
International Journal of Computer Vision 38(3), 231–245, 2000

Probabilistic Detection and Tracking of Motion Boundaries

MICHAEL J. BLACK
Xerox Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, CA 94304, USA;
Department of Computer Science, Brown University, Box 1910, Providence, RI 02912, USA
black@cs.brown.edu

DAVID J. FLEET
Xerox Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, CA 94304, USA;
Department of Computing Science, Queen's University, Kingston, K7L 3N6, Canada
fleet@cs.queensu.ca

Abstract. We propose a Bayesian framework for representing and recognizing local image motion in terms of two basic models: translational motion and motion boundaries. Motion boundaries are represented using a non-linear generative model that explicitly encodes the orientation of the boundary, the velocities on either side, the motion of the occluding edge over time, and the appearance/disappearance of pixels at the boundary. We represent the posterior probability distribution over the model parameters given the image data using discrete samples. This distribution is propagated over time using a particle filtering algorithm. To efficiently represent such a high-dimensional space we initialize samples using the responses of a low-level motion discontinuity detector. The formulation and computational model provide a general probabilistic framework for motion estimation with multiple, non-linear, models.

Keywords: motion discontinuities, occlusion, optical flow, Bayesian methods, particle filtering

1. Introduction

A particularly rich source of visual motion information is found at surface boundaries where optical flow is typically discontinuous due to motion parallax or the independent motion of objects. This gives rise to *motion boundaries* between adjacent image regions having different image velocities. These motion boundaries provide information about the position and orientation of surface boundaries in the scene. Moreover, analysis of the occlusion or disocclusion of pixels at motion boundaries can provide information about the relative depth ordering of the neighboring surfaces. In turn, information about surface boundaries and depth ordering may be useful for tasks as diverse as navigation, structure from motion, video compression, perceptual organization, and object recognition.

While motion boundaries provide valuable information about the scene, they also cause problems for optical flow techniques that assume the image motion is spatially smooth. Therefore, the detection of motion boundaries is often seen as a means for improving optical flow estimation. This, combined with the salience of motion boundaries for inferring scene properties, has made the detection and analysis of motion boundaries an important research topic in computer vision. Despite this, approaches reported to date have produced somewhat disappointing experimental results and motion boundary detection remains problematic.

In this paper we formulate a probabilistic, model-based, approach to image motion analysis; that is, the image motion in each local neighborhood of an image is estimated and represented using one of several possible models. This approach allows us to use different motion models that are suited to the diverse types of optical flow that occur in natural scenes. In this paper we consider two models, namely, smooth motion and motion boundaries. Regions of smooth motion may be

modeled using conventional constant or affine models while the complex phenomena that occur at motion boundaries are accounted for by an explicit, non-linear, boundary model.

To cope with image noise, matching ambiguities, and model uncertainty, we adopt a Bayesian probabilistic framework that integrates information over time and represents multiple, competing, hypotheses. Our goal is to compute the posterior probability distribution over models and model parameters, conditioned on image measurements. The computation of the posterior distribution is expressed in terms of a likelihood function and a prior probability distribution. The likelihood represents the probability of observing the current image data given the model parameters. The prior represents our belief about models at the current time based on previous observations. This "temporal" prior embodies our assumptions about the temporal dynamics of how the models and model parameters evolve over time.

Our Bayesian formulation rests on the specification of *generative* models for smooth motion and motion boundaries. These generative models define our probabilistic assumptions about the spatial structure of the motion within a region, the temporal evolution of the model parameters, and the probability distribution over the image measurements that one would expect to observe given a particular instance of the model parameters.

The motion boundary model, illustrated in Fig. 1, encodes the orientation of the boundary, the image velocities of the pixels on each side of the boundary, the foreground/background assignment for the two sides, and an offset of the boundary from the center of the region. With this explicit model, we can predict the visibility of occluded and disoccluded pixels so that these pixels may be excluded when estimating the probability of a particular model. Moreover, the explicit offset parameter allows us to predict the location of the edge within the region of interest, and hence track its movement through the region. Tracking the motion of the edge allows foreground/background ambiguities to be resolved. Explicit generative models such as this have not previously been used for detecting motion discontinuities due to the non-linearity of the model and the consequent difficulty of estimating the model parameters.

The use of multiple models, including the non-linear motion boundary model, means that the posterior probability distribution will be non-Gaussian. In most cases, we expect it to be multi-modal. Therefore, rather than represent and propagate the posterior in time using a simple parametric form (as in a Kalman filter), here we represent the posterior distribution using factored sampling, and we propagate it through time using a particle filter (Gordon et al., 1993; Isard and Blake, 1998a; Liu and Chen, 1998).

The parameter space we must represent is of a relatively high dimension. It includes a discrete parameter to encode the type of motion model (smooth motion or motion boundary), and a vector of continuous parameters (2 parameters for the smooth motion model, and 6 for the motion boundary model). Given the dimensionality of the parameter space, naive sampling methods will be extremely inefficient. But if the samples can be directed to the appropriate portion of the parameter space, small numbers of samples can well characterize

