

# IN-FOCUS IMAGING BY MOSAICKING AND SUPER-RESOLUTION\*

Jieyong Lv<sup>1</sup> and Pengwei Hao<sup>1,2</sup>

<sup>1</sup>National Laboratory on Machine Perception, Peking University, Beijing, 100871, China

<sup>2</sup>Department of Computer Science, Queen Mary, University of London, E1 4NS, UK

E-mail: lvjieyong@gmail.com, phao@cis.pku.edu.cn

## ABSTRACT

*Although an all-in-focus image can be generated by mosaicking all the best focused regions of a sequence of acquired images taken under different focal settings, an even better-resolved result can be obtained with those discarded less-focused regions which actually contain more useful information. This paper presents a method to estimate an all-in-focus and super-resolved image of a scene from a sequence of such images. The proposed method applies mosaicking and super-resolution techniques simultaneously to iteratively reconstruct a better image by blind trellis searching. Partitioning images into small blocks makes the new method parallel and fast over blocks. With our regularized approach, a high resolution image can be efficiently estimated both in the underdetermined case and in the case with registration errors.*

**Index Terms**— Super-resolution, omnifocus, mosaicking

## 1. INTRODUCTION

Because of the limitation of imaging device, all imaging cameras have finite depth of field and a digital view of a scene is generally deteriorated and the imaged objects are not all within the depth ranges, thus imaging is not always all-in-focus. Although an all-in-focus image can be generated by mosaicking all focused regions in the acquired images together, an even better-resolved result can be obtained with those discarded defocused regions which actually contain a lot of useful information.

Because of the limited depth of field of CCD cameras, a sequence of images taken with different focal settings must have different regions that differently focused. The infinite depth of field problem has been investigated by many researchers [1, 2, 12]. The basic idea is to extract those best focused regions of all the images first and then mosaic them into a single complete image. Such techniques are impressive because they succeed in recording more high frequency information in the resultant digital image. However, even such an all-in-focus image may not yet meet the demand. For example, people tend to look for more details of a scene in a picture. The

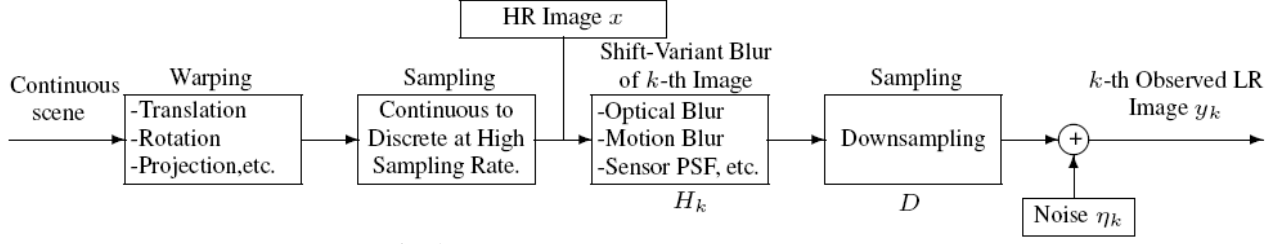
mosaicking algorithms only keep the focused regions of the images and discard all others, and this must be a waste of information. In fact, those discarded regions of the scene are very helpful for recovering the even higher frequency components. In this paper, we try to re-use those discarded regions and eventually generate an all-in-focus image of higher resolution by mosaicking and super-resolution techniques.

A super-resolution (SR) technique has been developed known as restoring the high frequency or details of the image from a sequence of low resolution images(LRI). In fact, SR can only be performed on different views of the same scene. This idea was first investigated by several researchers in the frequency domain [3, 4] and then by using DCT representation [5]. Correspondingly, some spatial domain methods [6] were proposed with much more flexibility, where a regularization method is applied to reconstruct a reasonable result image. All these methods were categorized into reconstruction based SR algorithms, which were most focused in enhancing the resolution of differently translated images. The concrete theoretic limitation and some other mathematic properties were extensively studied in [7].

