Pseudo Zernike Moment-based Multi-frame Super Resolution

Sara Salkhordeh*, Hamidreza Rashidy Kanan

Department of Electrical, Biomedical and Mechatronic Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran

Abstract

The goal of multi-frame Super Resolution (SR) is to fuse multiple Low Resolution (LR) images to produce one High Resolution (HR) image. The major challenge of classic SR approaches is accurate motion estimation between the frames. To handle this challenge, fuzzy motion estimation method has been proposed that replaces value of each pixel using the weighted averaging all its neighboring pixels in all LR images which carries the degree of similarity between image blocks centered on two pixels. Since in case of rotation between LR images, comparing the gray level of blocks around the pixels is not a suitable criterion for calculating weight, so, magnitude of Zernike Moments (ZM) has been used as a rotation invariant feature. Due to the lower sensitivity of Pseudo Zernike Moments (PZM) to noise and the higher discrimination capability of it for the same order compared to ZM, in this paper, we propose a new method based on magnitude of PZM of the blocks as a rotation invariant descriptor for representation of pixels in weight calculation. Experimental results on several image sequences show that the performance of the proposed algorithm is better than the existing and new techniques from the aspect of PSNR and visual image quality.

Keywords: Super Resolution, Resolution enhancement, Zernike Moments, Pseudo Zernike Moments, Fuzzy motion estimation, Rotation invariant.

1. Introduction

Spatial resolution enhancement of the images has been an active research field for three decades due to the importance of these images in various applications such as remote sensing, diagnosis and video surveillance. The term Super Resolution (SR) or resolution enhancement refers to the image processing algorithms overcoming the limits and defects of inexpensive image acquisition systems. SR algorithms are divided into single-frame and multiple-frame categories. In the single-frame, the image resolution is enhanced using information of a single image, while in the multiple-frame, information of multiple Low Resolution (LR) images are combined to produce a High Resolution (HR) image. In other words, the main idea of the multi-frame SR is a fusion of several blurred, noisy and LR images to reconstruct an HR one. The first step for analysing the multi-frame SR task is providing a model which relates the main HR image to LR images. Figure (1) shows this model.

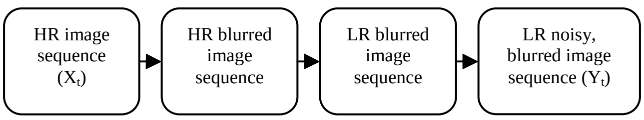


Fig. 1. The model for obtaining LR images from HR images

If we define the t^{th} HR image as X_t and the t^{th} LR image as Y_t, then this model can be defined as follows:

$$Y_t = DBM_t X_t + n \quad (1)$$

In the above equation, considering down sampling factor L_1 in the horizontal direction and L_2 in the vertical direction, the HR image size is equal to $L_1N_1 \times L_2N_2$ and LR image size is equal to $N_1 \times N_2$. T is the number of observed LR images and M_t is wrap matrix that could include local or global translation or rotation and so on. The blur function B represents atmospheric, sensors, or lens' effects during image acquisition process. Matrix D is a down sample matrix which make down samples out of blurred and distorted HR image equal to L_1, L_2 factors in the horizontal and vertical directions. *n* is additive white Gaussian noise. The goal of SR is the restoration of X_t from the input set of images Y_t, reversing the above process.

SR techniques have three steps including registration, interpolation, and restoration. These techniques can be classified to frequency domain-based approaches, interpolation-based approaches, regularization-based approaches, and learning-based approaches.

The problem of multi-frame super resolution reconstruction was first proposed by Tsai and Huang [1] in the frequency-domain. They considered LR images without noise and they transformed LR image data into the Discrete

* Corresponding author. Email: s.salkhordeh@qiau.ac.ir

Fourier Transform (DFT) domain. Kim et al [2] extended their approach, considering the same noise and the same blurring properties for all LR images. Rhee and Kang [3] exploited discrete cosine transform for LR images.

