A Variational Approach to Scene Reconstruction and Image Segmentation from Motion-Blur Cues

P. Favaro S. Soatto

Technical Report CSD-TR 040011

Dept. of Computer Science, UCLA, Los Angeles, CA 90095
e-mail: *favaro,soatto@ucla.edu*

Abstract

In this paper we are interested in the joint reconstruction of geometry and photometry of scenes with multiple moving objects from a collection of motion-blurred images. We make simplifying assumptions on the photometry of the scene (we model each object in the scene as self-luminous) and infer the motion field of the scene, its depth map, and its radiance. In particular, we choose to partition the image into regions where motion is well approximated by a simple planar translation. We model motion-blurred images as the solution of an anisotropic diffusion equation, whose initial conditions depend on the radiance and whose diffusion tensor encodes the depth map of the scene and the motion field. We propose an algorithm to infer the unknowns of the model. Inference is performed by minimizing the discrepancy between the measured images and the ones synthesized via diffusion. Since the problem is ill-posed, we also introduce additional Tikhonov regularization terms.

1. Introduction

Motion blurring is a common distortion of images that becomes perceivable when objects in the scene move at a speed higher than the speed of the shutter of the camera [1]. Given motion blurred images, one may be interested in recovering a sharp or deblurred image of the scene. In order to do so, one needs to recover both the deblurred image and some description of the motion of the scene. For example, one can assume that the motion characterizing a motion blurred image can be represented by a two dimensional velocity vector. This assumption, however, is not realistic when multiple objects are simultaneously moving with different speed and/or along different directions. In this case, the complexity of motion cannot be captured by a single two dimensional vector. In order to model a complex motion one can choose a very rich global model, that explains the motion of the entire image, or a very simple model, selected from a small parametric class, together with a segmentation of the regions of the images where the model is satisfied within a prescribed accuracy. In this paper we choose the latter approach. For simplicity, we adopt the simplest possible model, i.e. that each region moves with constant, purely translational motion. Notice that while this would be a severe restriction for a *global* motion model, any motion field can be approximated locally by a pure translation to an arbitrary degree of accuracy. Naturally, the price we pay for such a model is that the partitioning process may result in very fine segments, hence over-segmenting the scene. Subsequent aggregation can be performed based on richer motion models, but we do not address this issue in this paper.

1.1. Existing Work and Contributions of this Paper

As we mentioned in the previous section, motion blur is a phenomenon that becomes perceivable whenever we capture images of an object that is moving faster than the shutter of the camera. Given a description of the geometry and the appearance of a scene, and its motion, one may be interested in *simulating* (or *rendering*) images with motion blur [2, 3]. This problem is also called a *direct problem* since it aims at mimicking the physical process as it happens in nature [4]. One may also be interested in the *inverse problem*, i.e. in the problem of inferring a description of the scene (geometry and appearance) and of its motion, given motion-blurred images [1, 5, 6, 7, 8, 9]. This problem is called *motion deblurring*, or *motion smear* [10, 11], or *super-resolution* [12, 13] when the deblurred image is reconstructed at a resolution higher than the resolution of the input images.

Most of the approaches for motion deblurring are based on using a single image in input [14, 5, 6, 7, 8, 1]. In this case, one has to introduce strong assumptions on the scene and/or the blurring (see Remark 1). For example, in [14] the blurring kernel support is assumed known. In [6], the motion of the scene is known. In [8] the radiance is assumed to be isotropic. These assumptions are unavoidable due to the severe lack of data, and introduce constraints that are too restrictive.

Some work has also been done when multiple images are used [9, 15]. [9] uses images captured while the scene is moving along different motion directions. They consider blur is shift-invariant and that a single object is moving in the scene. In [15] deblurring is also performed by using multiple images of the same scene captured for different shutter intervals and pixel resolutions. They propose an innovative hybrid camera that can also estimate the path of the moving scene, up to the resolution of the fastest camera.

We propose a novel approach to motion-deblurring and scene reconstruction when multiple objects are simultaneously moving in a scene. Some work has been done along this direction [6, 7], although only one image is considered as an input. In this paper, instead, we consider multiple images so as to avoid introducing additional assumptions on the unknowns. We assume we are made available images captured for different shutter intervals. In addition, we model the motion of the objects on the images by considering their three dimensional geometry and their three dimensional motion, which has not been done in the context of motion-deblurring. Some work has been done towards recovering depth information from motion-blurred images, but restricted to the specific motion generated by lens zooming and for a single object in the scene [16]. Also, our approach differs from most of the previous approaches in that we pose our inference problem as an optimization procedure in a variational framework. To the best of our knowledge, the only other work in this framework is the recent paper in [6]. However, in [6] only a single image is used, the motion on the image plane of the objects is assumed known, and there is no geometric model of the scene.

2. Notation and Problem Formulation

We represent an image with a function $I : \Omega \subset \mathbb{R}^2 \mapsto [0, \infty)$, that assigns an energy value to each pixel on the image plane. We assume that Ω is a bounded domain with piecewise smooth boundary $\partial\Omega$. The intensity of the measured energy depends on the reflectivity properties of the surfaces of the objects in the scene, which we describe with a function $r : \mathbb{R}^2 \mapsto [0, \infty)$; r assigns an energy value at each point on the surface of the objects and it is called, with an abuse of terminology[1], *radiance* of the scene.

