Unscented Transformation for Depth from Motion-Blur in Videos

C. Paramanand and A.N. Rajagopalan

Image Processing and Computer Vision Laboratory, Department of Electrical Engineering,
Indian Institute of Technology Madras, Chennai 600 036, India

`paramanand@gmail.com,raju@ee.iitm.ac.in`

Abstract

In images and videos of a 3D scene, blur due to camera shake can be a source of depth information. Our objective is to find the shape of the scene from its motion-blurred observations without having to restore the original image. In this paper, we pose depth recovery as a recursive state estimation problem. We show that the relationship between the observation and the scale factor of the motion-blur kernel associated with the depth at a point is nonlinear and propose the use of the unscented Kalman filter for state estimation. The performance of the proposed method is evaluated on many examples.

1. Introduction

Motion-blur is a commonly observed phenomena and occurs due to relative motion between the camera and objects while capturing the scene. If the extent of blur is constant at all points, it is referred to as space-invariant blur, while it is called space-variant blur if it varies. In scenes with depth variations, the amount of blur induced is related to the 3D structure of the scene.

Traditionally, motion-blur is regarded as an undesirable artifact in images and videos. A large number of approaches that address the problem of deblurring exist in the literature. A motion-blurred image can be related to the original reference image of the scene using a point spread function (PSF). In [2], Ezra and Nayar propose the use of a hybrid camera for estimating the PSF which is used for restoration. This approach is extended in [18] to restore space-variantly motion-blurred images and videos. Another approach for the removal of motion-blur in videos is proposed in [1]. The method uses video frames blurred with different exposure times and having different PSFs to restore a frame.

Motion-blurred images have been used to estimate depth information in [7] and [11] by determining the extent of blurring. These methods assume that the scene can be approximated by a set of planar patches. Depth estimation from space-variantly blurred images is performed us-

ing variational methods in [5] and [15]. Favaro and Soatto [5] consider scenes in which different objects move along different directions. They estimate the motion field, depth-map and the restored image from the motion-blurred observations which are captured with different exposure times. The motion-blurred images are modeled as solutions of an anisotropic diffusion equation. This model constrains the shape of the PSF to be Gaussian. In real scenarios, the shape of PSF can be arbitrary and the Gaussian model is appropriate only when the extent of blur is small [15]. Sorel and Flusser have proposed a technique to estimate the shape and the restored image using two observations which are blurred in different ways [15]. They consider the PSF to be of arbitrary shape (due to non-uniform velocity of the camera). In their method, the PSF is initially determined from the blurred observations by choosing regions of constant depth. With this PSF, an initial estimate of the depth-map is obtained using a window-based technique. From this initial estimate, the image and the depth-map are simultaneously estimated. The method requires many iterations to converge.

We propose a recursive filtering framework for depth estimation from motion-blurred images. The camera motion is considered to be in-plane translations, following the work in [15]. Our method uses two blurred observations of a scene captured in such a manner that one of the images can be treated to be a space-variantly blurred version of the other. Such observations can be obtained from an image sequence when the camera shake is lesser in one of the frames as compared to that in other frames. These can also be obtained by a hybrid camera which captures videos at two different frame rates [18], or by time averaging image sequences [5]. The proposed method is applicable even when the camera translation is arbitrary and does not require the knowledge of camera calibration for determining the shape of the scene.

The relative blur between the two observations is represented using a PSF which can be of arbitrary shape (due to non-uniform camera motion). We relate the PSF and depth at a point in the image using a parameter which we call as

1

38

scale factor. The relationship between the pixel intensities of the blurred image and the scale factor turns out to be non-linear. The scale factor at each instant (pixel) is regarded as the state which is determined from the observation using an unscented Kalman filter (UKF) [8]. UKF is a recursive state estimation technique which can handle even nonlinear transformations. It has been applied in computer vision for different problems such as image restoration [17] and tracking [20]. To preserve discontinuities in depth, we incorporate an edge-adaptive prior model for the state. The first two moments of the prior distribution are estimated using importance sampling [12]. Based on the blurred observation, the Bayesian estimate of the state is obtained at every pixel using the UKF. The depth of the scene can be obtained from this estimate of the scale factor.

