Simultaneous Estimation of Super-Resolved Scene and Depth Map from Low Resolution Defocused Observations

Deepu Rajan and Subhasis Chaudhuri, *Senior Member*, *IEEE*

Abstract—This paper presents a novel technique to simultaneously estimate the depth map and the focused image of a scene, both at a super-resolution, from its defocused observations. Super-resolution refers to the generation of high spatial resolution images from a sequence of low resolution images. Hitherto, the super-resolution technique has been restricted mostly to the intensity domain. In this paper, we extend the scope of super-resolution imaging to acquire depth estimates at high spatial resolution simultaneously. Given a sequence of low resolution, blurred, and noisy observations of a static scene, the problem is to generate a dense depth map at a resolution higher than one that can be generated from the observations as well as to estimate the true high resolution focused image. Both the depth and the image are modeled as separate Markov random fields (MRF) and a maximum a posteriori estimation method is used to recover the high resolution fields. Since there is no relative motion between the scene and the camera, as is the case with most of the super-resolution and structure recovery techniques, we do away with the correspondence problem.

Index Terms—Super-resolution, depth from defocus, space-variant blur identification, restoration, Markov random field.

—————————— ◆ ——————————

1 INTRODUCTION

A VAILABILITY of high spatial resolution images is often desirable in most computer vision applications. Be it remote sensing, medical imaging, robot vision, industrial inspection, or video enhancement (to name a few), operating on high resolution images leads to a better analysis in the form of fewer misclassifications, better fault detection, more true-positives, etc. However, acquisition of high resolution images is severely constrained by the drawbacks of sensors that are commercially readily available. Thus, images acquired through such sensors suffer from aliasing and blurring. Aliasing occurs as a consequence of insufficient density of the detector array, which causes frequencies above the Nyquist rate to alias. An imaging sensor having a dense detector array is too expensive to be considered as an alternative to generate high resolution images. Blurring occurs due to integration of the sensor point spread function (PSF) at the sensor surface. Blurring due to the relative motion between the camera and the scene and the optical out-of-focus blur add to the degradations. Reducing the pixel size in order to pack more photodetectors results in shot noise. The minimum pixel size beyond which the existence of shot noise becomes visible is estimated to be approximately 50 μm^2 [1]. Hence, one must resort to image processing methods to construct a high resolution image from one or more available low resolution images. *Super-resolution* refers to the process of producing a high spatial resolution image from several low resolution images. It

includes upsampling the image, thereby increasing the maximum spatial frequency that can be represented and removing degradations that arise during the image capture, viz., aliasing and blurring. Many researchers have tackled the super-resolution problem, both for still and video images, e.g., [2], [3], [4], [5] (see [6] for details). In this paper, we present a new approach to super-resolution generation of both the intensity map and the scene structure (depth), simultaneously, using the depth related defocus as the cue.

It was Pentland who first suggested that measuring the amount of blurring at a given point in the image could lead to computing the depth to the corresponding point in the scene, provided the parameters of the lens system, like aperture, focal length and lens-to-image plane distance, are known [7]. Given two images of a scene recorded with different camera settings, we obtain two constraints on the spread parameters of the point spread functions corresponding to the two images. One of the constraints is obtained from the geometry of image formation, while the other is obtained from the intensity formation in the defocused images. These two constraints are simultaneously solved to determine distances of objects in the scene [8].

In this paper, we expand the scope of super-resolution to include high resolution depth information in a scene in addition to recovering intensity values. As mentioned earlier, one of the degradations in a low resolution image is the sensor related blur, which appears as a consequence of the low resolution point spread function of the camera. Blurring can also arise due to relative motion between the camera and the scene. In the case of real aperture imaging, we know that the blur at a point is a function of the depth of the scene at that point. Thus, we notice that blur is a *natural* cue in a low resolution image formed by a real-aperture camera. We exploit this blur to recover the depth map through the depth from defocus formulation; moreover, we

————————————————
- D. Rajan is with the School of Computer Engineering, Nanyang Technological University, Singapore 639798.
  E-mail: asdrajan@ntu.edu.sg.
- S. Chaudhuri is with the Department of Electrical Engineering, Indian Institute of Technology, Bombay, Mumbai 400 076, India.
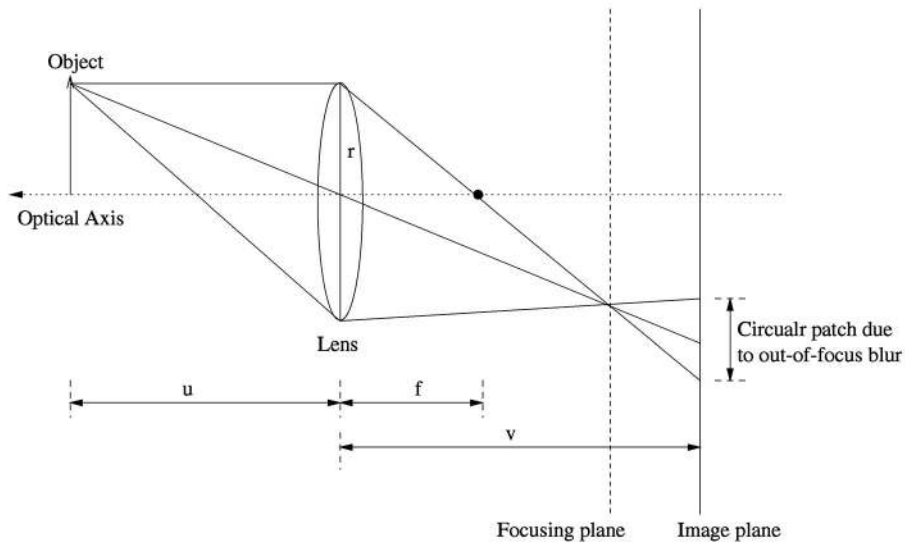 E-mail: sc@ee.iitb.ac.in.

Fig. 1. Illustration of out-of-focus image formation.

propose how the depth map can be estimated at a higher resolution than one that can be normally extracted from such observations. We call such a dense spatial depth map a *super-resolved depth*. In addition to this, we propose how to simultaneously estimate the true, high resolution focused image of the scene. This process may be called super-resolved space varying restoration. The two stage process of identifying the blur and deconvolving the observed image with the corresponding PSF performs unsatisfactorily in the presence of noise [9]. In this paper, we present a new concept where these two tasks, namely, the super-resolved depth recovery and the super-resolved space varying image restoration, have been combined through the interplay of two separate Markov random fields (MRFs)—one representing the depth map and the other representing the intensity field.

This paper is organized as follows: In Section 2, we briefly review the prior work in depth from defocus (DFD) and the super-resolution techniques. We model the formation of the low resolution depth and image fields in Section 3 and describe the proposed method to simultaneously extract the super-resolved depth and the image fields in Section 4. Experimental results and conclusions are presented in Sections 5 and 6, respectively.

2 RELATED WORK

The basic premise of depth from defocus is that, since the degree of defocus is a function of lens setting and the depth of the scene, it is possible to recover the latter if the amount of blur can be estimated, provided the lens setting is known. An out-of-focus point light source images into a blurred circle [10], whose radius is described by a blur parameter σ defined as

$$\sigma = \rho \, r \, v \left(\frac{1}{f} - \frac{1}{v} - \frac{1}{u} \right), \qquad (1)$$

where f is the focal length, u is the distance of the object point from the lens, v is the distance between the lens and the image detector, r is the radius of the lens aperture, and ρ is a camera constant that depends on its optics and the CCD array resolution. The above relationship is valid primarily

in geometric optics and when the lens suffers from no abberations. Hence, these assumptions are implied throughout in this paper. Fig. 1 illustrates the formation of the image of an object point as a circular patch due to the dislocation of the image plane from the focusing plane. Since the depth at various points in the scene may be varying continuously, σ would also vary all over the image accordingly. The shift-varying PSF of the optical system is modeled as a circularly symmetric 2D Gaussian function

$$h(i, j; m, n) = \frac{1}{2\pi\sigma^2(m,n)} \exp\left(-\frac{i^2 + j^2}{2\sigma^2(m,n)} \right). \qquad (2)$$

In the literature, we encounter two kinds of blur, viz., the Gaussian blur and the pillbox (circular) blur. We have used the Gaussian blur here for computational ease, although the circular blur will work equally as well. As a matter of fact, any single-parameter class of PSF can be handled under the current formulation.

Early investigations of the DFD problem were carried out by Pentland [10], where he compared two images locally, one of which was formed with a pin-hole aperture and then recovered the blur parameter through deconvolution in the frequency domain. In [11], Subbarao removed the constraint of one of the images being formed with a pin-hole aperture by allowing several camera parameters like aperture, focal length, and lens-to-image plane distance to vary simultaneously. Prasad et al. [12] formulate the DFD problem as a 3D image restoration problem. The defocused image is modeled as the combinatorial output of the depths and intensities of the volume elements (voxels) of an opaque 3D object. Klarquist et al. propose a maximum likelihood (ML)-based algorithm that computes the amount of blur as well as the deconvolved images corresponding to each subregion [13]. In [14], Gökstorp estimates blur by obtaining local estimates of the instantaneous frequency, amplitude, and phase, using a set of Gabor filters in a multiresolution framework. In [15], Watanabe and Nayar describe a class of broadband rational operators which are capable of providing invariance to the scene texture. Schechner and Kiryati address the similarities in DFD and stereo techniques on the basis of the geometric triangulation

principle in [16]. An active ranging device that uses an optimized illumination pattern to obtain an accurate and high resolution depth map is described by Nayar et al. in [17]. In [18], Rajagopalan and Chaudhuri use the complex spectrogram and the pseudo-Wigner distribution for recovering the depth within the framework of the space-frequency representation of the image. In [19], they extend this approach to impose smoothness constraints on the blur parameter to be estimated and use a variational approach to recover the depth. In [20], a MAP-MRF framework is used for recovering the depth as well as the focused image of a scene from two defocused images. However, the recovered depth map and the scene image are at the same resolution as the observations. Part of the current proposal derives its origin from this paper. Other techniques for depth recovery and issues related to optimal camera settings are described in [8]. The DFD problem can also be viewed as a special case of the space-variant blur identification problem since, eventually, it is the blur at a point that acts as the cue for determining the depth in the scene at that point. Space variant blur identification methods have been described in [21], [22], and [23].

