DCT based Learning Approach for Image Super-Resolution from Zoomed Observations

Pinkesh G. Kalariya,
EC Department, L.D. College of Engineering,
Ahmedabad, India.

Prakash P. Gajjar, Ph.D.
EC Department, Government Girls Polytechnic,
Surat, India.

Priti J. Muliya
EC Department, L.D. College of Engineering,
Ahmedabad, India.

ABSTRACT
In this paper we propose a zoom based technique to super resolve static scene using observation captured at different camera zoom factor. We capture a static scene at different camera zoom and obtain super resolved image of entire scene at resolution of most zoom observation. Minimum absolute error criteria is used to model image features such as edges, corners etc. We utilize the fact that the local geometry of these features in the low resolution image is similar to their corresponding high resolution version. Missing high frequency details of low resolution observation are learnt in the form of discrete cosine transform coefficient from high resolution training images in the database. The experiments are conducted on real world scene and results are compared with standard interpolation techniques.

Keywords
Learning based method, discrete cosine transforms (DCT), minimum absolute error criteria and super resolution.

1. INTRODUCTION
Most of consumer cameras have single imaging sensor which include photo detecting elements that sense the optical signal. The required size of photo detecting elements reduces with increase in the desired resolution. Although there is significant advance in the fabrication process of the imaging sensors, their performance still remains limited that compromise the achievable spatial resolution. Therefore, signal processing techniques are needed to improve quality of the final image. Image interpolation technique like pixel replication, bilinear interpolation etc.has limited use because of aliasing presented in the low resolution (LR) image. As compare to edges and other sharp details, such techniques are more effective in smooth regions in the image.However more effective approach is algorithmic approach called *Super-resolution* (SR) which produce single high resolution (HR) image from several LR images.The super resolved image can be reconstructed from LR observation using training database comprising of LR imagesand corresponding HR images (LR-HR pairs). It is an important area in the field of image processing having broad range of applications such as medical imaging, computer vision, remote sensing, forensic science etc. Because of application specific limitations like cost, storage capability and transmission bandwidth, it may not feasible to use HR image acquisition system. For example in wild life sensor network, large numbers of cameras are required to install in the forest to observe wild life activities. Due to cost and security issues, LR cameras are chosen. Images captured by these cameras are transmitted to the base stations for detailed analysis. Due to continuous transmission activities, transmission bandwidth is also a major constraint. At base station SR technique can be applied to the received LR images to obtain HR image.

2. RELATED WORK
Many authors proposed various algorithms to solve the super-resolution problem. S. Park et al. [1] classified super resolution techniques as non uniform interpolation approach, frequency domain approach, regularized SR reconstruction approach, iterative back projection approach and learning based approach. Tsai and Huang et al [2] were the first researchers that presented the technique to improve the resolution of an image. They presented the relationship between HR and LR images in the form of Discrete Fourier Transform (DCT) coefficient. R. C. Hardie et al. [3] use maximum a posteriori (MAP) criteria to estimate the registration parameters and the HR image jointly from severely aliased observations. Elad and Feueret al. [4] proposed a technique to super-resolve an image from geometrically warped, blurred, noisy and down sampled observations. They use maximum likelihood (ML) and MAP criteria with projection onto convex sets approach. H. Shen and L. Zhang et al. [5]proposeda joint MAP formulation which combines motion estimation,segmentation, and super-resolution (SR) together. They developed a cyclic coordinate decentprocess that considers the HR image, the motion and the segmentation fieldsas unknowns and estimates them jointlyusing the available data.K. V. Suresh and G. M. Kumar et al. [6] proposed a technique to enhance license plate numbers of moving vehicles in real traffic videos. They obtain a HR image of the license plate by combining the information obtained from multiple, sub-pixel shifted and noisy LR observations. They model theHR image as a Markov random field and estimatedusing a graduated nonconvexity optimization procedure. Recently learning based super resolution algorithms have drawn much attention which arequite different from conventional signal processing approaches. In learning based approaches HR image is reconstructed using image database involving LR images and corresponding HR images (LR-HR pairs). The first learning based approach was introduced by Freeman et al. [7].Authors in [8]-[9] presented learning based methods that use training database of LR-HR pairs all captured using same camera. P. Gajjaret al. [8] developed wavelet based learning approach in which HR estimation is obtained by learning high frequency details of observed LR image from training database in the form of wavelet coefficients. P. Pithadia et al. [9] developed a DCT based learning method for super resolution in which missing high frequency details of LR observation are extracted using Local Binary Pattern (LBP) feature model.H. Yue and X. Sun et al. [10] use scale invariant image features for high frequency approximation. M. Joshi et al. [11] presented a technique for super-resolution from observations at different camera zooms. Markov Random Field (MRF) and Simultaneous Autoregressive (SAR) are used to model super resolved image.

