Published in IET Computer Vision
Received on 1st December 2007
Revised on 23rd April 2008
doi: 10.1049/iet-cvi:20070072

In Special Issue on Visual Information Engineering

ISSN 1751-9632

Harnessing defocus blur to recover high-resolution information in shape-from-focus technique

R.R. Sahay A.N. Rajagopalan

Image Processing and Computer Vision Lab, Department of Electrical Engineering, Indian Institute of Technology, Madras, Chennai 600 036, India
E-mail: sahayiitm@gmail.com

Abstract: Traditional shape-from-focus (SFF) uses focus as the singular cue to derive the shape profile of a 3D object from a sequence of images. However, the stack of low-resolution (LR) observations is space-variantly blurred because of the finite depth of field of the camera. The authors propose to exploit the defocus information in the stack of LR images to obtain a super-resolved image as well as a high-resolution (HR) depth map of the underlying 3D object. Appropriate observation models are used to describe the image formation process in SFF. Local spatial dependencies of the intensities of pixels and their depth values are accounted for by modelling the HR image and the HR structure as independent Markov random fields. Taking as input the LR images from the stack and the LR depth map, the authors first obtain the super-resolved image of the 3D specimen and use it subsequently to reconstruct a HR depth profile of the object.

1 Introduction

One of the main objectives of computer vision is to extract structural information from images. Techniques used for computing shape can be classified broadly into two categories: passive and active. Although active methods rely on specialised illumination generated from an artificial light source, passive schemes use the ambient light energy. A wide range of cues have been exploited to infer shape of a 3D object from its 2D images. Under passive methods, with monocular images, structure can be derived using occlusion cues, texture gradient analysis [1, 2], photometric methods [3, 4], focus [5] and defocus [6] information. Multiple images captured with different relative positions of the camera are used in stereopsis [7] and structure-from-motion techniques [8]. Shape-from-focus (SFF) [5] is a passive ranging technique used to estimate the structure of a 3D object. It involves capturing a sequence of observations with different portions of the 3D object coming into focus in different images. The degree of focus is the only cue used to estimate shape. Depth maps of a 3D scene are useful in widely disparate areas such as medical imaging based applications [9, 10], video matting [11], 3DTV [12, 13] and depth image-based rendering. In these applications, one is also interested in viewing finer image details and subtle variations in the shape profile of the 3D specimen. However, in SFF, space-variant blurring degrades the high-frequency information in the captured images. Also, traditional SFF computes the structure of the 3D object at the same resolution as the observations in the stack. Hence, there exists a need to obtain a high-resolution (HR) image of the 3D specimen as well as its shape profile at a higher resolution than the captured observations.

The resolution of an image captured by a camera is limited by the resolution of the sensor array. HR cameras are expensive and the spatial density of photodetectors cannot be increased beyond a limit when shot noise begins affecting the images. Aliasing, blurring and noise degrade images captured with a low-resolution (LR) camera. However, if a sequence of LR images of the scene is available, using additional information embedded in such multiple LR images, super-resolution algorithms yield an HR image by performing de-aliasing, deblurring and denoising. Several cues in LR images have been explored by different works in the past for super-resolution. Relative motion between the camera and

the scene which gives rise to subpixel displacements among the LR images has been used in [14–16]. In contrast, motion-free super-resolution does not have any relative displacements between the LR observations. The classic paper of Papoulis [17] provides the theoretical foundation for the techniques used in motion-free super-resolution. Interestingly, it is shown in [18, 19] that blur in a sequence of LR images can be used as a cue for super-resolution. Recently, several researchers [20–22] have used learning-based algorithms for super-resolution. Here, statistical relationships between corresponding image regions in LR and HR images are learned during the training phase and these learned relationships are then used to predict minute details for enlarging other LR images.

In recent literature, researchers have attempted to recover the HR structure of 3D objects. However, most of these works fall in the category of active ranging using laser scanners [23–25]. Under passive methods, it has been shown in [26] that it is possible to enhance the spatial resolution of the recovered structure using photometric cues. In [19], the depth from defocus technique is used to simultaneously extract the HR image and the HR depth profile. To the best of our knowledge, there has been no prior attempt to estimate both the HR image and the HR depth profile in SFF.

We now make several observations which can be exploited to obtain the HR image and the structure of the 3D object at a higher spatial resolution in SFF. First, the rich information embedded in the space-variant blur of the stack of LR images is unutilised for shape reconstruction. Second, in SFF, the sum modified Laplacian (SML) operator [5] is used to obtain the depth estimates of every point independent of the depth values of the neighbouring points. Third, we note that real-world objects generally have smooth depth profiles, and hence, there exist local spatial dependencies of depth values. Traditional SFF does not take this factor into account. Moreover, the behaviour of the focus measure profile is sensitive to the texture in the local region of the object. This results in erroneous estimates of depth at smooth regions where the scene lacks texture. Fourth, we observe that the depth estimates in traditional SFF are obtained by Gaussian interpolation of a few values near the peak of the focus measure profile, which can also introduce errors.

In this paper, we expand the scope of the traditional SFF method to enable reconstruction of an HR image of the underlying 3D object as well as the HR depth map, given its LR depth profile and a few LR observations chosen from the stack. As in motion-free super-resolution, we take advantage of the natural defocus cue in the stack of LR observations to perform super-resolution. To incorporate the local smoothness of the greyscale intensity values and the depth values in the reconstruction of the HR image and the HR depth map, we model them as separate and independent Markov random fields (MRFs). Depth computation in traditional SFF is vulnerable to low-texture regions in the images of the stack. However, since we model the HR image and the HR depth profile of the 3D specimen with MRFs, we are able to incorporate spatial dependencies and hence mitigate this effect provided the scene lacks texture only locally. An algorithm which computes the HR image and the HR depth map simultaneously would be computationally very expensive. In SFF, the depth map of the 3D object can be constructed at the same resolution as the captured observations in the stack. This LR depth map can be used to obtain the HR image in the first step, which is subsequently used to obtain the HR depth map of the underlying 3D object. Thus, we can super-resolve the focused image of the 3D specimen and its depth map in separate steps. We use the same LR observations from the stack to compute the HR image and the HR depth map. Since, in SFF, a large number of LR observations are readily available, we attempt super-resolution by higher magnification factors by using more images from the stack in the proposed framework.