In this paper, our aim is not only to recover the depth from defocused images, but also to do so at a higher spatial resolution, besides generating the super-resolved image of the scene. Thus, given a sequence of low resolution blurred observations of size $M_1 \times M_2$, we wish to generate a dense depth map of size, say, $qM_1 \times qM_2$, where q is the upsampling factor. We call this the *super-resolved depth* estimate. Clearly, by doing this, we get a more accurate description of the depth in the scene, which eventually leads to a better performance of the task at hand.

Now, we review some of the prior works on super-resolution imaging. Tsai and Huang [2] were the first to propose a frequency domain approach to reconstruction of a high resolution image from a sequence of undersampled low resolution, noise-free images. Kim et al. discuss a recursive algorithm, also in the frequency domain, for the restoration of super-resolution images from noisy and blurred images [24]. A minimum mean squared error approach for multiple image restoration, followed by interpolation of the restored images into a single high resolution image is presented in [25]. Ur and Gross use the Papoulis-Brown generalized sampling theorem to obtain an improved resolution picture from an ensemble of spatially shifted pictures [26]. However, these shifts are assumed to be known by the authors. An iterative backprojection method is used in [3] wherein a guess of the high resolution output image is updated according to the error between the observed and the low resolution images obtained by simulating the imaging process. But, back-projection methods can be used only for those blurring processes for which such an operator can be calculated. A projection onto convex sets (POCS)-based method is described in [27]. A MAP estimator with Huber-MRF prior is described by Schultz and Stevenson in [28]. Elad and Feuer [29] propose a unified methodology for super-resolution restoration from several geometrically warped, blurred, noisy, and downsampled images by combining ML, MAP, and POCS approaches. An adaptive filtering approach to super-resolution restoration is described by the same authors in [30]. Chiang and Boult [31] use edge models and a local blur estimate to develop an edge-based super-resolution algorithm. Recently, Rajan and Chaudhuri have proposed a
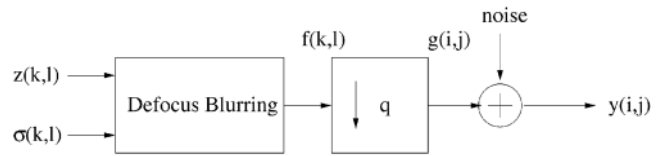

Fig. 2. Low resolution image formation from high resolution image and due to defocus blur.

generalized interpolation scheme and used it to generate super-resolution images from photometric stereo [32], [33].

Shekarforoush et al. use MRFs to model the images and obtain a 3D high resolution visual information (albedo and depth) from a sequence of displaced low resolution images [4]. The effect of sampling a scene at a higher rate is acquired by having interframe subpixel displacements. But, they do not consider blurred observations. Cheeseman et al. describe another Bayesian approach for constructing a super-resolution surface by combining information from a set of images of the given surface [34]. However, their model includes registration parameters, the PSF, and camera parameters that are estimated first and, subsequently, the surface reconstruction is carried out. In both these cases, the issue of registration has to be addressed since they involve camera displacement. Errors in registration are reflected in the quality of the super-resolved image generated as well as in the depth estimate. Hence, if we can avoid any relative motion between the camera and the scene, we would be able to do away with the correspondence problem. This is precisely what is achieved by resorting to using the defocus cue as is commonly done in the depth from defocus approach. However, the restoration problem becomes a space varying one and the accuracy would depend on how well we can estimate the blur parameter.

3 LOW RESOLUTION IMAGE AND BLUR FORMATION

We now discuss the formation of a low resolution image from a high resolution description of the scene. Note that the problem we solve here is actually the inverse. It should be noted here that, through the choice of the term super-resolved depth or the intensity, we mean the enhancement in the spatial resolution and not that of the quantization levels of the depth or the intensity map.

The actual model of the low resolution observation is as follows: A high resolution image of the scene is formed by the camera optics and this image is defocused due to varying depth components in the scene. The defocused scene is now sensed optically by the low resolution CCD elements. A sensor noise is now added to these measurements and one obtains the observed image.

Let us now refer to Fig. 2 for an illustration of the above model. Let $z(k,l)$ and $\sigma(k,l)$ be the true high resolution intensity and blur (parameterized) maps, respectively. Due to depth related defocus, one obtains the blurred but high spatial resolution intensity map $f(k,l)$. Suppose the low resolution image sensor plane is divided into $M_1 \times M_2$ square sensor elements and $\{g(i,j)\}$, $i = 0, \ldots, M_1 - 1$ and $j = 0, \ldots, M_2 - 1$ are the low resolution intensity values. For a decimation ratio of q, the high resolution grid (images z and f) will be of size $qM_1 \times qM_2$. The forward process of obtaining $\{g(i,j)\}$ from $\{f(k,l)\}$ is written as [28]
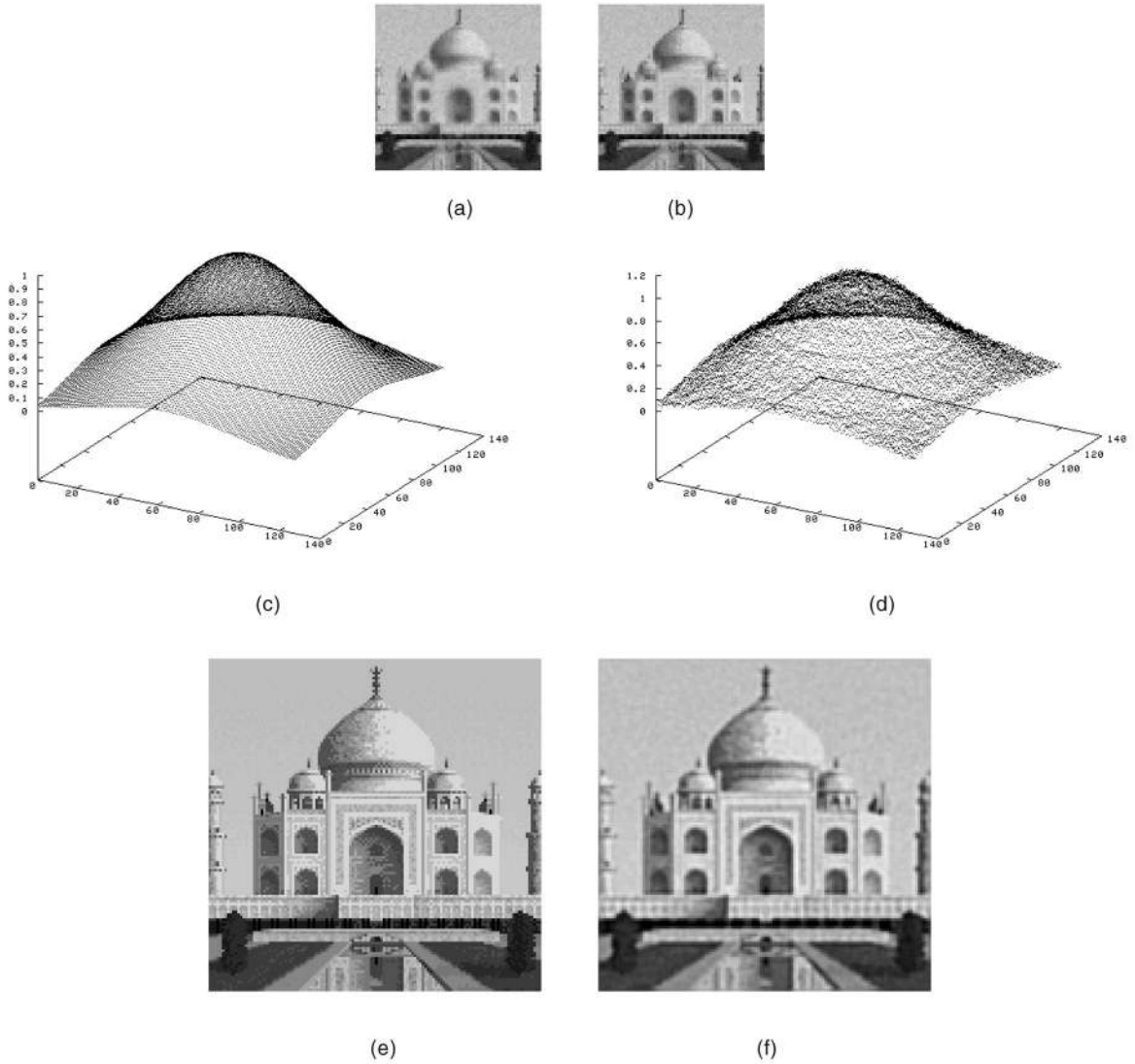
Fig. 3. (a) and (b) Two of the defocused low resolution Taj images. (c) The true high resolution blur. (d) The estimated super-resolved blur. (e) The super-resolved Taj image.

$$g(i,j) = \frac{1}{q^2} \sum_{k=qi}^{(q+1)i-1} \sum_{l=qj}^{(q+1)j-1} f(k,l), \qquad (3)$$

i.e., the low resolution intensity is the average of the high resolution intensities over a neighborhood of q^2 pixels. This decimation model simulates the integration of light intensity that falls on the high resolution detector. It should be mentioned here that the above relationship assumes that the entire area of the pixel is used for light sensing and nothing is used for electrical wiring or insulation. Thus, we assume the fill-factor for the CCD array to be unity.

The process of blurring the high resolution image $z(k,l)$ due to defocus is modeled by

$$f(k,l) = \sum_u \sum_v z(u,v) h(k,l;u,v), \qquad (4)$$

where $f(\cdot)$ is the defocused version of the high resolution image and $h(\cdot;\cdot)$ is the space variant blurring function given in the previous section.

The space varying blurring function is dependent only on a single blur parameter σ. However, in the present context, this blur describes a high resolution description of the depth field compared to the spatial resolution at which the scene is observed. The addition of white Gaussian noise at the CCD sensor completes the low resolution observation model and is illustrated in Fig. 2.