In this paper we proposed a zoom based approach to recover a HR image of a static scene from observations taken at different camera zoom levels. Image of an entire scene is obtained at the resolution of most zoom image by learning high frequency details from training database of LR-HR pairs. Missing high frequency details are learnt as Discrete Cosine Transform (DCT) coefficient.While learning, minimum absolute error criterion is used to model image features. In the DCT transform domain, fine details related to high resolution information reconstructed using high frequency coefficients. We have compared our approach with Bicubic interpolation and bilinear interpolation methods.

3. LOW-RESOLUTION IMAGE FORMATION

Let $\{X_m\}_{m=1}^{p}$ is a set of p images of static scene taken at different camera zoom. Each image in this set is of same size (M×M). Zooming level is assumed in increasing order so that X_1 be the least zoom (LZ) image and X_p be the most zoom (MZ) image. Also known zoom factor was assumed between successive observations. Fig. 1 illustrate the relationship between LR observations of the static scene and corresponding HR image for p=3 with X_1 be LZ image and X_3 be MZ image.

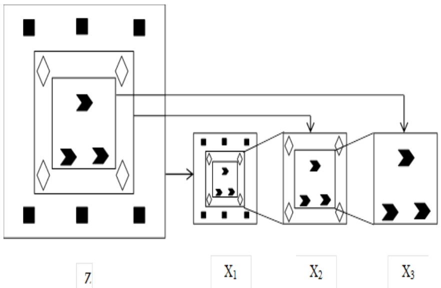


Figure 1 relationship between LR observations of the static scene and corresponding HR image for p=3

4. THE PROPOSED APPROACH

We proposed a novel approach for image super resolution in which entire scene in captured at three different camera zoom. The super resolved image has resolution correspond to the most zoom image. We divide this approach into two parts: zoom based approach and DCT based SR technique for single image enhancement.

4.1. Zoom based Approach

The block diagram of zoom based approach is shown in fig. 2. In this approach the observed scene is captured at three different camera zooms. We got three images denoted as Least Zoom (LZ) image, Zoom (Z) image and Most Zoom (MZ) image. All three images have same size. LZ image cover entire scene but it has least resolution because entire scene is represented by limited number of pixels. As compare to LZ image, Z image has better resolution but it cover less area of the scene. MZ image has best resolution as compare to both but it covers only few area of the scene.

First LZ image is enhanced by a factor of four using proposed SR technique which we discuss next. Resultant enhanced LZ image is denoted as LZ_E image. Similarly Z image is enhanced by a factor of two and resultant image is denoted as Z_E image. Finally, LZ_E, Z_E, MZ images are fused as shown in fig. 2.