2 Defocus blur, focus measure and LR depth map

Blurring occurs naturally because of the point spread function (PSF) of a real aperture camera. In the literature, the PSF of a camera is usually modelled by a 2D Gaussian function [27, 28] with standard deviation σ (also called the blur parameter). A real aperture camera follows the lens law and thus brings object portions at a certain depth into focus. A 3D object induces space-variant blurring and different points on it will come into focus at different depths from the camera lens. The spatial distribution of the blur in relation to the depth at a point on the 3D object is given by $\sigma = \rho R v((1/w_{\mathrm{d}}) - (1/D))$ where w_{d} is the working distance of the camera, D the distance of the object point from the lens, v the distance of the lens from the image plane, R the radius of the aperture of the lens and ρ a camera constant.

The basic idea behind the SFF method is to locate the respective positions of best focus of all points on the 3D object as it is translated relative to a real aperture camera, in order to derive its shape. The working principle of the traditional SFF scheme [5] is shown in Fig. 1. The 3D object, whose shape is to be estimated, is placed on a translational stage which moves in the vertical direction in steps of Δd. The initial position of the stage is denoted by the reference plane. The optics of the camera defines a 'focused plane' wherein all the points will be perfectly focused on the sensor plane. The distance of the focused plane from the lens plane is w_{d}. The separation between the reference and the focused planes is d_{r}. The distance between the translating stage and the reference plane is d. Distances w_{d}, d_{r} and d are known. As the object is moved vertically, at each step an image is captured in which barring a small portion of the object, other regions are

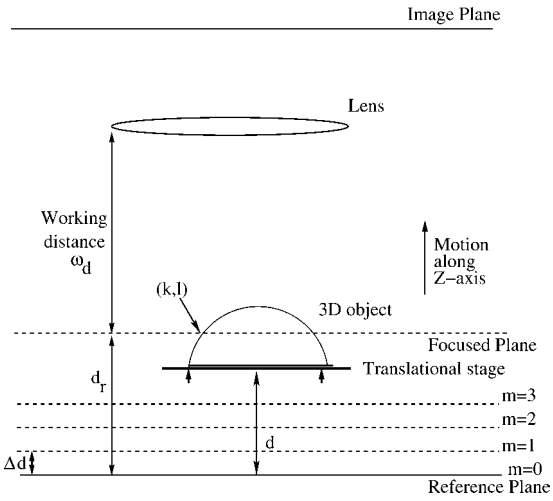

Figure 1 *Schematic of traditional SFF*

defocused to varying degrees because a real aperture camera cannot bring all the points of a 3D object into focus at the same time.

When the 3D object is placed on the translational stage which is at the reference plane as shown in Fig. 1, the blur induced by a point on the object in the reference frame is governed by blur parameter σ_0 which is given by $\sigma_0 = \rho Rv((1/w_d) - (1/D_0))$ where D_0 is the distance of the object point from the lens when the stage is at the reference position and w_d is the working distance of the camera, that is, $1/w_d = (1/f) - (1/v)$. The stage is moved vertically by a distance of $m\Delta d$ to capture the mth LR frame. For the same point on the 3D object, the blurring induced in the mth frame can be expressed by the blur parameter σ_m which is given by $\sigma_m = \rho Rv((1/w_d) - 1/(D_0 \pm m\Delta d))$. The change in magnification across the stack of LR observations is assumed to be negligible so that there are no errors because of registration. Eliminating the common term w_d from the above expressions for σ_m and σ_0 we obtain

$$\sigma_m = \sigma_0 + \rho Rv\left(\frac{1}{D_0} - \frac{1}{D_0 \pm m\Delta d}\right) \tag{1}$$

The product ρRv, in the above equation can be found using an appropriate calibration procedure.

As the translating stage is moved vertically in steps of Δd, for the mth frame we can also express the blur parameter for a 3D point whose image pixel coordinates are (k, l) as $\sigma_m(k, l) = \rho Rv((1/w_d) - (1/(w_d - m\Delta d + \overline{d}(k, l))))$. Here, $\overline{d}(k, l)$ is the distance by which the stage must be moved from the reference plane to bring point (k, l) into focus. If the separation between the stage and the reference plane is such that $d = \overline{d}(k, l)$, then this point satisfies lens law, and will appear in perfect focus. Note that under this condition, the blur parameter becomes zero, as expected (assuming the effects of optical aberrations to be negligible).

In order to find where a point on the 3D specimen comes into focus, in traditional SFF, a focus measure profile [5] is computed, using the SML focus measure operator, for every pixel in the image sequence as the translational stage moves in finite steps of Δd. The focus measure at a point (k, l) in an image I is usually computed using the SML operator defined as

$$F(k, l) = \sum_{m=k-W}^{k+W} \sum_{n=l-W}^{l+W} O_L(m, n) \quad \text{for } O_L(m, n) \geq T_1 \tag{2}$$

where T_1 is a threshold value, $2W+1$ the size of the window around the point (k, l) and O_L the modified Laplacian defined in the discrete domain as $O_L(m, n) = |2I(m, n) - I(m - \alpha, n) - I(m + \alpha, n)| + |2I(m, n) - I(m, n - \alpha) - I(m, n + \alpha)|$. Here, α is a variable spacing between the image pixels used to compute the derivatives. For all our experiments, we chose the value of α as 1. The focus measure profile for the pixel at (k, l) is obtained by plotting the value of $F(k, l)$ computed for every image captured in the stack of observations starting from the reference frame. The final estimate of the depth $\overline{d}$ at (k, l) is arrived at by using Gaussian interpolation of a few values near the peak value of the focus measure profile. It is to be noted that the blurring induced at the point (k, l) is zero when the translating stage is at a height of $\overline{d}$ from the reference plane. Thus, the focus measure profile is helpful in knowing when the point (k, l) comes into focus. The values of $\overline{d}$ calculated for all the points of the 3D object and denoted as $\overline{\boldsymbol{d}}$ yields the LR depth map, that is, the shape of the object is at the same spatial resolution as the captured frames in the stack.