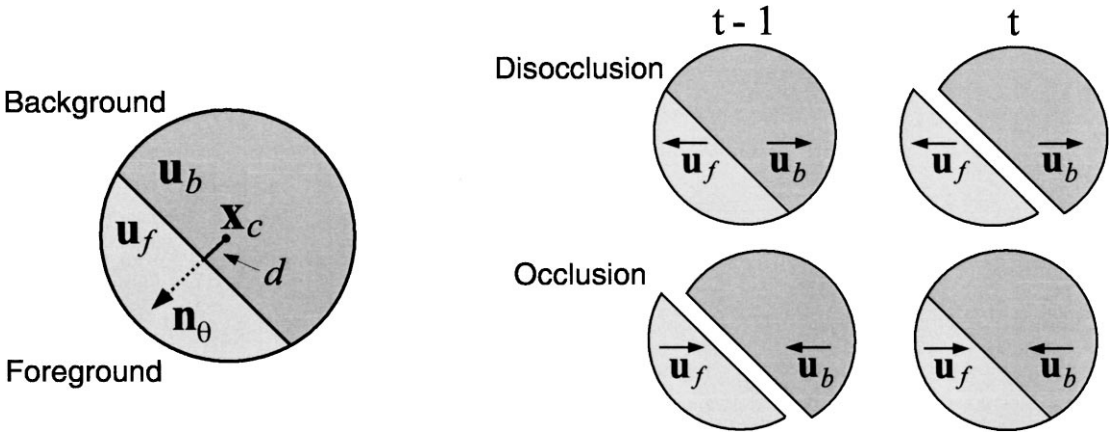

Figure 1. Model of an occlusion boundary, parameterized by foreground and background velocities, $\mathbf{u}_f$ and $\mathbf{u}_b$, an orientation θ with normal $\mathbf{n}_\theta$, and a signed distance d from the neighborhood center $\mathbf{x}_c$. With this model we predict which pixels are visible between frames at times $t-1$ and t.

such distributions (Isard and Blake, 1998b). It is for this purpose that, in addition to the temporal prior, we use an initialization prior. At the low level, a set of dense detectors signal the presence of potential motion boundaries and give estimates of the model parameters. These detectors are based on approximate linear models of motion boundaries, the coefficients of which can be estimated with robust optical flow techniques (Fleet et al., 2000). Neighborhoods of these filter outputs provide a *prior* distribution over model parameters that is sampled from when initializing the full non-linear models.

While the method described here can be thought of simply as a motion boundary detector, the framework has wider application. The Bayesian formulation and computational model provide a general probabilistic framework for motion estimation with multiple, non-linear, models. This generalizes previous work on recovering optical flow using linear models (Bergen et al., 1992; Fleet et al., 2000). Moreover, the Bayesian formulation provides a principled way of choosing between multiple hypothesized models for explaining the image variation within a region.

We illustrate the method on natural image sequences and show how the Bayesian formulation and the particle filtering method allow motion discontinuities to be detected and tracked over multiple frames. These experiments also show how depth ambiguities can be resolved by observing the motion of the boundary over time.

2. Previous Work

The detection of motion boundaries has been a long-standing problem in optical flow estimation, primarily because most approaches to computing optical flow fail to be reliable in the vicinity of motion discontinuities (Barron et al., 1994; Fleet, 1992; Otte and Nagel, 1994). In addition, it has long been acknowledged that motion boundaries provide useful information about the position and orientation of surface boundaries.

Most previous approaches for detecting occlusion boundaries have treated the boundaries as a form of "noise"; that is, as the violation of a smoothness assumption. This approach is taken in regularization schemes where robust statistics, weak continuity, or line processes are used to locally disable smoothing across motion discontinuities (Cornelius and Kanade, 1983; Harris et al., 1990; Heitz and Bouthemy, 1993; Konrad and Dubois, 1988; Murray and Buxton, 1987; Nagel and Enkelmann, 1986; Schunck, 1989; Shulman and Hervé, 1989; Thompson et al., 1985). Robust regression (Black and Anandan, 1996; Sawhney and Ayer, 1996) and mixture models (Ayer and Sawhney, 1995; Jepson and Black, 1993; Weiss and Adelson, 1996) have been used to account for the multiple motions that occur at motion boundaries but these methods fail to explicitly model the boundary and its spatiotemporal structure; in particular, they do not model the orientation of the boundary, the pixels that are occluded or disoccluded, or the depth ordering of the surfaces at the boundary.

Numerous methods have attempted to detect discontinuities in optical flow fields by analyzing local distributions of flow (Spoerri and Ullman, 1987) or by performing edge detection on the flow field (Potter, 1980; Schunck, 1989; Thompson et al., 1985). It has often been noted that these methods are sensitive to the accuracy of the optical flow and that accurate optical flow is hard to estimate without prior knowledge of the occlusion boundaries. Other methods have focused on detecting occlusion from the structure of a correlation surface (Black and Anandan, 1990), or of the spatiotemporal brightness pattern (Beauchemin and Barron, 2000; Chou, 1995; Fleet and Langley, 1994; Niyogi, 1995). Still others have used the presence of unmatched features to detect dynamic occlusions (Mutch and Thompson, 1985; Thompson et al., 1985).

None of these methods explicitly model the spatial structure of the image motion present in the immediate neighborhood of the motion boundary, and they have not proved sufficiently reliable in practice. Recent approaches have formulated explicit (approximate) models of motion boundaries using linear combinations of basis flow fields (Fleet et al., 2000). Estimating the image motion with these models is analogous to the estimation of motion from image derivatives using conventional linear parameterized models such as affine flow. Moreover, from the estimated linear coefficients, one can compute the orientation of the boundary and the velocities on either side. However, these linear models still provide only a coarse approximation to the motion at a boundary. For example, they do not explicitly model which image pixels are occluded or disoccluded between frames. This means that these pixels, which are not visible in both frames, are treated as noise. With our explicit non-linear model, these pixels can be predicted and therefore taken into account in the likelihood computation.

Additionally, most of the above methods have no explicit temporal model. With our generative model, we predict the motion of the occlusion boundary over time and hence integrate information over multiple frames. When the motion of the discontinuity is consistent with that of the foreground surface we can explicitly determine the foreground/background relationships (depth ordering) between the surfaces.