We now cast the super-resolution problem in a restoration framework. There are p observed images $y_m(i,j)$, each of size $M_1 \times M_2$. These images are defocused, decimated, and noisy versions of a single high resolution image $z(k,l)$ of size $qM_1 \times qM_2$. If y_m is the $M_1M_2 \times 1$ lexicographically ordered vector containing pixels from the low resolution image $y_m(i,j)$, then a vector f of size $q^2M_1M_2 \times 1$ containing pixels of the high resolution defocused image can be formed by placing each of the $q \times q$ pixel neighborhoods sequentially so as to maintain the relationship between a low resolution pixel and its corresponding high resolution pixel. This relationship can also be represented through a decimation matrix D of size $M_1M_2 \times q^2M_1M_2$ consisting of q^2 values of $\frac{1}{q^2}$ in each row, the exact form of which can be found in [28].
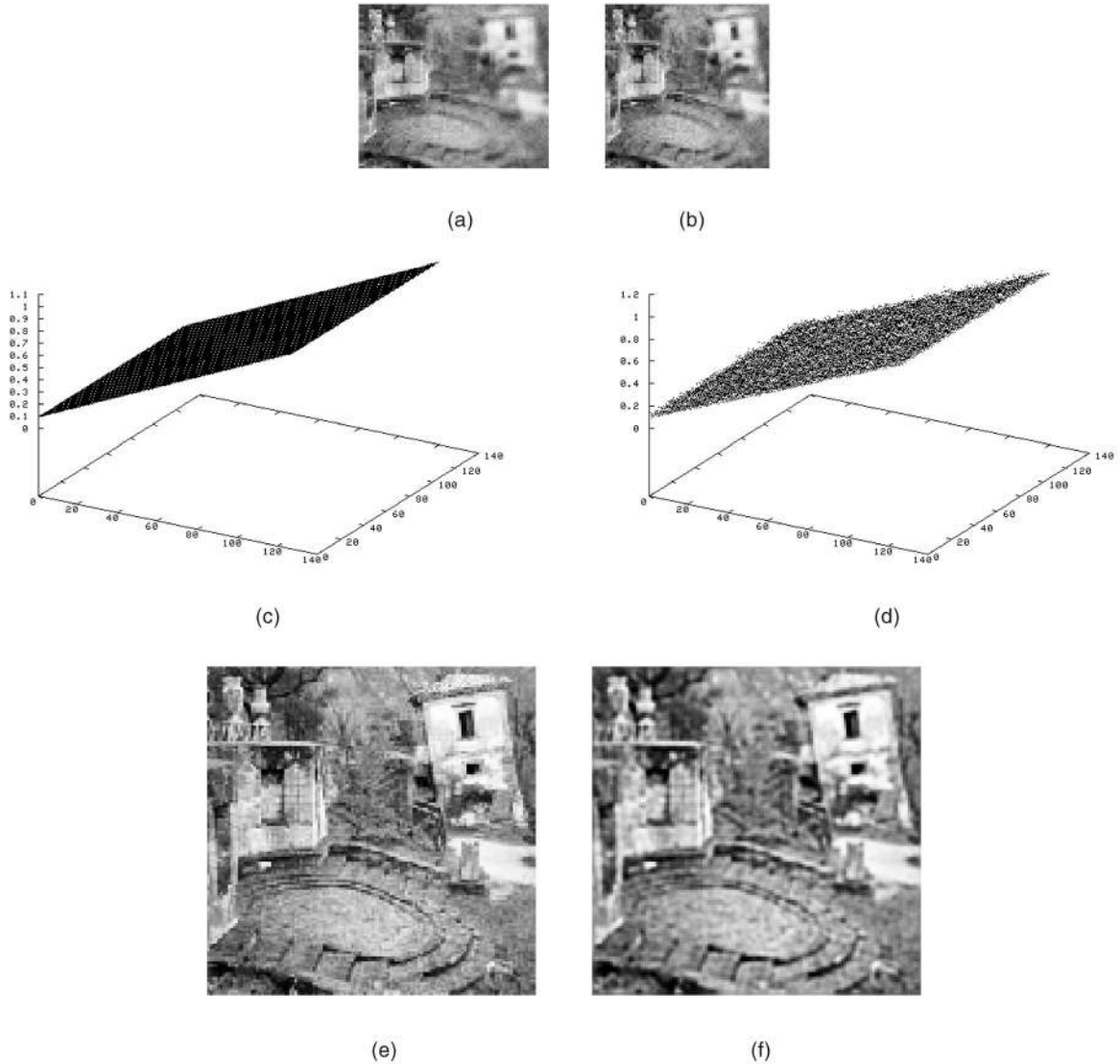
Fig. 4. (a) and (b) Two of the defocused low resolution Graveyard images. (c) The true high resolution blur. (d) The estimated super-resolved blur. (e) The original high resolution Graveyard image. (f) The super-resolved Graveyard image.

Let H be the blur matrix corresponding to the space variant blurring function $h(k, l; u, v)$ in (4). We reiterate that this function is governed by the high resolution blur $\sigma(k, l)$. Thus, for the mth low resolution observation, the blur matrix H_m is a function of $\sigma_m(k, l)$, where the earlier notation is modified to include a subscript for the observation number. The blur field is now lexicographically ordered to obtain a vector s_m. The blur matrix $H_m = H(s_m)$ corresponding to the mth observation can now be formed, but, due to the space varying nature of the blur, it does not possess a block Toeplitz structure. The image formation model is now compactly written as

$$y_m = DH(s_m)z + n_m, \quad m = 1, \ldots, p, \qquad (5)$$

where $H(s_m)$s are the high resolution space varying blurring matrix (PSF) of size $q^2 M_1 M_2 \times q^2 M_1 M_2$, D is the decimation matrix, n_m is the $M_1 M_2 \times 1$ noise vector, and p is the number of low resolution observations. Thus, the model consists of a collection of low resolution images, each of which differs from the others in the blur matrix, which is

akin to changing the focus of a stationary camera looking at a stationary scene.

It is interesting to note what happens when one interchanges the order of the two operators, D and H, given in (5). If one expresses the low resolution observation as

$$y_m = H(s_m)Dz + n_m,$$

the size of the blur matrix H is now much reduced. This allows a much faster computation of the fields s and z, albeit the corresponding blur (or depth) is estimated at a lower resolution. However, the interchanging of the two operators H and D may not always be mathematically valid as H and D may not commute. The operators would commute if the decimation does not introduce any aliasing, i.e., the super-resolved depth map should be a fairly smooth one except possibly at the points of discontinuity preserved through the introduction of line fields (see the next section). Experimental observations suggest that, for many practical applications, the above formulation yields quite an acceptable accuracy. In order to derive the computational benefit,
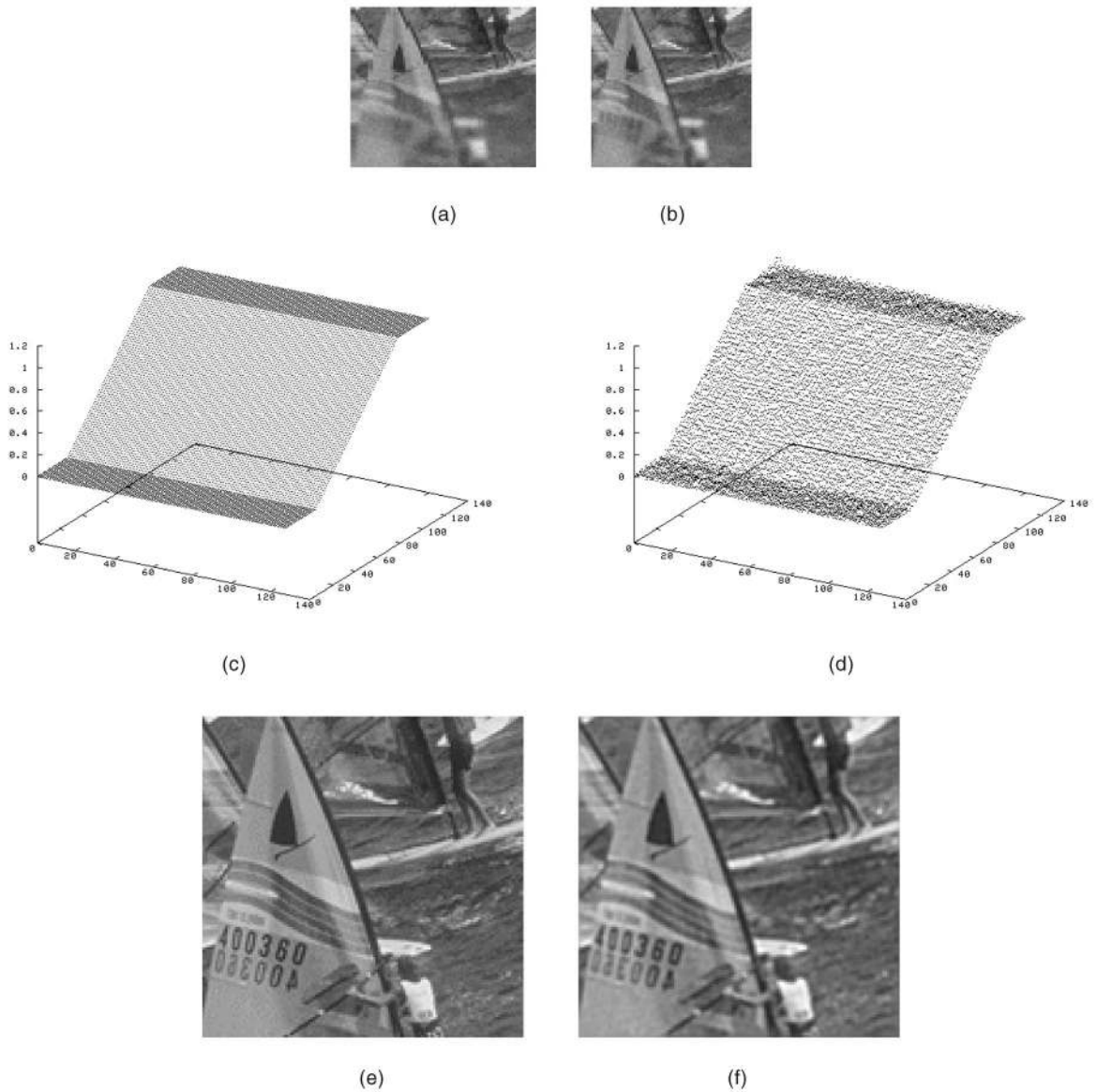
Fig. 5. Experimentation with discontinuous blur variation. (a) and (b) Two of the low resolution Sail images. (c) The true high resolution blur. (d) The estimated super-resolved blur. (e) The original high resolution Sail image. (f) The super-resolved Sail image.

one may start with the above equation, the solution of which will serve as a very good initial guess for the correct model given in (5).