3 Image super-resolution

In this section, we explore the possibility of obtaining an image of the underlying 3D object in SFF in which all pixels are simultaneously in focus and the spatial resolution is higher than the frames of the stack. The idea is to use the defocus information in the captured images of the 3D object as a cue for computing the HR image.

Let there be p number of LR observations $\{y_m(i, j)\}$, each of size $M \times M$ which are decimated, blurred and noisy versions of a single HR image $\{x(m, n)\}$ of size $qM \times qM$. If $\boldsymbol{y}_m$ is the lexicographically arranged vector containing pixels from the mth LR image of size $M^2 \times 1$ and $\boldsymbol{x}$ is the lexicographically arranged vector containing pixels from the HR image of size $q^2M^2 \times 1$ then they can be related in the following ways according to two commonly used observation models in super-resolution. According to one model used for super-resolution [29]

$$\boldsymbol{y}_m = \boldsymbol{D}\boldsymbol{H}_m\boldsymbol{x} + \boldsymbol{n}_m, \quad m = 1, \ldots, p \tag{3}$$

where $\boldsymbol{H}_m$ is the blur matrix of size $q^2M^2 \times q^2M^2$, $\boldsymbol{D}$ the decimation matrix of size $M^2 \times q^2M^2$ and $\boldsymbol{n}_m$ the noise

doi: 10.1049/iet-cvi:20070072

vector of size $M^2 \times 1$, assumed to be zero mean Gaussian with variance σ_η^2.

Another model that is also used in motion-free super-resolution [18, 29] is given as

$$\boldsymbol{y}_m = \boldsymbol{H}_m \boldsymbol{D}\boldsymbol{x} + \boldsymbol{n}_m, \quad m = 1, \ldots, p \tag{4}$$

where $\boldsymbol{H}_m$ is the blur matrix of size $M^2 \times M^2$, $\boldsymbol{D}$ the decimation matrix of size $M^2 \times q^2M^2$ and $\boldsymbol{n}_m$ the noise vector of size $M^2 \times 1$. The observation noise is assumed to be zero mean Gaussian with variance σ_η^2.

To recover the HR image and the HR depth map in SFF, we exploit both these models to our advantage. We observe that by using the degradation model of (4), it is possible to obtain the HR image first. In the previous section, we outlined the method of traditional SFF to obtain the LR depth map. The blur parameter at every point of the object in the reference frame image can be computed from this recovered LR depth map. Using the relationship in (1), the blur parameter σ at each point of a defocused image, at any position in the stack of frames, can then be computed. The blur matrix $\boldsymbol{H}_m$ can now be constructed from the depth map obtained using SFF and the relation between the blur parameter σ at each pixel across the stack of LR frames in (1). The availability of the LR depth map in SFF enables us to separate the tasks of super resolving the intensity field and obtaining the HR depth map. Thus, we avoid a simultaneous estimation of the super-resolved focused image and the HR depth map, reducing significantly the computational load.

3.1 MAP solution

The problem of reconstructing the HR image $\boldsymbol{x}$ is an ill-posed inverse problem and some form of regularisation to constrain the solution space is necessary. We propose to derive an optimal estimate of the HR image as the maximum a posteriori (MAP) estimate given by

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} P(\boldsymbol{x}|\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_p) \tag{5}$$

The MAP framework allows us to impose a priori constraints on the HR image. Since statistical models can encode contextual constraints in images in a natural way, we model the original HR image as a MRF [30]. Specifically, we model it as a Gauss Markov random field (GMRF). From the Hammersley–Clifford theorem the joint distribution is Gibbsian [31]. If the HR image is modelled as a GMRF, then the prior distribution is given by

$$P(\boldsymbol{x}) = \frac{1}{Z}\exp\left[-\sum_{c\in C} V_c(\boldsymbol{x})\right] = \frac{1}{Z}\exp\left[-\sum_{c\in C} (\boldsymbol{d}_c^{\mathrm{T}}\boldsymbol{x})^2\right] \tag{6}$$

where Z is the partition function, c a clique, C the set of all cliques, $V_c(\cdot)$ the clique potential function, $\boldsymbol{d}_c$ a coefficient vector for clique c such that $\boldsymbol{d}_c^{\mathrm{T}}\boldsymbol{x}$ provides a measure of smoothness of the image by computing discrete approximations for first or second derivatives at each image pixel.

The prior distribution of the HR image can also be written as

$$P(\boldsymbol{x}) = \frac{1}{(2\pi)^{N^2/2}|(1/\lambda_1)R_{\boldsymbol{x}}|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\left(\frac{R_{\boldsymbol{x}}}{\lambda_1}\right)^{-1}\boldsymbol{x}\right\} \tag{7}$$

The matrix $R_{\boldsymbol{x}}^{-1} = \boldsymbol{D}_r^{\mathrm{T}}\boldsymbol{D}_r$ where $\boldsymbol{D}_r$ represents a one-step forward difference operator. Matrix $\boldsymbol{D}_r^{\mathrm{T}}\boldsymbol{D}_r$ can be approximated as a circulant matrix which is the Laplacian operator and is related to the operator $\boldsymbol{d}_c$ discussed in (6) as $\sum_{c\in C}(\boldsymbol{d}_c^{\mathrm{T}}\boldsymbol{x})^2 = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{D}_r^{\mathrm{T}}\boldsymbol{D}_r\boldsymbol{x}$.

Using the degradation model in (4) and assuming that the the noise processes are independent, we have

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}}\left[\sum_{m=1}^{p}\frac{\|\boldsymbol{y}_m - \boldsymbol{H}_m\boldsymbol{D}\boldsymbol{x}\|^2}{2\sigma_\eta^2} + \lambda_1\sum_{c\in C}(\boldsymbol{d}_c^{\mathrm{T}}\boldsymbol{x})^2\right] \tag{8}$$

where a first-order MRF neighbourhood has been assumed. Here, λ_1 is the regularisation factor which is typically tuned to derive the best estimate of $\boldsymbol{x}$. The number of LR observations, p, used in the objective function depends upon the upsampling factor. For super-resolution by a factor of q, the number of frames chosen from the stack is q^2. We remark that in SFF, since a large number of LR frames are naturally available in the stack, we can attempt super-resolution by higher factors. The above cost function is convex and hence a simple minimisation technique such as gradient descent can be applied to find the estimate of $\boldsymbol{x}$. At the nth iteration, the gradient of the cost function is given by

$$\mathrm{grad}^{(n)} = \frac{1}{\sigma_\eta^2}\sum_{m=1}^{p}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{H}_m^{\mathrm{T}}(\boldsymbol{H}_m\boldsymbol{D}\boldsymbol{x}^{(n)} - \boldsymbol{y}_m) + \lambda_1 Q^{(n)} \tag{9}$$

where $Q^{(n)} = \sum_{i=1}^{qM}\sum_{j=1}^{qM} 2[4x^{(n)}(i,j) - x^{(n)}(i,j-1) - x^{(n)}(i,j+1) - x^{(n)}(i-1,j) - x^{(n)}(i+1,j)]$