3. Generative Models

For the purposes of this work, as suggested in Fig. 1, we decompose an image into a grid of circular neighborhoods in which we estimate motion information. We assume that the motion in any region can be modeled by one of several motion models. Here we consider two models, namely smooth motion and motion boundaries. Generative models of these motions are used to specify the model parameters of interest, the probabilistic relationship between these parameters and image observations, and the way in which we expect these parameters to vary over time (i.e. the temporal dynamics).

For the smooth motion model, we express the optical flow within the circular region as simple image translation; more complex models, such as affine motion, can also be used. The translational model has two parameters, namely, the horizontal and vertical components of the velocity, denoted $\mathbf{u}_0 = (u_0, v_0)$. Exploiting the common assumption of brightness constancy, the generative model states that the image intensity, $I(\mathbf{x}', t)$, of a point $\mathbf{x}' = (x', y')$ at time t in a region R is equal to the intensity at some location $\mathbf{x}$ at time $t-1$ with the addition of noise ν_n:

$$I(\mathbf{x}', t) = I(\mathbf{x}, t-1) + \nu_n(\mathbf{x}, t), \tag{1}$$

where $\mathbf{x}' = \mathbf{x} + \mathbf{u}_0$. Here, we are assuming that the noise, $\nu_n(\mathbf{x}, t)$, is white and Gaussian with a standard deviation of σ_n; that is, $\nu_n \sim \mathcal{N}(0, \sigma_n^2)$.

The motion boundary model is more complex and contains 6 parameters: the edge orientation, the velocities of the foreground ($\mathbf{u}_f$) and the background ($\mathbf{u}_b$), and the distance from edge to the center of the region $\mathbf{x}_c$. In our parameterization, shown in Fig. 1, the orientation, $\theta \in [-\pi, \pi)$, specifies the direction of a unit vector, $\mathbf{n} = (\cos(\theta), \sin(\theta))$, that is normal to the occluding edge. We represent the location of the edge by its signed perpendicular distance d from the center of the region (positive meaning in the direction of the normal). The edge is therefore normal to $\mathbf{n}$ and passes through the point $\mathbf{x}_c + d\,\mathbf{n}$. Relative to the center of the region, we adopt a convention that defines the foreground to be the side to which the normal $\mathbf{n}$ points. Therefore, a point $\mathbf{x}$ is on the foreground if $(\mathbf{x} - \mathbf{x}_c) \cdot \mathbf{n} > d$. Similarly, points on the background satisfy $(\mathbf{x} - \mathbf{x}_c) \cdot \mathbf{n} < d$.

At most motion boundaries some pixels will be either occluded or disoccluded and, as a consequence, one should not expect to find corresponding pixels in adjacent frames. It is therefore important that we identify these pixels when formulating the likelihood function. Towards this end, let us assume that the motion boundary edge moves with the same velocity as the pixels on the foreground side of the edge (i.e., the occluding side).[1] With this assumption, the occurrence of occlusion or disocclusion depends solely on the difference between the background and foreground velocities. Pixels are occluded from one frame to the next when the background moves faster than the foreground in the direction of the edge normal. More precisely, if $u_{fn} = \mathbf{u}_f \cdot \mathbf{n}$ and $u_{bn} = \mathbf{u}_b \cdot \mathbf{n}$ denote the two normal velocities, occlusion occurs when $u_{bn} - u_{fn} > 0$. Disocclusion occurs when $u_{bn} - u_{fn} < 0$. The width of the occluded/disoccluded region, measured normal to the occluding edge, is $|u_{bn} - u_{fn}|$.

With this model, parameterized by $(\theta, \mathbf{u}_f, \mathbf{u}_b, d)$, we can now specify how pixels move from one frame to the next. A pixel $\mathbf{x}$ at time $t-1$, that remains visible at time t, moves to location $\mathbf{x}'$ at time t given by

$$\mathbf{x}' = \begin{cases} \mathbf{x} + \mathbf{u}_f & \text{if } (\mathbf{x} - \mathbf{x}_c) \cdot \mathbf{n} > d \\ \mathbf{x} + \mathbf{u}_b & \text{if } (\mathbf{x} - \mathbf{x}_c) \cdot \mathbf{n} < d + w \end{cases} \tag{2}$$

where $w = \max(u_{bn} - u_{fn}, 0)$ is the width of the occluded region. Finally, with $\mathbf{x}'$ defined by (2), along with the assumptions of brightness constancy and white Gaussian image noise, the image observations associated with a motion edge are given by (1).

Referring to Fig. 1 (right), in the case of disocclusion, a circular neighborhood at time $t-1$ maps to a pair of regions at time t, separated by the width of the disocclusion region $|u_{bn} - u_{fn}|$. Conversely, in the case of occlusion, a pair of neighborhoods at time $t-1$, separated by $|u_{bn} - u_{fn}|$, map to a circular neighborhood at time t. Being able to look forward or backward in time in this way allows us to treat occlusion and disocclusion symmetrically.

So far we have focused on the spatial structure of the generative models. We must also specify the evolution of the model parameters through time since this will

be necessary to disambiguate which side of the motion boundary is the foreground. From optical flow alone one cannot determine the motion of the occlusion boundary using only two frames. The boundary must be observed in at least two separate instances (e.g., using three consecutive frames) to discern its motion. The image pixels whose motion is consistent with that of the boundary are likely to belong to the occluding surface. Thus, to resolve the foreground/background ambiguity, we propose to accumulate evidence over time.