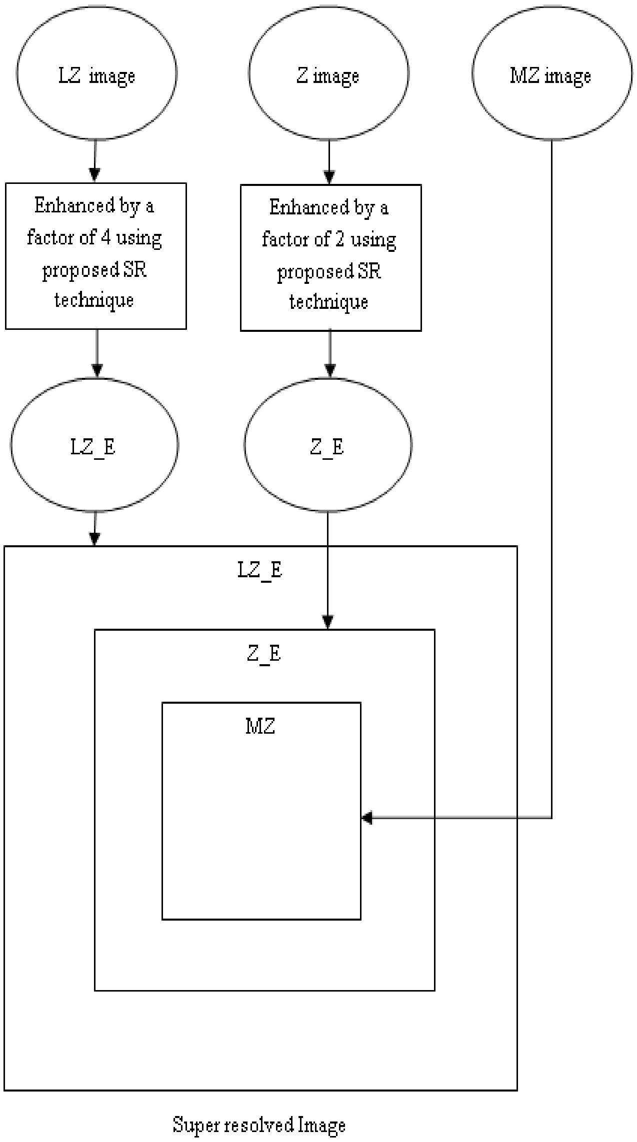


Figure 2 Block diagram of zoom based approach

4.2. SR TECHNIQUE FOR SINGLE IMAGE

In order to enhance images in our approach, we proposed a DCT based SR technique as shown in fig. 3. It can be divided into three steps: image feature modeling, searching matched LR training images and learning HR coefficients. In the first step, we first resize the LR test image and LR training images by a factor of two using pixel replication method then we take a block based DCT of test image and HR training images. In the second step we model image features using minimum absolute error criterion. Next we explore the training database for similar features that match to the features in the test image. After that we learn the high frequency details in the form of DCT coefficients and lastly inverse DCT is applied to obtain the super resolved image. We can also apply this technique to enhance the image by the factor of four.