In the spatial field and interpolation-based approaches Ur and Gross [4] performed a non-uniform interpolation of a set of spatially shifted LR images by utilizing the generalized multichannel sampling theorem. IBP [5] and POCS [6] are amongst these algorithms. Both techniques are iteratively approaches.

The main idea of the regularization-based approaches is to use the regularization strategy to incorporate the prior knowledge of the unknown high-resolution image. From the Bayesian point of view, the information that can be extracted from observations (i.e., LR images) about unknown signal (i.e., HR image) is contained in the probability distribution of unknown. Two most popular Bayesian-based SR approaches are ML [7] and MAP [8].

In learning-based approaches, the high frequency information of the given single LR image is enhanced, by retrieving the most likely high-frequency information from the given training image samples based on the local features of the input low-resolution image. Hertzmann [9] has suggested an approach which has two steps: the training step which is applied offline and SR reconstruction step. It should be noted that in the most of these approaches, the defined relocation between LR images is translation or Affine. So these methods don't have a favourable performance when we use actual scenes with local motion (e.g., a person talking) and their restored output image would have distortion. Some approaches were suggested for reducing the registration error, which examine the motion with fuzzy approach. One of these approaches which is a generalization of a denoising method [10], is the super resolution based on Non-local means (NLMSR) [11]. In this approach the amount of each pixel is replaced with a weighted average of its neighbour's pixels, which these weights are criteria for the similarity between the reference pixel and its neighbour's pixels. However, if we define the gray levels of two pixels as the similarity criteria, in case of rotation between LR input images, inappropriate weights will be allocated to the pixels. To solve this problem, Zernike Moments (ZM) of image blocks centered on two pixels has been employed for defining their weight in [12]. The rotation invariant property of the ZM, will lead to more similar blocks adaption and consequently a more appropriate restored image comparing to [11] approach. In order to make ZM invariant to the rotation, Gao et al [12] used absolute value of these moments as a feature vector.

In this paper, we propose a new method based on magnitude of PZM which is more robust to noise and has higher discrimination capability for the same order compared to ZM for describing the pixels in weight calculation. Experimental results on several image sequences indicate that the performance of the proposed algorithm is better than the existing methods from the aspect of PSNR and visual image quality.

The rest of the paper is organized as follows: Section 2 presents the preliminary concept of ZM. Section 3 describes ZM-based SR algorithm. Section 4 presents the proposed approach and its capabilities in details. The experimental results are presented in Section 5. Finally, the paper concludes in Section 6.

2. Zernike Moments

Zernike Moments (ZM) are orthogonal moments that due to its rotation invariant property and near zero value of redundancy, is widely used as a shape descriptor in various applications of image processing and pattern recognition. ZM are defined as a mapping function of $f(x,y)$ on a set of Zernike polynomials inside unit circle. The basic Zernike polynomials with order of n and repetition m can be defined as follows [13]:

$$V_{nm}(r,\theta)=V_{nm}(x,y)=R_{nm}(r)e^{jm\theta} \tag{2}$$

Where, n is a positive integer or zero, m is an integer with the $n-|m|=even$ and $|m|\le n$ conditions, In other words, $n=0,1,2,\ldots$, $m=-n,-n+2,\ldots,n$. Also $r=\sqrt{x^2+y^2}$ is the length of the vector from the origin to the pixel *(x, y)* and $\theta=\tan^{-1}\left(\frac{y}{x}\right)$ is the angle between the vector r and the x-axis. The radial polynomials $R_{nm}(r)$ in equation (2) are defined as:

$$R_{nm}(r)=\sum_{s=0}^{(n-|m|)/2}\frac{(-1)^s(n-s)!}{s!\left(\frac{n+|m|}{2}-s\right)!\left(\frac{n-|m|}{2}-s\right)!}r^{n-2s} \tag{3}$$

Assuming that f is a digital image $M\times M$ that is mapped to the unit circle, its Zernike Moment with order of n and repetition of m is as follows:

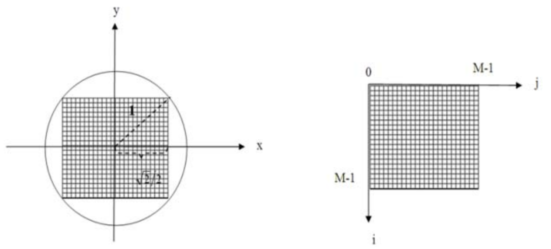
$$ZM_{nm}=\frac{n+1}{\pi}\sum_{x=0}^{M-1}\sum_{y=0}^{M-1}V_{nm}^{*}(x,y)f(x,y) \tag{4}$$

By omitting the condition of $n-|m|=even$ from Zernike Moments, Pseudo Zernike Moments will be obtained. Under this assumption, equation (3) is changed as follows:

$$R_{nm}(r)=\sum_{s=0}^{n-|m|}(-1)^s\frac{(2n+1-s)!}{s!(n+|m|+1-s)!(n-|m|-s)!}r^{n-s} \tag{5}$$

The amount of m and n parameters will be changed to $n=0,1,2,\ldots$; $m=-n,-n+1,\ldots,n$ since these orthogonal moments are not naturally the invariance to scaling changes, before calculating these moments, digital images should be mapped to the unit circle [14]. In the mapping method which we used in this paper the entire $N\times N$ image is bounded inside the unit circle. This method ensures that there is no pixel loss during PZM calculation. This linear mapping could be done using the following equations as shown in Figure (2):

$$x_j = -\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{M-1} j, j = 0,1,\ldots,(M-1)$$
$$y_i = \frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{M-1} i, i = 0,1,\ldots,(M-1) \tag{6}$$



(a) (b)

Fig. 2. pixel coordinates normalization after mapping into the unit circle

Assuming that digital image will be rotated with an angle of θ_0, the values of ZM change as:

$$ZM_{nm}^{R} = ZM_{nm} e^{-jm\theta_0} \tag{7}$$

Where, ZM_{nm} and ZM_{nm}^{R} are original *ZM* and the rotated version respectively, also:

$$\left|ZM_{nm}^{R}\right| = \left|ZM_{nm}^{e^{-jm\theta_0}}\right| = \left|ZM_{nm}\right| \tag{8}$$

Where, $\left|ZM_{nm}\right|$ represents magnitude of *ZM*. According to equation (8), magnitude of *ZM* with rotation will remain constant. This could be generalized to *PZM*.

3. Zernike Moments-based SR Method

As mentioned above, sometimes due to the abrupt changes in position of the camera, it may occur slight global rotation between successive frames, or due to local motion between frames, it may be found blocks in the candidate frame and the reference frame that they are similar to each other but with rotation with respect to each other. In these particular cases, to calculate the weight, if blocks centered on the pixels in the reference frame and the candidate frames were compared based on gray level, the output reconstructed image will be distorted.

Gao et al in [12], by utilizing this feature of ZM that are rotation invariant, has solved this problem partly. They examined magnitude of ZM of two blocks, instead of defining feature vector for two blocks based on gray level. Since moments of higher orders are more computationally expensive, they only extracted ZM till 3rd degree (n=3). In other words, they described each segment with 6 features. In fact for calculating the weight between two pixels *(k, l)*, *(i, j)* they considered 7×7 blocks around these two pixels, and then they extracted magnitude of ZM of these blocks till 3rd degree as equation (9):

$$M(k,l) = (M_{00}, M_{11}, M_{20}, M_{22}, M_{31}, M_{33})$$
$$M'(i,j) = (M'_{00}, M'_{11}, M'_{20}, M'_{22}, M'_{31}, M'_{33}) \tag{9}$$

Then they rewrote weight calculating equation as follows:

$$\omega_{ZER}(k,l,i,j,t) = \exp\left(-\frac{\sum \left\| M(k,l,t_0) - M'(i,j,t) \right\|_2^2}{h^2}\right) \tag{10}$$

Where h^2 is smoothing parameter and controls the effect of ZM differences between two image patches.

