

Decimation Estimation and Super-Resolution Using Zoomed Observations

Prakash P. Gajjar, Manjunath V. Joshi, Asim Banerjee, and Suman Mitra

Dhirubhai Ambani Institute of Information and Communication Technology,
Gandhinagar, India

Abstract. We propose a technique for super-resolving an image from several observations taken at different camera zooms. From the set of these images, a super-resolved image of the entire scene (least zoomed) is obtained at the resolution of the most zoomed one. We model the super-resolution image as a Markov Random Field (MRF). The cost function is derived using a Maximum a posteriori (MAP) estimation method and is optimized by using gradient descent technique. The novelty of our approach is that the decimation (aliasing) matrix is obtained from the given observations themselves. Results are illustrated with real data captured using a zoom camera. Application of our technique to multiresolution fusion in remotely sensed images is shown.

1 Introduction

In many of the imaging applications, images with high spatial resolution are desired and often required. The spatial resolution can be increased by using high density sensor for capturing the image. However, this is not possible as there exist a limit on pixel size. The resolution enhancement from a single observation using image interpolation is of limited application because of the aliasing present in the low resolution image. Super-resolution refers to the process of producing a high spatial resolution image from several low-resolution observations. When one captures the images with different zoom settings, the amount of aliasing is different in differently zoomed observations. This is because the least zoomed entire area of the scene is represented by a very limited number of pixels, i.e., it is sampled with a very low sampling rate and the most zoomed image with a higher sampling frequency. Therefore, larger scene coverage will have lower resolution with more aliasing effect. By varying the zoom level, one observes the scene at different levels of aliasing and blurring. Thus, one can use zoom as a cue for generating high-resolution images at the lesser zoomed area of a scene.

The super-resolution idea was first proposed by Tsai and Huang [1] using frequency domain approach and employing motion as a cue. In [2], the authors use a Maximum a posteriori framework for jointly estimating the registration parameters and the high-resolution image for severely aliased observations. A MAP estimator with Huber-MRF prior is described by Schultz and Stevenson in [3]. Lin and Shum determine the fundamental limits of reconstruction-based super-resolution algorithms using the motion cue and obtain the magnification limits

from the conditioning analysis of the coefficient matrix [4]. Capel and Zisserman [5] have proposed a technique for automated mosaicing with super-resolution zoom by fusing information from several views of a planar surface in order to estimate its texture. The authors in [6] integrate the tasks of super-resolution and recognition by directly computing a maximum likelihood parameter vector in high-resolution tensor space for face recognition.

Most of the methods of super-resolution proposed in literature use motion cue for estimating high resolution image. This requires registration of images with sub-pixel accuracy. The non-redundant information can also be obtained by using different camera parameters or different lighting conditions while capturing the scene. The authors in [7] describe an MAP-MRF based super-resolution technique using blur cue. They recover both the high-resolution scene intensity and the depth fields simultaneously using the defocus cue. The authors in [8] recover the super-resolution intensity field from a sequence of zoomed observations. The resolution of entire scene is obtained at the resolution of the most zoomed observed image which consists of only a portion of the actual scene. For more details refer to [9].

In this paper, we obtain super-resolution by using zoom as a cue. We model the super-resolution image as an MRF and assume that the high resolution image at the most zoom setting is super-resolved. In our image formation model, we learn the decimation (aliasing) matrix from the most zoomed observation and use MAP-MRF formulation to obtain super-resolved image for the entire scene. We are also assuming that the images are registered while zooming. However after registering the images we need to estimate the aliasing accurately so that the model fits well. It may be interesting to see that our approach generates a super-resolved image of the entire scene, although only a part of the observed scene has multiple observations.

2 Image Formation Model

The zoom based super-resolution problem can be cast in a restoration framework. There are p observed images Y_i , $i = 1$ to p , each captured with different zoom settings and are of size $M_1 \times M_2$ pixels each. Fig. 1 illustrates the block schematic of how the low-resolution observations of a scene at different zoom settings are related to the high-resolution image. Here we consider that the most zoomed observed image of the scene Y_p ($p = 3$) has the highest spatial resolution. We are assuming that there is no rotation about the optical axis between the observed images taken at different zooms. Since different zoom settings give rise to different resolutions, the least zoomed scene corresponding to entire scene needs to be upsampled to the size of $(q_1 q_2 \dots q_{p-1}) \times (M_1 \times M_2)$ pixels ($= N_1 \times N_2$ pixels), where q_1, q_2, \dots, q_{p-1} are the corresponding zoom factors between two successively observed images of the scene $Y_1 Y_2, Y_2 Y_3, \dots, Y_{p-1} Y_p$ respectively. Given Y_p , the remaining $(p - 1)$ observed images are then modeled as decimated and noisy versions of this single high-resolution image of the appropriate region in the scene. The most zoomed observed image will have no decimation. The low resolution image observation model is shown in Fig. 2.