4   SUPER-RESOLVED SCENE AND DEPTH RECOVERY

4.1   Stochastic Models of Fields:

Markov random fields (MRFs) have emerged as a popular stochastic model for images due to its ability to capture local dependencies and its equivalence to the Gibbs random field (GRF) [35]. Let X be a random field over an arbitrary $N \times N$ lattice of sites $L = \{(i,j)|1 \le i, j \le N\}$. From the Hammersley-Clifford theorem [36], which proves the equivalence of an MRF and a GRF, we have $P(X = x) = \frac{1}{K}e^{-U(x)}$, where x is an instance of X, K is the partition function given by $\sum_x e^{-U(x)}$, and $U(x)$ is the energy function given by $U(x) = \sum_{c \in \mathcal{C}} V_c(x)$.

$V_c(x)$ denotes the potential function for clique c and $\mathcal{C}$ is the set of all cliques.

The presence or absence of discontinuities conveys important information such as change in surface orientation, depth, texture, etc. The concept of line fields on a dual lattice consisting of sites corresponding to vertical and horizontal line fields was introduced in [37]. The horizontal line field element $l(i,j)$ connecting site (i,j) to $(i,j-1)$ aids in detecting a horizontal edge, while the vertical line field element $v(i,j)$ connecting site (i,j) to $(i-1,j)$ helps in detecting a vertical edge. We have chosen $l(i,j)$ and $v(i,j)$ to be binary variables over the line fields L and V. The on-state of the line-process variable indicates that a discontinuity, in the form of a high gradient, is detected between neighboring points, e.g.,

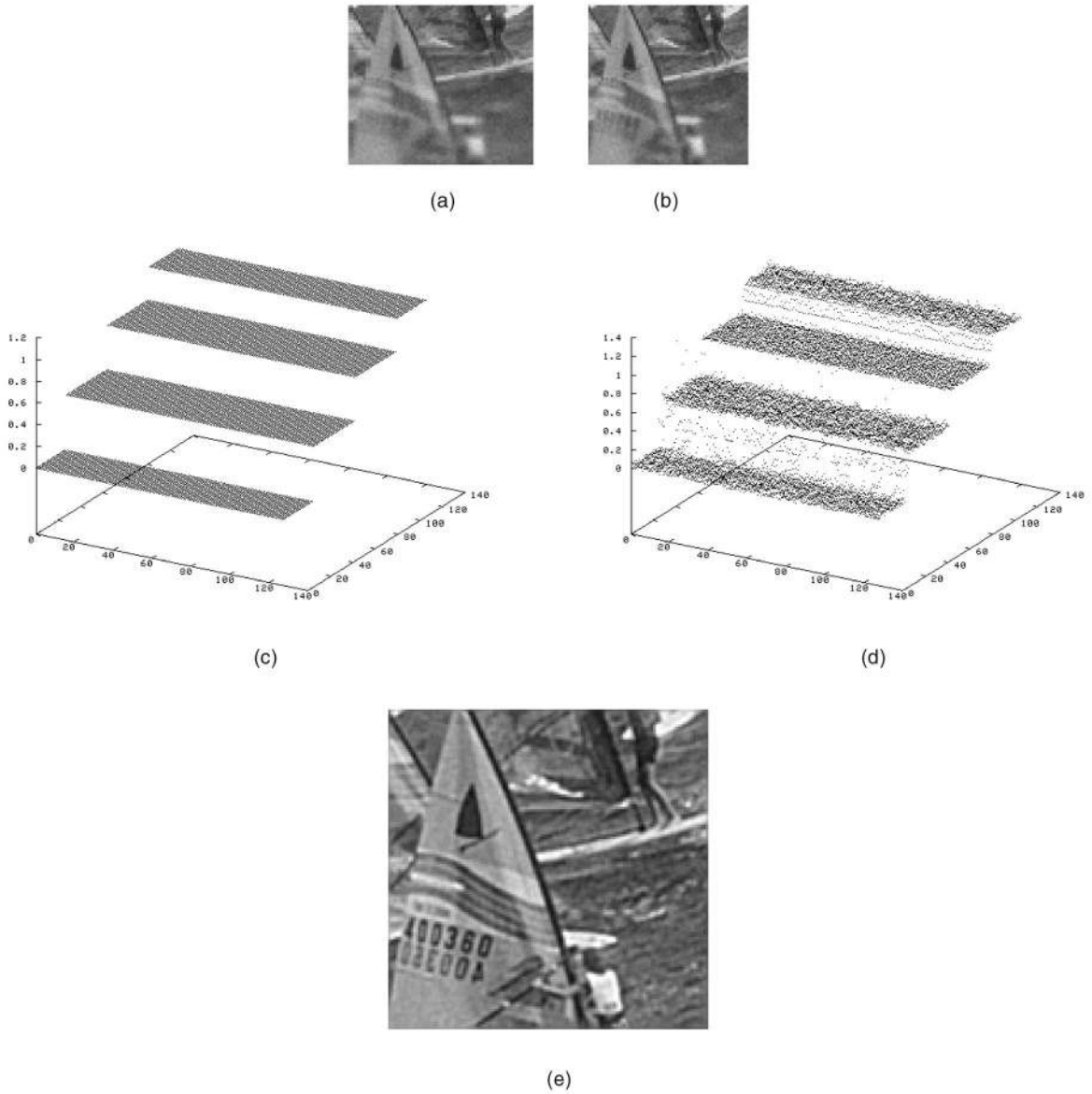$$l(i,j) = 1 \text{ if } |x(i,j) - x(i-1,j)| > \text{Threshold, else } l(i,j) = 0.$$

Fig. 6. Another experiment with step-discontinuous blur. (a) and (b) Two of the low resolution Sail images with step variation in blurs. (c) The true high resolution blur. (d) The estimated super-resolved blur. (e) The super-resolved Sail image.

Each turn-on of a line-process variable is penalized by a quantity γ so as to prevent spurious discontinuities. Thus, for the so-called *weak membrane* model [38], the energy function for the random process X with discontinuity fields L and V is written as
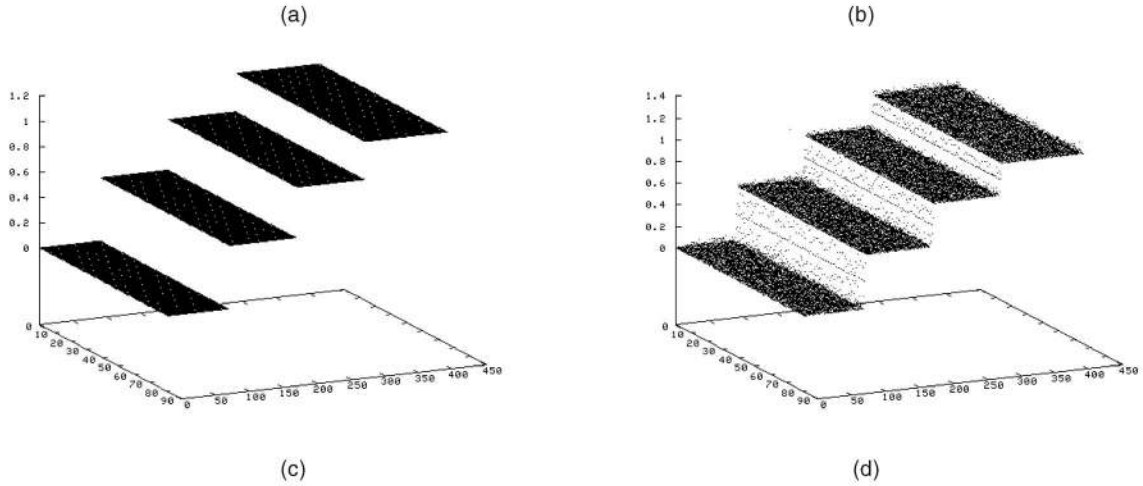
$$
\begin{aligned}
U(x,l,v) &= \sum_{c \in \mathcal{C}} V_c(x,l,v) \\
&= \sum_{i,j} [(x(i,j)-x(i,j-1))^2(1-v(i,j)) + (x(i,j+1) \\
&\quad - x(i,j))^2(1-v(i,j+1)) + (x(i,j)-x(i-1,j))^2 \\
&\quad (1-l(i,j))+(x(i+1,j)-x(i,j))^2(1-l(i+1,j))] \\
&\quad + \gamma[l(i,j)+l(i+1,j)+v(i,j)+v(i,j+1)].
\end{aligned}
\tag{6}
$$

We use this particular energy function in our studies. Any other form of energy function can also be used without changing the solution modality proposed here.

4.2 Maximum a posteriori (MAP) Solution:

In [39] and [40], the depth of a scene is modeled as an MRF. This is justified because the change in depth of a scene is usually gradual and, hence, depth can be said to exhibit a local dependency. Since the space varying blurring parameter σ in (1) is a function of the scene depth, we expect it to exhibit similar local dependencies and model it by an MRF. Thus, we model both the high resolution image $z(k,l)$ and the blur process $\sigma(k,l)$ as separate Markov random fields. Let S and Z denote the random fields corresponding to the high-resolution space-variant blur parameter $\sigma(k,l)$ and the high-resolution focused image $z(k,l)$ over the $qM_1 \times qM_2$ lattice of sites L, respectively. One may notice there that we are modeling the blur as an MRF and not the depth field because the DFD problem is essentially a shift-varying blur estimation problem. The modeling of the blur field simplifies the mathematical relationship. One may, however, model the depth field directly in conjunction with (1) and is expected to obtain similar results. We assume that S

Fig. 7. Experimentation for readability of text image. (a) and (b) Two of the low resolution Text images. (c) The true high resolution blur. (d) The estimated super-resolved blur. (e) The original high resolution Text image. (f) The super-resolved Text image.

can take B possible levels and Z can take C possible levels. Although the fields S and Z are actually continuous, the blur field is quantized to reduce the number of acceptable configurations in the combinatorial minimization, while the intensity field is usually quantized to 256 gray levels. One may use a nonlinear quantization scheme for the levels of S for better results, but this is not pursued in this exercise.