We capture images from scenes where a number of objects are moving in different directions, possibly with different speed. For now, assume the scene is made of a single object. If the camera shutter remains open while the object is moving with velocity v for a time interval ΔT, then the image I that we measure on the image plane can be modeled by:

$$I(x) = \frac{1}{\Delta T} \int_{-\frac{\Delta T}{2}}^{\frac{\Delta T}{2}} r(x + vt)dt = \int_{-\frac{1}{2}}^{\frac{1}{2}} r(x + \Delta T vt)dt \tag{1}$$

which we approximate with the following

$$I(x) \simeq \int \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} r(x + \Delta T vt)dt. \tag{2}$$

Now consider the scene is composed of M objects that are moving simultaneously in front of the camera. Denote with $\{\Omega_j\}_{j=1\ldots M}$ the regions on the image plane occupied by the projections of each of the moving objects. We assume that $\{\Omega_j\}_{j=1\ldots M}$ is a *partition* of Ω, i.e. that $\Omega = \bigcup_{j=1}^{M} \Omega_j$ and that $\Omega_j \bigcap \Omega_i = \emptyset$ for $\forall i, j = 1 \ldots M$, $i \neq j$. In this case, the image model becomes

$$I(x) = \int \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} r(x + \Delta T v_j t)dt \quad \forall x \in \Omega_j. \tag{3}$$

Consider that we are made available N images $\{J_1, \ldots, J_N\}$ collected while the shutter remains open for different spans of time $\{\Delta T_1, \ldots, \Delta T_N\}$. Then, one can pose the problem of inferring velocities $\{v_j\}_{j=1\ldots M}$, partition $\{\Omega_j\}_{j=1\ldots M}$ and radiance r of the scene as the following minimization:

$$\hat{v}_j, \hat{\Omega}_j, \hat{r} = \arg \min_{v_j, \Omega_j, r} \sum_{i=1}^{N} \int_{\Omega} (J_i(x) - I_i(x))^2 dx \tag{4}$$

[1] In the context of *radiometry*, the term *radiance* refers to a more complex object that describes energy emitted along a certain direction, per solid angle, per foreshortened area and per time instant. Here we are implicitly assuming that scene radiance and image irradiance are the same, which is an approximation that is only valid for Lambertian scenes under uniform illumination.

where $\{J_i\}_{i=1\ldots N}$ are images measured on the image plane, while $\{I_i\}_{i=1\ldots N}$ are images synthesized by using the image model eq. (3).

Remark 1. *The problem in eq. 4 is an inverse problem and it is known to be* ill-posed*. One of the main factors that cause the ill-posedness of this problem is the lack of data. For the sake of example, let us consider the simpler case where a single object is moving in the scene. In this case, the problem amounts to recovering the velocity v of the scene and to restoring the radiance r. It is immediate to see that there are infinite solutions to the problem when only a single image J is used. For example, $\{\hat{r}, \hat{v}\} = \{J, 0\}$ and $\{\hat{r}, \hat{v}\} = \{r, v\}$ are both valid solutions. More in general, the following is also a valid (infinite) set of solutions:*

$$\begin{aligned} \hat{r} &= \int \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} r\left(x + \frac{(1-\alpha)}{\sqrt{(1-\alpha)^2+\alpha^2}} v\tau\right) d\tau \\ \hat{v} &= \frac{\alpha}{\sqrt{(1-\alpha)^2+\alpha^2}} v \end{aligned} \tag{5}$$

for all $\alpha \in [0, 1]$.

Remark 2. *One may raise the concern that capturing images of the same scene but with different shutter intervals might present some technical difficulties. Here we propose two ways to perform this operation. The most straightforward method is to use different cameras that can capture simultaneously images with different shutter intervals. However, in this case one might encounter some difficulties in registering the images with each other and in synchronizing the cameras. [15] describes hardware that can be used in this modality. Another way to capture images with different shutter intervals is to collect a sequence of images. Time averaging the sequence simulates a long shutter interval. For example, one could collect three motion-blurred images $[\bar{J}_1, \bar{J}_2, \bar{J}_3]$, and then consider $J_1 = \bar{J}_2$ as one input image and $J_2 = \frac{1}{3}\sum_{i=1}^{3} \bar{J}_i$ as a second input image. The shutter interval for the second image J_2 is 3 times the shutter interval of the first image J_1. In this case, the data collection is rather simple, since no alignment and no synchronization is required, but it is based on the assumption that motion does not change among the three frames. Due to its simplicity, in this manuscript we choose this second modality for data collection.*

3. Modeling Motion-Blur of Multiple Objects

In the previous section we briefly introduced a model for motion blurred images in eq. (3). The model was described by a certain motion v_j, a region Ω_j corresponding to the motion v_j, and the radiance r of the scene. In the next subsection, we will specify more in detail how the motion v_j depends on the surfaces in the scene and the 3D motion of the scene. Then, in subsection 3.2 we will introduce an alternative model to eq. (3) based on anisotropic diffusion.