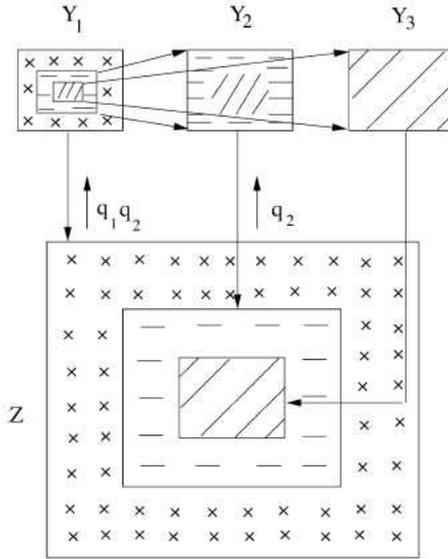


Fig. 1. Illustration of observations at different zoom levels, Y_1 corresponds to the least zoomed and Y_3 to the most zoomed images. Here z is the high resolution image of the scene.

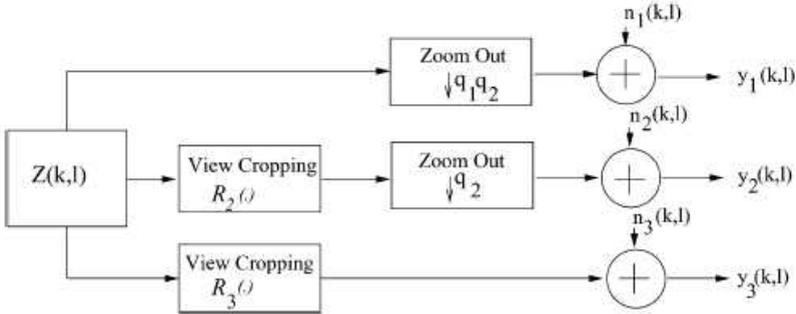


Fig. 2. Low-resolution image formation model for three different zoom levels. View cropping block just crops the relevant part of the high resolution image Z as the field of view shrinks with zooming.

Let \mathbf{y}_m represent the lexicographically ordered vector of size $M_1 M_2 \times 1$, which contains the pixels from differently zoomed images Y_m and \mathbf{z} be the super-resolved image. The observed images can be modeled as

$$\mathbf{y}_m = D_m \mathbf{C}_m (\mathbf{z} - z_{\alpha_m}) + \mathbf{n}_m, \quad m = 1, \dots, p, \quad (1)$$

where D is the decimation matrix which takes care of aliasing present while zooming. The subscript m in D denotes that the amount of decimation depends on the amount of zoom for m^{th} observation, size of which depends on the zoom factor. For an integer zoom factor of q , the decimation matrix D consists of

The aliased pixel intensity at a location (i, j) for a zoom factor of $q = 2$ is now given by

$$y(i, j) = a_1 z(2i, 2j) + a_2 z(2i, 2j + 1) + a_3 z(2i + 1, 2j) + a_4 z(2i + 1, 2j + 1) + n(i, j). \quad (9)$$

In [11], authors discuss the spatial interaction model and choice of neighbors and use the same for texture synthesis. They model every pixel in an image as a linear combination of neighboring pixels considering neighborhood system. They estimate the model parameters using the Least Squares (LS) estimation approach as the initial estimates. In this paper, we use their approach for estimating the weights in decimation matrices for different zoom factors.

4 Super-Resolving a Scene

4.1 MRF Prior Model for the Super-Resolved Image

In order to obtain a regularized estimate of the high-resolution image, we define an appropriate prior term using an MRF modeling of the field. The MRF provides a convenient and consistent way of modeling context dependent entities. This is achieved through characterizing mutual influence among such entities using conditional probabilities for a given neighborhood. The practical use of MRF models is largely ascribed to the equivalence between the MRF and the Gibbs Random Fields (GRF). We assume that the high-resolution image can be represented by an MRF. This is justified because the changes in intensities in a scene is gradual and hence there is a local dependency. Let Z be a random field over an regular $N \times N$ lattice of sites $L = \{(i, j) | 1 < i, j < N\}$. From the Hammersley-Clifford theorem for MRF-GRF equivalence, we have,