The a posteriori conditional joint probability of S and Z is given by $P[S = s, Z = z | Y_1 = y_1, \ldots, Y_p = y_p]$, where the Y_ms denote the random fields corresponding to the mth observed image. From Bayes' rule,

$$P[S = s, Z = z | Y_1 = y_1, \ldots, Y_p = y_p] =$$
$$\frac{P[Y_1 = y_1, \ldots, Y_p = y_p | S = s, Z = z] P[S = s, Z = z]}{P[Y_1 = y_1, \ldots, Y_p = y_p]}. \quad (7)$$

The random fields S and Z are assumed to be statistically independent in this study as they refer to two independent processes, namely, the depth and intensity processes. It may be worth mentioning that the assumption of statistical independence of the two fields S and Z may not always be valid. In many cases, the intensity and the depth maps are related, for example, in shape from shading or texture applications where the shading may depend on perspective effects or object geometry. In the absence of any knowledge of the cross covariance matrix between the two fields, we assume them to be independent. Since the denominator in (7) is not a function of s or z, the MAP problem of simultaneous estimation of high resolution space-variant blur identification and super-resolved image can be posed as:
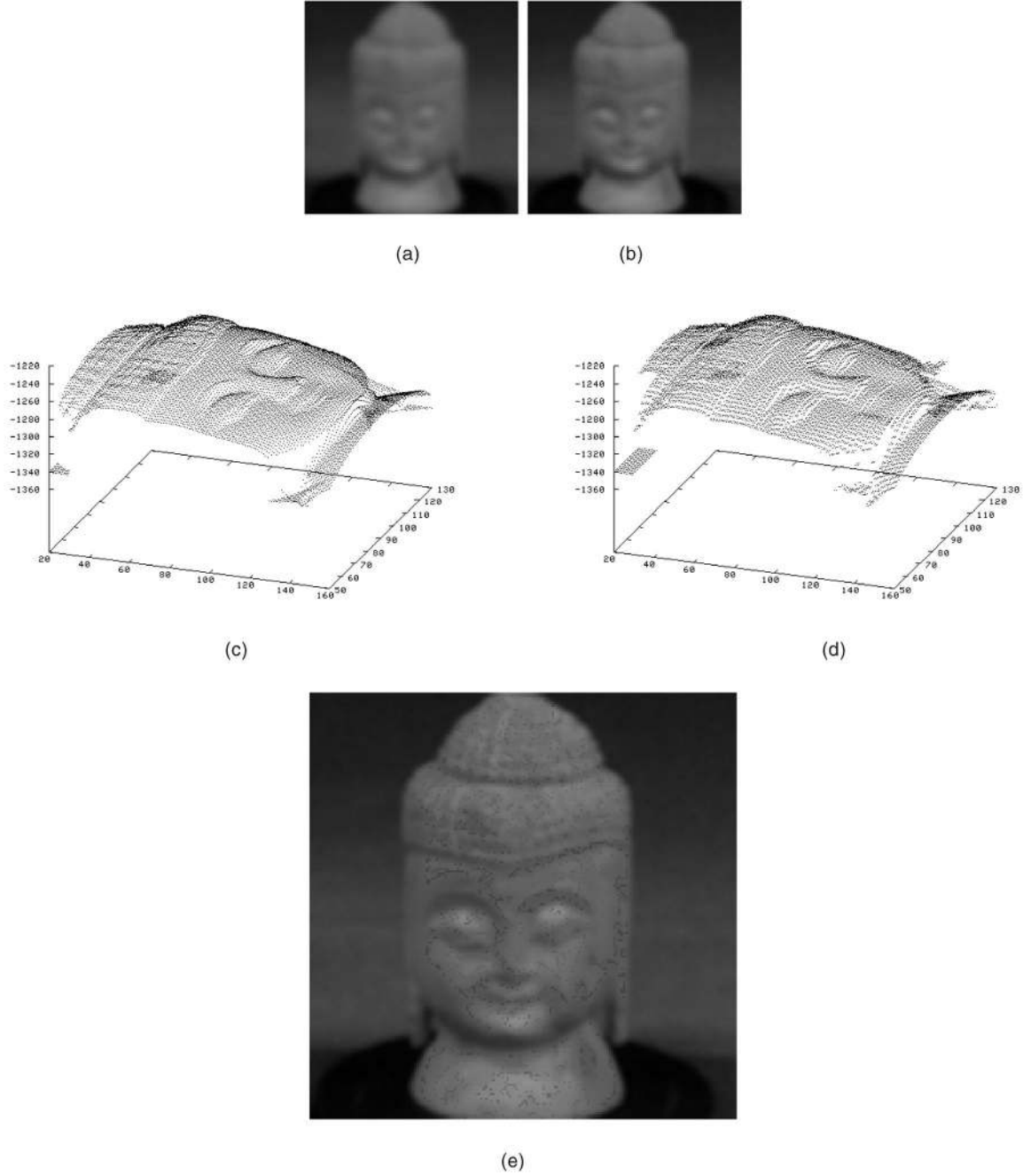
Fig. 8. (a) and (b) Two of the low resolution Buddha images. (c) The true high resolution depth. (d) The estimated super-resolved depth. (e) The super-resolved Buddha image.

$$\max_{s,z} P[Y_1 = y_1, \ldots, Y_p = y_p | S = s, Z = z] P[S = s] P[Z = z]. \tag{8}$$

Note that the random fields S and Z are high resolution, while the observations are low resolution. Since S and Z are both modeled as MRFs, the priors $P[S = s]$ and $P[Z = z]$ have a Gibbs distribution given by

$$P[S = s] = \frac{1}{K^s} \exp\left\{ -\sum_{c \in \mathcal{C}^s} V_c^s(s) \right\} \tag{9}$$

and

$$P[Z = z] = \frac{1}{K^z} \exp\left\{ -\sum_{c \in \mathcal{C}^z} V_c^z(z) \right\}, \tag{10}$$

where K^s and K^z are normalizing constants known as partition functions, $V_c(.)$ is the clique potential, and $\mathcal{C}^s$ and $\mathcal{C}^z$ are the set of all cliques in S and Z, respectively. Thus, the posterior energy function to be minimized is obtained by taking the log of posterior probability and assuming the noise to be i.i.d. Gaussian

$$U = \sum_{m=1}^{p} \frac{||y_m - DH(s_m)z||^2}{2\sigma_\eta^2} + \sum_{c \in \mathcal{C}^s} V_c^s(s) + \sum_{c \in \mathcal{C}^z} V_c^z(z), \tag{11}$$

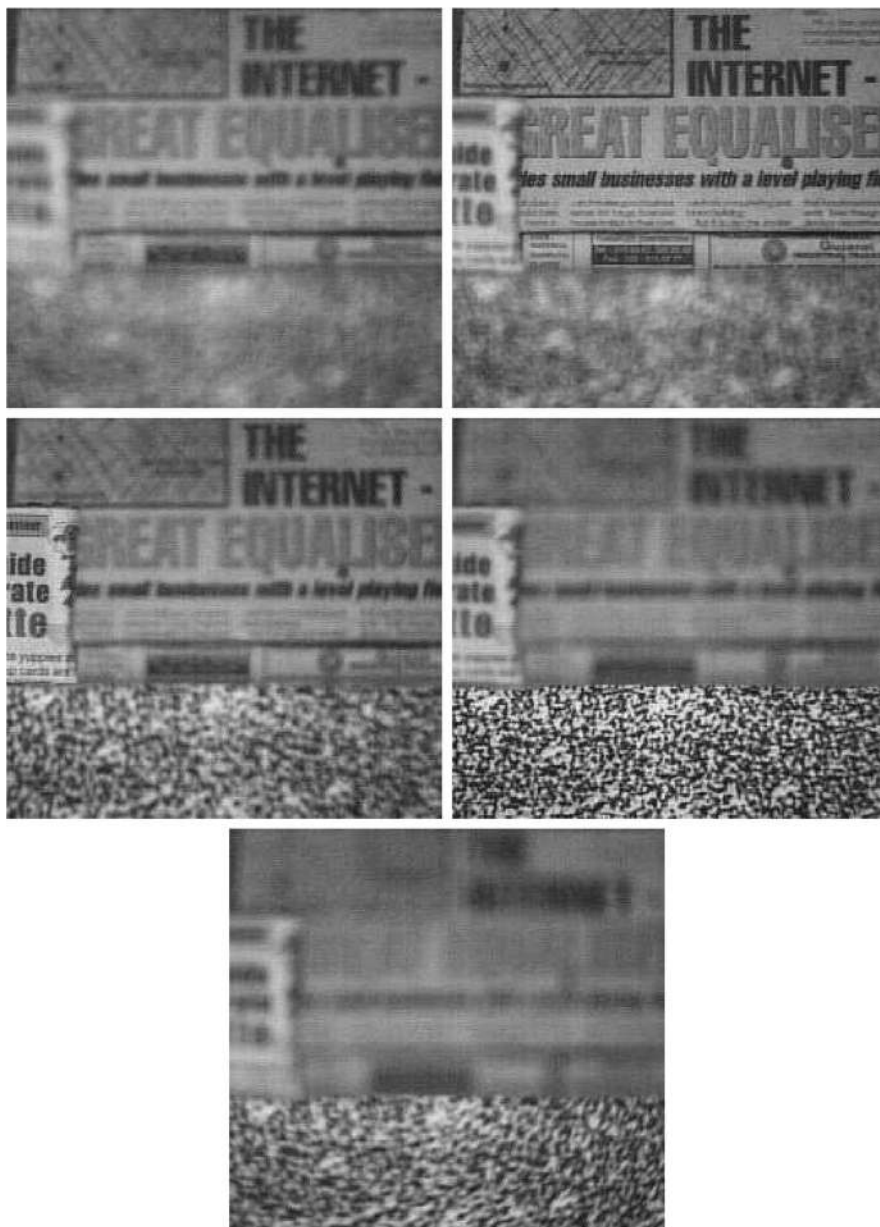where σ_η is the noise variance.

Fig. 9. Five low resolution observations of the "blocks world" captured in the laboratory by varying the focus setting of the camera.

