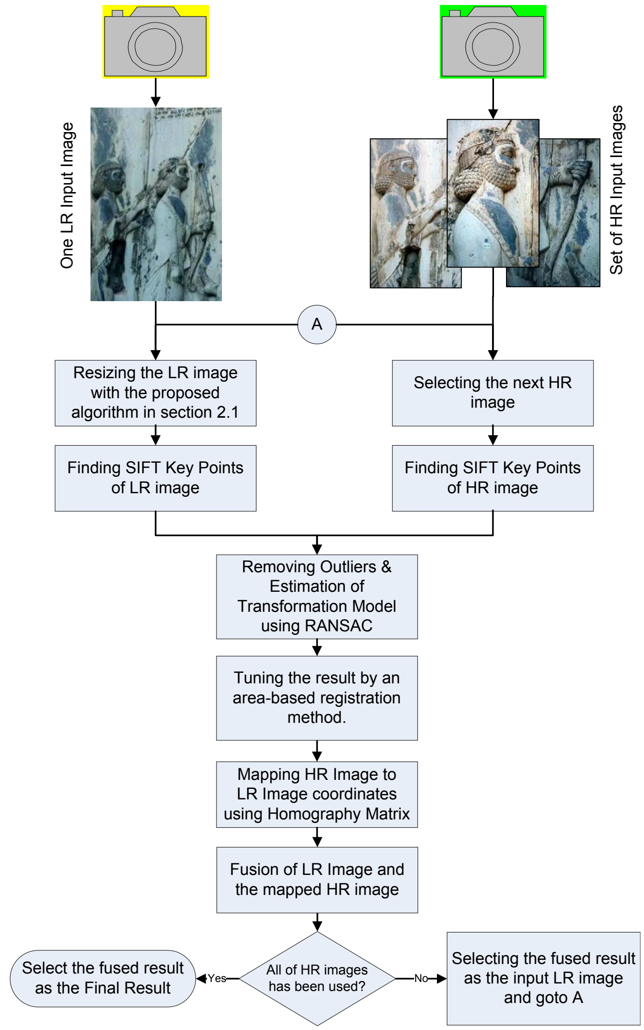


Figure 2. The overall framework of LR image enhancement using HR Images.

of the proposed method are resizing the LR image, estimation of the transformation model, mapping HR image to LR image, and fusion of mapped HR image and resized LR image. In the following sub-sections each stage illustrated in figure 2 will be explained with more details.

2.1. Resizing the LR image

As [6] we apply an initial analytic interpolation, such as replication or cubic spline, for enlarging the LR image. This generates an image of the desired number of pixels that lacks high-resolution details. But in contrast to usual super-resolution methods, which is supposed that magnification factor is an integer number; here this factor is dependent to the relative resolutions of LR and HR images. A roughly estimation of scale factor may be taken by visually investigation of two images; but here an automated method for estimating the resizing factor, based on their common key-points is proposed. Suppose that $\{(P_1, Q_1), \ldots, (P_N, Q_N)\}$ are N corresponding key-point pairs of two images, where $P_i = (x_i, y_i)^T$ is the coordinate of the i^{th} key-point in HR image, and Q_i is its corresponding point in LR image. Let $\{r_1, \ldots, r_n\}$ and $\{s_1, \ldots, s_n\}$ be two sets containing $n \leq N$ random indices ($1 \leq r_i, s_i \leq N$). The distance $d_i^{HR} = norm(P_{r_i} - P_{s_i})$ is the Euclidian distance between two points P_{r_i} and P_{s_i} in HR image; and $d_i^{LR} = norm(Q_{r_i} - Q_{s_i})$ is the distance between the corresponding points of P_{r_i} and P_{s_i} in LR image. The value $\frac{1}{n}\sum_{i=1}^{n} d_i^{HR}/d_i^{LR}$ is an estimation of the scale of two images. For robustness against the outliers, instead of averaging, the median of the proportion of key-points distances has been used as follows:

$$scale = Median\{d_1^{HR}/d_1^{LR}, \ldots, d_n^{HR}/d_n^{LR}\} \qquad (1)$$

and resize the LR image with the computed $scale$, only if $scale > 1$.

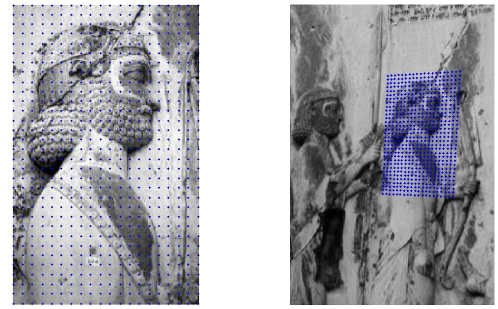
Our experimental results show that resizing the LR image, with the mentioned method produces larger number of inlying matches.