One can use a more complicated model such as the discontinuity adaptive MRF but a carefully chosen low value of λ_1 in the GMRF model will preserve edges reasonably well. Matrix $\boldsymbol{D}^{\mathrm{T}}$ spreads equally the LR pixel intensity value at corresponding pixel locations in the HR image. Matrix $\boldsymbol{H}_m$ is computed using the LR depth map obtained by the traditional SFF method [5] and the relation among the σ values across the stack of LR images given in (1). The estimate of the HR image at the $(n+1)$th iteration is obtained as $\boldsymbol{x}^{(n+1)} = \boldsymbol{x}^{(n)} - \beta\,\mathrm{grad}^{(n)}$ where β is the step size. The iterations continue until $\|\boldsymbol{x}^{(n+1)} - \boldsymbol{x}^{(n)}\| <$ threshold.

4 HR depth map

The fundamental motivation for obtaining the HR structure is to enable visibility of finer details of the shape of the 3D object. Recently, there has been a need to obtain the HR depth map in applications such as video matting [11]. Laser scanners which produce high-quality range maps are expensive and slow. The ability to obtain an HR depth map in near real-time could aid several applications. In medical applications, an HR depth map could prove useful in creating virtual reality-based models of complex surgical procedures [9, 10]. An accurate HR depth map could also benefit non-invasive diagnostic procedures in automated vision-based inspection applications. In graphics and animation, HR depth maps can aid in realistic depiction of the scene.

The LR depth map of the 3D object computed in traditional SFF suffers from errors because of various reasons. This method does not take into account spatial dependencies of the depth estimates. It is expected that natural objects have depth variations which are locally smooth. However, in SFF, depth estimates are computed using the SML focus measure operator, independent of depth values of neighbouring points. The behaviour of the focus measure profile is dependent on local texture and can yield erroneous shape estimates at smooth regions. Gaussian interpolation of a few values near the peak of the focus measure profile is used to obtain the final shape estimate; but this is not an ideal way to fill up missing information. Note that data from the captured frames are not used to guide the estimation of depth in SFF. In this work, we seek to obtain an HR depth map of the 3D object. As described in Section 2, a point on the 3D object comes into focus when the separation between the translating stage and the reference plane is $\overline{d}$. The structure of the 3D object, in SFF, is computed by estimating the quantity $\overline{d}$ for all points on the specimen. The spatial resolution of the depth profile, $\overline{\boldsymbol{d}}$, estimated in SFF, is the same as the LR images of the stack. Here, we choose to obtain the depth profile at a higher spatial resolution.

Consider p number of LR observations $\{y_m(i,j)\}$, each of size $M \times M$ which are decimated, blurred noisy versions of a single HR image $\{x(m, n)\}$ of size $qM \times qM$, obtained by the method outlined in the previous section. If $\boldsymbol{y}_m$ is the lexicographically arranged vector containing pixels from the mth LR image of size $M^2 \times 1$ and $\boldsymbol{x}$ is the lexicographically arranged vector containing pixels from the HR image of size $q^2M^2 \times 1$ then the degradation model that we use to obtain the HR depth map is given by

$$\boldsymbol{y}_m = \boldsymbol{D}\boldsymbol{H}_m(\overline{\boldsymbol{d}})\boldsymbol{x} + \boldsymbol{n}_m, \quad m = 1, \ldots, p \tag{10}$$

where $\boldsymbol{H}_m(\overline{\boldsymbol{d}})$ is the blur matrix of size $q^2M^2 \times q^2M^2$, $\boldsymbol{D}$ the decimation matrix of size $M^2 \times q^2M^2$ and $\boldsymbol{n}_m$ the zero mean noise vector of size $M^2 \times 1$. The observation noise is assumed to be zero mean Gaussian with variance σ_η^2. Since we seek to compute the HR depth map, we use the above degradation model, where the HR blur map acts on the HR image.

Construction of the HR depth map is also an ill-posed inverse problem. Regularisation in the form of a priori constraints on the solution is used to obtain an estimate of the HR shape profile. Real-world objects, in general, have smooth depth variations. To incorporate spatial dependencies of the depth estimates on neighbouring points, we model the HR depth map $\overline{\boldsymbol{d}}$ by a GMRF and obtain an MAP estimate of the HR depth map. Using Bayes' rule, we can write

$$P(\overline{\boldsymbol{d}}|\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_p) = \frac{P(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_p|\overline{\boldsymbol{d}})P(\overline{\boldsymbol{d}})}{P(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_p)} \tag{11}$$

where $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_p$ are the p chosen observations from the stack. Note that the same LR observations which were used to obtain the HR image are chosen for reconstruction of the HR depth map. By the Markovian property of MRF, the probability of a pixel being assigned a particular depth value depends only on the depth estimates of pixels in its neighbourhood. We restrict ourselves to a first-order neighbourhood. If the HR shape profile $\overline{\boldsymbol{d}}$ is modelled as an MRF, then the prior distribution is given by

$$P(\overline{\boldsymbol{d}}) = \frac{1}{Z}\exp\left[-\sum_{c \in C} V_c(\overline{\boldsymbol{d}})\right] \tag{12}$$

where Z is the partition function, c a clique, C the set of all cliques and $V_c(\cdot)$ the potential function associated with clique c. For a first-order neighbourhood, we propose the following potential function

$$\begin{aligned}\sum_{c \in C} V_c(\overline{\boldsymbol{d}}) = \sum_{i=1}^{qM}\sum_{j=1}^{qM} [&(\overline{d}(i,j) - \overline{d}(i,j-1))^2 \\ &+ (\overline{d}(i,j+1) - \overline{d}(i,j))^2 \\ &+ (\overline{d}(i+1,j) - \overline{d}(i,j))^2 \\ &+ (\overline{d}(i,j) - \overline{d}(i-1,j))^2]\end{aligned} \tag{13}$$