3.1. A Model for Motion of Multiple Objects

We denote the surfaces of the objects with a function $s : \mathbb{R}^2 \mapsto [0, \infty)$ that assigns a depth value to each pixel coordinate and it is called *depth map*. A point on the depth map s at time t can be written as

$$X(t) = [x(t)\ 1]^T\ s(x(t)) \tag{6}$$

where $x(t) \in \mathbb{R}^2$ are the 2D coordinates of a pixel. We denote with $V = [V_X(t)\ V_Y(t)\ V_Z(t)]^T \in \mathbb{R}^3$ the translational velocity and with $\omega \in \mathbb{R}^3$ the rotational velocity of one of the objects in the scene. Then, it is well known that the time derivative of the coordinates x satisfies (see [17] for more details):

$$\begin{aligned} \dot{x}(t) &= \tfrac{1}{s(x(t))} \begin{bmatrix} F & 0 & -x_1(t) \\ 0 & F & -x_2(t) \end{bmatrix} V + \\ &\quad + \begin{bmatrix} -x_1(t)x_2(t) & 1+x_1^2(t) & -x_2(t) \\ -1-x_2^2(t) & x_1(t)x_2(t) & x_1(t) \end{bmatrix} \omega. \end{aligned} \tag{7}$$

We define $v \doteq \dot{x}(t)$ and call it the *velocity field*.

As we have anticipated, we restrict ourselves to a crude motion model that only represents sideways translations parallel to the image plane

$$v(t) = F\frac{V_{X,Y}(t)}{s(x(t))} = \frac{\bar{V}_{X,Y}(t)}{s(x(t))} \tag{8}$$

where $\bar{V}_{X,Y} = FV_{X,Y}$ is the velocity in focal length units. From now on we will not make a distinction between $\bar{V}_{X,Y}$ and $V_{X,Y}$, and use V to denote $V_{X,Y}$ for simplicity. Although we derived the velocity field only for the case of translational motion, it is straightforward, in principle, to extend it to the general case of eq. (7). Conceptually, however, both cases correspond to a chosen motion model, and given that the scene in general will violate it, we will have to segment it into regions that satisfy the model. Therefore, we concentrate on the simplest possible model, aware of the fact that simpler models will generate finer partitions and therefore more fragmentation of the image.

When we have M objects moving in the scene, or even when we have a single object that is moving with a more general motion, such as general rigid motion, or piecewise rigid, or even non-rigid, we decompose the scene into segments each of which corresponds to a portion that is well-modeled by pure translational motion. Now, assume we have M objects moving in the scene with constant velocities $V_1 \dots V_M$, the velocity field v can be partitioned into a number of regions $\{\Omega_j\}_{j=1\dots M}$ each corresponding to a different velocity V_j (see also section 2). Since there is a *scale ambiguity* between the magnitude of the velocity V_j of a region Ω_j and the magnitude of the corresponding depth map s (see eq. (8)), objects that are moving along the same direction are clustered together. In other words, we can only partition the velocity field into regions with *uniform motion direction*.

In our implementation, we represent the regions implicitly using signed distance functions [18, 19]. The regions $\{\Omega_j\}_{j=1\dots M}$ are implicitly represented by levelset functions. For simplicity, we consider the case of two regions, so that a single levelset function suffices. However, the extension to more than two regions is straightforward and can be achieved by considering more levelset functions. The levelset function ϕ is a map $\phi : \Omega \mapsto \mathbb{R}$, so that

$$\begin{array}{rcl} \Omega_1 & = & \{x \in \Omega : \phi(x) \geq 0\} \\ \Omega_2 & = & \{x \in \Omega : \phi(x) < 0\} = \Omega \backslash \Omega_1. \end{array} \tag{9}$$

Using the Heaviside function H

$$H(z) = \begin{cases} 1, & \text{if } z \geq 0 \\ 0, & \text{if } z < 0 \end{cases} \tag{10}$$

we can equivalently write

$$\begin{array}{rcl} \Omega_1 & = & \{x \in \Omega : H(\phi(x)) = 1\} \\ \Omega_2 & = & \{x \in \Omega : H(\phi(x)) = 0\}. \end{array} \tag{11}$$

This notation will be useful later when we will define more explicitly the cost functional introduced in eq. (4).

3.2. A Model for Motion-Blurred Images

Under the assumption that the depth map s is smooth, we can substitute the model in eq. (2) with a PDE whose solution $u : \mathbb{R}^2 \times [0, \infty) \mapsto \mathbb{R}$, $(x, t) \mapsto u(x, t)$, at each time t represents an image with a certain amount of blurring. In formulas, we have that $J(y) = u(y, T)$, where T is related to the amount of blurring of J. We use the following *anisotropic diffusion* partial differential equation:

$$\begin{cases} \dot{u}(y,t) = \nabla \cdot (D(y) \nabla u(y,t)) & t > 0 \\ u(y,0) = r(y) & \forall y \in \Omega \end{cases} \tag{12}$$