2.2. Finding Interest-points

For mapping HR image to LR image we have to find the homography matrix between them. A feature based registration method has been used here. Among the usual feature points, we tried Harris corners[8] and Lowe SIFT key-points[13] as interest-points and found that SIFT key-points produce better transformation models and are more suitable than Harris corners. Many of the huge amount of possible key-points pairs, are removed based on the suggestion of D. Lowe[13]: the matches are identified by finding the 2 nearest neighbors of each key-point from the first image among those in the second image, and only accepting a match if the distance to the closest neighbor is less than a specified threshold of that to the second closest neighbor. But this set contains some false matches yet, hence we have to use a better outlier detection method for achieving a reliable transformation model between images.

2.3. Outlier Removal and Transformation Model Estimation

RANdom SAmple Consensus (RANSAC) algorithm of Fischler and Bolles[5] partitions the data set into inliers (the largest consensus set) and outliers (the rest of the data set), and also delivers an estimate of the model, computed from the minimal set with greatest support [9]. In this paper, the model is a planar homography, and the data is a set of 2D point correspondences, hence the minimal subset consists of four correspondences. In the next section mapping the HR image onto LR image has been done using the estimated model by RANSAC.



(a) Some points on HR Image.

(b) The mapping of the specified points on LR image.

Figure 3. Mapping some points on HR image to LR image coordinates using the estimated homography matrix by RANSAC.

2.4. Mapping HR image on LR image using the Homography Matrix

By definition[9] a planar projective transformation is a linear transformation on homogeneous 3-vectors represented by a non-singular 3x3 matrix H:

$$\begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \qquad (2)$$

so that the point X can be transformed to X' by $X' = HX$. Now, HR image can be mapped to LR image with the homography matrix estimated by RANSAC. Figure 3 shows the result of the mapping some points in HR image to LR image using the estimated H. In transforming an image, a situation which may be happen frequently, is transforming two or more points in source image to one point in transformed image. It causes some holes in transformed image. If X' be a hole point in transformed image, we estimated its corresponding point from source image by $H^{-1}X'$ and filled the hole by the value of X.

2.5. Tuning the Homography with an Area-based Registration Method

Some times the estimated transformation model is not accurate enough. Usually the motion parameters estimated by the feature-based method, are refined by an area-based method [15]. One of the famous registration methods which is used here is the pioneering work of Lucas and Kanade [14, 1]. This is an area-based method which is based on using of a Taylor series approximation of the images. The motion parameters are the unknowns in the approximation, and they can be computed from the set of equations that can be derived from this approximation.

2.6. Fusion

Image fusion is the process by which two or more images are combined into a single image retaining the important features from each of the original images. Fusion techniques include the simplest method of pixel averaging to more complicated methods such as principal component analysis and wavelet transform fusion[7, 11].

In the first approach we used wavelet transform fusion as the most common form of transform image fusion. Wavelet transform fusion is more formally defined by considering the wavelet transforms ω of the two registered input images $I_1(x, y)$ and $I_2(x, y)$ together with the fusion rule ϕ. Then, the inverse wavelet transform ω^{-1} is computed, and the fused image $I(x, y)$ is reconstructed:

$$I(x, y) = \omega^{-1}\Big(\phi\big(\omega(I_1(x, y)), \omega(I_2(x, y))\big)\Big) \qquad (3)$$

Since wavelet coefficients having large absolute values contain the information about the salient features of the images such as edges and lines, a good fusion rule is to take the maximum of the corresponding wavelet coefficients. But the situation encountered in this paper made us to take a different rule. We want to amplify the high frequency details of LR image using another HR image which represents the salient features better than LR image, obviously. This leads us to use the low frequency information of LR image and high frequency details of HR image as fusion rules for reconstructing with inverse wavelet transform. When the LR and HR images have illumination differences this approach performs well like blending procedure in Panorama context [4].

Also we tested the pixel averaging fusion method. Our experimental results on some other images (not shown here) showed that fusion with wavelet produces better results than pixel averaging method.

As was shown in figure 2, the above procedure repeated for each of the training HR images. The result of each iteration considered as the new input LR image for the next iteration.

3. Experimental Results

Comparison with criteria such as mean square error impeded by lack of a high resolution version of our LR image. Hence we used the method introduced by Wang *et al.*[18]