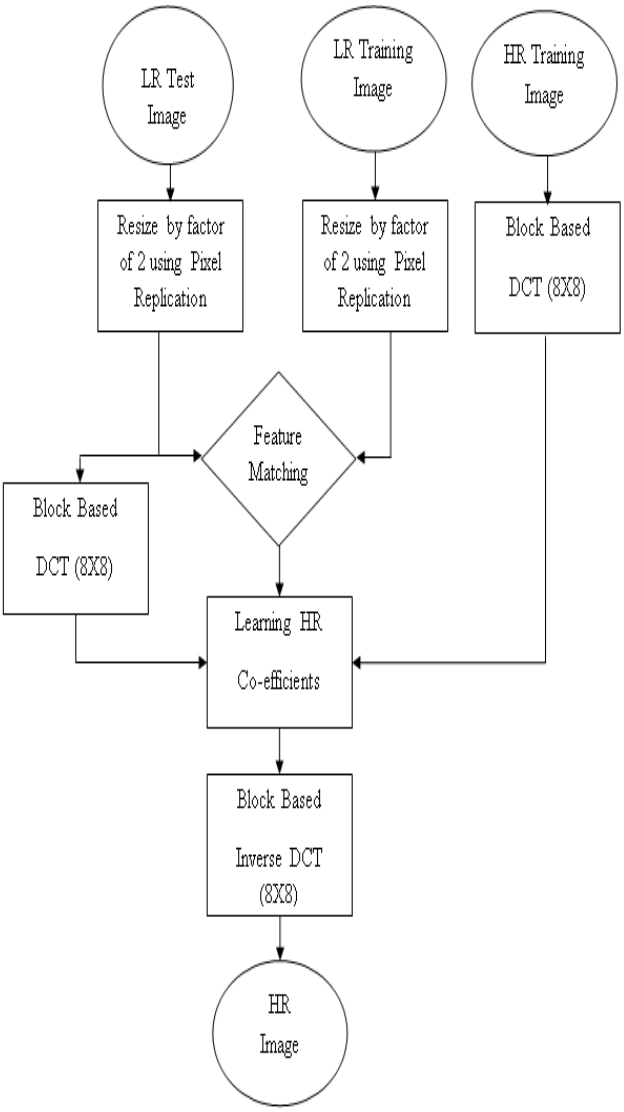


Figure 3 DCT based SR technique

4.2.1. Discrete Cosine Transform

DCT is a well-known signal analysis tool in the field of image processing [13]-[15]. DCT attempts to decorrelate the image data. Then each transform coefficient can be encoded independently. DCT coefficients reproduce the pixel intensities of the image as DC coefficients with low frequency regions and AC coefficients with high frequency regions that can be used to retrieve image details in the region of interest without losing compression efficiency.

Let the size of the LR test image and LR training images in the training database be M×M and that of the HR training images be 2M×2M. First of all, LR test image and LR training images are resized to 2M×2M. Then after this resized LR test image and HR training images of training database are divided into patches of 8×8 pixels. At this stage DCT is applied. So, we got DCT coefficient matrices of size 2M×2M corresponding to all the HR images in the training database and the test image. These matrices characterize the image data in the frequency domain, we consider 8×8 block of these matrices and the high frequency coefficients are used to extract fine details analogous to high resolution information. The coefficient at location (0,0) represents the DC level of the image.

4.2.2. Image Feature Modeling

LR test image and LR training images are resize by a factor of two and image feature model is applied to search the similar features from LR training images corresponding to features of LR test image. Here, we use minimum absolute error criterion model as image feature. Mathematically, it can be expressed for an 8×8 patch as

$$e = \sum_{x=1}^{x=8} \sum_{y=1}^{y=8} |f(x,y) - g(x,y)| \qquad \dots 1$$

In above equation $f(x,y)$ is LR test image and $g(x,y)$ is LR training image. Here, absolute error is computed for 8×8 image patches. HR training image is selected corresponding to the LR training image for which this error will be minimum.

4.2.3. Learning HR Feature Coefficients

Once we find the best matching LR training image corresponding to each patch of the test image having minimum absolute error, we learn the high resolution information in the form of detailed DCT coefficients from the corresponding HR training image in the training database. We extract the high frequency details from HR training image in the database because it is the true high resolution image. Fig. 4 illustrates the learning process. It shows an 8×8 block of DCT coefficients of the test image and its best matching HR training image from the database. Darken boxes represent the high frequency coefficients. Here, high frequency coefficients of test image are replaced with that of the HR training image. As soon as all the DCT coefficients are learned, we apply inverse DCT to the learned image to obtain HR test image.

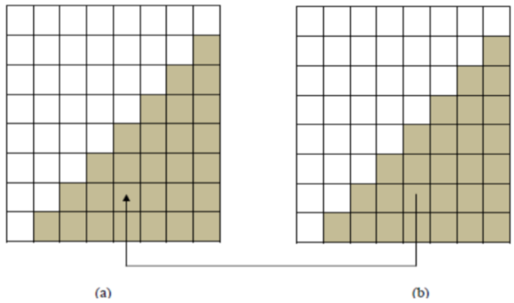


Figure 4. Replacing the high frequency DCT coefficients: (a) An 8×8 block of resized test Image (b) An 8×8 block of best matching HR training image

5. EXPERIMENTAL RESULTS

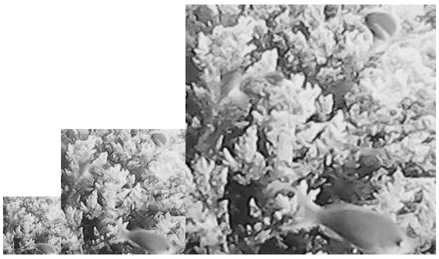
In this section, we evaluate the performance of the proposed approach to super resolve a static scene using zoom based observation. We capture static scene at three different camera zoom. So, we have three images (LZ, Z and MZ) of same size 64×64 as input test image and we obtain a single high resolution image. LZ image is enhanced to 256×256 (LZ_E) and Z image is enhanced to 128×128 (Z_E) using proposed approach. First 64×64 and corresponding 128×128 training images in training database are used as LR-HR pair to enhance LZ image to 128×128. Then 128×128 and corresponding 256×256 training images are used as LR-HR pair to obtain enhanced image of 256×256. So, training database has images at three different resolutions: 64×64, 128×128 and 256×256. However database does not have upsampled or downsmapled images. All the images in the database are captured using computer controlled camera. The camera has facility to capture the scene at different resolution. In image capturing procedure a stable setup was used to avoid motion of the camera. The training database contains images of 400 various scenes. Therefor total images in the database are 400×3=1200. Fig. 5 shows set of training images in the database. Fig. 5(a) shows 64×64 image, fig. 5(b) shows corresponding 128×128 image and fig. 5(c) shows 256×256 image.