Assuming the noise process $\boldsymbol{n}_m$ s to be independent in (10), and from (11) and (13), the posterior energy function to be minimised is given by

$$U^p(\overline{\boldsymbol{d}}) = \sum_{m=1}^{p} \frac{\|\boldsymbol{y}_m - \boldsymbol{D}\boldsymbol{H}_m(\overline{\boldsymbol{d}})\boldsymbol{x}\|^2}{2\sigma_\eta^2} + \lambda_2 \sum_{c \in C} V_c(\overline{\boldsymbol{d}}) \tag{14}$$

The MAP estimate of the HR depth map can be obtained by minimising $U^p(\overline{\boldsymbol{d}})$. The parameter λ_2 must be tuned to arrive at a good estimate of the HR depth profile of the 3D object.

Assignment of depth values to pixel locations is a combinatorial optimisation problem. In recent literature, graph cuts have been used to minimise objective functions

doi: 10.1049/iet-cvi:20070072

arising in a wide variety of vision problems [32]. However, they are proven to converge only on singly-connected graphs. Also, it is not possible to compute the gradient with respect to $\overline{\boldsymbol{d}}$ of the cost function in (14). Hence, we use simulated annealing (SA), a well-known optimisation technique that is known to converge to the global minimum with probability one, to minimise the above objective function. We present the computational steps to obtain the MAP estimate of the HR depth profile using the SA algorithm in Fig. 2.

We obtain the LR shape profile from the traditional SFF [5] method and interpolate it to the size of the HR grid. This forms the initial estimate of the HR depth map $\overline{\boldsymbol{d}}$. The initial temperature in the SA algorithm is denoted by T_0. A linear cooling schedule was used in our simulations. The parameter deciding the cooling schedule is δ. The number of iterations used for the annealing and the Metropolis loops are given by A_l and M_l, respectively. The variance of the Gaussian sampler for $\overline{d}_{i,j}$ is denoted by $\sigma_{\overline{d}}^2$.

Importantly, it is to be noted that, when $\overline{d}_{i,j}$ is perturbed, it is not necessary to compute the new posterior energy function $U^p[\overline{\boldsymbol{d}}(\text{new})]$ over the entire HR grid, but only over a small posterior neighbourhood around the point (i, j) as shown in [28]. By using this well-known fact of locality of the posterior neighbourhood, it is possible to successfully employ the SA algorithm.

5 Experimental results

To demonstrate the effectiveness of the proposed algorithm, we present results obtained using images acquired by an LV150 Nikon industrial microscope. The lens objective was 2.5×, for which the working distance $w_d = 8.8$ mm, focal length $f = 80$ mm and depth of field = 48.9 μm.

For the first experiment, we selected a gold ring which had a human face engraved on its surface as the 3D object. A stack of 150 LR frames was captured by translating the stage in finite steps of $\Delta d = 0.025$ mm. The LR images captured were of size 75 × 75 pixels. The LR depth map of the 3D specimen is derived using the traditional SFF [5] method and the blur maps corresponding to the LR frames of the stack are computed. Initially, we obtain the super resolved image of the 3D specimen by selecting q^2 LR observations from the stack when the chosen magnification factor is q. The initial estimate of the HR image is obtained using bicubic interpolation of one of the LR images chosen from the stack and is shown in Fig. 3*a*. Choosing four LR images from the stack, frame numbers 85, 95, 130 and 140 and using $\beta = 1$, $\lambda_1 = 0.001$ and $\sigma_\eta^2 = 5$, in the proposed algorithm, we super-resolve the focused image by a factor of 2, which is shown in Fig. 3*b*. The super-resolved image of Fig. 3*b* is much superior than the interpolated image of Fig. 3*a*. Various high-frequency components that were lost in Fig. 3*a* are recovered in Fig. 3*b*.

begin

1. Initialization: Choose $\overline{\mathbf{d}}(initial)$ to be estimates obtained from the traditional SFF [5] method, interpolated to the size of the HR grid. Choose values for T_0, δ, M_l, A_l and $\sigma_{\overline{d}}$. $\overline{\mathbf{d}}(old) = \overline{\mathbf{d}}(initial)$. $k = 0$.

2. repeat (annealing loop)

for $l = 1$ to M_l, do (Metropolis loop)

begin

for $i = 1$ to qM, $j = 1$ to qM, do

begin

Get $\overline{d}_{i,j}(new)$ from Gaussian sampler with mean $\overline{d}_{i,j}(old)$ and variance $\sigma_{\overline{d}}^2$.

if $U^p(\overline{\mathbf{d}}(new)) \leq U^p(\overline{\mathbf{d}}(old))$ then $\overline{\mathbf{d}}(old) = \overline{\mathbf{d}}(new)$,

else

if $\exp\left(\frac{U^p(\overline{\mathbf{d}}(old)) - U^p(\overline{\mathbf{d}}(new))}{T_k}\right) > \text{rand}[0, 1]$ then $\overline{\mathbf{d}}(old) = \overline{\mathbf{d}}(new)$.

end

end

$k = k + 1$.