Smoothness is an important assumption underlying a wide range of physical phenomena. However, careless imposition of the smoothness criterion can result in undesirable, oversmoothed solutions. This could happen at points of discontinuities either in the image or in the depth map. Hence, it is necessary to take care of discontinuities. Smoothness constraints on the estimates of the space-variant blur parameter and the intensity process are encoded in the potential function. As mentioned earlier, preservation of discontinuities is done through line fields [37]. We introduce separate line fields for these two processes. After incorporating the first order weak-membrane model, the posterior energy function to be minimized is now defined as

$$U = \sum_{m=1}^{p} \frac{||y_m - DH(s_m)z||^2}{2\sigma_\eta^2} + \sum_{i,j}[\mu_s e_{ss} + \gamma_s e_{sp}]$$
$$+ \sum_{i,j}[\mu_z e_{zs} + \gamma_z e_{zp}], \tag{12}$$

where the smoothness (e_{ss} and e_{zs}) and the penalty (e_{sp} and e_{zp}) terms have been defined in (6) earlier. Parameters μ and γ correspond to the relative weights of the smoothness term and the penalty term necessary to prevent occurance of spurious discontinuities.

When the energy function is nonconvex, there is a possibility of the steepest descent type of algorithms getting trapped in a local minima. Hence, simulated annealing is used to minimize the energy function and to obtain the MAP estimates of the high resolution space-variant blur and the super-resolved image simultaneously. Simulated annealing applies a sampling algorithm such as the Metropolis
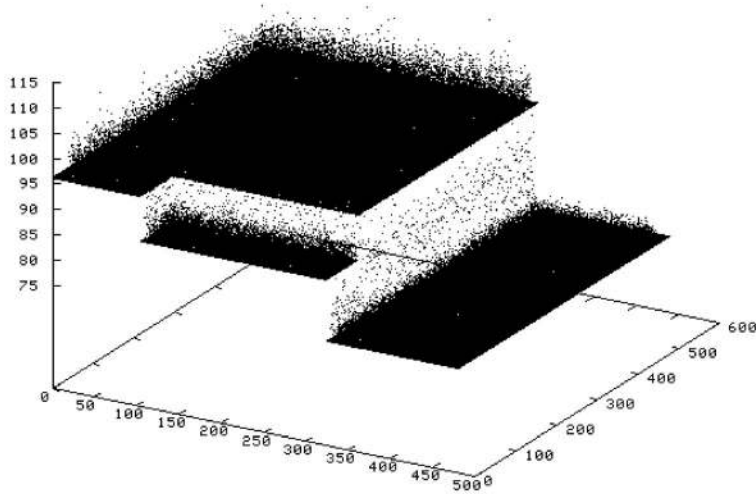
Fig. 10. The super-resolved depth estimates in the "blocks world."

algorithm or Gibbs sampler, successively, at decreasing values of a temperature variable T. In this work, we have chosen a linear cooling schedule, i.e., $T^{(t)} = \delta \, T^{(t-1)}$, where δ is typically between 0.8 and 0.99. The parameters for the MRF models are chosen by trial and error and the optimization is done through sampling the configuration spaces Z and S alternately. The details of the optimization process can be found in [41].

It is interesting to note the effect of the value of p in (12) in super-resolving the fields. For an upsampling factor of q, one requires estimating $2q^2$ parameters (intensity and depth values) per pixel. Hence, one would ideally like to have $p \geq 2q^2$. However, this would be tantamount to using only the first (data fitting) term of the equation and it does not exploit the power of model-based restoration techniques. Due to the punctuated smoothness terms, one can obtain a very good estimate of both the fields, even when $p < 2q^2$. Imposition of the penalty for the line detection saves the algorithm from the possibility of excessive smoothing. This is demonstrated in the next section.

5 EXPERIMENTAL RESULTS

In this section, we illustrate the efficacy of the proposed algorithm for simultaneous super-resolved blur identification and image reconstruction through several examples of simulation and real data.

5.1 Simulation Experiments

We note that, since the blur is a function of depth, it suffices to recover the distribution of the blur parameter over the image. In all simulation experiments, only five low resolution images were considered ie, $p < 2q^2$ for $q = 2$. In addition to the space variant blur, each low resolution observation was also corrupted with additive white Gaussian noise of variance 5.0. The low resolution blur was estimated from any two of the five observations using the complex spectrogram method described in [18] which offers a good initial estimate of the blur with a very little computation. A square window of size 16×16 was used for the purpose. A zero order hold expansion of the estimated

low resolution blur yields the initial estimate of the high resolution blur. The bilinear interpolation of the least blurred image was chosen as the initial estimate of the true focused image. The number of discrete levels for the space variant blur parameter (σ) was taken as 128. For the intensity process, 256 levels were used, which is the same as the CCD dynamic range. Further, in all experiments, we have used a upsampling factor of $q = 2$. The parameters involved in the simulated annealing algorithm while minimizing (12) are as follows:

- μ_s and μ_z—relative weights of the smoothness term for the space-variant blur parameter and the super-resolved image processes, respectively.
- γ_s and γ_z—penalty term for presence of discontinuities in blur and super-resolved image processes, respectively.
- θ_s and θ_z—thresholds for the line fields corresponding to the blur parameter and intensity processes, respectively.
- T_0—initial temperature.
- δ—cooling schedule (rate of cooling).
- σ_s and σ_z—variances of the Gaussian samplers for the blur and the intensity processes, respectively.

In order to carry out the simulation experiments, we adopt the following strategy: First, we consider the image to have been taken with a pin-hole camera, implying that we obtain a focused image of the scene. Next, we assign an arbitrary depth map to the scene. Since the depth at a point in the scene is a function of the amount of blur at that point, the depth and the blur are deemed to be analogous. Using the space varying blur, we carry out a space varying convolution with the scene map to obtain a defocused image. What we are doing, in essence, is mapping a texture to a particular depth map. Appropriate noise sequences are added to obtain the observations.

Figs. 3a and 3b show two of the five low resolution observations of the Taj image. In general, we display two of the least blurred observations during the simulation experiments. The blur parameter in these defocused images

Fig. 11. The super-resolution image of the blocks world.

is related through $\sigma_{m+1}(i,j) = 0.75\,\sigma_m(i,j)$, $m = 1, \ldots, 4$. We recall that such a linear relationship between the blurs exists when defocused images of a scene are obtained using different values of the camera parameters. The true high resolution blur $\sigma(k,l) = a \exp(-\frac{(k-64)^2 + (l-64)^2}{2b^2})$, $0 \leq k, l < 128$, which is plotted in Fig. 3c. In our experiment, the values are $a = 1.0$ and $b = 35.0$. As mentioned earlier, we have chosen a decimation factor of $q = 2$. The original Taj image is blurred using the space-varying Gaussian blurring kernels formed from the blurs given above and then sampled down. White Gaussian noise is now added to the observations. The estimated values of the super-resolved blur parameters are shown in Fig. 3d. The values of the parameters used in the simulated annealing (SA) algorithm are $\mu_s = 1,000.0$, $\gamma_s = 15.0$, $\theta_s = 0.15$, $\sigma_s = 1.2$, $\mu_z = 0.005$, $\gamma_z = 5.0$, $\theta_z = 25.0$, $\sigma_z = 3.0$, $\delta = 0.975$, and $T_0 = 3.0$. The rms error in the estimate of the blur is only 0.033. It is to be noted here that no attempt has been made in this study to obtain the best parameter set for the optimization purpose. The algorithm has been able to determine the super-resolved blur parameter quite accurately. The super-resolved Taj image is shown in Fig. 3f. Compare this to the original high resolution Taj image given in Fig. 3e. We also observe that the quality of the estimated super-resolved image is also good, especially in the region of the main entrance and near the minarets. Note that the technique has

worked well even in the case of a fairly nontextured image such as the Taj image.

We next consider a blur profile in the form of a ramp function. The blur varies linearly from a value of 0.02 at the left edge of the image to 0.97 at the right edge. Two of the least blurred low resolution Graveyard images generated using our observation model are shown in Figs. 4a and 4b. Once again the blurs are related through $\sigma_{m+1}(i,j) = 0.75\,\sigma_m(i,j)$, $m = 1, \ldots, 4$, and the true high resolution blur $\sigma(k,l)$ is plotted in Fig. 4c. As before, the initial estimate of the super-resolved blur is the zero order hold expansion of the low resolution estimate of the blur determined using the complex spectrogram method. The bilinearly interpolated, least blurred image is the initial estimate of the super-resolved focused image. The super-resolved blur parameters and the super-resolved Graveyard image are shown in Figs. 4d and 4f, respectively. The rms error in the estimate for the super-resolved blur is again only 0.019, yielding an average estimation error of about 4 percent. The parameters of the SA algorithm are kept the same as in the last experiment. We observe that the degradations in the observations are eliminated in the super-resolved image. Note that steps in the far end are now clearly visible. Compare the reconstructed image with the original high resolution image given in Fig. 4e.

In both the above experiments, the field representing the blur process can be assumed to be sufficiently smooth so as to preclude the use of line fields for the blurring process altogether. In order to see how the method performs for discontinuities in the blur process, we consider next the case
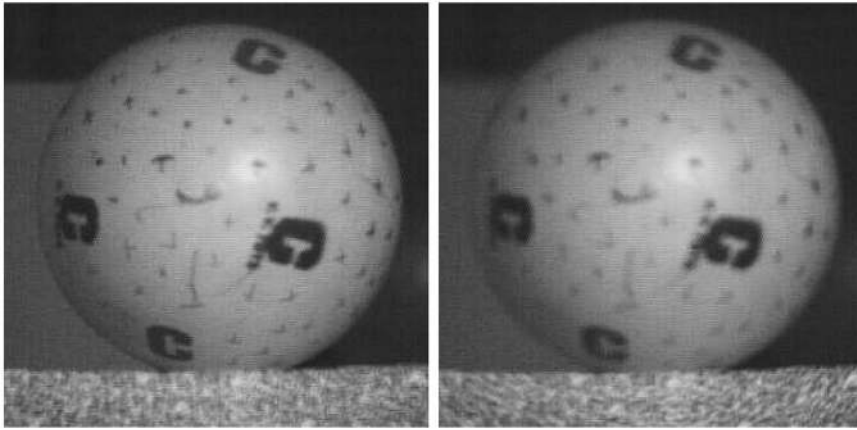
Fig. 12. Two of the low resolution Ball images.

where the blurring is constant over a certain contiguous region of the image and then varies linearly over a second region and, finally, is constant again over the remaining part of the image. Two such blurred observations of the Sail image are shown in Figs. 5a and 5b. The true depth profile is shown in Fig. 5c, where the constant blur regions are of widths 32 pixels each. The estimated super-resolved blur parameters are shown in Fig. 5d and the super-resolved image in Fig. 5e. The rms error in the estimation of super-resolved blurs is only 0.020. The super-resolved image recovery technique has performed quite well. The numerals on the sail as well as the thin lines are clearly discernible. Two sailors are clearly identifiable in the super-resolved image, which compares very favorably with the original picture given in Fig. 5e.