where $D \doteq \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix}$ with $d_{ij} : \mathbb{R}^2 \mapsto \mathbb{R}$ for $i, j = 1, 2$ and $d_{12} \equiv d_{21}$, is called *diffusion tensor*. We assume that $d_{ij} \in C^1(\mathbb{R}^2)$ (i.e. the space of functions with continuous partial derivatives in $\mathbb{R}^2$) for $i, j = 1, 2$, and[2] $D(y) \geq 0 \; \forall y \in \mathbb{R}^2$. The symbol ∇ is the gradient operator $\left[\frac{\partial}{\partial y_1} \;\; \frac{\partial}{\partial y_2}\right]^T$ with $y = [y_1 \;\; y_2]^T$, and the symbol $\nabla \cdot$ is the divergence operator $\sum_{i=1}^{2} \frac{\partial}{\partial y_i}$. Notice that there is a scale ambiguity between the time T and the diffusion tensor D. We will set $T = \frac{1}{2}$ to resolve this ambiguity.

When the motion field is constant, it is easy to show that $2tD = \Delta T^2 v v^T$. In particular, at time $t = T = \frac{1}{2}$ we have $D = \Delta T^2 v v^T$. Now, in the space-varying case we let

$$D(y) = \Delta T^2 v(y) v(y)^T. \tag{13}$$

In particular, when eq. (8) is satisfied, we have

$$D(y) = \Delta T^2 \frac{V V^T}{s^2(y)}. \tag{14}$$

[2] Since D is a tensor, the notation $D(y) \geq 0$ means that $D(y)$ is positive semi-definite.

Notice that the diffusion tensor just defined is guaranteed to be always positive semi-definite.

Remark 3. *The advantage of using the PDE-based model just introduced in eq.* (12) *versus using the integral-based model of eq.* (3) *becomes more evident at the algorithmic implementation level. The two models yield (approximately) the same solutions, but behave differently in the cases of motion-blurring due to small velocities and motion-blurring due to large velocities. The integral-based model is more efficient in the latter case, but less efficient in the former one. Vice versa, the PDE-based model is more efficient for small velocities, but more inefficient for large ones. The range of velocities for which our problem yields a sensible solution is more biased towards small velocities, thus favoring the PDE-based model.*

3.3. Motion-Blur Segmentation and Image Restoration

We infer the radiance r, the depth map s, the velocities $\{V_1, V_2\}$ and the partition $\{\Omega_1, \Omega\backslash\Omega_1\}$ of the scene by minimizing the following least-squares functional with Tikhonov regularization (cf. [20])

$$\begin{aligned} E \quad = \quad & \sum_{i=1}^{N} \int_{\Omega_1} \left(u_i(x,T,V_1) - I_i(x)\right)^2 dx + \\ & + \int_{\Omega\backslash\Omega_1} \left(u_i(x,T,V_2) - I_i(x)\right)^2 dx + \\ & + \alpha \left\| r - r^* \right\|^2 + \beta \left\| \nabla s \right\|^2 + \\ & + \gamma \left(\int_{\Omega} s(x) dx - M \right)^2 + \nu \left\| \nabla H(\phi(x)) \right\|^2, \end{aligned} \tag{15}$$

i.e. we seek for

$$\hat{\Omega}_1, \hat{r}, \hat{s}, \hat{V}_1, \hat{V}_2 \quad = \quad \arg \min_{\Omega_1, r, s, V_1, V_2} E \tag{16}$$

where α, β, γ and ν are positive regularization parameters, r^* is a prior[3] for r and M is a suitable positive number[4]. The last term imposes a length constraint on the boundary of Ω_1. One can choose the norm $\| \cdot \|$ depending on the desired space of solutions. We choose the L_2 norm for the radiance and the components of the gradient of the depth map.

In this functional, the first two terms take into account the discrepancy between the model and the measurements; the third and fourth term are classical regularization functionals, penalizing large deviations of the radiance from the prior and imposing some regularity on the estimated depth map. The fifth term fixes the scale ambiguity between the depth map s and the velocity field v. To fix the scale ambiguity we choose the mean of the depth map s to be equal to a constant M, so that small changes of s will not result in sensible variations of this term. Finally, the last term imposes a length constraint on the boundary of Ω_1 thus penalizing boundaries that are too fragmented or irregular.

To minimize the cost functional (16) we employ a gradient descent flow. For each unknown we compute a sequence converging to a local minimum of the cost functional, i.e. we have sequences $\hat{r}(x,\tau), \hat{s}(x,\tau), \hat{V}_1(\tau), \hat{V}_1(\tau), \hat{\phi}(\tau)$, such that

$$\begin{aligned} \hat{r}(x) &= \lim_{\tau \mapsto \infty} \hat{r}(x,\tau), \\ \hat{s}(x) &= \lim_{\tau \mapsto \infty} \hat{s}(x,\tau), \\ \hat{V}_1 &= \lim_{\tau \mapsto \infty} \hat{V}_1(\tau), \\ \hat{V}_2 &= \lim_{\tau \mapsto \infty} \hat{V}_2(\tau), \\ \hat{\phi}(x) &= \lim_{\tau \mapsto \infty} \hat{\phi}(x,\tau), \end{aligned} \tag{17}$$

At each iteration we update the unknowns by moving in the opposite direction of the gradient of the cost functional with

[3] We do not have a preferred prior for the radiance r. However, it is necessary to introduce this term to guarantee that the estimated radiance does not diverge. In practice, one can use as a prior r^* one of the input images, or a combination of them, and choose a very small α.