(a) (b) (c)

Set 1

(a) (b) (c)

Set 2



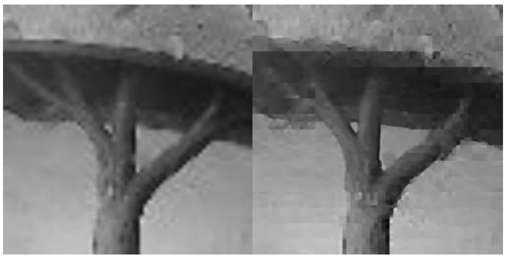
(a) (b) (c)

Set 3

Figure 1 Set of training images: (a) 64×64 image, (b) 128×128 and (c) 256×256

Finally, LZ_E which is enhancement of LZ image, Z_E which is enhancement of Z image and MZ image are fused as shown in figure 2. The resulting super resolved image is of size 256×256. Fig. 6 shows the experimental results conducted on gray scale scenes. Fig. 6(a) shows LZ image of size 64×64, fig. 6(b) shows Z image of size 64×64, fig. 6(c) shows MZ image of size 64×64, fig. 6(d) shows interpolated image of size 256×256 using bicubic interpolation, fig. 6(e) shows interpolated image of size 256×256 and fig. 6(f) shows super resolved image of size 256×256 using proposed approach.

(a) (b) (c) (d)



(e) (f)

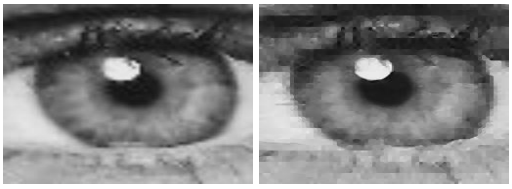
Image 1

(a) (b) (c) (d)

(e) (f)

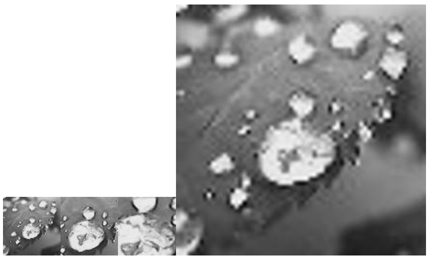
Image 2

(a) (b) (c) (d)

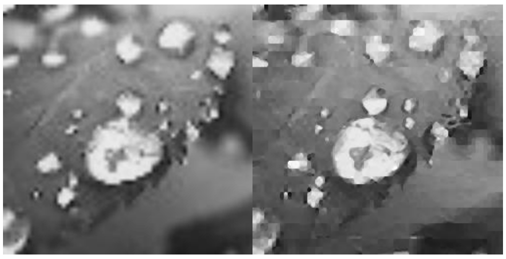


(e) (f)

Image 3



(a) (b) (c) (d)



(e) **(f)**

Image 4

Figure 2 Results on gray scale scene. (a) LZ images, (b) Z images, (c) MZ images, (d) interpolated images using bicubic interpolation, (e) interpolated images using bilinear interpolation and (f) Super resolved image using proposed approach

We compare our results with standard interpolation techniques named bicubic interpolation and bilinear interpolation. We use Mean Squared Error (MSE) as measurement index for quantitative measure. MSE is defined as,

$$MSE = \frac{\sum_{x,y}[f(x,y) - g(x,y)]^2}{\sum_{x,y}[f(x,y)]^2}$$

Where $f(x, y)$ represents original HR image and $g(x, y)$ represents obtained HR image. We use LR camera to capture images at different zoom factor. All the experiments were conducted on a computer with Intel Pentium Dual M, 2.00-GHz processor and 2 GB RAM.