Generally, the SR problem can be viewed as an inverse-problem, and the forward model can be recognized as warping first and then blurring and downsampling [5]. So the blurring factor, which is determined by the point spread function (PSF), plays an essential role in the model. For a sequence of differently focused images, all the PSFs are supposed to be Gaussian [8], or more accurately, Bezier, which is determined by the depth of the scene. Fortunately, another technique named Depth from Defocus (DFD) can serve this information and some researchers combine this technique and Markov Random Field(MRF) model to derive a super-resolved image [9]. However, these algorithms are not perfect in both theory and practice. First, the model supposed the blurring factor and the downsampling factor to be inter-exchanged with little errors, which is not guaranteed in practice. Second, the Gaussian blur model is noise-sensitive and nonlinear, and the simulated annealing algorithm should be applied to solve the problem, which is generally slow and not stable.

What if we give up the Gaussian restriction and get down to seeking another representation of the model? Because defocus is typical Finite Impulse Filter (FIR) [8], we can incor-

\*This work was supported by FANEDD of China (200038).



**Fig. 1:** Imaging model: from scene to digital images

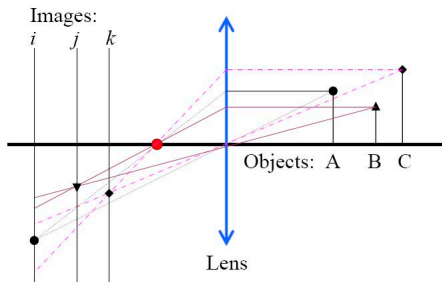
porate the blind identification technique [10, 11] to derive our model. Although the blind identification problem has been extensively investigated and various methods have been proposed, it still encounters much more difficulty in SR problem because of the relatively limited data source. Anyway, some prior knowledge on image acquisition could be taken to regularize the problem in order to achieve a meaningful result.

The outline of this paper is as follows. *Sec.2* is for the imaging model. In *Sec.3*, the new approach for in-focus imaging by mosaicking and super-resolution is proposed. In *Sec.4*, the results with simulated images and naturally acquired images are presented. In *Sec.5*, we come to a conclusion of our paper.

## 2. IMAGING MODEL

For simplicity, we suppose that the imaging camera and the scene are fixed and all the acquired images are only of different focal settings. Thus, our in-focus imaging is taken as an inverse problem based on a forward imaging model as shown in *Fig.1*. In a CCD camera system, the observed low resolution images (LRI) are recognized as resulting from warping, blurring and downsampling operators performed on a high resolution image (HRI) [5]. Due to the limitation of cameras, when we take a picture for a scene with some objects at various distances, only some regions in the picture can be in-focus and other regions out of focus. Both the depth of field and the in-focus regions of the scene in the images change with the distance of the image planes, as shown in *Fig.2*.

In this paper, we assume that the LRIs are acquired with different blurs and at the same downsampling rate. Let  $x$  denote the vectorized original image to be recovered and  $h_i$  denote the vectorized FIR coefficients of the blurs, we can for-



**Fig. 2:** Imaging model: Objects at different distances are focused in different images.

mulate the problem as:

$$\begin{cases} y_1 = DH_1x + \eta_1 \\ y_2 = DH_2x + \eta_2 \\ \vdots \\ y_n = DH_nx + \eta_n \end{cases} \quad (1)$$

where  $y_i$  denote vectorized LRIs and  $\eta_i$  the i.i.d Gaussian white noise,  $H_i = H_i(h_i)$  are the matrix form of the blurs and  $D$  is the downsampling matrix.