Towards this end, we assume that the parameters of the motion models obey a first-order Markov process and, hence, parameter values at time t depend only on the parameter values at $t-1$. For the smooth motion model, we assume that the expected image translation remains constant from one time to the next. More precisely, we assume that the image translation at time t, $\mathbf{u}_{0,t}$, is given by

$$\mathbf{u}_{0,t} = \mathbf{u}_{0,t-1} + \nu_u, \quad \nu_u \sim \mathcal{N}\big(0, \sigma_u^2 \mathbf{I}_2\big), \tag{3}$$

where $\mathbf{I}_2$ is the 2D identity matrix. Here, the Gaussian noise represents the modeling uncertainty (error) implicit in this simple first-order dynamical model.

For the motion boundary model, we assume that the expected velocities on either side of the boundary, along with the expected orientation of the boundary, remain constant. Moreover, as discussed above, we assume that the expected location of the boundary translates with the foreground velocity. The dynamics are then governed by

$$\mathbf{u}_{\mathrm{f},t} = \mathbf{u}_{\mathrm{f},t-1} + \nu_{u,\mathrm{f}}, \quad \nu_{u,\mathrm{f}} \sim \mathcal{N}\big(0, \sigma_u^2 \mathbf{I}_2\big) \tag{4}$$

$$\mathbf{u}_{\mathrm{b},t} = \mathbf{u}_{\mathrm{b},t-1} + \nu_{u,\mathrm{b}}, \quad \nu_{u,\mathrm{b}} \sim \mathcal{N}\big(0, \sigma_u^2 \mathbf{I}_2\big) \tag{5}$$

$$\theta_t = [\theta_{t-1} + \nu_\theta] \bmod 2\pi, \quad \nu_\theta \sim \mathcal{N}\big(0, \sigma_\theta^2\big) \tag{6}$$

$$d_t = d_{t-1} + \mathbf{n}_{t-1} \cdot \mathbf{u}_{\mathrm{f},t-1} + \nu_d, \quad \nu_d \sim \mathcal{N}\big(0, \sigma_d^2\big). \tag{7}$$

Here we use a wrapped-normal distribution over angles; therefore, orientation, θ_{t-1}, is propagated in time by adding Gaussian noise and then removing an integer multiple of 2π so that $\theta_t \in [-\pi, \pi)$. The location of the boundary moves with the velocity of the foreground, and therefore its expected location at time t is equal to that at time $t-1$ plus the magnitude of the component of the foreground velocity in the direction of the boundary normal. As above, Gaussian noise is included to represent the modeling errors implicit in this simple dynamical model. Note that more sophisticated models of temporal dynamics (e.g., constant acceleration) could be used.

4. Probabilistic Framework

Given the generative models described above, we are now ready to formulate our state description and the computation of the posterior probability density function over models and model parameters. First, let *states* be denoted by $\mathbf{s} = (\mu, \mathbf{p})$, where μ is the model type (translation or motion boundary), and $\mathbf{p}$ is a parameter vector appropriate for the model type. For the translational model the parameter vector is two-dimensional, $\mathbf{p} = (\mathbf{u}_0)$. For the motion boundary model it is 6-dimensional, $\mathbf{p} = (\theta, \mathbf{u}_\mathrm{f}, \mathbf{u}_\mathrm{b}, d)$. Our goal is to find the posterior probability distribution over states at time t given the measurement history up to time t; i.e., $p(\mathbf{s}_t \mid \vec{\mathbf{Z}}_t)$. Here, $\vec{\mathbf{Z}}_t = (\mathbf{z}_t, \ldots, \mathbf{z}_0)$ denotes the measurement history.

From the generative models above, it follows that the temporal dynamics of the motion models form a Markov chain for which states at time t depend only on states at the previous time instant:

$$p(\mathbf{s}_t \mid \vec{\mathbf{S}}_{t-1}) = p(\mathbf{s}_t \mid \mathbf{s}_{t-1}),$$

where $\vec{\mathbf{S}}_t = (\mathbf{s}_t, \ldots, \mathbf{s}_0)$ denotes the state history. In other words, $\mathbf{s}_t$ and $\vec{\mathbf{S}}_{t-2}$ are conditionally independent given $\mathbf{s}_{t-1}$. The generative models also assume conditional independence of the observations and the dynamics. In other words, given $\mathbf{s}_t$, the current observation $\mathbf{z}_t$ and the previous observations $\vec{\mathbf{Z}}_{t-1}$ are independent. With these assumptions one can show that the posterior distribution $p(\mathbf{s}_t \mid \vec{\mathbf{Z}}_t)$ can be factored and reduced using Bayes' rule to obtain

$$p(\mathbf{s}_t \mid \vec{\mathbf{Z}}_t) = k p(\mathbf{z}_t \mid \mathbf{s}_t)\, p(\mathbf{s}_t \mid \vec{\mathbf{Z}}_{t-1}) \tag{8}$$

where k is a constant used to ensure that the distribution integrates to one. Here, $p(\mathbf{z}_t \mid \mathbf{s}_t)$ represents the likelihood of observing the current measurement given the current state, while $p(\mathbf{s}_t \mid \vec{\mathbf{Z}}_{t-1})$ is referred to as a temporal prior, the prediction of the current state given all previous observations.

The specific form of the likelihood function $p(\mathbf{z}_t \mid \mathbf{s}_t)$ follows from the generative models. In particular, the state specifies the motion model and the mapping from visible pixels at time $t-1$ to those at time t. The observation equation, derived from the brightness constancy

assumption (1), specifies that the intensity differences between corresponding pixels at times t and $t-1$ should be white and Gaussian, with zero mean and standard deviation σ_{n}.