Kalman filtering framework has been used in earlier works to estimate depth from a sequence of images [13, 14]. To the best of our knowledge, this is the first work of its kind that uses UKF for estimating depth from motion-blurred observations. The proposed recursive implementation involves a single pass across the image and does not require any initial estimate. Another advantage of our scheme is in terms of memory usage while implementing on hardware. In the recursive filtering framework, computing depth at a pixel requires data only from its neighboring pixels. Consequently, the memory requirements for processing is less and the technique can be readily implemented on DSPs which have limited memory. In contrast, in iterative techniques, image intensities and gradients corresponding to all of the image pixels must be computed and stored while estimating depth.

2. Blur model and problem formulation

In this section, we discuss the modeling of a motion-blurred image. The shape of the PSF remains constant across the image and the only variation it undergoes is scaling. The extent of scaling at any point is related to the depth of the scene.

Consider a 3D scene captured by a moving camera with motion being restricted to in-plane translations. The velocity vector of the camera at time τ is denoted by $\mathbf{T}(\tau) = [T_x(\tau)\ T_y(\tau)\ 0]^T$ with respect to the coordinate system having origin at the optical center of the camera, the X and Y axes parallel to the image plane, and the Z axis along the optical axis of the camera. Let f denote the image of the scene captured by a still camera. The velocity $\mathbf{v}$ at a point (x, y) is given [4] as

$$\mathbf{v}(x, y, \tau) = \begin{bmatrix} v_x(x, y, \tau) \\ v_y(x, y, \tau) \end{bmatrix} = \frac{-\nu}{d(x, y)} \begin{bmatrix} T_x(\tau) \\ T_y(\tau) \end{bmatrix} \quad (1)$$

where ν is the "focal length" of projection and $d(x, y)$ is the depth of the image point (x, y). This shows that the velocity of points near to the camera will be higher than those

which are farther. The blurred image g is the average of the light intensities experienced by the image sensors during the exposure time. Consequently, in the blurred image, the extent of smearing of a scene point nearer to the camera is more than that for a point which is farther. Ignoring occlusion effects, the blurred image is modeled [15] using the space-variant PSF $h(x, y, s, t)$ as

$$f *_v h(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x - s, y - t) h(x - s, \\ y - t; s, t)\, ds dt \quad (2)$$

where $*_v$ denotes space-variant blurring operation. The noisy blurred observation g can be written as

$$g(x, y) = f *_v h(x, y) + e(x, y) \quad (3)$$

where $e()$ is the observation noise. Sorel and Flusser [15] have shown that the shape of the PSF $h(x, y, s, t)$ remains the same throughout the image except for a scale factor which is related to depth. Consider a reference point (o_x, o_y) with depth $d(o_x, o_y) = d_o$. If the PSF at (o_x, o_y) is $h_o(o_x, o_y; s, t)$, then the PSF at any other point (x, y) is a scaled version of $h_o(o_x, o_y; s, t)$ and is given by

$$h(x, y; s, t) = k^2(x, y) h_o(o_x, o_y; sk(x, y), tk(x, y)) \quad (4)$$

where $k(x, y) = \frac{d(x, y)}{d_o}$ is a scale factor and represents the relative depth with respect to the depth at the point (o_x, o_y). Discretization of Eqns. (2) and (3) yields

$$f *_{v_d} h(i, j) = \sum_{m, n} f(i - m, j - n) h(i - m, j - n; m, n)$$

$$g(i, j) = f *_{v_d} h(i, j) + e(i, j)$$

$$(5)$$

2.1. State estimation of scale

As mentioned in section 1, for depth estimation from motion-blurred images, we use two observations g_1 and g_2 such that g_2 can be obtained by space-variant blurring of g_1. i.e., we can write $g_2(i, j) = g_1 *_{v_d} h_r(i, j) + e(i, j)$. Here, g_1 is regarded as the reference image, g_2 as the observation, and the PSF h_r denotes the relative blur.

The initial step of our algorithm is to estimate the shape of the relative blur h_r between the two observations. We crop a small region of constant depth from the two images and estimate the blur between these regions. There are many algorithms which can be used for PSF estimation [6, 16, 19]. We employ an energy minimization approach for finding the PSF. It is not discussed in this paper since our focus is on depth estimation. This blur kernel is considered as the reference PSF $h_{r_o}(, ; m, n)$ from which the relative blur at any other image point (i, j), denoted by $h_r(i, j; m, n)$, can be obtained from its scale factor $k(i, j)$

through Eqn. (4). In order to facilitate recursive state estimation, we locally approximate the blurring model of Eqn. (5) as