In the above formula, both  $x$  and  $H_i$  are unknowns and  $H_i$  is determined by the blurs (including the sensor PSF) when taking  $i$ -th picture. The blurs are shift variant filters (SV), and the PSF at a point is a Gaussian function determined by the depth of that point, if the inner parameters of the camera do not change. To simplify the model, we take the blurs as locally linear shift invariant (LSI) filters in small regions, as it is reasonable to assume that the depth of a small block does not change owing to the area coherence or the smoothness of the objects in the scene. Thus, the relations of (1) still hold, with the difference that  $x, y_i, D, H_i$  denote the corresponding HRI, LRI, downsampling, blurring for the  $i$ -th small LRI blocks. Then, the formula can be rewritten as

$$y = Px + \eta \quad (2)$$

where  $P = ((DH_1)^T, (DH_2)^T, \dots, (DH_n)^T)^T$ ,  $y = (y_1^T, y_2^T, \dots, y_n^T)^T$ ,  $\eta = (\eta_1^T, \eta_2^T, \dots, \eta_n^T)^T$ .

Because the coefficients of matrix  $DH_i$  are a linear combination of  $h_i$ , we can construct matrix  $T_i$  from  $x$  so that  $T_i h_i = DH_i x$ , thus the formula can be re-formulated as:

$$y = Th + \eta \quad (3)$$

where  $T = (T_1^T, T_2^T, \dots, T_n^T)^T$  and  $h = (h_1^T, h_2^T, \dots, h_n^T)^T$ .

## 3. IN-FOCUS MOSAICKING AND SUPER-RESOLUTION

### 3.1. In-Focus Mosaicking

As mentioned above, a set of pictures with different focal settings may be in-focus at different regions. Intuitively, we can extract all the best in-focus pixels in all images and mosaic them together into a single all-in-focus image. The essential of this method is the criterion to select the best focused pixels.

A few measures have been proposed from different points of view, as summarized in [2].

The basic idea of the focus measures is to evaluate the smoothness in a neighborhood or a patch around a pixel by computing the strength of edges, variances, etc. of the patch. The best focused patch at the same location of all the images tends to have sharpest edges, biggest variance, etc. The pixel whose patch is the least smooth at the same location of all the LRIs is recognized as the selected pixel at that location.

Afterwards, these pixels are mosaicked together to form the final all-in-focus image. Simultaneously, we can also obtain a corresponding label map for all the best focused regions among all the LRIs. The mosaicked image can be zoomed and taken as the initial value to our SR algorithm, and the label map gives the information of the relative depth and helps determine the weights of images for SR reconstruction.

### 3.2. Super-Resolution

When the image source is not enough, the problem becomes ill-posed and the prior knowledge should be incorporated. Following the least squares principle, we have:

$$\{x, h\} = \arg \min_{x, h} \{\|y - Px\|_2 + \lambda \|Cx\|_2\} \quad (4)$$

where  $C$  is a Laplacian filter,  $\lambda$  is a user defined regularization parameter. The bigger  $\lambda$  is, the smoother the result image is.

The whole system is not linear, but the HRI pixels and the blur coefficients can be supposed to be locally linear. Trellis search method [11] can be used to solve such a problem in an iterative way, in which the filter coefficients and the HRI pixels are updated alternatively. Because the inverse problem is generally ill-posed, we have to solve the problem by regularization with an additional term  $C^T C$ :

$$T = T(x^{(k)}) \quad (5a)$$

$$h^{(k+1)} = T^+ y \quad (5b)$$

$$P = P(h^{(k+1)}) \quad (5c)$$

$$x^{(k+1)} = (P^T P + \lambda C^T C)^{-1} P^T y \quad (5d)$$

Generally, the iterative algorithm can be accelerated by feeding a good initial input. In our experiments, we first build an infinite depth-of-field image by mosaicking all the in-focus regions into an image [2, 1], and then interpolate it to some size as an initial input.

For color images, it is not necessary to restore all the three channels. Because human eyes are more sensitive to the illuminance of images, we apply our reconstruction algorithm only to the Y channel and finally interpolate the Cb and Cr channels to make the all-in-focus images together with the super-resolved Y channel.

The practical algorithm can be applied to small blocks, thus can be implemented in parallel for blocks of partitioned images.