For a higher degree of discontinuity, we consider a step profile for the variation in blur/depth in the same Sail image. Two of the defocused sail images resulting from the space varying convolution of the step form of blur variation with the original scene map are shown in Figs. 6a and 6b. The true variation of blur is plotted in Fig. 6c. The estimated super-resolved blur and image are shown in Figs. 6d and 6e, respectively. Since the blur variation is highly discontinuous, we observed slightly reduced values of $\mu_s = 500$ and $\gamma_s = 10$ (i.e., less demand for smoothness and lowering the penalty for introducing a discontinuity in the depth field). The rms error in the blur estimates is 0.068, which is slightly on the higher side compared to the previous cases. Still the image reconstruction is very good.

Next, we present the results of our technique on low resolution observations of a Text image. The purpose of the experimentation is to subjectively judge the improvement in readability after the super-resolution restoration. Each observation of size 41×207 is blurred by a space varying blur, which has a similar variation as in the previous example, viz., step variation. Two of the five low resolution images are shown in Figs. 7a and 7b. The true high resolution blur parameters are shown in Fig. 7c. Due to the step-like variation in the blur profile, we notice the text getting progressively blurred from the left edge to the right edge of the input images. The estimated super-resolved blur parameters and the super-resolved text image are shown in Figs. 7d and 7e, respectively. The rms error for the blur

parameters in this case is 0.051. The same parameter set used in the previous experiment is used for the optimization purpose. Since the image field is also very discontinuous, a similar change in the values of μ_z and γ_z tends to yield partly improved results. The super-resolved blur recovery is very encouraging. The text in the super-resolved image is easily readable. All these experiments substantiate our claim that both these fields can, indeed, be super-resolved.

We also experimented on the efficacy of the proposed technique by varying the number of available observations (p). It was found that, with further increase in p, there is barely any improvement in the quality of the restored fields. However, when we reduced the number of observations, there was some degradation in the quality of the output. The improvement tends to saturate for a value of $p = 4$ or 5 when $q = 2$.

Finally, we test our algorithm on data sets containing ground truth of the depth values in the image. Specifically, we used the Buddha image from the SAMPL database of Ohio State University, which contains the range image as well as the actual intensity image. Low resolution observations were generated by defocusing the original intensity image with blurs that were calculated from the ground truth depth values and the camera parameters used in the experiments on real data (see Section 5.2 below). Two of the five low resolution images are shown in Figs. 8a and 8b. Also shown in Fig. 8c are the true depth values obtained from the range scanner. The estimated super-resolved depth map and the super-resolved intensity image are shown in Figs. 8d and 8e, respectively. The rms error for the depth estimate in this case is 9.72, yielding a ranging error less than just 1 percent. The super-resolved depth recovery is quite satisfactory except around the region of the mouth, where a small amount of detail appears to be lost. Also, the super-resolved intensity image of Buddha has been recovered quite well.

5.2 Experiments with Real Data

The performance of the proposed method was next tested on real data captured in our laboratory under controlled conditions. The first experimental setup was the "blocks world," where three concrete blocks were arranged at different depths, the nearest one at a distance of 73 cm,
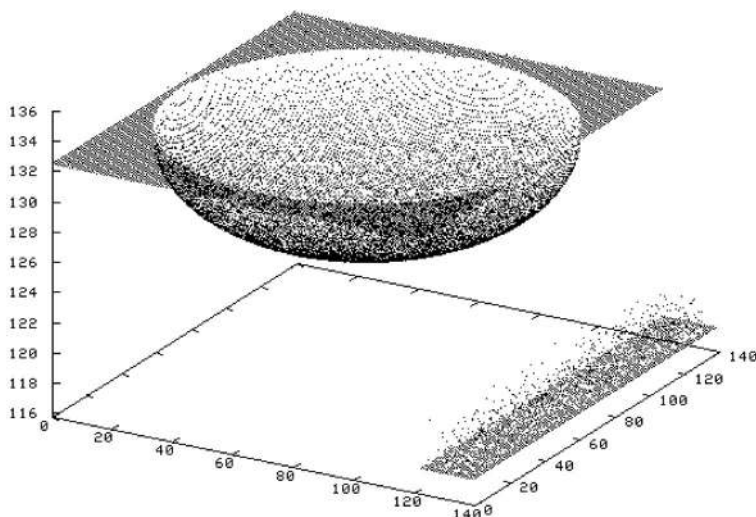
Fig. 13. The super-resolved depth map of the Ball image (only one out of every four points is plotted in order to avoid clutter).

another at 82.7 cm, and the farthest block at 96.6 cm. All the blocks are placed perpendicular to the optical axis of the camera and, hence, there is no depth variation for a particular face of a block. Newspaper cuttings were pasted on the blocks to provide some texture as well as for ease of evaluation. A Pulnix CCD camera fitted with a Fujinon HF35A-2 lens of focal length 3.5 cm was used to grab the images. The lens aperture was kept at F1.7. The camera was coarsely calibrated using an object at a known depth. Five low resolution images, each of size 240×260, were captured. These are shown in Fig. 9. Depending on the selection of lens-to-image plane distances, one obtains different amounts of defocus in different observations. The estimated super-resolved depths are shown in Fig. 10 and the super-resolved image is shown in Fig. 11. As we can see, the proposed method has been able to capture the depth variation quite satisfactorily. The root mean square error in the estimation of depth is 1.768 cm, which is equivalent to a ranging error of just 1.96 percent. Also, the super-resolved image has been recovered quite well as is evident from the readability of the text on both the blocks. This is not so in the captured images where the texts on either or both the blocks are always out-of-focus. The random dot pattern pasted on the lower block has also been recovered well.

The second experimental setup consisted of a ball resting on a block. The selection of the ball as the scene was motivated primarily to verify the performance of the proposed method when the scene does not have much textural content. The block was at a distance of 117 cm from the camera. The point on the ball nearest to the camera was at 121.8 cm, while the farthest points, viz., the points lying on the occluding boundary of the ball, were at 132.3 cm from the camera. Two of the five low resolution images each of size 280×280 are shown in Fig. 12. We have changed the lens-to-image plane distance in our experiments to obtain the differently defocused observations. This introduces a small amount of changes in magnification in successive observations. We neglect the effect in our studies. The super-resolved depth is shown in Fig. 13 in which one out of every four points is plotted in order to avoid clutter. The super-resolved Ball image is shown in

Fig. 14. The proposed technique has been able to capture the spherical depth variation well together with the depth of the block on which the ball is resting. The restored image is also of very good quality as the characters on the spherical surface are clearly visible.

All these experiments now substantiate our claim that it is indeed possible to super-resolve both the intensity and depth maps using the depth related defocus as a natural cue.

6   CONCLUSIONS

We have described a MAP-MRF framework for simultaneously generating the super-resolved depth map and the super-resolved image from low resolution defocused observations. This method avoids the correspondence and warping problems inherent in current super-resolution techniques involving the motion cue in the low resolution observations and uses a more natural depth related defocus as a natural cue in real aperture imaging. Both the super-resolved blur parameter and the image are modeled as separate MRFs. It is interesting to note that a large class of problems in computer vision, such as DFD, super-resolution, optical flow, shape from shading, etc., can all be solved in a similar MAP framework. The basic structure of the solution remains the same. The equations of image formation are written in conjunction with appropriate priors and the solution is obtained by optimizing the resultant energy function. The proposed method is no exception.

The use of line fields preserves discontinuities in the super-resolved depth and image fields. We have chosen the line fields to be binary variables in this study. However, one can use continuous variables as well without much changing the problem formulation. The advantage of using continuous variable line fields lies in having a differentiable cost function when a gradient-based optimization method can be used. The super-resolved depth maps have been generated with a very high accuracy. The quality of the super-resolved images is also quite good. Through the restoration process, we are able to obtain a pin-hole equivalent image (i.e., there is no depth of field) of an
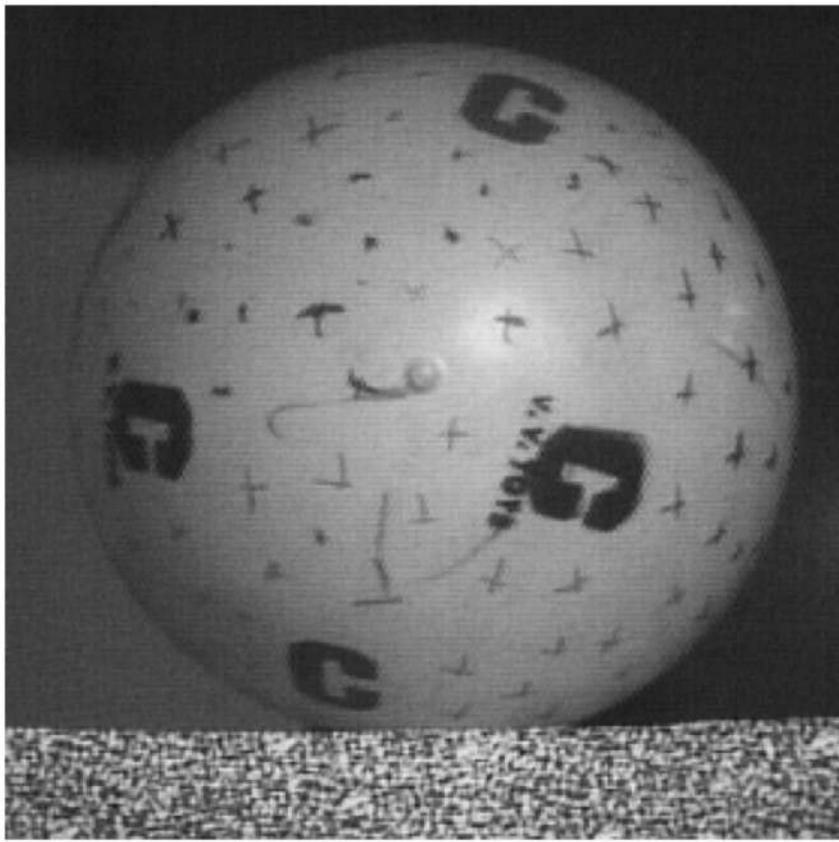
Fig. 14. The super-resolution Ball image.

arbitrary scene using commercially available real-aperture cameras. Our future research effort would involve integrating additional cues from multiple cameras to obtain even better estimates of both the fields.