$$g_2(i,j) = \sum_{m,n} g_1(i-m, j-n) h_r(i,j;m,n) + e(i,j) \quad (6)$$

where $e()$ is assumed to be additive Gaussian noise with zero mean and variance σ_e^2. According to this model, the blurred image intensity $g_2(i,j)$ is dependent on the reference image g_1, the reference kernel $h_{r_o}(,;m,n)$, and the scale factor $k(i,j)$. We address the problem of finding $k(i,j)$ with the knowledge of g_1, g_2 and h_{r_o}. We formulate it as a recursive state estimation problem with $k(i,j)$ as the state. The relationship between the state $k(i,j)$ and the observation $g_2(i,j)$ is nonlinear. For illustrating this nonlinearity, let us consider the reference kernel $h_{r_o}(,;m,n)$ as a Gaussian with standard deviation σ. The PSF $h_r(i,j;m,n)$ will be a Gaussian with standard deviation $\frac{\sigma}{k(i,j)}$. If $k(i,j)$ is multiplied by a constant c_k, then the PSF $h_r(i,j;m,n)$ will have standard deviation $\frac{\sigma}{c_k k(i,j)}$. From Eqn. (6) we see that, the change in the value of the scale factor does not necessarily alter the value of $g_2(i,j)$ to $c_k g_2(i,j)$. Hence, the relationship between $k(i,j)$ and $g_2(i,j)$ is nonlinear. For recursively estimating $k(i,j)$ despite the nonlinearity, we propose to use the unscented Kalman filter [8]. The UKF avoids the problem of estimating the Jacobians of the nonlinear relation which is necessary in other linearization-based approaches.

The scale-map (k evaluated at all pixels) represents the shape of the scene. The absolute value of depth d can be obtained from k at all pixels if the depth is known at any one image point.

3. Estimation of moments

The scale factor $k(i,j)$ is considered as the state at pixel (i,j). Its mean and covariance are denoted by $\mu_{k(i,j)}$ and $\mathbf{P}_{k(i,j)}$, respectively. These moments are estimated based on the blurred observation through the Kalman filter. For the system model, we use a discontinuity adaptive Markov random field (DAMRF) prior [10]. The use of a statistical model is advantageous over a linear model since for a linear model to work well, its coefficients must be known correctly which is difficult in practice. Also, the DAMRF prior helps in adapting the estimation process at the edges.

Kalman filtering requires propagation of mean and covariance of the state estimate through the system and observation models. We predict the first two moments of the state distribution using importance sampling [12]. The mean and covariance of the measurement is predicted by unscented transformation [8]. We discuss these in the following subsections.

3.1. Prediction of state moments

The previously estimated states are related to the current state $k(i,j)$ through a DAMRF prior with a non-symmetric half-plane support (NSHP) as the neighborhood system [17]. Let $\bar{k}_{i,j}$ denote the first order NSHP neighbors of $k(i,j)$. i.e., $\bar{k}(i,j) = \{k(i,j-1), k(i-1,j-1), k(i-1,j), k(i-1,j+1)\}$. The conditional pdf of $k(i,j)$ takes the form

$$P\left(k(i,j)/\bar{k}(i,j)\right) =$$
$$\frac{1}{Z} \exp\left(-\gamma \log\left(1 + \frac{\eta^2\left(k(i,j), \bar{k}(i,j)\right)}{\gamma}\right)\right) \quad (7)$$

where γ is a parameter of the discontinuity-adaptive model, Z is a normalization constant, and

$$\eta^2\left(k(i,j), \bar{k}(i,j)\right) = \frac{1}{\rho^2(i,j)} \sum_{k \in \bar{k}(i,j)} (k(i,j) - k)^2 \quad (8)$$

Parameter $\rho^2(i,j)$ controls the variations of $k(i,j)$ with respect to its neighbors. The mean $\widehat{\mu}_p$ and variance $\widehat{\sigma}_p^2$ of the conditional pdf correspond to the predicted mean and covariance based on the previously estimated scale factors, and are denoted by $\mu_{k(i,j)/(i,j-1)}$ and $\mathbf{P}_{k(i,j)/(i,j-1)}$, respectively. The conditional density function is non-Gaussian and it is not straightforward to estimate its mean and covariance. We employ importance sampling, a Monte Carlo method [12] to estimate these moments. We follow a procedure similar to the one given in [17] for implementing importance sampling.