Using Bayes' rule and the conditional independence assumed above, it is straightforward to show that the temporal prior can be written in terms of the temporal dynamics that propagate states from time $t-1$ to time t and the posterior distribution over states at time $t-1$. In particular,

$$p(\mathbf{s}_t \mid \vec{\mathbf{Z}}_{t-1}) = \int p(\mathbf{s}_t \mid \mathbf{s}_{t-1})\, p(\mathbf{s}_{t-1} \mid \vec{\mathbf{Z}}_{t-1})\, d\,\mathbf{s}_{t-1}, \quad (9)$$

where the conditional probability distribution $p(\mathbf{s}_t \mid \mathbf{s}_{t-1})$ embodies the temporal dynamics, and $p(\mathbf{s}_{t-1} \mid \vec{\mathbf{Z}}_{t-1})$ is the posterior distribution over the state space at time $t-1$.

This completes our description of the state space, and the mathematical form of the posterior probability distribution over the possible interpretations of the motion within an image region.

5. Computational Model

We now describe the details of our computational embodiment of the probabilistic framework outlined above. First, we consider the representation of the posterior distribution and its propagation through time using a particle filter. We then address the computation of the likelihood function and discuss the nature of the prior probability distribution that facilitates the state space search for the most probable models and model parameters.

5.1. Particle Filter

The first issue concerns the representation of the posterior distribution, $p(\mathbf{s}_t \mid \vec{\mathbf{Z}}_t)$. Because of the non-linear nature of the motion boundary model, the existence of multiple models, and because we expect foreground/background and matching ambiguities, we can assume that $p(\mathbf{s}_t \mid \vec{\mathbf{Z}}_t)$ will be non-Gaussian, and often multi modal. For this reason we represent the posterior distribution in a non-parametric way, using factored sampling. We then use a particle filter (with sequential importance resampling) to propagate the posterior through time (Black, 1999; Gordon et al., 1993; Isard and Blake, 1998a; Liu and Chen, 1998).

The posterior is represented with a discrete, weighted set of S random samples $\{(\mathbf{s}_t^{(i)}, w_t^{(i)})\}_{i=1,\ldots,S}$. At each time step, following the Condensation algorithm (Isard and Blake, 1998a), the posterior is computed by drawing a fair set of state samples from the prior probability distribution and then evaluating the likelihood of each sample. Normalizing the likelihoods of the state samples, so that they sum to one, produces the weights $w_t^{(i)}$:

$$w_t^{(i)} = \frac{p\left(\mathbf{z}_t \mid \mathbf{s}_t^{(i)}\right)}{\sum_{n=1}^{S} p\left(\mathbf{z}_t \mid \mathbf{s}_t^{(n)}\right)}.$$

These weights ensure that our sample set $\{(\mathbf{s}_t^{(i)}, w_t^{(i)})\}_{i=1,\ldots,S}$ contains properly weighted samples with respect to the desired posterior distribution $p(\mathbf{s}_t \mid \vec{\mathbf{Z}}_t)$. A sufficiently large number of independent samples then provides a reasonable approximation to the posterior.

5.2. Likelihood

The next issue concerns the detailed computation of the likelihood. To evaluate the likelihood $p(\mathbf{z}_t \mid \mathbf{s}_t^{(i)})$ of a particular state, we draw a uniform random sample $\mathcal{R}$ of visible image locations (as constrained by the generative model and the current state). Typically we sample 50% of the pixels in the region. Given this subset of pixels, we compute the un-normalized likelihood as

$$p\left(\mathbf{z}_t \mid \mathbf{s}_t^{(i)}\right) = \left(\exp\left[\frac{-1}{2\sigma_{\mathrm{n}}^2} \sum_{\mathbf{x}\in\mathcal{R}} E\left(\mathbf{x}, t; \mathbf{s}_t^{(i)}\right)^2\right]\right)^{1/T} \quad (10)$$

where $E(\mathbf{x}, t; \mathbf{s}_t^{(i)}) = I(\mathbf{x}', t) - I(\mathbf{x}, t-1)$, $T = |\mathcal{R}|$ is the number of sampled pixel locations, and the warped image location $\mathbf{x}'$ is a function of the state $\mathbf{s}_t^{(i)}$ (as, for example, defined in (2)). The warped image value $I(\mathbf{x}', t)$ is computed using bi-linear interpolation. Note that sampling a fraction of the pixels gives some measure of robustness to outliers (Bab-Hadiashar and Suter, 1997).

Note that this likelihood function is equivalent to that specified by the generative model, but raised to the power $1/T$. This is computationally, rather than probabilistically, motivated. A large value of T has the effect of smoothing the posterior distribution making the peaks broader. Within a sampling framework, this allows a more effective search of the parameter space, reducing the chances of missing a significant peak.

5.3. Prior

The prior probability distribution serves to constrain our samples to relevant portions of the parameter space. We are seeking solutions (drawing state samples) from within a seven-dimensional state space. This is a relatively high dimensional space and naive approaches for representing or searching it will be infeasible. Therefore, unlike a conventional particle filter for which the prior is derived solely by propagating the posterior from the previous time instant, we also exploit an initialization prior that provides a form of bottom-up information to initialize new states. This is useful at time 0 when no posterior is available from the previous time instant. It is also useful to help avoid getting trapped at local maxima thereby missing the occurrence of novel events that might not have been predicted from the posterior at the previous time. For example, it helps to detect the sudden appearance of motion edges in regions where only translational state samples existed at the previous time instant. This use of bottom-up information, along with the prediction from the temporal prior, allows us to effectively sample the most interesting portions of the state-space.

The actual prior used here is a linear mixture of a temporal prior and an initialization prior. In the experiments that follow we use constant mixture proportions of 0.8 and 0.2 respectively; that is, 80% of the samples are drawn from the temporal prior. Importance sampling (Gordon et al., 1993; Isard and Blake, 1998b; Liu and Chen, 1998) provides an alternative way of achieving similar results.