[4] As mentioned in subsection 3.1, there is a scale ambiguity between the velocity field v and the depth map of the scene. We choose to fix a scale quantity of the depth map rather than fixing the velocity field. We choose the mean of the depth map s to be equal to the constant M.

respect to the unknowns. In other words, we let

$$
\begin{aligned}
\frac{\partial \hat{r}(x,\tau)}{\partial \tau} &\doteq -\nabla_{\hat{r}} E(x), \\
\frac{\partial \hat{s}(x,\tau)}{\partial \tau} &\doteq -\nabla_{\hat{s}} E(x), \\
\frac{\partial \hat{V}_1(\tau)}{\partial \tau} &\doteq -\nabla_{\hat{V}_1} E, \\
\frac{\partial \hat{V}_2(\tau)}{\partial \tau} &\doteq -\nabla_{\hat{V}_2} E, \\
\frac{\partial \hat{\phi}(x,\tau)}{\partial \tau} &\doteq -\nabla_{\hat{\phi}} E(x).
\end{aligned} \tag{18}
$$

It can be shown that the above iterations decrease the cost functional as τ increases. The computation of the above gradients is rather involved due to the fact that the explicit solution u of eq. (12) is not available, but yields the following formulas that can be easily implemented:

$$
\begin{aligned}
\nabla_r E(x) &= \sum_{i=1}^{N} w_i(x,T,V_1) H(\phi(x)) + \\
&\quad + \sum_{i=1}^{N} w_i(x,T,V_2)(1 - H(\phi(x))) + \\
&\quad + 2\alpha \left(r(x) - r^*(x)\right) \\
\nabla_s E(x) &= 2\frac{V_1^T e_1(x) V_1}{s^3(x)} H\left(\phi(x)\right) + \\
&\quad + 2\frac{V_2^T e_2(x) V_2}{s^3(x)} \left(1 - H\left(\phi(x)\right)\right) + \\
&\quad + 2\gamma \left(\int_\Omega s(x)dx - M\right) \\
\nabla_{V_1} E &= -\int_{\Omega_1} \frac{[1\ 0] e_1(x) V_1 + V_1^T e_1(x) \begin{bmatrix} 1 \\ 0 \end{bmatrix}}{s^2(x)} dx \\
\nabla_{V_2} E &= -\int_{\Omega \setminus \Omega_1} \frac{[1\ 0] e_2(x) V_2 + V_2^T e_2(x) \begin{bmatrix} 1 \\ 0 \end{bmatrix}}{s^2(x)} dx \\
\nabla_\phi E(x) &= \left(g_1(x) - g_2(x) - \nabla \cdot \frac{\nabla \phi(x)}{|\nabla \phi(x)|}\right) \delta(\phi(x))
\end{aligned} \tag{19}
$$

where for $j = 1, 2$ we define

$$
\begin{aligned}
e_j(x) &= \sum_{i=1}^{N} \int_0^T \nabla u_i(x,t,V_j) \nabla w_i(x,T-t,V_j)^T dt \\
g_j(x) &= \sum_{i=1}^{N} \left(u_i(x,T,V_j) - I_i(x)\right)^2
\end{aligned} \tag{20}
$$

and $w_i(x,t,V_j)$ satisfies the following adjoint parabolic equation

$$
\begin{cases}
\dot{w}(y,t) = \nabla \cdot (D(y) \nabla w(y,t)) \\
w(y,0) = u(y,T) - I_i(y) \\
(D(y) \nabla w(y,t)) \cdot n = 0
\end{cases} \tag{21}
$$

with $D(x) = \Delta T_i^2 \frac{V_j V_j^T}{s^2(x)}$. Similarly, the notation $u_i(x,t,V_j)$ denotes the solution of the PDE eq. (12) where the diffusion tensor $D(x) = \Delta T_i^2 \frac{V_j V_j^T}{s^2(x)}$.

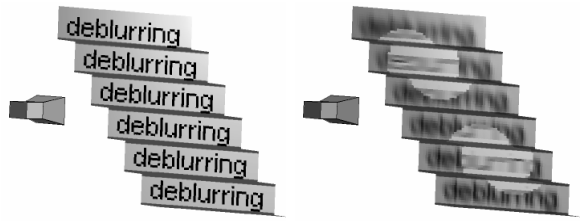

Figure 1: Left: setup of the scene without motion-blur (static scene and static camera). The depth map is stair-shaped. The steps on the top are closer to the camera than the steps on the bottom. Right: setup of the scene with motion-blur. The two disks on the stair move from left to right, while the remaining part of the stair moves along the top-left to bottom-right diagonal. The texture of the two disks has been brightened to make them more visible.