4. The Proposed Super Resolution Algorithm

Since PZM extract more features than ZM for certain n_{max}, and PZM perform better than ZM in the presence of noise, so in this paper, we will define feature vector based on magnitude of PZM. Value of n_{max} parameter is adjusted on 3 and then two vectors consisting of ten PZM features are built as relation (11):

$$A(k,l) = (A_{00}, A_{10}, A_{11}, A_{20}, A_{21}, A_{22}, A_{30}, A_{31}, A_{32}, A_{33})$$
$$A'(i,j) = (A'_{00}, A'_{10}, A'_{11}, A'_{20}, A'_{21}, A'_{22}, A'_{30}, A'_{31}, A'_{32}, A'_{33}) \tag{11}$$

So, the equation for calculating weight between two pixels will be changed as follows:

$$\omega_{abs}(k,l,i,j,t) = \exp\left(\frac{-\sum \left\| A(k,l,t_0) - A'(i,j,t) \right\|_2^2}{h^2}\right) \tag{12}$$

Where, $A(k,l,t_0)$ is extracted feature vector of the block around pixel *(k,l)* in reference frame of t_0 and $A'(i,j,t)$ is extracted vector of the block around pixel *(i,j)* candidate frame of t-th. Using calculated weight, new value of pixel *(k,l)* in high resolution image, will be calculated as follows:

$$X(k,l) = \frac{\sum_{t\in[1,\ldots,T]} \sum_{i,j\in N(k,l)} \omega_{abs}(k,l,i,j,t) y_t(i,j)}{\sum_{t\in[1,\ldots,T]} \sum_{i,j\in N(k,l)} \omega_{abs}(k,l,i,j,t)} \tag{13}$$

It should be mentioned that before extracting magnitude of PZM from each block, we need to do a preprocessing on the input block. In fact, we need to define $R_{a,b}$ operator circular. In order to understanding this subject, please see Figure (3).

(a) (b)

Fig. 3. (a) Cameraman image, (B) its rotated version

Figure (3-a) is the cameraman image and (3-b) is its rotated version with 20 degrees. If we extract magnitude of PZM of these two images with order and repetition 3 ($\left|A_{3,3}\right|$) as a sample, we will have the amount of 5.9082 for Figure (3-a) and 3.1024 for Figure (3-b).

Considering Figure (3), this difference could be attributed to the black areas in the rotated image. To solve this problem and omitting the black areas, as it can be seen in Figure (4), a circular mask is applied to the image so that the main image and its rotated version become similar. After this modification, magnitude of Pseudo Zernike Moments in the original image and the rotated image will be equal.

(a) (b) (c)

Fig. 4. (a) Circular mask, (b) the original image after applying the mask, (c) the rotated image after applying the mask

According these explanations, first we apply circular mask to all segments around the pixels ($R_{a,b}$) and then extract feature vector of these circular segments. In the next section, the reconstructed image based on magnitude of PZM will be compared together visually and based on PSNR criteria and also will be compared to NLM-based SR method [11] and ZMNLM-based SR algorithm [12]. It should be mentioned that the peak signal-to-noise ratio (PSNR) is defined as:

$$PSNR = 10 \times \log_{10} \frac{(255)^2}{MSE} \tag{14}$$

Where MSE is the mean-square error obtained per pixel.

5. Experimental Results

In this section two types of image sequence will be used for validation of the proposed algorithm. The first image sequence (synthetic sequence) contains 9 images (frames) with global motion. The cameraman image is used for building this image sequence and this image has been shifted zero, one and two pixels in both horizontal and vertical direction (dx={0,1,2} , dy={0,1,2}). Thus, we will have 9 HR frames with global shifts to each other. Then this sequence is blurred using a 3×3 uniform mask (mean filter), decimated by a factor of 1:3 (in each axis) and then contaminated by an additive noise (white Gaussian noise) with standard deviation 3 and mean zero. After applying these operators, a sequence of 9 frames LR, blurred and noisy, with global shift is obtained.

The second type sequence contains images with local shifts, for example “Miss America” and “mobile”. As mentioned before, the operators down sample, noise and blur is applied to all images of this sequence.