$$P(Z = z) = \frac{1}{Z_p} e^{-U(z)}, \quad (10)$$

where \mathbf{z} is a realization of Z , Z_p is a partition function given by $Z_p = \sum_{\mathbf{z}} e^{-U(\mathbf{z})}$ and $U(\mathbf{z})$ is energy function given by $U(\mathbf{z}) = \sum_{c \in \mathbf{C}} V_c(\mathbf{z})$. $V_c(\mathbf{z})$ denotes the potential function of clique c and \mathbf{C} is the set of all cliques. The lexicographically ordered high resolution image \mathbf{z} satisfying Gibbs density function is now written as

$$P(\mathbf{z}) = \frac{1}{Z_p} e^{-\sum_{c \in \mathbf{C}} V_c(\mathbf{z})}. \quad (11)$$

We consider pair wise cliques on a first-order neighborhoods consisting of the four nearest neighbors for each pixel and impose a quadratic cost which is a function of finite difference approximations of the first order derivative at each pixel location. i.e.,

$$\sum_{c \in \mathbf{C}} V_c(\mathbf{z}) = \lambda \sum_{k=1}^{N_1} \sum_{l=1}^{N_2} [(z_{k,l} - z_{k,l-1})^2 + (z_{k,l} - z_{k-1,l})^2], \quad (12)$$

where λ represents the penalty for departure from the smoothness in \mathbf{z} .

4.2 Maximum a Posteriori (MAP) Estimation

Having defined the MRF prior, we use the MAP estimator to restore the high-resolution field \mathbf{z} . Given the ensemble of images \mathbf{y}_i , $i = 1$ to p , at different resolutions, the MAP estimate $\hat{\mathbf{z}}$, using Bayesian rule, is given by

$$\hat{\mathbf{z}} = \underset{\mathbf{z}}{\operatorname{argmax}} P(\mathbf{z}|\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_p) = \underset{\mathbf{z}}{\operatorname{argmax}} P(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_p|\mathbf{z})P(\mathbf{z}). \quad (13)$$

Taking the log of the posterior probability we can write,

$$\hat{\mathbf{z}} = \underset{\mathbf{z}}{\operatorname{argmax}} \left[\sum_{m=1}^p \log P(\mathbf{y}_m|\mathbf{z}) + \log P(\mathbf{z}) \right], \quad (14)$$

since n_m are independent. Now using Eqs. (1) and (2), we get

$$P(\mathbf{y}_m|\mathbf{z}) = \frac{1}{(2\pi\sigma_n^2)^{\frac{M_1 M_2}{2}}} e^{-\frac{\|\mathbf{Y}_m - D_m C_m(\mathbf{z} - \mathbf{z}_{\alpha_m})\|^2}{2\sigma_n^2}}. \quad (15)$$

The final cost function is obtained as

$$\hat{\mathbf{z}} = \underset{\mathbf{z}}{\operatorname{argmin}} \left[\sum_{m=1}^p \frac{\|\mathbf{y}_m - D_m C_m(\mathbf{z} - \mathbf{z}_{\alpha_m})\|^2}{2\sigma_n^2} + \sum_{c \in \mathcal{C}} V_c(\mathbf{z}) \right]. \quad (16)$$

The above cost function is convex and is minimized using the gradient descent technique. The initial estimate $\mathbf{z}^{(0)}$ is obtained as follows. Pixels in the zero order hold of the least zoomed observation corresponding to the entire scene is replaced successively at appropriate places with zero order hold of the other observed images with increasing zoom factors. Finally, the most zoomed observed image with the highest resolution is copied at the appropriate location (see Fig. 1.) with no interpolation.

5 Experimental Results

In this section, we present the results of the proposed method of obtaining super-resolution by estimating the decimation. All the experiments were conducted on real images taken by a zoom camera and known integer zoom factors. It assumed that the lateral shift during zooming is known. In each experiment, we consider three low resolution observations Y_1, Y_2, Y_3 of an image. Each observed image is of size 72×96 . Zoom factor q between Y_1 and Y_2 is 2 and that between Y_1 and Y_3 is 4. The super-resolved images for the entire scene are of size 288×384 . We obtain super-resolution and compare the results obtained using decimation matrix of the form in Eq. (4) consisting of equal weights. We used the quantitative measures Mean Square Error (MSE) and Mean Absolute Error (MAE) for comparison of the results. The MSE used here is

$$MSE = \frac{\sum_{i,j} [f(i,j) - \hat{f}(i,j)]^2}{\sum_{i,j} [f(i,j)]^2} \quad (17)$$

and MAE is

$$MAE = \frac{\sum_{i,j} |f(i,j) - \hat{f}(i,j)|}{\sum_{i,j} |f(i,j)|}, \quad (18)$$

where $f(i,j)$ is the original high resolution image and $\hat{f}(i,j)$ is estimated super-resolution image. In order to use high resolution image for the entire scene the most zoomed image was captured with entire scene content. However, while experimenting only a portion of it was used. The estimated D matrices are used in the cost function given by Eq. (16).