Our iterative algorithm is summarized as follows:

1. Construct an all-in-focus LRI by mosaicking, and zoom the LRI to a larger image by interpolation so that it can be the initial HRI input  $x^{(0)}$  for this iterative algorithm.
2. Transform the interpolated image from RGB space into YCbCr space.
3. Transform each LRI into YCbCr space.
4. Partition the Y channel of all LRIs into small blocks in the same way.
5. Apply the following SR algorithm for each block to all the blocks of the Y channel:
  - (a) Construct the regularization matrix  $C$ . Set  $\lambda$  as an appropriate value, e.g. 0.5.  $k = 0$ .
  - (b) Construct  $T$  from  $x^{(k)}$ ,  $T = T(x^{(k)})$ , and calculate its Moore-Penrose inverse  $T^+$ , then update  $h$ :  $h^{(k+1)} = T^+ y$
  - (c) Construct  $P$  from  $h^{(k+1)}$ ,  $P = P(h^{(k+1)})$ , and update  $x$ :  $x^{(k+1)} = (P^T P + \lambda C^T C)^{-1} P^T y$
  - (d) If  $(x^{(k+1)} - x^{(k)}) < \epsilon$  or  $k >$  the max number of iterations, exit for next block.
  - (e)  $k = k + 1$ , goto 5c.
6. Mosaic all the restored blocks in an image.
7. Interpolate Cb and Cr channels, and combine with Y channel by transforming the image back into RGB space.

## 4. EXPERIMENTAL RESULTS

### 4.1. Simulation Experiments

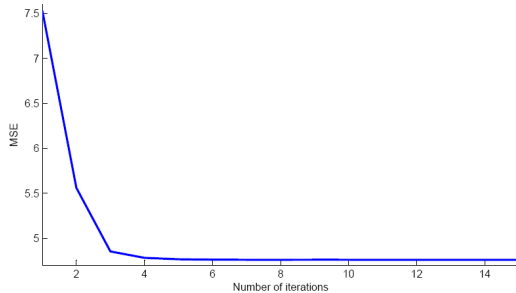
In order to verify the validity and the convergence of our block SR algorithm, we first try it with some simulated LRIs from a natural image, which we can compare the reconstructed HRI with the ground truth of the true HRI.

The block size we use for partitioning the image is  $12 \times 12$ . The original image is first blurred with nine different Gaussian filters and Gaussian noise, respectively. Then, each blurred image is downsampled at 2/3 of the original sampling rate so that the block size becomes  $8 \times 8$ . From these  $8 \times 8$  blocks, we try our algorithm to restore the  $12 \times 12$  block image.

In our experiments, a convergence curve for a block is given in Fig.3, which shows that the error between the ground truth and the estimated image converges within 10 iteration steps. The results show that the algorithm is stable and the filter coefficients are also well estimated.

Image	airplane	lena	baboon	camera	peppers
<b>M</b>	31.22	32.56	34.29	27.38	33.35
<b>M+SR</b>	34.09	35.98	35.31	28.93	35.10

**Table 1:** PSNR of the reconstructed images (dB): (M) by mosaicking only, (M+SR) by mosaicking and superresolution



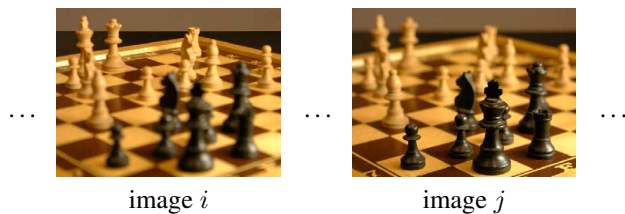
**Fig. 3:** Convergence curve of the iterations over a block

In *Table 1*, we give PSNR (dB) of each reconstructed image relative to the original image. The table shows that our algorithm of mosaicking and superresolution with 9 differently-blurred LR images (M) is better than the direct enlargement with just one blur-free LR image (M+SR).

#### 4.2. Experiments on Real Images

In our experiments for acquired natural images, we use the *Chessboard* image sequence [12] as shown in *Fig.4*.