3.2. Unscented transformation

Unscented transformation (UT) is a deterministic sampling approach which can be used to calculate the statistics of a random variable that undergoes a nonlinear transformation [8]. Consider an n_x dimensional random vector $\mathbf{x}$. Let $\mathbf{F} : \Re^{n_x} \to \Re^{n_y}$ be a nonlinear function acting on $\mathbf{x}$ i.e., $\mathbf{y} = \mathbf{F}(\mathbf{x}) \in \Re^{n_y}$. Let $\bar{\mathbf{x}}$ and $\mathbf{P}_x$ denote the mean and covariance of $\mathbf{x}$, respectively. The statistical moments of $\mathbf{x}$ are captured using $2n_x + 1$ samples or sigma points. The moments of $\mathbf{y}$ can be evaluated from these samples and weighting factors which are given by the following equations [9].

$$\mathbf{X}_0 = \bar{\mathbf{x}}; \; \mathbf{X}_i = \bar{\mathbf{x}} + (\sqrt{(n_x + \lambda)\mathbf{P}_x})_i, i = 1,..,n_x$$
$$\mathbf{X}_i = \bar{\mathbf{x}} - (\sqrt{(n_x + \lambda)\mathbf{P}_x})_i, i = n_x+1,....,2n_x;$$
$$w_0^{(m)} = \frac{\lambda}{(n_x + \lambda)}; \; w_0^{(c)} = w_0^{(m)} + (1 - \alpha_{UT}^2 + \beta_{UT});$$
$$w_i^{(m)} = w_i^{(c)} = \frac{1}{2(n_x + \lambda)}, i = 1,....,2n_x;$$
$$\text{where } \lambda = \alpha_{UT}^2(n_x + \kappa) - n_x \quad (9)$$

40

where $\mathbf{X}_i$ denotes the ith sigma point, and $w_i^{(c)}$ and $w_i^{(m)}$ denote its weights for evaluating mean and covariance, respectively. Including $\mathbf{X}_0$, the number of sigma points is $2n_x + 1$. The term $(\sqrt{(n_x + \lambda)\mathbf{P}_x})_i$ is the ith column of the matrix square root $(\sqrt{(n_x + \lambda)\mathbf{P}_x})_i$ [8].

The nonlinear mapping is applied on each $\mathbf{X}_i$ to get $\mathbf{Y}_i = \mathbf{F}(\mathbf{X}_i)$, $i = 0, 1, ..., 2n_x$. The mean and covariance of $\mathbf{y}$ are estimated as

$$\overline{\mathbf{y}} = \sum_{i=0}^{2n_x} w_i^{(m)}\mathbf{Y}_i, \quad \mathbf{P}_y = \sum_{i=0}^{2n_x} w_i^{(c)}(\mathbf{Y}_i - \overline{\mathbf{y}})(\mathbf{Y}_i - \overline{\mathbf{y}})^T \quad (10)$$

The sigma points completely capture the distribution of $\mathbf{x}$ up to the second moment. This results in the estimates of the moments of $\mathbf{y}$ being accurate to the second-order of the Taylor series expansion for any nonlinear function [8].

4. Recursive depth estimation

Initially, we estimate the reference PSF $h_{r_o}(,;m,n)$. At a pixel (i,j), the state mean $\mu_{k(i,j)/(i,j-1)}$ (denoted by $\widehat{\mu}_p$) and covariance $\mathbf{P}_{k(i,j)/(i,j-1)}$ $(\widehat{\sigma}_p^2)$ are predicted using previous states as discussed in section 3. At the first pixel, the previous states can be assigned to any arbitrary constant. The observation and its covariance at (i,j) are predicted from the observation model using predicted state mean and covariance, $h_{r_o}(,;m,n)$, and g_1 through UT. Based on the observation g_2, the estimates of the state are updated.

4.1. Observation model

Eqn. (6) relates the blurred image intensity $g_2(i,j)$ to the scale factor $k(i,j)$. It can be used as the measurement model for state estimation. However, for accurately estimating the state $k(i,j)$, in addition to $g_2(i,j)$, we consider the blurred intensities at some more points as the observation. The scale factor $k(i,j)$ represents the PSF at (i,j) and it affects those locations in the blurred image where the scaled PSF has higher values. Consequently, the blurred image at these locations contain information about $k(i,j)$. We scale the reference kernel $h_{r_o}(,;m,n)$ by a scale factor $\mu_{k(i,j)/(i,j-1)}$ and place it around (i,j). Depending on where the peaks of this PSF occur, we pick n_o blurred image pixels around (i,j) including $g_2(i,j)$ and regard it as the observation $\mathbf{g}(i,j)$. With the knowledge of g_1 and $h_{r_o}(,;m,n)$, the nonlinear relationship between the state and the observation can be written as

$$\mathbf{g}(i,j) = \mathbf{H}_{i,j}(k(i,j)) + e(i,j) \quad (11)$$

Typically, we choose n_o to be three or four pixels. When there are color images, we consider that the state and in turn the PSF at a point is the same for all channels. We stack the blurred image pixels of all the three channels and regard it as the observation $\mathbf{g}(i,j)$. Since the number of observations is more, state estimation can be done more accurately.