$T_k = \delta\, T_{k-1}$.

until (k equals A_l).

end

Figure 2 *SA algorithm for computation of the HR depth map*

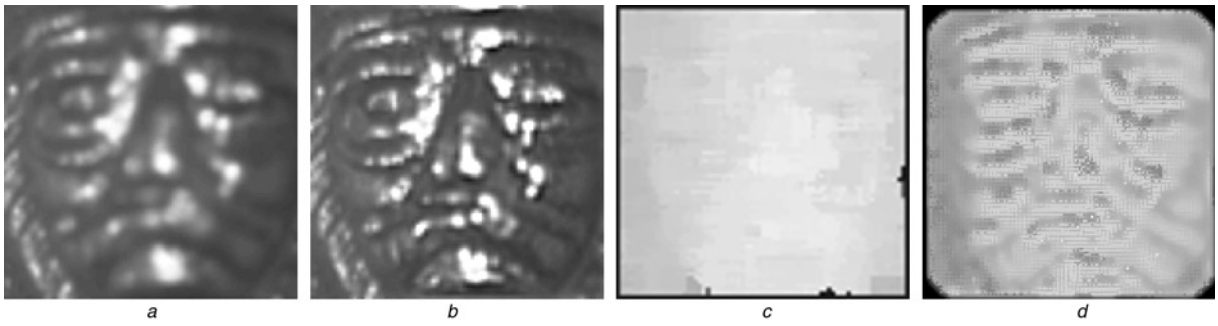

Figure 3 *Super-resolution of face image and depth map using the proposed method*

a Initial estimate of the HR image obtained using bicubic interpolation for $q = 2$
b Super-resolved image obtained using the proposed algorithm
c Greyscale image of the interpolated depth map from traditional SFF [5]
d Greyscale image of the HR depth map

Next, we obtain the HR depth map of the above specimen for a magnification factor of $q = 2$. Since we have already obtained the HR image, we can use it to obtain the HR structure. The initial estimate of the HR depth map, shown in Fig. 3*c*, was obtained by upsampling the LR depth map obtained from the traditional SFF method [5], by a factor of 2, using bicubic interpolation. The interpolated depth map is not very good because the image size is small (only 75×75 pixels) and the depth variations are very fine. We next show how this result can be significantly improved using the proposed approach. The objective function proposed in (14) is minimised using the SA algorithm, wherein, the values of the parameters are chosen as $T_0 = 6$, $A_1 = 200$, $M_l = 10$, $\sigma_\eta^2 = 5$ and $\lambda_2 = 1 \times 10^8$. The corresponding greyscale image of the estimated HR depth map is shown in Fig. 3*d*. Comparing Figs. 3*c* and 3*d*, we note that the proposed method is able to reconstruct the details of the depth variations of the 3D specimen well. The

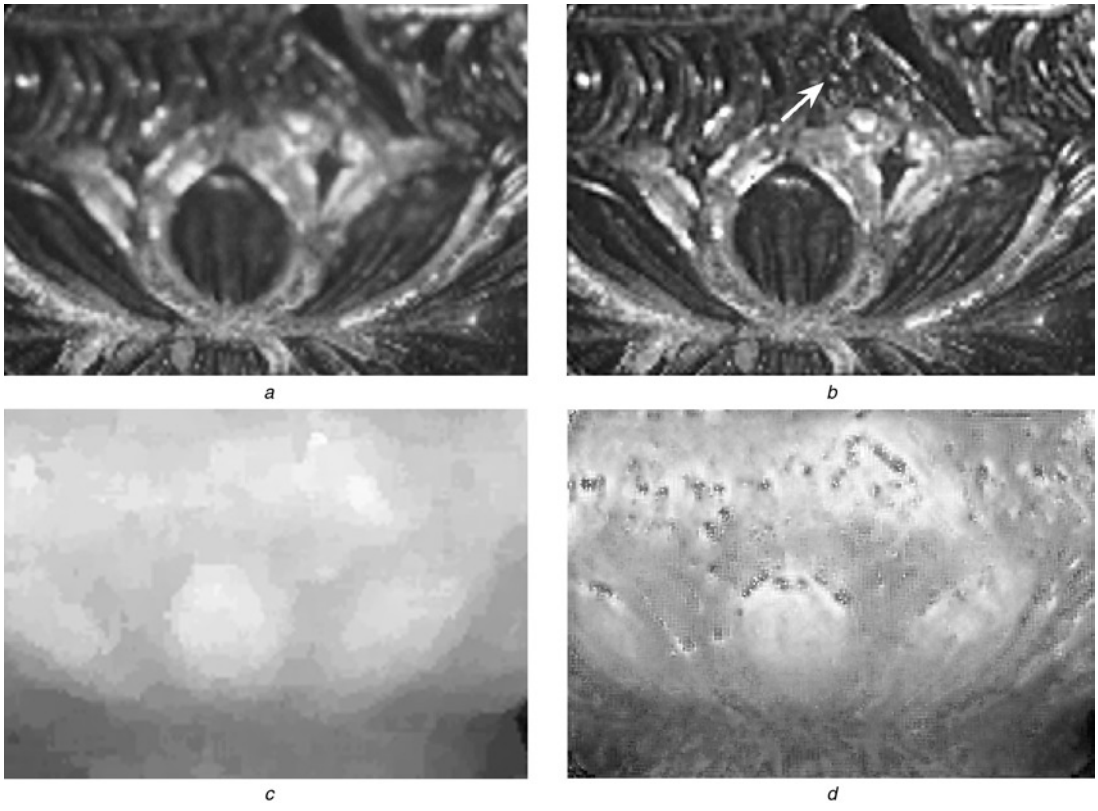

Figure 4 *HR image and depth map for the specimen with floral design*

a HR image obtained using bicubic interpolation for $q = 2$
b Super-resolved image obtained using the proposed algorithm
c Greyscale image of the interpolated depth map from traditional SFF [5]
d Corresponding greyscale image of the HR depth map

doi: 10.1049/iet-cvi:20070072

regions on the face which are closer to the camera appear brighter in the greyscale image of Fig. 3*d*. Various features of the face can be discerned from this depth map. In particular, note the eyebrows, eyelids, lips and the wrinkles on the cheeks.

We next selected a metal object with a floral design engraved on its surface. The object also contained several intricately sculpted grooves indicating the dress covering the leg of a goddess depicted as sitting on a lotus flower. This 3D specimen was kept on the translating stage under the microscope and a stack of 150 LR frames was captured by moving the stage in steps of $\Delta d = 0.025$ mm, relative to the camera. The size of each captured LR image is 95×135 pixels. Using these LR images, the LR depth map of the specimen was reconstructed by the traditional SFF algorithm [5]. As explained in Section 2, using (1) the blur maps corresponding to the LR frames captured in the stack can be computed. In order to obtain an HR image for an upsampling factor of $q = 2$, four LR images, namely frame numbers 30, 40, 50 and 100, were chosen from the stack. Since we have earlier computed the blur maps corresponding to these four LR space-variantly blurred images, matrix $\boldsymbol{H}_m$ in (4) can be constructed to relate the chosen LR images to the super-resolved image that we seek to obtain. The initial estimate of the HR image is obtained by upsampling one of the selected LR frames by a factor of $q = 2$. This is shown in Fig. 4*a*. Using the values of the parameters as $\beta = 1$, $\sigma_\eta^2 = 5$ and $\lambda_1 = 0.003$ in the gradient descent algorithm described in Section 3.1, we obtain the super-resolved image as shown in Fig. 4*b*. Notice that in Fig. 4*b* several features that were invisible in Fig. 4*a* become discernible. The separation between the grooves on the petals of the lotus flower can be clearly observed. The white dots in the region pointed out by a white arrow in Fig. 4*b* also become visible.