5.3.1. Temporal Prior. According to the generative model for translational motion, the temporal dynamics (3) yield

$$p(\mathbf{s}_t \mid \mathbf{s}_{t-1}) = G_{\sigma_u}(\Delta \mathbf{u}_0) \tag{11}$$

where G_{σ_u} denotes a multivariate mean-zero Gaussian with covariance matrix $\sigma_u^2 \mathbf{I}_2$, and $\Delta \mathbf{u}_0 = \mathbf{u}_{0,t} - \mathbf{u}_{0,t-1}$ denotes the temporal velocity difference. Similarly, the generative model for the motion boundary ((4)–(7)) specifies that

$$p(\mathbf{s}_t \mid \mathbf{s}_{t-1}) = G_{\sigma_u}(\Delta \mathbf{u}_f)\, G_{\sigma_u}(\Delta \mathbf{u}_b) G_{\sigma_d}(\Delta d - \mathbf{n} \cdot \mathbf{u}_{f,t-1})\, G^w_{\sigma_\theta}(\Delta\theta) \tag{12}$$

where G^w denotes a wrapped-normal (for circular distributions) and, as above, $\Delta\theta = \theta_t - \theta_{t-1}$ and $\Delta d = d_t - d_{t-1}$.

Because the posterior, $p(\mathbf{s}_t \mid \vec{\mathbf{Z}}_{t-1})$, at time $t-1$ is represented as a weighted sample set and the conditional priors, ((11) and (12)), are Gaussian, the temporal prior given by (9) can be viewed as a Gaussian mixture model. To see this, note that the posterior is being approximated as a weighted sum of Dirac delta functions (at the sample states), so (9) can be viewed as a convolution of this sum of delta functions with the Gaussian temporal dynamics. The result of the convolution is a sum of Gaussians. There is one Gaussian component for each sample $\mathbf{s}_{t-1}^{(i)}$ at time $t-1$ and the mixture probabilities are equal to the weights, $w_{t-1}^{(i)}$.

To draw a fair sample from a Gaussian mixture, one first draws one of the components according to the mixture probabilities (the weights), and then one draws a sample from that Gaussian component. To sample a component from the mixture, one can construct a cumulative probability distribution using the weights $w_{t-1}^{(i)}$, and then draw a sample from it (Isard and Blake, 1998a). Let the cumulative probabilities be

$$c_{t-1}^{(0)} = 0$$
$$c_{t-1}^{(i)} = c_{t-1}^{(i-1)} + w_{t-1}^{(i)}.$$

We sample this distribution by uniformly choosing a value, r, between zero and one. We then find the smallest $c_{t-1}^{(i)}$ such that $c_{t-1}^{(i)} > r$. The state $\mathbf{s}_{t-1}^{(i)}$ is then selected for propagation. Given $\mathbf{s}_{t-1}^{(i)}$, we then sample from the dynamics, $p(\mathbf{s}_t \mid \mathbf{s}_{t-1}^{(i)})$, which, as explained above, is a multivariate Gaussian. This is repeated for every sample drawn from the temporal prior.

Thus far we have assumed that the motion model type remains constant as we propagate states from one time to the next. However, when a boundary passes through a region and out the other side, it is natural to switch model types from a motion boundary model to a smooth motion model. Accordingly, given a motion boundary state at time $t-1$, we let the probability of switching to a translational model at time t be given by the probability that the temporal dynamics will place the boundary outside the region of interest at time t. This can be computed as the integral of $p(\mathbf{s}_t \mid \mathbf{s}_{t-1})$ over boundary locations d that fall outside of the region. In practice, we accomplish this by sampling from the temporal prior as described above. But whenever we sample a motion boundary state $\mathbf{s}_t^{(i)}$ for which the edge is outside the circular neighborhood, we simply change model types, sampling instead from a translational model whose velocity is consistent with whatever

side of the motion boundary would have remained in the region of interest.

5.3.2. Initialization Prior

Low-Level Motion Boundary Detection. To initialize new states and provide a distribution over their parameters from which to sample, we use a method described by Fleet et al. (2000) for detecting motion discontinuities. This approach uses a robust, gradient-based optical flow method with a linear parameterized motion model. Motion edges are expressed as a weighted sum of basis flow fields, the coefficients of which are estimated using an area-based regression technique. Fleet et al. then solve for the parameters of the motion edge that are most consistent (in a least squares sense) with the linear coefficients.

Figure 2 shows an example of applying this method to an image sequence in which a Pepsi can translates horizontally relative to the background. The method provides a mean velocity estimate at each pixel (i.e., the average of the velocities on each side of the motion edge). This is simply the translational velocity when no motion edge is present. As explained in Fleet et al. (2000) a confidence measure, $c(\mathbf{x}) \in [0, 1]$ can be used to determine where edges are most likely, and is computed from the squared error in fitting a motion edge using the linear coefficients (Fig. 2, "Confidence"). The bottom images in Fig. 2 show estimates for the orientation of the edge as well as the horizontal and vertical velocity differences across the edge at all points where $c(\mathbf{x}) > 0.5$.

While the method provides good approximate estimates of motion boundaries, it produces false positives and the parameter estimates are corrupted by noise, with estimates of disocclusion being more reliable than those of occlusion. Also, the localization of the boundary using the confidence measure is relatively crude, and since the detector does not provide a foreground assignment, it does not predict the velocity of the occluding edge. Despite these weaknesses, it is a relatively straightforward, but sometimes error prone, source of information about the presence of motion discontinuities. This information can be used to significantly constrain the regions of the state space that we need to sample with the particle filter.