In the experiments, an all-in-focus LRI and an all-in-focus HRI are reconstructed from the sequence of natural images, see *Fig.5*. We partition the original LRI images into  $8 \times 8$  blocks, and set the magnification factor as 1.5. From the result, we can see that the detailed information in the all-in-focus HRI by mosaicking and superresolution has been better restored than that in the all-in-focus LRI by mosaicking only, especially near the edges.



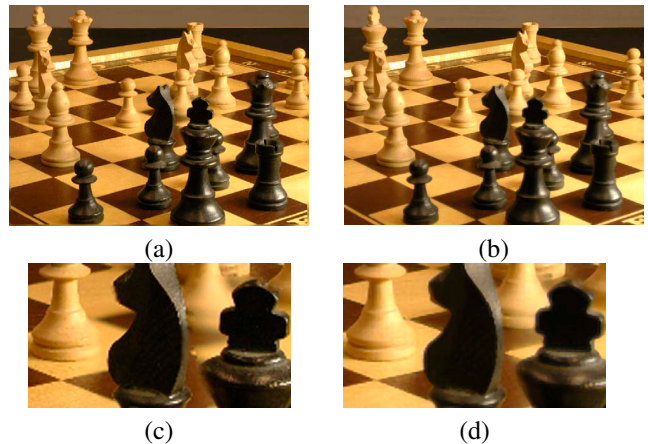
**Fig. 4:** A sequence of 29 real images. Different regions are blurred differently.

### 5. CONCLUSION

In this paper we have introduced an all-in-focus method for digital imaging with super-resolution by using a sequence of images with various focal settings. When the initial enhancement is accurate with a zoomed mosaicked all-in-focus LRI, the iterative algorithm converges fast. Even if the initial estimate is not so accurate, the iterative result is still largely improved. Compared to the DFD-based algorithms, our new method is potentially fast.

### 6. REFERENCES

[1] H. Eltoukhy and S. Kavusi, "A computationally efficient algorithm for multi-focus image reconstruction," in *Proc. SPIE. Electronic imaging*, 2003.



**Fig. 5:** The reconstructed all-in-focus images (a) LRI by mosaicking only, (b) HRI by mosaicking and super-resolution, (c) part of image (a), (d) part of image (b).

[2] S. Ball, "Generating images with enhanced depth of field by creating a seamless mosaic from a sequence of images with different focal settings," Tech. Rep., 2004.

[3] R.Y. Tsai and T.S. Huang, "Multiframe image restoration and registration," in *Advances in Computer Vision & Image Proc.*, vol. 1, pp. 317–339. JAI Press Inc., 1984.

[4] S.P. Kim, N.K. Bose, and H.M. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframes," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 38, no. 6, pp. 1013–1027, June 1990.

[5] S. Rhee and M.G. Kang, "Dct-based regularized algorithm for high-resolution image reconstruction," in *Proc. Int. Conf. Image Proc.*, 1999, vol. 3, pp. 184–187.

[6] 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 Proc.*, vol. 6, no. 12, pp. 1621–1633, 1997.

[7] Z. Lin and H.-Y. Shum, "Fundamental limits of reconstruction-based superresolution algorithms under local translation," *IEEE Trans. PAMI*, vol. 26, no. 1, pp. 83–97, 2004.

[8] T. Choi M. Subbarao and A. Nikzad, "Focusing techniques," in *Proc. SPIE. Electronic imaging*, 1992.

[9] D. Rajan and S. Chaudhuri, "Simultaneous estimation of super-resolved scene and depth map from low resolution defocused observations," *IEEE Trans. PAMI*, vol. 25, no. 9, pp. 1102–1117, 2003.

[10] G.H. Harikumar and Y. Bresler, "Perfect blind restoration of images blurred by multiple filters: Theory and efficient algorithms," *IEEE Trans. Image Proc.*, vol. 8, no. 2, pp. 202–219, 1999.

[11] N. Seshadri, "Joint data and channel estimation using fast blind trellis search techniques," *IEEE Trans. Commun.*, vol. 42, pp. 1000–1011, Feb.-Apr. 1994.

[12] "http://www.cs.cornell.edu/erenrich/dof/," .