Table-I shows the quantitative performance using MSE as measurement index for gray scale super resolution. The numbers for the proposed approach show significant improvement compared with Bicubic interpolation and bilinear interpolation methods. The super resolved images are less noisy leading to smaller values of MSE.

Table I Performance Comparison for Gray Scale Super-Resolution

Image	Bicubic Interpolation	Bilinear Interpolation	Proposed approach
	MSE		
1	0.0290	0.0257	0.0218
2	0.0404	0.0353	0.0301
3	0.0271	0.0265	0.0202
4	0.0585	0.0508	0.0394

6. CONCLUSION

We have presented a novel approach to super-resolve an image using zoomed observations. The super-resolved image is obtained at the resolution of MZ image which covers only few area of the scene. Images are enhanced using proposed SR technique in which image features are recovered by computing absolute errors between test image and training images. The missing high frequency details are learned from a database in form of discrete cosine transform coefficients. As presented in results section, the proposed approach of super-resolution yield better result as compared to the bilinear and Bicubic interpolation method. The proposed approach outperforms the widely used interpolation techniques.

7. REFERENCES

[1] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang, "Super-Resolution Image Reconstruction: A Technical Overview", IEEE Signal Processing Magazine, pp.21-36, May 2003.

[2] R. Y. Tsai and T. S. Huang, "Multiframe image restoration and registration," Advances in Computer Vision and Image Processing, pp. 317-339, 1984.

[3] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint map registration and high-resolution image estimation using a sequence of undersampled images," IEEE Trans. Image Process., vol. 6, pp. 1621–1633, 1997.

[4] M. Elad and A. Feuer, "Restoration of a single super-resolution image from several blurred, noisy and undersampled measured images," IEEE Trans. Image Process, vol. 6, pp. 1646–1658, 1997.

[5] H. Shen, L. Zhang, B. Huang, and P. Li, "A MAP approach for joint motion estimation, segmentation, and super resolution," IEEE Trans. Image Process., vol. 16, pp. 479–490, 2007.

[6] K. V. Suresh, G. M. Kumar, and A. N. Rajagopalan, "Superresolution of license plates in real traffic videos," IEEE Trans. Intell. Transport. Syst., vol. 8, no. 2, pp. 321–331, Feb. 2007.

[7] W. T. Freeman, T. R. Jones and E. C. Pasztor, 'Example-based super-resolution", IEEE Computer Graphics and Applications, pp. 56-65, 2002.

[8] P.P. Gajjar and M.V. Joshi, "New Learning Based Super-Resolution: Use of DWT and IGMRF Prior" in IEEE transaction on image processing, vol.19, no 5, May 2010, pp. 1201-1213.

[9] P.V. Pithadia, P.P. Gajjar, J.G. Dave, "Super Resolution using DCT based Learning with LBP as Feature Model", in IEEE International conference on Computing, Communication & Networking Technologies (ICCC NT 2012).

[10] H. Yue, J. Yang, X. Sun, F. Wu, "SIFT-Based Image Super-Resolution" in IEEE 2013, pp. 2896-2899.

[11] M.V. Joshi, S. Chaudhuri, and R. Panuganti, "A Learning-Based Method for Image Super-Resolution From Zoomed Observations" in IEEE transactions on system, man, and cybernetics-part B: cybernetics, Vol. 35, NO. 3, JUNE 2005, pp. 527-537.

[12] S. Dabbaghchian, M. P. Ghaemmaghami and Ali Aghagolzadeh, "Feature extraction using discrete cosine transform and discrimination power analysis with a face recognition technology",Journal of Pattern Recognition, 2010, pp. 1431-1440.

[13] Nuno Roma and Leonel Sousa, "A tutorial overview on the properties of the discrete cosine transform for encoded image and video processing", Journal of Signal Processing, 2011, pp. 2443-2464.

[14] Wei Zhang and Wai-Kuen Cham, "Hallucinating Face in the DCT Domain", IEEE Transactions on Image Processing, 2011, pp. 2769-2779.