Formulating the Initialization Prior. When initializing a new state we use the distribution of confidence values $c(\mathbf{x})$ within a neighborhood to first decide on

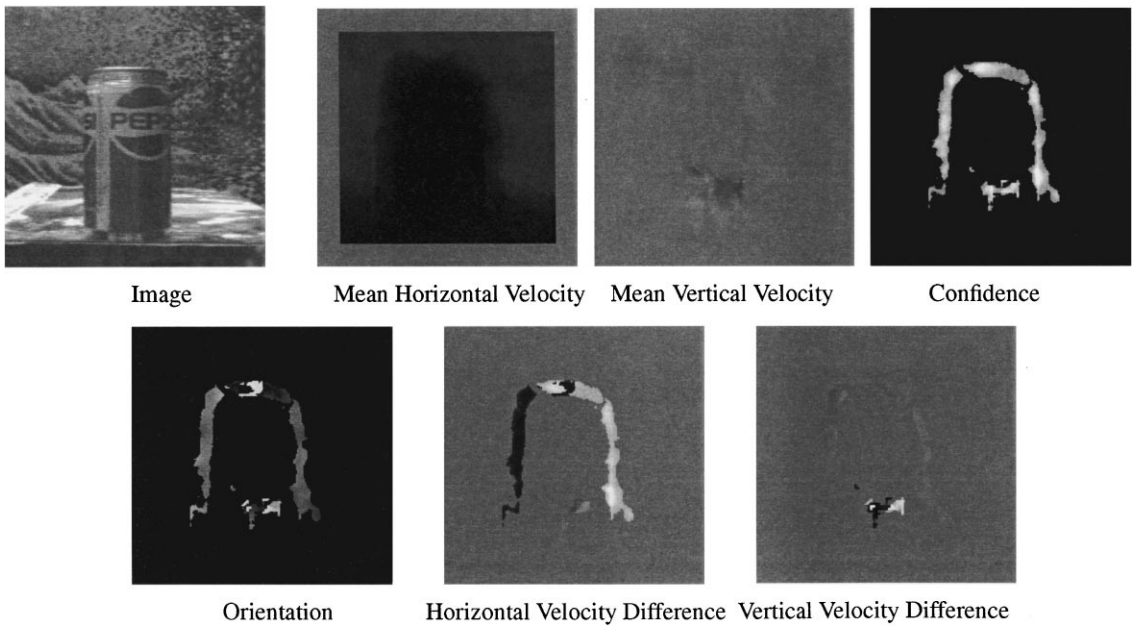

Figure 2. One frame of the Pepsi Sequence, with responses from the low-level motion edge detector, which feeds the initialization prior. The image velocities and the velocity differences are primarily horizontal. In the orientation image, grey denotes vertical orientations, while white and dark grey denote near horizontal orientations.

the motion type (translation or motion boundary). If a motion boundary is present, then we expect some fraction of confidence values, $c(\mathbf{x})$, within our region of interest, to be high. We therefore rank order the confidence values within the region and let the probability of a motion boundary state be the 95th percentile confidence value, denoted C_{95}. Accordingly, the probability of initializing a translation model is $1 - C_{95}$.

When we wish to initialize (sample) a motion boundary state, we assume that actual boundary locations are distributed according to the confidence values in the region; i.e., the boundary is more likely to pass through pixel locations with large $c(\mathbf{x})$. Sampling from the confidence values gives potential boundary locations. Given a boundary position, the low-level detector parameters at that position provide estimates of the edge orientation and the image velocity on each side, but they do not specify which side is the foreground. Thus, the probability distribution over the state space, conditioned on the detector parameters and boundary location, will have two distinct modes, one for each of the two possible foreground assignments. We take this distribution to be a mixture of two Gaussians which are separable with covariance matrices $2.25\sigma_u^2\mathbf{I}_2$ for the velocity axes, and variances $16\sigma_\theta^2$ for the orientation axis and $4\sigma_d^2$ for the position axis. The variances are larger than those used in the temporal dynamics described in Section 3 because we expect greater noise from these low-level estimates.

In generating a translational model, we first sample a spatial position according to the distribution of $1 - c(\mathbf{x})$. Locations within the region that are sampled in this way are likely to correspond to translation rather than a motion boundary model. The distribution over translational velocities, given the detector estimate at the sampled spatial position, is then taken to be a Gaussian distribution centered at the mean velocity estimate of the detector at that location. The Gaussian distribution has a covariance matrix of $2.25\sigma_u^2\,\mathbf{I}_2$.

5.4. Algorithm Summary and Model Comparison

Initially, at time 0, a set of S samples is drawn from the initialization prior. Their likelihoods are then computed and normalized to give the weights $w_0^{(i)}$. At each subsequent time, as shown in Fig. 3, the algorithm repeats the process of sampling from the combined prior, computing the likelihoods, and normalizing.

From the non-parametric, sampled approximation to the posterior distribution, $p(\mathbf{s}_t \mid \vec{\mathbf{Z}}_t)$, we can compute

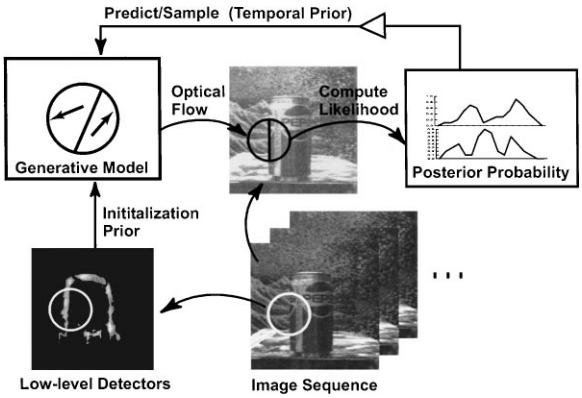

Figure 3. Particle filtering algorithm. State samples are drawn from a mixture of the temporal prior and the initialization prior. The temporal prior combines information from the posterior probability distribution at the previous time instant with the temporal dynamics of the motion models. The initialization prior is derived from the responses of low-level motion boundary detectors within an image region. The parameters of a state determine the image motion within a neighborhood as specified by the generative models for each type of motion. These generative models assume brightness constancy and hence specify how to compute the likelihood of a particular state in terms of the pixel intensity differences between the image region at one time instant and a warped version of the image at the next time instant. Normalizing the likelihood values for S states gives an approximate, discretely sampled, representation of the posterior probability distribution at the next time instant. In this way the posterior distribution is predicted and updated over time, integrating new information within the Bayesian framework.