The *h* parameter is set to 30. Also for methods [11], [12] and the proposed method, the search area is considered 5×5 pixels and block size 11×11 pixels. For comparison purposes, the same value for n_{max} of the ZM-based method (i.e. n_{max}=3) was used for each block in the proposed algorithm.

To compare the proposed method with [11], [12], and Lanczos interpolation method [15,16] , three experiments are conducted. In the first experiment, cameraman sequence is used and each time one of sequence frames is rotated 5 degrees and is reconstructed utilizing 8 other frames which are without rotation. The second experiment is performed on sequence of 30 frames of “Miss America” and similar to the first experiment, each time one of the sequence frames is rotated 5 degrees and is reconstructed utilizing 29 other frames which are free of rotation. The third experiment is applied on “mobile” sequence and is done without any rotation.

In the first experiment, the second frame of “cameraman” sequence is rotated 5 degrees and then this frame is reconstructed using 8 other frames with up scaling factor 3. This process will be repeated for frames 5 and 8. The obtained results are indicated in Table (1).

Form Table (1), it can be seen that in case of rotation of even one of sequence images, NLM-based SR method will reach to even a worse result than single-frame Lanczos interpolation method. Also, for all studied frames, the results of the proposed method are better than other methods.

In Figure 5, the first row from left to right respectively shows rotated LR image, rotated HR image and reconstructed image by Lanczos method. The second row from left to right respectively shows reconstructed image by [11] method, output image of method [12] and the output image of the proposed method.

In the second experiment, first, the 8th frame of “Miss America” sequence is rotated 5 degrees and then the rotated frame is reconstructed using other 29 free of rotation images, with up scaling factor 3. This process will be repeated for frames 18th and 28th of the sequence. The PSNR results of this experiment are shown in Table 2.

In this experiment, the results of both ZMNLM-based SR and NLM-based SR methods are worse than Lanczos interpolation method. In other words, due to the rotation between frames, the results of Lanczos interpolation method are better than the results of NLM-based SR method. On the other hand, since in addition to global rotation in these sequences, there are some local shifts, so, the number of Zernike Moment features (n_{max}=3) are not enough to describe a segment in the size of 11×11 pixels.

In Figure 6, from left to right respectively shows rotated HR image, reconstructed image by Lanczos interpolation method, reconstructed image by [11] method, output image of method [12], the output image of the proposed method.

The third experiment is performed on 30 frames sequence of “mobile”. The frames 8, 18, 28 are reconstructed without rotation and with up scaling factor 3 which their results are shown in Table 3.

The goal of designing this experiment and reconstructing frames of a natural sequence without rotation is to prove that absolute value of Pseudo Zernike Moments in reconstruction of frames of this sequence will provide better results than the above mentioned methods, not only in rotation condition, but also in natural condition and it is more accurate descriptor than absolute values of Zernike Moments and gray level of segments.

In Figure (7), the first row from left to right respectively shows rotated HR image, reconstructed image by Lanczos interpolation method and reconstructed image by [11] method. The second row from left to right respectively shows output image of method [12] and the output image of the proposed method.

6. Conclusion

Pseudo Zernike Moments are more robust to noise and have higher discrimination capability for the same order compared to ZM. Hence, in this paper, absolute values of Pseudo Zernike Moments are used as rotation invariant features in super resolution field. The proposed algorithm has been evaluated and compared with some previously popular existing algorithms from the viewpoint of PSNR and the reconstructed image quality. It is a very encouraging finding that the performance of the proposed approach is better than all the compared benchmark approaches.