Using the super-resolved image obtained above and the same four LR images from the stack that were chosen to obtain it, we reconstruct the HR depth map of this 3D object for a magnification factor of $q = 2$. We used identical values of the parameters in the SA algorithm as described for the previous experiment. The initial estimate of the HR depth map, shown in Fig. 4*c*, was obtained by upsampling the LR depth map of traditional SFF method [5], by a factor of 2, using bicubic interpolation. The HR depth map is obtained using the proposed algorithm and its corresponding greyscale image is shown in Fig. 4*d*. The depth variations over the petals of the flower are more clearly visible in Fig. 4*d*, as compared with Fig. 4*c*. Note that in Fig. 4*d*, the regions on the 3D specimen which are closer to the camera, such as the petal in the central portion, appear brighter. In contrast, the regions between the petals of the lotus, which are further away from the lens, appear darker.

Next, we show the performance of the proposed algorithm for obtaining the super-resolved image of the 3D specimen by a higher magnification factor of $q = 3$. Since the entire stack of LR images is readily available to us, we select nine LR frames from it. Using the parameter values as $\beta = 1$, $\sigma_\eta^2 = 5$ and $\lambda_1 = 0.03$, we minimise the cost function in (8). One of the nine selected LR images is upsampled by

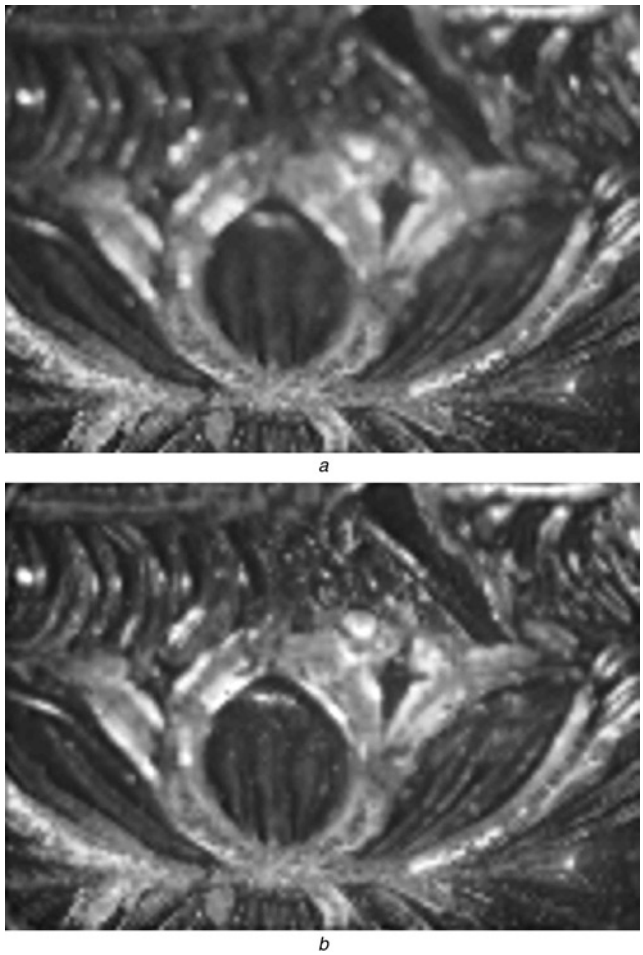

Figure 5 *HR image obtained for a higher upsampling factor*
a HR image obtained using bicubic interpolation for $q = 3$
b Super-resolved image obtained using the proposed algorithm

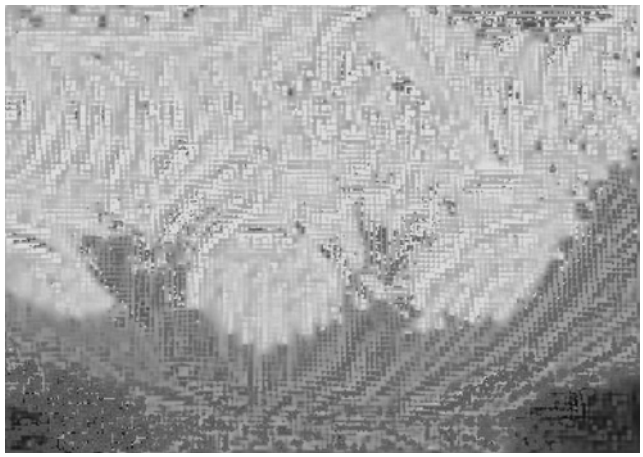

Figure 6 *Greyscale image of the HR depth map*

$q = 3$ (using bicubic interpolation) to obtain the initial estimate of the HR image and is shown in Fig. 5*a*. The super-resolved image is depicted in Fig. 5*b*. The intricate details on the petals of the flower and the sculpted grooves are much sharper and more clearly visible in Fig. 5*b* than in Fig. 5*a*. The proposed algorithm for obtaining the HR depth map was used with the same parameter values that were used for the earlier specimen. The greyscale image corresponding to the estimated HR depth map is given in Fig. 6. The depth variations on the surface of the 3D specimen can be clearly seen here. Since the magnification factor of $q = 3$ is high, the depth map estimated is slightly jagged. However, it is interesting to see how the proposed method is able to reconstruct the fine undulations in depth even for this large upsampling factor.

6 Conclusions

We have proposed a method for obtaining a super-resolved image and HR depth map given the LR space-variantly defocused and noisy images in SFF. Using a two-step approach and by modelling the HR image and the HR shape profile of the 3D object as MRFs, we first obtained an MAP estimate of the HR image and used it subsequently to derive the HR depth map. Exploiting the fact that in SFF the stack consists of several LR frames, we also showed that it is possible to super-resolve the image and the 3D structure even at higher magnification factors.