Fig. 3. Observed images of 'Nidhi' captured with three different integer zoom settings. The zoom factor between (a) and (b) is 2 and between (b) and (c) is also 2.

In the first experiment, we considered three low resolution observations of a girl image 'Nidhi' shown in Fig. 3, where the observed images have less intensity variations. Fig. 4(a) and (b) shows zoomed 'Nidhi' image obtained by successive pixel replication and successive bicubic interpolation respectively. In both the images the seam is clearly visible. Fig. 5(a) shows super-resolved 'Nidhi' image obtained by using the decimation matrix of the form in Eq. (4) and Fig. 5(b) shows super-resolved 'Nidhi' image obtained by proposed method by using the estimated decimation matrix of the form in Eq. (7). The comparison of the



Fig. 4. Zoomed 'Nidhi' image (a) using successive pixel replication and (b) using successive bicubic interpolation



Fig. 5. Super-resolved 'Nidhi' image (a) using equal weights decimation matrix and (b) using estimated weights for decimation matrix.

images show more clear details in the regions like cheeks and forehead in the image obtained by the proposed method.

In the second experiment, we considered low resolution observations of a house shown in Fig. 6. Zoomed house images obtained by successive pixel replication and successive bicubic interpolation are shown in Fig. 7 (a) and (b) respectively. Fig. 8 shows super-resolved house images obtained using the two different methods. The comparison of the figures show that there is less blockiness in the super-resolved image obtained by the proposed method. Branches of trees opposite to windows are more clearly visible.

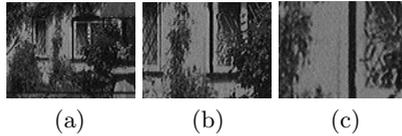


Fig. 6. Observed images of a house captured with three different integer zoom settings



Fig. 7. Zoomed house image (a) using successive pixel replication and (b) using successive bicubic interpolation



Fig. 8. Super-resolved house image. (a) using equal weights decimation matrix and (b) using estimated weights for decimation matrix.

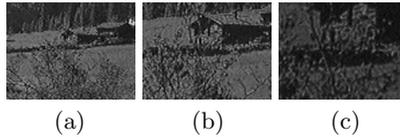


Fig. 9. Observed images of a scene captured with three different integer zoom settings



Fig. 10. Super-resolved scene image. (a) using equal weights decimation matrix and (b) using estimated weights for decimation matrix.

In order to consider images with significant texture, we experimented by capturing zoomed images of a natural scene. The observed images are displayed in Fig. 9. Fig. 10 shows super-resolved scene images. The small house near the center of image appear sharper in the image super-resolved using the proposed approach.

Table 1. shows the quantitative comparison of the our results with the one obtained using equal weights for decimation matrix. It can be seen that for all the three experiments, MSE and MAE of the super-resolved images obtained by using estimated decimation matrices is lower than those obtained by fixed decimation matrix entries showing improvement in the quantitative measures.

Table 1. Comparison of performance of the two methods of super-resolution

Image	MSE		MAE	
	Estimated decimation	Fixed decimation (Equal weights)	Estimated decimation	Fixed decimation (Equal weights)
Nidhi	0.0484	0.0514	0.0489	0.0525
House	0.6671	0.6733	0.6678	0.6751
Scene	0.2732	0.3056	0.2741	0.3082

6 Application of Zoom Based Super-Resolution to Multiresolution Fusion in Remotely Sensed Images

In this section we show the application of the proposed zoom based super-resolution to multiresolution fusion in remotely sensed images. The process of combining panchromatic (Pan) and multispectral (MS) data to produce images characterized by both high spatial and spectral resolutions is known as multiresolution fusion. Because of the technological limitations, MS images are generally acquired with a lower spatial resolution. With a fusion of different images, we can overcome the limitations of information obtained from individual sources and obtain a better understanding of the observed scene. Since the Pan image has high spatial resolution and MS images have lower spatial resolution, we estimate the aliasing on MS images by using the Pan image. The same Pan image is used to estimate the aliasing on each of MS images. Available Pan image can be used for estimating aliasing matrices for all the MS images as the aliasing depends on difference in spatial resolution between high resolution and low resolution images. For the experiment, we consider LANDSAT-7 Enhanced Thematic