ACKNOWLEDGMENTS

REFERENCES

[1] T. Komatsu, T. Igarashi, K. Aizawa, and T. Saito, "Very High Resolution Imaging Scheme with Multiple Different-Aperture Cameras," *Signal Processing: Image Comm.,* vol. 5, pp. 511-526, Dec. 1993.

[2] R.Y. Tsai and T.S. Huang, "Multiframe Image Restoration and Registration," *Advances in Computer Vision and Image Processsing,* pp. 317-339, JAI Press Inc., 1984.

[3] M. Irani and S. Peleg, "Improving Resolution by Image Registration," *Computer Vision, Graphics, and Image Processing: Graphical Models and Image Processing,* vol. 53, pp. 231-239, Mar. 1991.

[4] H Shekarforoush, M. Berthod, J. Zerubia, and M. Werman, "Sub-Pixel Bayesian Estimation of Albedo and Height," *Int'l J. Computer Vision,* vol. 19, no. 3, pp. 289-300, 1996.

[5] R.R. Schultz and R.L. Stevenson, "Extraction of High-Resolution Frames from Video Sequences," *IEEE Trans. Image Processing,* vol. 5, pp. 996-1011, June 1996.

[6] S. Chaudhuri, *Super-Resolution Imaging.* Boston: Kluwer Academic, 2001.

[7] A.P. Pentland, "Depth of Scene from Depth of Field," *Proc. Image Understanding Workshop,* pp. 253-259, 1982.

[8] S. Chaudhuri and A.N. Rajagopalan, *Depth from Defocused Images: A Real Aperture Imaging Approach.* New York: Springer-Verlag, 1999.

[9] M. Subbarao, "Efficient Depth Recovery Through Inverse Optics," *Machine Vision for Inspection and Measurement,* H. Freeman, ed., Academic Press, 1989.

[10] A.P. Pentland, "A New Sense for Depth of Field," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 9, no. 4, pp. 523-531, July 1987.

[11] M. Subbarao, "Parallel Depth Recovery by Changing Camera Parameters," *Proc. IEEE Conf. Vision and Pattern Recognition,* pp. 149-155, 1988.

[12] K.V. Prasad, R.J. Mammone, and J. Yogeshwar, "3D Image Restoration Using Constrained Optimization Techniques," *Optical Eng.,* vol. 29, no. 4, pp. 277-288, Apr. 1990.

[13] W.N. Klarquist, W.S. Geisler, and A.C. Bovik, "Maximum-Likelihood Depth-from-Defocus for Active Vision," *Proc. IEEE Int'l Conf. Intelligent Robots and Systems,* 1995.

[14] M. Gökstorp, "Computing Depth from Out-of-Focus Blur Using a Local Frequency Representation," *Proc. Int'l Conf. Pattern Recognition,* pp. 153-158, 1994.

[15] M. Watanabe and S.K. Nayar, "Minimal Operator Set for Passive DFD," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition,* pp. 153-158, 1996.

[16] Y.Y. Schechner and N. Kiryati, "Depth from Defocus vs. Stereo: How Different Really Are They?" Technical Report EE PUB No. 1155, Technion—Israel Inst. of Technology, May 1998.

[17] S.K. Nayar, M. Watanabe, and M. Noguchi, "Real-Time Focus Range Sensor," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 12, pp. 1186-1198, Dec. 1996.

[18] A.N. Rajagopalan and S. Chaudhuri, "Space-Variant Approaches to Recovery of Depth from Defocused Images," *Computer Vision and Image Understanding,* vol. 68, no. 3, pp. 309-329, Dec. 1997.

[19] A.N. Rajagopalan and S. Chaudhuri, "A Variational Approach to Recovering Depth from Defocused Images," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 10, pp. 1158-1165, Oct. 1997.

[20] A.N. Rajagopalan and S. Chaudhuri, "An MRF Model Based Approach to Simultaneous Recovery of Depth and Restoration from Defocused Images," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 7, pp. 577-589, July 1999.

[21] M.K. Ozkan, A.M. Tekalp, and M.I. Sezan, "POCS Based Restoration of Space Varying Blurred Images," *IEEE Trans. Image Processing,* vol. 3, no. 4, pp. 450-454, July 1994.

[22] S. Koch, H. Kaufmann, and J. Biemond, "Restoration of Spatially Varying Blurred Images Using Multiple Model-Based Extended Kalman Filter," *IEEE Trans. Image Processing,* vol. 4, no. 4, pp. 520-523, Apr. 1995.

[23] H. Trussel and S. Fogel, "Identification and Restoration of Spatially Variant Motion Blurs in Sequential Images," *IEEE Trans. Image Processing,* vol. 1, no. 1, pp. 123-126, Jan. 1992.

[24] S.P. Kim, N.K. Bose, and H.M. Valenzuela, "Recursive Reconstruction of High Resolution Image from Noisy Undersampled Multiframes," *IEEE Trans. Accoustics, Speech, and Signal Processing,* vol. 18, no. 6, pp. 1013-1027, June 1990.

[25] C. Srinivas, M.D. Srinath, "A Stochastic Model Based Approach for Simultaneous Restoration of Multiple Mis-Registered Images," *Int'l Soc. Optical Eng.,* vol. 1360, pp. 1416-1427, 1990.

[26] H. Ur and D. Gross, "Improved Resolution from Sub-Pixel Shifted Pictures," *Computer Vision, Graphics, and Image Processing: Graphical Models and Image Processing,* vol. 54, pp. 181-186, Mar. 1992.

[27] A.M. Tekalp, M.K. Ozkan, and M.I. Sezan, "High Resolution Image Reconstruction from Lower-Resolution Image Sequences and Space-Varying Image Restoration," *Proc. Int'l Conf. Acoustics Speech and Signal Processing,* pp. 169-172, 1992.

[28] R.R. Schultz and R.L. Stevenson, "A Bayesian Approach to Image Expansion for Improved Definition," *IEEE Trans. Image Processing,* vol. 3, no. 3, pp. 233-242, May 1994.

[29] M. Elad and A. Feuer, "Restoration of a Single Super-Resolution Image from Several Blurred, Noisy and Undersampled Measured Images," *IEEE Trans. Image Processing,* vol. 6, no. 12, pp. 1646-1658, Dec. 1997.

[30] M. Elad and A. Feuer, "Super-Resolution Restoration of an Image Sequence: Adaptive Filtering Approach," *IEEE Trans. Image Processing,* vol. 8, no. 3, pp. 387-395, Mar. 1999.

[31] M.-C. Chiang and T.E. Boult, "Local Blur Estimation and Super-Resolution," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 821-826, 1997.

[32] D. Rajan and S. Chaudhuri, "A Generalized Interpolation Scheme for Image Scaling and Super-Resolution," *Proc. Erlangen Workshop Vision, Modeling, and Visualization,* Univ. of Erlangen-Nuremberg, Germany, pp. 301-308, Nov. 1999.

[33] D. Rajan and S. Chaudhuri, "Generalized Interpolation and Its Application in Super-Resolution Imaging," *Image and Vision Computing,* vol. 19, no. 13, pp. 957-969, Nov. 2001.

[34] P. Cheeseman, B. Kanefsky, R. Kraft, J. Stutz, and R. Hanson, "Super-Resolved Surface Reconstruction from Multiple Images," Technical Report FIA-94-12, NASA Ames Research Center, Moffet Field, Calif., Dec. 1994.

[35] R.C. Dubes and A.K. Jain, "Random Field Models in Image Analysis," *J. Applied Statistics,* vol. 16, no. 2, pp. 131-164, 1989.

[36] J. Besag, "Spatial Interaction and the Statistical Analysis of Lattice Systems," *J. Royal Statistical Soc., Series B,* vol. 36, pp. 192-236, 1974.

[37] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Image," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 6, no. 6, pp. 721-741, 1984.

[38] A. Blake and A. Zisserman, *Visual Reconstruction.* MIT Press, 1987.

[39] J. Subrahmonia, Y.P Hung, and D.B. Cooper, "Model Based Segmentation and Estimation of 3D Surfaces from Two or More Intensity Images Using Markov Random Fields," *Proc. IEEE Int'l Conf. Pattern Recognition,* pp. 390-397, 1990.

[40] D.B. Cooper, J. Subrahmonia, Y.P. Hung, and B. Cernuschi-Frias, "The Use of Markov Random Fields in Estimating and Recognizing Objects in 3D Space," *Markov Random Fields: Theory and Applications,* R. Chellappa and A. Jain, eds. Academic Press, 1993.

[41] D. Rajan, "Some New Approaches to Generation of Super-Resolution Images," PhD thesis, School of Biomedical Eng., Indian Inst. of Technology, Bombay, 2001.

Deepu Rajan received the BE degree in electronics and communication engineering from Birla Institute of Technology, Ranchi, the MS degree in electrical engineering from Clemson University, and the PhD degree from the Indian Institute of Technology, Bombay. From April 1992 until May 2002, he was a lecturer in the Department of Electronics at the Cochin University of Science and Technology, India. Since June 2002, he has been an assistant professor in the School of Computer Engineering at Nanyang Technological University, Singapore. His research interests include image and video processing, computer vision, and neural networks.

Subhasis Chaudhuri received the BTech degree in electronics and electrical communication engineering from the Indian Institute of Technology (IIT), Kharagpur in 1985. He received the MS and PhD degrees, both in electrical engineering, from the University of Calgary, Canada and the University of California, San Diego, respectively. He joined the IIT, Bombay in 1990 as an assistant professor and is currently serving as a professor. He has also served as a visiting professor at the University of Erlangen-Nuremberg, Germany, and the University of Paris XI. He is a fellow of the Alexander von Humboldt Foundation, Germany, Indian National Academy of Engineering, and the IETE, India. He is the recipient of the Dr. Vikram Sarabhai Research Award for the year 2001. He is the coauthor of the book *Depth from Defocus: A Real Aperture Imaging Approach*, (Springer). He has also edited a book on super-resolution imaging, published by Kluwer Academic in 2001. His research interests include image processing, computer vision, and multimedia. He is a senior member of the IEEE.