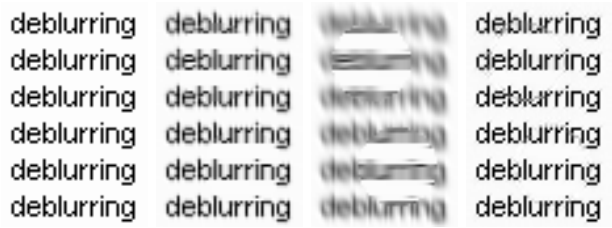

Figure 2: First from the left: synthetically generated radiance. It is the image of the scene when both scene and camera are static. Second and third from the left: motion-blurred images captured with different shutter intervals. The motion blur of the third image is three times the motion blur of the second image. Rightmost: final deblurred radiance estimated from the two input images. The reconstruction presents some artifacts at the locations corresponding to the boundaries of the disks.

4. Experiments with Synthetic Data

In this set of experiments we synthetically generate a scene whose depth map is a stair-shaped object (see Figure 1). Two disks at opposite corners (see image on the right in Figure 1) move sideways (left to right) while the remaining part of the object moves along the top-left to bottom-right diagonal. When the scene is static, the image we capture coincides with the radiance of the object (see leftmost image in Figure 2). The second and third image from the left of Figure 2 show the two input images captured for different shutter intervals. The shutter interval of the third image is three times the shutter interval of the second image. Also notice that the amount of motion blur is larger on the top of the image than on the bottom. This effect is due to the depth map of the scene. The rightmost image of Figure 2 is the resulting deblurred image that we restored from the given input. Notice that the reconstruction is fairly close to the original radiance (leftmost image in Figure 1), although there are some artifacts at locations corresponding to the boundary of the two disks. This is due to the error between the correct segmentation of the scene and the estimated segmentation (see Figure 3). In Figure 3 we show a few snapshots of the segmentation evolution of the two moving objects. The motion field direction is correctly estimated. Also, notice that the levelset representation easily handles topological changes of the represented contour. In Figure 4 we show some snapshots of the deblurring evolution. More precisely, the first three snapshots from the left correspond to the first three steps

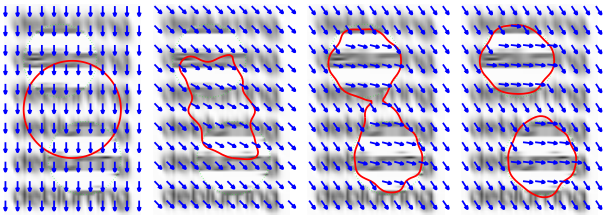

Figure 3: Snapshots of the evolution of segmentation together with motion estimation on synthetic data. Motion is initialized with vertical direction.

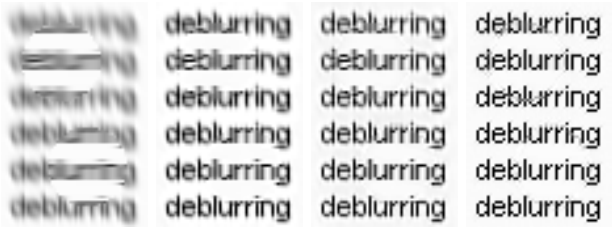

Figure 4: Snapshots of the evolution of the deblurring of the radiance. Leftmost: the radiance is initialized with the most motion-blurred image. Second and third from the left: at the second and third iteration, the radiance sharpness improves dramatically. Rightmost: the recovered radiance compares well with the original radiance (see leftmost image in Figure 2).

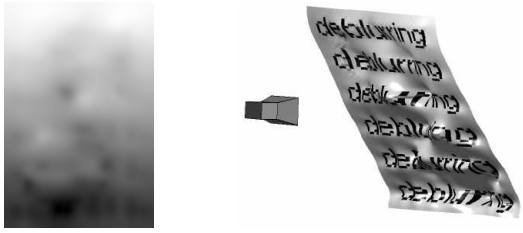

Figure 5: Estimated depth map. Left: visualization of the depth map as a gray level intensities image. Light intensities correspond to points that are close to the camera, while dark intensities correspond to points that are far from the camera. Right: reconstruction of the setup of the scene by using the recovered radiance and the reconstructed depth map.

in the iterative scheme, while the rightmost snapshot corresponds to the last estimation step of the radiance. Although the radiance is initialized with the most blurred image (the third image from the left in Figure 2), it converges rather quickly to the deblurred image. Finally, in Figure 5 we show the reconstructed scene with the estimated depth map. On the left we have a gray level image of the estimated depth map. Light intensities correspond to points that are close to the viewer, while dark intensities correspond to points that are far from the viewer. On the right we show the reconstructed setup using the estimated depth map and the recovered radiance.