4.2. State update

Observation mean $\mu_{\mathbf{g}(i,j)/(i,j-1)}$ and covariance $\mathbf{P}_{\mathbf{g}(i,j)/(i,j-1)}$ are predicted from the moments of the state and the observation model using UT. We augment the state $k(i,j)$ with n_o independent observation noise terms to get a $n_o + 1$ dimensional vector $\mathbf{x}$ on which UT is applied. The sigma point selection scheme (Eqn. (9)) is applied as follows:

$$\overline{\mathbf{x}}_{(i,j)/(i,j-1)}^a = [\widehat{\mu}_p \quad 0.. \quad ..0]^T$$

$$\mathbf{P}_{(i,j)/(i,j-1)}^a = \begin{bmatrix} \widehat{\sigma}_p^2 & 0. & .. & .0 \\ 0 & \sigma_e^2 & 0. & .0 \\ . & . & . & . \\ 0 & 0 & 0 & \sigma_e^2 \end{bmatrix}$$

$$\mathbf{X}_{(i,j)/(i,j-1)}^a = \begin{bmatrix} \overline{\mathbf{x}}_{(i,j)/(i,j-1)}^a & \overline{\mathbf{x}}_{(i,j)/(i,j-1)}^a \end{bmatrix}$$

$\pm\sqrt{(n_a + \lambda)\mathbf{P}_{(i,j)/(i,j-1)}^a}$ for $n_a = 1,..n_o + 1$. $\mathbf{X}^a$ is the augmented sigma point matrix. Let $\mathbf{X}^a = [(\mathbf{X}^k) \ (\mathbf{X}^e)]^T$. where $\mathbf{X}^k$ and $\mathbf{X}^e$ contain the $2n_o + 3$ sigma points of state $k(i,j)$ and measurement noise, respectively. We propagate each of these sigma points through the measurement model given in Eqn. (11) to obtain the sigma points of the observation, denoted by $\mathbf{Y}$.

$$\mathbf{Y}_l = \mathbf{H}_{i,j}(\mathbf{X}_l^k) + \mathbf{X}_l^e, \quad l = 1...2n_o + 3 \quad (12)$$

i.e., at the pixel (i,j), for each sigma point, the reference PSF $h_{r_o}(,;m,n)$ is scaled by a factor $\mathbf{X}_l$. The sigma point of the observation $\mathbf{Y}_l$ is obtained through the observation model using this scaled PSF and the intensity values of the reference image g_1. From the sigma points $\mathbf{Y}$ and the weights, the statistics $\overline{\mathbf{y}}$ and $\mathbf{P}_y$ which correspond to $\mu_{\mathbf{g}(i,j)/(i,j-1)}$ and $\mathbf{P}_{\mathbf{g}(i,j)/(i,j-1)}$, respectively, are estimated by Eqn. (10). The state-measurement cross-covariance $\mathbf{P}_{xy}$ is obtained using the relation

$$\mathbf{P}_{xy} = \sum_{l=0}^{2n_o+3} w_l^{(c)}(\mathbf{X}_l^k - \widehat{\mu}_p)(\mathbf{Y}_l - \overline{\mathbf{y}})^T \quad (13)$$

Based on the observation $\mathbf{g}(i,j)$, the Bayesian estimate of the state conditional mean and its covariance are obtained from the Kalman filter update equations as

$$\mu_{k(i,j)} = \mu_{k(i,j)/(i,j-1)} + \mathbf{K}_{(i,j)}(\mathbf{g}_{(i,j)} - \mu_{\mathbf{g}(i,j)/(i,j-1)})$$
$$\mathbf{P}_{k(i,j)} = \mathbf{P}_{k(i,j)/(i,j-1)} - \mathbf{K}_{(i,j)}\mathbf{P}_y\mathbf{K}_{(i,j)}^T \quad (14)$$

where the Kalman gain $\mathbf{K}_{(i,j)}$ is given by $\mathbf{K}_{(i,j)} = \mathbf{P}_{xy}\mathbf{P}_y^{-1}$. The updated mean $\mu_{k(i,j)}$ is regarded as the estimated scale factor $k(i,j)$. In the next step, $k(i,j+1)$ is predicted similarly and the covariance $\mathbf{P}_{k(i,j)}$ is used as the parameter $\rho^2(i,j+1)$ in the prior.