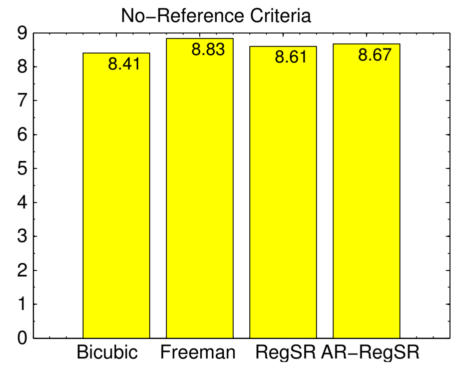


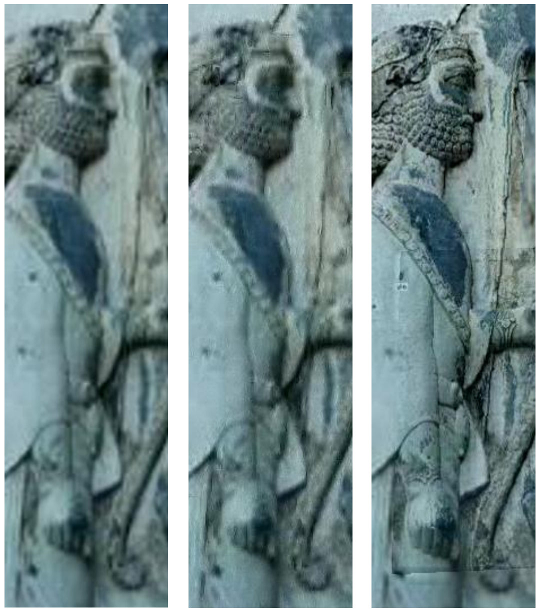
Figure 4. Quantitative comparison using No-Reference Perceptual Quality Assessment method [18].

which is a no-reference quality measurement algorithm for JPEG compressed images. Figure 4 shows the comparison results for various methods. In figure 4 RegSR stands for the proposed method (Regional Super-Resolution) without the area-based tuning stage and AR-RegSR stands for the proposed method with the mentioned refining step. The score returned by Wang's method typically has a value between 1 and 10 (10 represents the best quality, 1 the worst). As can be seen in figure 4 the two forms of the proposed method have better scores with respect to classical bicubic interpolation; but the Freeman method achieved the highest score.

Figure 5 shows a subjective comparison between different methods on a magnified portion of their results. The proposed method (5(c)) produced the best result. Because we have not applied a seamless blending approach, the border of the mapped regions are obvious in figure 5(c)). Although based on the quantitative comparison (figure 4) Freeman's method is better than our approach, but inspection of the results shown in figure 5 indicates that the proposed method is very better than the Freeman method, visually. In the experiments, the parameter n in equation 1 was set to 10.

4. Conclusion

In this paper, we proposed a single image super-resolution method that, in principle, can be used for enhancing the resolution of some parts of LR image with arbitrary magnification factors up to HR training images. We accepted a few differences between LR and HR images including: slightly differ in view point, illumination and zooming differences. The high frequncies details of the input LR image are amplified by fusing a mapped version of HR image to it. In contrast to previous works, which enhance the whole LR image, since the used train-



(a) Bicubic (b) Freeman[6] (c) AR-RegSR

Figure 5. Close-up of bicubic resizing method, Freeman's example-based method [6], and the proposed method for enhancing the image shown in figure 1(a) using HR images 1(b)-1(d).

ing HR images did not covered the entire LR image plane, the resolution of some part of LR image was enhanced. It is why we named our method *Regional Varying Super-Resolution*. Many methods use integer magnifying factor but our method can be used with any magnification factor up to HR images, due to the proposed resizing algorithm. The proposed method can be used in Panorama context when the images in hand are non-overlapped with each other, but a low resolution image from the entire scene is available. The LR image can be used as a canvas for stitching the non-overlapping images to it and producing the desired Panorama. Our experiments [1] showed that this algorithm outperforms in quality the classical bicubic interpolation and the example based method of Freeman. Dealing with parallax, moving objects and shadows are the subject of the future researches in this area.

Acknowledgment. The authors wish to thank Dr. Peter Kovesi[2] and Dr. Simon Baker and his co-workers [1] for providing many useful MATLAB functions.

[1] All of the images shown in the paper, in addition to some intermediate results are available online at:
http://webpages.iust.ac.ir/mamintoosi/DataSets/CSO09Regional.zip

[2] http://www.csse.uwa.edu.au/∼pk/research/matlabfns/

References

[1] S. Baker, R. Gross, and I. Matthews. Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56:221–255, 2004.

[2] S. Baker and T. 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.

[3] S. Borman. *Topics in Multiframe Superresolution Restoration*. PhD thesis, University of Notre Dame, Notre Dame, IN, May 2004.

[4] M. Brown and D. G. Lowe. Recognising panoramas. In *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*, page 1218, Washington, DC, USA, 2003. IEEE Computer Society.

[5] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981.

[6] W. T. Freeman, T. R. Jones, and E. C. Pasztor. Example-based super-resolution. *IEEE Comput. Graph. Appl.*, 22(2):56–65, 2002.

[7] A. A. Goshtasby and S. Nikolov. Image fusion: Advances in the state of the art. *Inf. Fusion*, 8(2):114–118, 2007.

[8] C. Harris and M. Stephens. A combined corner and edge detector. In *Proc. Fourth Alvey Vision Conf.*, volume 15, pages 147–151, 1988.

[9] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[10] A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin. Image analogies. In E. Fiume, editor, *SIGGRAPH 2001, Computer Graphics Proceedings*, pages 327–340. ACM Press / ACM SIGGRAPH, 2001.

[11] P. Hill, N. Canagarajah, and D. Bull. Image fusion using complex wavelets. In *BMVC2002*, pages 487–496, 2002.

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

[13] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.

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

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

[16] R. Szeliski. Image alignment and stitching: a tutorial. *Found. Trends. Comput. Graph. Vis.*, 2(1):1–104, 2006.

[17] 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.

[18] 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.