5. Experiments with Real Data

To capture real images with different shutter intervals we use the modality described in section 2, i.e. we capture three motion-blurred images $[\bar{J}_1, \bar{J}_2, \bar{J}_3]$, and then consider $J_1 = \bar{J}_2$ as one input image and $J_2 = \frac{1}{3}\sum_{i=1}^{3} \bar{J}_i$ as a second input image. As in the previous section, the shutter interval for the second image J_2 is 3 times the shutter interval of the first image J_1. In Figure 6 at the top-left corner we show an image of the scene when static. This image coincides with the radiance of the scene. At the top-right corner we show the recovered image obtained by using our algorithm. As in the experiments with synthetic data, the reconstruction is fairly close to the radiance of the scene, although there are some artifacts at locations corresponding to the boundary of the segmented regions. In input we use the image at the bottom-left corner (which corresponds to J_1) and the image at the bottom-right corner (which corresponds to J_2) of Figure 6. The background is moving vertically, while the foreground (the cup and the banana) are moving horizontally. In Figure 7 we show a few snapshots of the segmentation evolution. To make the contour more visible in the illustrations, we changed the original brightness of the image underneath. Notice that the motion field direction of the scene is correctly estimated. In Figure 8 we show some snapshots of the deblurring evolution. We use as initial radiance the most motion-blurred image (top-left). The final estimate of the radiance (bottom-right) is also shown in Figure 6 for comparison with the original radiance. Finally, in Figure 9 we show the reconstructed depth map of the scene. The images on the top show both the estimated background and foreground depth map. Notice that the relative position of the two depth maps does not correspond to the depth map of the original scene. This inconvenience is due to the scale ambiguity between the depth map and the velocity of the scene (see section 3.1). The left and right bottom images show two views of the estimated depth map of the foreground. Notice that the qualitative shape of the cup and the banana has been captured.

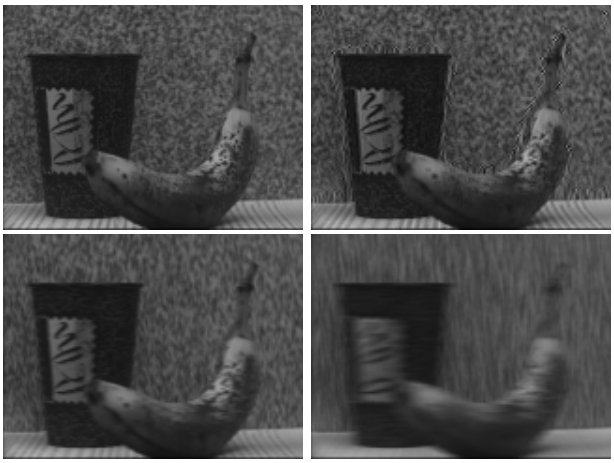

Figure 6: Top-left: original radiance. This image has been captured when the scene and the camera were static. Top-right: recovered radiance. Bottom-left and Bottom-right: input motion-blurred images. The image on the right has been obtained by averaging 3 motion-blurred images similar to the image on the left. The shutter interval of the image on the right is three times the shutter interval of the image on the left.

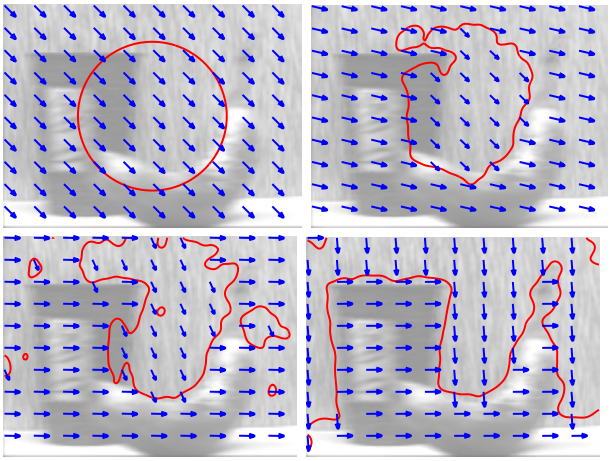

Figure 7: Snapshots of the evolution of the segmentation. The brightness of one of the motion-blurred images has been changed to enhance the contrast between the image and the contour evolution.

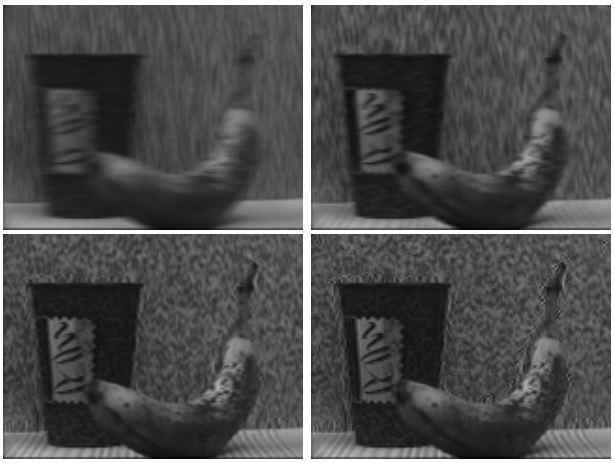

Figure 8: Snapshots of the evolution of deblurring. The radiance is initialized with the most motion-blurred image. The final estimate of the radiance compares well with the original radiance in Figure 6. However, at locations corresponding to the boundary of the segmented regions the reconstruction introduces some artifacts.