5. Experimental results

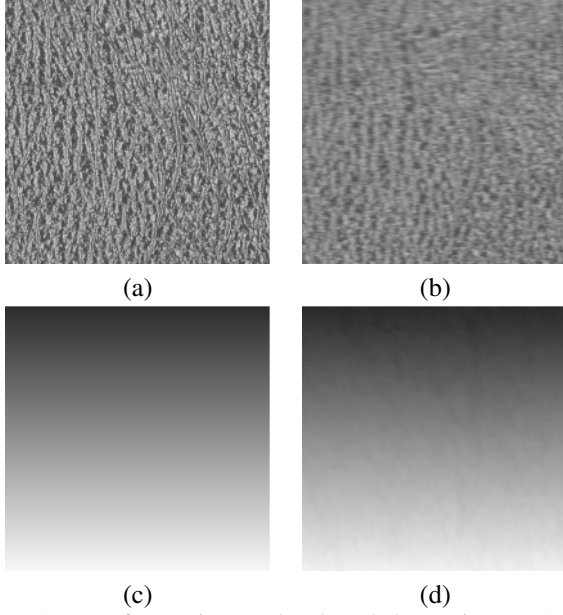For the purpose of validation, we tested the proposed method on synthetic and real examples. We initially applied

41

(a)

(b)

(c)

(d)

Figure 1. (a) Reference image. (b) Blurred observation. (c) Actual scale-map (d) Estimated scale-map.

(a)

(b)

(c)

(d)

Figure 2. (a) Reference image. (b) Blurred observation. (c) Estimated PSF, and (d) Estimated structure.

the proposed recursive filtering technique on synthetic data. The calf leather image ([3]) shown in Fig. 1 (a) was used as the reference frame. For this synthetic experiment, we assumed a Gaussian form for the reference PSF. The blurred image shown in Fig. 1 (b) was generated by space-variantly blurring the original image by scaling the reference PSF according to the depth-map shown in Fig. 1 (c). The estimated depth-map from the proposed method is shown in Fig. 1 (d). We evaluated the error in estimation ERR using the formula $ERR = \sqrt{\text{Avg}\left[\left(\frac{\hat{k}}{k} - 1\right)^2\right]}$, where Avg $[]$ denotes averaging, and $\hat{k}$ and k denote the estimated and actual scale factors, respectively. The value of the estimation error was found to be 0.015, which is quite small.

To test our algorithm on real data, we captured images with a compact digital camera operating in manual mode. The scene contained different objects at different depths. The reference image shown in Fig. 2 (a) was captured when the camera was stationary (although this is not a requirement). While capturing the blurred observation shown in Fig. 2 (b), the camera was arbitrarily shaken in a plane during the exposure time. The reference PSF estimated from a small patch in the bottom left corner of the image is shown in Fig. 2 (c). We see that the shape of the PSF which denotes the camera motion during the exposure is non-Gaussian. Fig. 2 (d) shows the estimated scale factors using the proposed technique. Farther objects have a higher scale factor and are represented by higher intensities. In Fig. 2 (d), we observe that the shape is captured correctly even at the depth
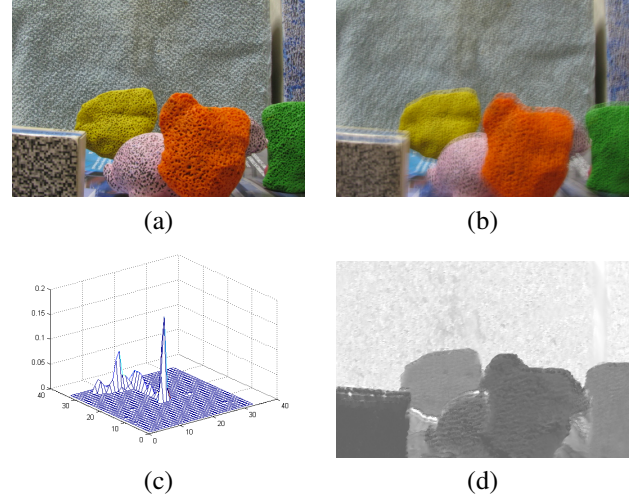
discontinuities.