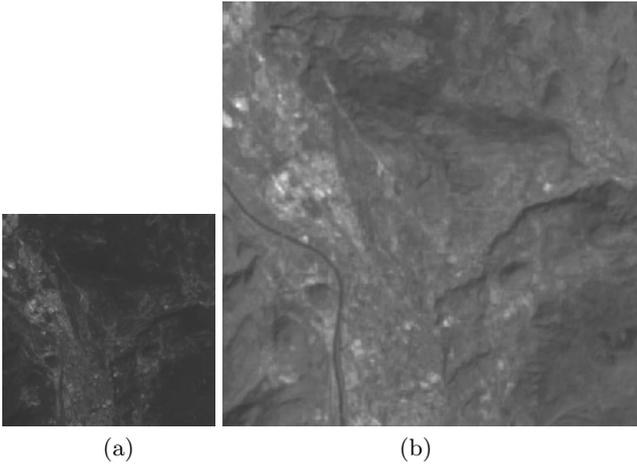


Fig. 11. (a) MS image (Band 1) and (b) fused image using estimated weights for decimation matrix

Mapper Plus (ETM+) images acquired over a city. We use available Pan image and six MS images having decimation factor of $q = 2$ between the Pan image and MS images. We learn the decimation matrix from the Pan image and minimize the cost function given by Eq. (16) for each of MS images separately. It may be noted that the \mathbf{z} in the equations has to be replaced by \mathbf{z}_m , where $m = 1, 2, \dots, 5$ (One of the MS images, Band 6, is not used.). Due to space limitation, we show results for only one MS image. Fig. 11(a) shows observed MS image (Band 1). The fused image obtained by using estimated weights is shown in in Fig. 11(b). From the figure it is clear that the fused image has high spatial resolution with negligible spectral distortion. It may be mentioned that we have not compared the performance of this method with other methods of fusion available in the literature due to space limitation.

7 Conclusion

We have presented a technique to recover the super-resolution intensity field from a sequence of zoomed observations by using decimation matrices derived from the observations. The resolution of the entire scene is obtained at the resolution of the most zoomed observed image that consists of only a small portion of the actual scene. The high-resolution image is modeled as an MRF and the decimation matrix entries are estimated using appropriate regions in the lesser zoomed image and the most zoomed image. Our future work involves incorporating the line fields in MRF model so that the discontinuities can be better preserved in the super-resolved image and also to extend the proposed approach for fractional zoom settings.

References

1. Tsai, R.Y., Huang, T.S.: Multiframe image restoration and registration. *Advances in Computer Vision and Image Processing*. (1984) 317 – 339
2. Hardie, R.C., Barnard, K.J., Armstrong, E.E.: Joint map registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Trans. Image Process.* **6** (1997) 1621 – 1633
3. Schultz, R.R., Stevenson, R.L.: A Bayesian approach to image expansion for improved definition. *IEEE Trans. Image Process.* **3** (1994) 233 – 242
4. Lin, Z., Shum, H.Y.: Fundamental limits of reconstruction-based super-resolution algorithms under local translation. *IEEE Trans. Pattern Anal. Machine Intell.* **26** (2004) 83 – 97
5. Capel, D., Zisserman, A.: Automated mosaicing with super-resolution zoom. *Proc. IEEE Int. Conf. Comput. Vision Pattern Recogn.* (1998) 885 – 891
6. Jia, K., Gong, S.: Multi-modal tensor face for simultaneous super-resolution and recognition. *Proc. of IEEE Int. Conf. on Computer Vision* **2** (2005) 1683 – 1690
7. Rajan, D., Chaudhuri, S.: Generation of super-resolution images from blurred observations using an MRF model. *J. Math. Imag. Vision* **16** (2002) 5 – 15
8. Joshi, M.V., Chaudhuri, S., Rajkiran, P.: Super-resolution imaging: Use of zoom as a cue. *Image and Vision Computing* **22** (2004) 1185 – 1196

9. Chaudhuri, S., Joshi, M.V.: Motion-free Super-resolution. Springer (2005)
10. Park, S.C., Park, M.K., Kang, M.G.: Super-resolution image reconstruction: A technical overview. *IEEE Signal Processing Magazine* **20** (2003) 21 – 36
11. Kashyap, R., Chellappa, R.: Estimation and choice of neighbors in spatial-interaction models of images. *IEEE Trans. Inf. Theory* **IT-29** (1983) 60 – 72