7 References

[1] GIBSON J.J.: 'The perception of the visual world' (Houghton-Mifflin, Boston, 1950)

[2] BAJCSY R.K., LIEBERMAN L.I.: 'Texture gradient as a depth cue', *Comput. Graph. Image Process.*, 1976, **5**, (1), pp. 52–67

[3] WOODHAM R.J.: 'A cooperative algorithm for determining surface orientation from a single view'. Proc. Int. Joint Conf. Artificial Intelligence, Cambridge, MA, 1977, pp. 635–641

[4] HORN B.K.P.: 'Obtaining shape from shading information' in WINSTON P.H. (ED.): 'The psychology of computer vision' (McGraw-Hill, New York, 1975), pp. 115–155

[5] NAYAR S.K., NAKAGAWA Y.: 'Shape from focus', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1994, **16**, (8), pp. 824–831

[6] SUBBARAO M.: 'Parallel depth recovery by changing camera parameters'. Proc. IEEE Int. Conf. Computer Vision, FL, USA, 1988, pp. 149–155

[7] MARR D., POGGIO T.: 'Cooperative computation of stereo disparity', *Science*, 1976, **194**, (4262), pp. 283–287

[8] JEBARA T., AZARBAYEJANI A., PENTLAND A.: '3D structure from 2D motion', *IEEE Signal Process. Mag.*, 1999, **16**, (3), pp. 66–84

[9] CARR J.C., FRIGHT W.R., BEATSON R.K.: 'Surface interpolation with radial basis functions for medical imaging', *IEEE Trans. Med. Imaging*, 1997, **16**, (1), pp. 96–107

[10] HAIGRON P., BELLEMARE M.E., ACOSTA O., *ET AL.*: 'Depth-map-based scene analysis for active navigation in virtual angioscopy', *IEEE Trans. Med. Imaging*, 2004, **23**, (11), pp. 1380–1390

[11] WANG O., FINGER J., YANG Q., DAVIS J., YANG R.: 'Automatic natural video matting with depth'. Proc. 15th Pacific Conf. Computer Graphics and Applications, Maui, Hawaii, 2007, pp. 469–472

[12] FEHN C.: 'A 3D-TV approach using depth-image-based rendering (DIBR)'. Proc. 3rd IASTED Conf. Visualization, Imaging and Image Processing, Spain, 2003, pp. 482–487

[13] KAUFF P., ATZPADIN N., FEHN C., *ET AL.*: 'Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability', *Signal Process., Image Commun.*, 2007, **22**, (2), pp. 217–234

[14] ELAD M., FEUER A.: 'Restoration of a single superresolution image from several blurred, noisy, and under-sampled measured images', *IEEE Trans. Image Process.*, 1997, **6**, (12), pp. 1646–1658

[15] ELAD M., HEL-OR Y.: 'A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur', *IEEE Trans. Image Process.*, 2001, **10**, (8), pp. 1187–1193

[16] FARSIU S., ROBINSON M.D., ELAD M., MILANFAR P.: 'Fast and robust multiframe super resolution', *IEEE Trans. Image Process.*, 2004, **13**, (10), pp. 1327–1344

[17] PAPOULIS A.: 'Generalized sampling expansion', *IEEE Trans. Circuits Syst.*, 1977, **24**, (11), pp. 652–654

[18] RAJAN D., CHAUDHURI S.: 'An MRF-based approach to generation of super-resolution images from blurred observations', *J. Math. Imaging Vis.*, 2002, **16**, (1), pp. 5–15

[19] RAJAN D., CHAUDHURI S.: 'Simultaneous estimation of super-resolved scene and depth map from low resolution defocused observations', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2003, **25**, (9), pp. 1102–1117

[20] BAKER S., KANADE T.: 'Limits on super-resolution and how to break them', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, (9), pp. 1167–1183

doi: 10.1049/iet-cvi:20070072

[21] FREEMAN W., PASZTOR E., CARMICHAEL O.: 'Learning low-level vision', *Int. J. Comput. Vis.*, 2000, **40**, (1), pp. 25–47

[22] HERTZMANN A., JACOBS C., OLIVER N., CURLESS B., SALESIN D.: 'Image analogies'. Proc. ACM SIGGRAPH, Los Angeles, USA, 2001, pp. 327–340

[23] DIEBEL J., THRUN S.: 'An application of MRFs to range sensing'. Proc. Neural Information Processing Systems Conf., Vancouver, Canada, 2005, pp. 291–298

[24] YANG Q., YANG R., DAVIS J., NISTER D.: 'Spatial depth super resolution for range images'. Proc. IEEE Conf. Computer Vision and Pattern Recognition, Minneapolis, USA, 2007, pp. 1–8

[25] KIL Y.J., MEDEROS B., AMENTA N.: 'Laser scanner super-resolution'. Proc. Eurographics Symp. Point-based Graphics, Boston, USA, 2006, pp. 9–16

[26] JOSHI M., CHAUDHURI S.: 'Simultaneous estimation of super-resolved depth map and intensity field using photometric cue', *Computer Vis. Image Underst.*, 2006, **101**, (1), pp. 31–44

[27] PENTLAND A.P.: 'A new sense for depth of field', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1987, **9**, (4), pp. 523–531

[28] CHAUDHURI S., RAJAGOPALAN A.N.: 'Depth from defocus: a real aperture imaging approach' (Springer-Verlag, New York, 1999)

[29] BOSE N.K., NG M.K., YAU A.C.: 'A fast algorithm for image super-resolution from blurred observations', *EURASIP J. Appl. Signal Process.*, (2006) Article ID 35726, 14 pp., 2006. doi: 10.1155/ASP/2006/35726

[30] LI S.Z.: 'Markov random field modeling in computer vision' (Springer-Verlag, Tokyo, 1995)

[31] BESAG J.E.: 'Spatial interaction and the statistical analysis of lattice systems', *J. R. Stat. Soc. B*, 1974, **36**, (2), pp. 192–236

[32] BOYKOV Y., VEKSLER O., ZABIH R.: 'Fast approximate energy minimization via graph cuts', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001, **23**, (11), pp. 1222–1239