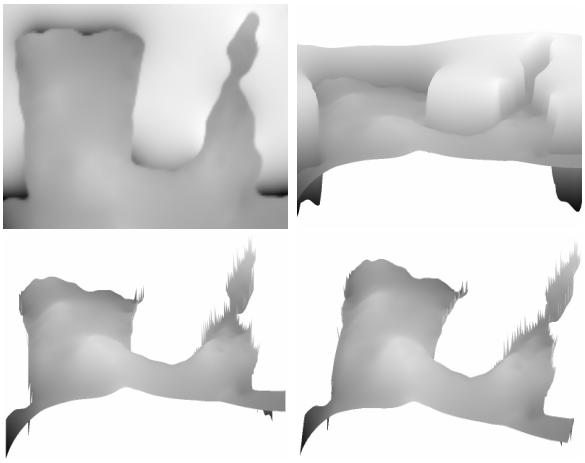

Figure 9: Estimated depth map. Top-left: depth map visualized as gray level intensities. Lighter intensities correspond to points close to the camera, while darker intensities correspond to points far from the camera. Top-right: different view of the estimated depth map. Bottom-left and bottom-right: views of the estimated depth map of the foreground. Notice that the qualitative shape of the cup and of the banana have been captured.

6. Summary and Conclusions

We presented a solution to the problem of jointly reconstructing scenes and restoring images from images affected by motion blurring due to multiple moving objects. We inferred motion field of a scene, depth map and radiance from a collection of motion-blurred images obtained for different shutter intervals. The presence of multiple objects in the scene, that are moving along different directions, induces a complex motion field on the blurred images. We found that a good tradeoff between complexity of the model and accuracy of the representation is to segment the motion field into regions with uniform translational motion. In addition, we proposed to model motion-blurred images as the solution of an anisotropic diffusion equation, whose initial conditions depend on the radiance and whose diffusion tensor encodes the depth map of the scene and the motion field. Finally, an algorithm to infer the unknowns of the model in presented. Inference is performed by minimizing the discrepancy between the measured images and the ones synthesized via diffusion, which we regularize via additional Tikhonov regularization terms.

References

[1] Kang, S., Min, J., Paik, J.: Segmentation-based spatially adaptive motion blur removal and its application to surveillance systems. In: ICIP01. (2001) I: 245–248

[2] Brostow, G.J., Essa, I.: Image-based motion blur for stop motion animation. In Fiume, E., ed.: SIGGRAPH 2001, Computer Graphics Proceedings. (2001) 561–566

[3] Kubota, A., Aizawa, K.: Arbitrary view and focus image generation: rendering object-based shifting and focussing effect by linear filtering. In: ICIP02. (2002) I: 489–492

[4] Bertero, M., Boccacci, P.: Introduction to inverse problems in imaging. Institute of Physics Publishing, Bristol and Philadelphia (1998)

[5] Tull, D., Katsaggelos, A.: Regularized blur-assisted displacement field estimation. In: ICIP96. (1996) III: 85–88

[6] Kim, J., Tsai, A., Cetin, M., Willsky, A.: A curve evolution-based variational approach to simultaneous image restoration and segmentation. In: ICIP02. (2002) I: 109–112

[7] Kang, S., Choung, Y., Paik, J.: Segmentation-based image restoration for multiple moving objects with different motions. In: ICIP99. (1999) I:376–380

[8] Yitzhaky, Y., Mor, I., Lantzman, A., Kopeika, N.: Direct method for restoration of motion blurred images. JOSA-A **15** (1998) 1512–1519

[9] Rav-Acha, A., Peleg, S.: Restoration of multiple images with motion blur in different directions. In: IEEE Workshop on Applications of Computer Vision (WACV). (Palm-Springs, 2000) 22–28

[10] Hammett, S., Georgeson, M., Gorea, A.: Motion blur and motion sharpening: temporal smear and local contrast non-linearity. In: Vision Research. Volume 38. (1998) 2099–2108

[11] Chen, W., Nandhakumar, N., Martin, W.: Image motion estimation from motion smear: A new computational model. PAMI **18** (1996) 412–425

[12] Zomet, A., Rav-Acha, A., Peleg, S.: Robust super-resolution. In: CVPR01. (2001) I:645–650

[13] Borman, S., Stevenson, R.L.: Super-Resolution from Image Sequences - A Review. In: Proceedings of the 1998 Midwest Symposium on Circuits and Systems, Notre Dame, IN (1998)

[14] You, Y., Kaveh, M.: Blind image restoration by anisotropic diffusion. IEEE Trans. on Image Processing **8** (1999) 396–407

[15] Ben-Ezra, M., Nayar, S.: Motion deblurring using hybrid imaging. In: Computer Vision and Pattern Recognition. Volume 1. (2003) 657–664

[16] Ma, J., Olsen, S.I.: Depth from zooming. In: JOSA A. Volume 7. (1990) 1883–90

[17] Heeger, D., Jepson, A.: Subspace methods for recovering rigid motion i. Int. J. of Computer Vision **7** (1992) 95–117

[18] Osher, S., Sethian, S.: Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulation. J. of Comput. Phys. **79** (1998) 12–49

[19] Chan, T., Vese, L.: Active contours without edges. IEEE Transactions on Image Processing **10** (2001) 266–77

[20] Engl, H., Hanke, M., Neubauer, A.: Regularization of Inverse Problems. Kluwer Academic Publishers, Dordrecht (1996)