In our next experiment, both the reference image as well as the observation were captured with a camera shake. The reference image and the blurred observation are shown in Figs. 3 (a) and (b), respectively. The PSF denoting the relative blur between the two images was estimated from a small region in the yellow clay model, and is shown in Fig. 3 (c). The scale factors estimated from the proposed UKF-based technique is shown in Fig. 3 (d). In the scene, the yellow and the pink clay models were kept at the same distance from the camera, and the model of the tree was kept behind the two objects. In Fig. 3 (d), we see that the estimated scale factors correctly reflect the scene. The scale factors in the hole of yellow clay model correctly correspond to that of the background. We measured the distance of the pink clay model from the camera and obtained the complete depth-map. A novel view rendering of the scene shown in Fig. 3 (e) correctly depicts the 3D structure.

We next tested our algorithm on images used by Sorel and Flusser in [15]. We used the original image of the scene (Fig. 4 (a)) as the reference image and one motion-blurred observation (Fig. 4 (b)). The estimated PSF is shown in the top left corner of Fig. 4 (b). The scale-map estimated by the proposed algorithm is shown in Fig. 4 (c). Fig. 4 (d) shows the scale factors reproduced from [15]. The convention adopted in [15] to indicate scale factors is opposite to ours. i.e., dark intensities indicate farther objects. In Fig. 4 (d), we observe that the estimated scale factors do not depict the depth correctly at a few places. For instance, on the carton box which is near to the camera and is a planar surface, we see a few dark patches which indicate a higher distance from the camera. In the shape estimated by the proposed algorithm (Fig. 4 (c)), we do not observe such errors. In Fig. 4 (c), we see that the region corresponding to the carton box is assigned darker intensities in the scale-map since
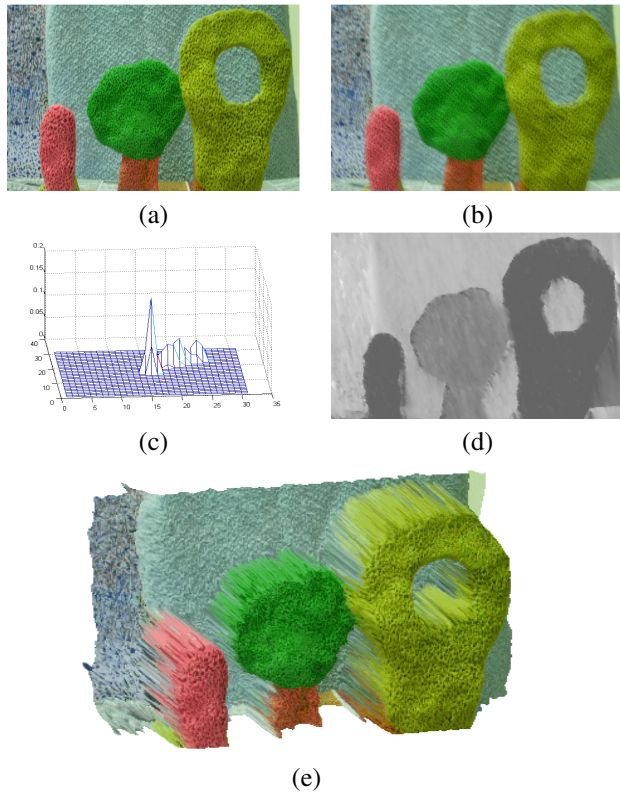
42

(a) (b)

(c) (d)

(e)

Figure 3. (a) and (b) Blurred observations. (c) Estimated PSF. (d) Scale-map, and (e) Novel view rendering of the scene.
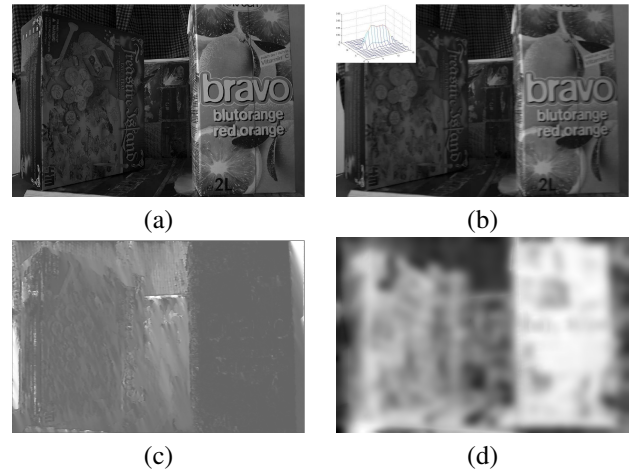

(a) (b)

(c) (d)

Figure 4. (a) Reference image. (b) Blurred observation and estimated PSF. Estimated shape from (c) proposed UKF-based technique, and (d) algorithm in [15].

it is near to the camera. Regions that are farther from the camera result in a higher intensity, as expected. The depth estimates are not as sharp as in the previous examples due to lack of texture and poor illumination in the scene. It must be mentioned that the algorithm in [15] uses two blurred observations whereas we used the original image and one blurred observation. This is because one of the two blurred observations in this experiment was captured with horizontal camera motion while the other was captured with a vertical motion.