Fig. 5. Results of reconstructing the 5th frame “Cameraman” sequence

Table 1

Results of cameraman sequence with 5 degrees rotation of frames (PSNR)

Frame number	Lanczos Interpolation Method [15, 16]	The NLM-based SR Method [11]	The ZMNLM-based SR Method [12]	The Proposed PZMNLM-based SR Method
2	24.0916	22.5330	24.3226	26.9165
5	24.1229	22.6830	24.4884	26.8936
8	24.1774	22.6039	24.5820	26.9145

Table 2

Results of Miss America sequence with 5 degrees rotation of frames (PSNR)

Frame number	Lanczos Interpolation Method [15,16]	The NLM-based SR Method [11]	The ZMNLM-based SR Method [12]	The Proposed PZMNLM-based SR Method
8	34.1683	33.1894	33.2828	34.9425
18	34.0795	33.6937	34.0790	35.7732
28	33.8079	33.4371	33.7706	35.4718

Table 3

Results of Mobile sequence without rotation of frames (PSNR)

Frame number	Lanczos Interpolation Method [15,16]	The NLM-based SR Method [11]	The ZMNLM-based SR Method [12]	The Proposed PZMNLM-based SR Method
8	19.5717	20.1925	20.6893	21.5237
18	20.2740	20.8337	21.29	22.3271
28	20.3658	20.7505	21.1737	22.1930

Fig. 6. Results of reconstructing frame 28th of “Miss America” sequence

Fig. 7. Results of 8th frame of “Mobile” sequence

Acknowledgments

The authors would like to thank the Islamic Azad University, Qazvin Branch, Qazvin, Iran for financially supporting this research under contract No. 91-121.

References

[1] R. Y. Tsai, T. S. Huang. "Multiframe image restoration and registration." *Advances in computer vision and Image Processing* 1.2 (1984): 317-339.

[2] S. P. Kim, K. B. Nirmal, H. M. Valenzuela. "Recursive reconstruction of high resolution image from noisy undersampled multiframes." *Acoustics, Speech and Signal Processing, IEEE Transactions on* 38.6 (1990): 1013-1027.

[3] R. Seunghyeon, M. G. Kang. "Discrete cosine transform based regularized high-resolution image reconstruction algorithm." *Optical Engineering* 38.8 (1999): 1348-1356.

[4] U. Hanoch, D. Gross. "Improved resolution from subpixel shifted pictures." *CVGIP: Graphical Models and Image Processing* 54.2 (1992): 181-186.

[5] M. Irani, S. Peleg. "Improving resolution by image registration." *CVGIP: Graphical models and image processing* 53.3 (1991): 231-239.

[6] A. J. Patti, M. Ibrahim Sezan, A. Murat Tekalp. "Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time." *Image Processing, IEEE Transactions on* 6.8 (1997): 1064-1076.

[7] C. Tom Brian, K. Katsaggelos Aggelos. "Reconstruction of a high-resolution image by simultaneous registration, restoration, and interpolation of low-resolution images." *Image Processing, 1995. Proceedings., International Conference on*. Vol. 2. IEEE, 1995.

[8] R. C. Hardie, , J. Barnard Kenneth, E. Armstrong Ernest. "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images." *Image Processing, IEEE Transactions on* 6.12 (1997): 1621-1633.

[9] A. Hertzmann, et al. "Image analogies." *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM, 2001.

[10] A. Buades, C. Bartomeu, J.M Morel, "Denoising image sequences does not require motion estimation." *Advanced Video and Signal Based Surveillance, 2005. AVSS 2005. IEEE Conference on*. IEEE, 2005.

[11] M. Protter, et al. "Generalizing the nonlocal-means to super-resolution reconstruction." *Image Processing, IEEE Transactions on* 18.1 (2009): 36-51.

[12] X. Gao, et al. "Zernike-moment-based image super resolution." *Image Processing, IEEE Transactions on* 20.10 (2011): 2738-2747.

[13] C-H. Teh, T. C. Roland "On image analysis by the methods of moments." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 10.4 (1988): 496-513.

[14] J. Flusser, Z. Barbara, Suk. Tomas *Moments and moment invariants in pattern recognition*. John Wiley & Sons, 2009.

[15] G. Wolberg, Digital Image Warping. Washington, DC: IEEE Computer Soc. Press, 1990.

[16] K. Turkowski, “Filters for common resampling tasks,” in Graphics Gems. New York: Academic, 1990.