6. Conclusions

We proposed a new recursive filtering approach for estimating the shape of a 3D scene from its motion-blurred observations. The shape of the scene is determined from arbitrary camera translations without the knowledge of camera calibration. The proposed framework involves computations only once across the image and requires less memory in the estimation process. As with other methods, our algorithm requires textured images to achieve good performance. Possible extensions of our work include considering general camera and/or object motion, and handling occlusions.

References

[1] A. Agrawal, Y. Xu, and R. Raskar. Invertible motion blur in video. *Proceedings of SIGGRAPH*, pages 1–8, 2009. 1

[2] M. Ben-Ezra and S. K. Nayar. Motion-based motion deblurring. *IEEE Trans. Patt. Anal. Mach. Intell.*, 26(6):689–698, 2004. 1

[3] P. Brodatz, 1966. Textures; a photographic album for artists and designers. Dover Publications, New York. 5

[4] P. Favaro and S. Soatto. *3-D Shape Estimation and Image Restoration Exploiting Defocus and Motion-blur*. Springer Verlag, 2006. 2

[5] P. Favaro and S. Soatto. A variational approach to scene reconstruction and image segmentation from motion-blur cues. *IEEE Conference on Computer Vision and Pattern Recognition*, 1:631–637, 2004. 1

[6] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman. Removing camera shake from a single photograph. *ACM Transactions on Graphics*, 25(3):787–794, 2006. 2

[7] J. S. Fox. Range from translational motion blurring. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 360–365, 1988. 1

[8] S. Julier and J. Uhlmann. A new extension of the Kalman filter to nonlinear systems. *The 11th International Symposium on Aerospace/Defense Sensing, Simulation and Controls*, pages 182–193, 1997. 2, 3, 4

[9] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Transactions on Automatic Control*, 45(3):477–482, 2000. 3

[10] S. Z. Li. *Markov random field modeling in computer vision*. Springer Verlag, 1995. 3

[11] H. Y. Lin and C. H. Chang. Depth recovery from motion blurred images. *International Conference on Pattern Recognition*, 1:135–138, 2006. 1

[12] D. J. C. Mackay. Introduction to Monte Carlo methods. In M. I. Jordan, Ed., "Learning in graphical models", NATO Science Series, 175-204, Kluwer, 1998. 2, 3

[13] L. Matthies, T. Kanade, and R. Szeliski. Kalman filter-based algorithms for estimating depth from image sequences. *International Journal of Computer Vision*, 3(3), 1989. 2

[14] G. Qian, R. Chellappa, and Q. Zheng. Robust structure from motion estimation using inertial data. *J. Opt. Soc. Am. A*, 18(12):2982–2997, 2001. 2

[15] M. Sorel and J. Flusser. Space-variant restoration of images degraded by camera motion blur. *IEEE Trans. Imag. Proc.*, 17(2):105–116, 2008. 1, 2, 5, 6

[16] F. Sroubek and J. Flusser. Multichannel blind deconvolution of spatially misaligned images. *IEEE Trans. Imag. Proc.*, 14(7):874–883, 2005. 2

[17] G. R. K. S. Subrahmanyam, A. N. Rajagopalan, and R. Aravind. A recursive filter for despeckling SAR images. *IEEE Trans. Imag. Proc.*, 17(10):1969–1974, 2008. 2, 3

[18] Y. Tai, H. Du, M. Brown, and S. Lin. Image/video deblurring using a hybrid camera. *IEEE Conference on Computer Vision and Pattern Recognition*, 2008. 1

[19] L. Yuan, J. Sun, L. Quan, and H. Y. Shum. Image deblurring with blurred/noisy image pairs. *ACM Transactions on Graphics*, 26(3):1–10, 2007. 2

[20] J. Ziegler, K. Nickel, and R. Stiefelhagen. Tracking of the articulated upper body on multi-view stereo image sequences. *IEEE Conference on Computer Vision and Pattern Recognition*, 1:774–781, 2006. 2