moments and marginalize over various parameters of interest. In particular we can compute the expected value for some state parameter, $f(\mathbf{s}_t)$, as

$$E[f(\mathbf{s}_t) \mid \vec{\mathbf{Z}}_t] = \sum_{n=1}^{S} f\left(\mathbf{s}_t^{(n)}\right) w_t^{(n)}.$$

However, in doing so, care needs to be taken because the posterior will often be multimodal, in which case such expectations are often not very useful. With the motion models used here, it is common to find three distinct modes in the posterior distribution. One mode is often associated with the best fitting smooth motion model. The other two modes are associated with the motion boundary model. These two boundary models typically differ in orientation by π, reflecting two opposite foreground assignments. Accordingly, for model comparison and for displaying the results, we first isolate these three modes.

For displaying results, we compute the mean state for each principal mode by computing the expected value of the parameters for the mode divided by the sum of

all normalized likelihoods for that mode. These mean states can be overlaid on the image, as shown below. Deciding which model type is most likely within a region can be performed by comparing the sum of the likelihoods for each model type. Given the way the particle filter allocates samples, this is not necessarily the most reliable measure of how well each model fits the data. If the likelihood of a model drops rapidly between two frames, the distribution may temporarily have many low likelihood states allocated to that portion of the state space. The combined likelihood of these states may easily be greater than the likelihood of a new model that does a much better job of fitting the data. Instead, we therefore compute and compare the likelihoods of the mean models to determine which model type is most likely.

6. Experimental Results

We illustrate the method with experiments on 8-bit natural image sequences. For these experiments, the standard deviation of the image noise was $\sigma_n = 7.0$. The standard deviations for the temporal dynamics were empirically determined and remained the same in all experiments. We used circular image regions with a 16 pixel radius and used 3500 state samples to represent the posterior probability distribution in each region. A few regions were chosen to illustrate the performance of the method and its failure modes.

Because the particle filter provides us with an approximation to the posterior distribution over models and model parameters, rather than a single best state, it can be difficult to visualize the results. Here, we rely on marginalized distributions to show how probability distributions over specific state parameters vary through time. In addition, as shown in Fig. 4, on each image frame we display the mean state of the most likely motion models. The smooth motion (translation) models are shown as empty circles (e.g., Fig. 4, region *B*). For motion boundary models we sample pixel locations from the generative model of the mean state; pixels that lie on the foreground are white and background pixels are black. The position and orientation of the edge are depicted by the boundary between the white and black sides of the region. The occluded pixels are not color-coded (e.g., Fig. 4, region D).

These experiments are designed to show the general behavior of the method. More work must be done to integrate these techniques into a complete system for motion estimation and analysis.

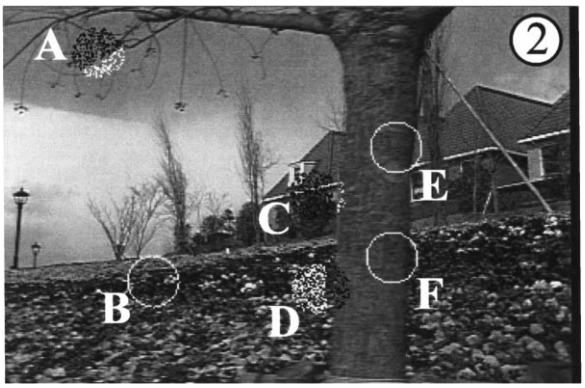

Figure 4. Flower Garden results at frame 2. Most likely mean models are overlaid on the image. Translational models are shown as empty circles (as in region B). Motion boundary models are shown as filled disks (as in region D). In this case, the position and orientation of the boundaries are depicted by the edges between the white and black sides. The white and black sides denote the foreground and background sides of the model respectively.

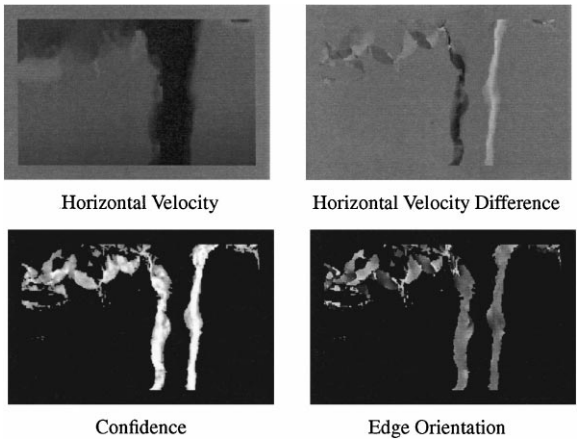

Figure 5. Low level detector responses for one pair of frames in the Flower Garden sequence.

6.1. *Flower Garden Sequence*

The Flower Garden Sequence shown in Fig. 4 contains a fast moving tree in front of a slower moving, complex, background. The low-level detector responses for the initialization prior are shown in Fig. 5. The detectors find the occluding and disoccluding sides of the tree and provide reasonable estimates of the edge orientation and the velocities on either side of the boundary. One can see from the confidence map in Fig. 5, however, that their localization is not precise.

Results of the particle filter from frames 2 through 7 are shown in Fig. 6. Regions C, D, E, and F correctly

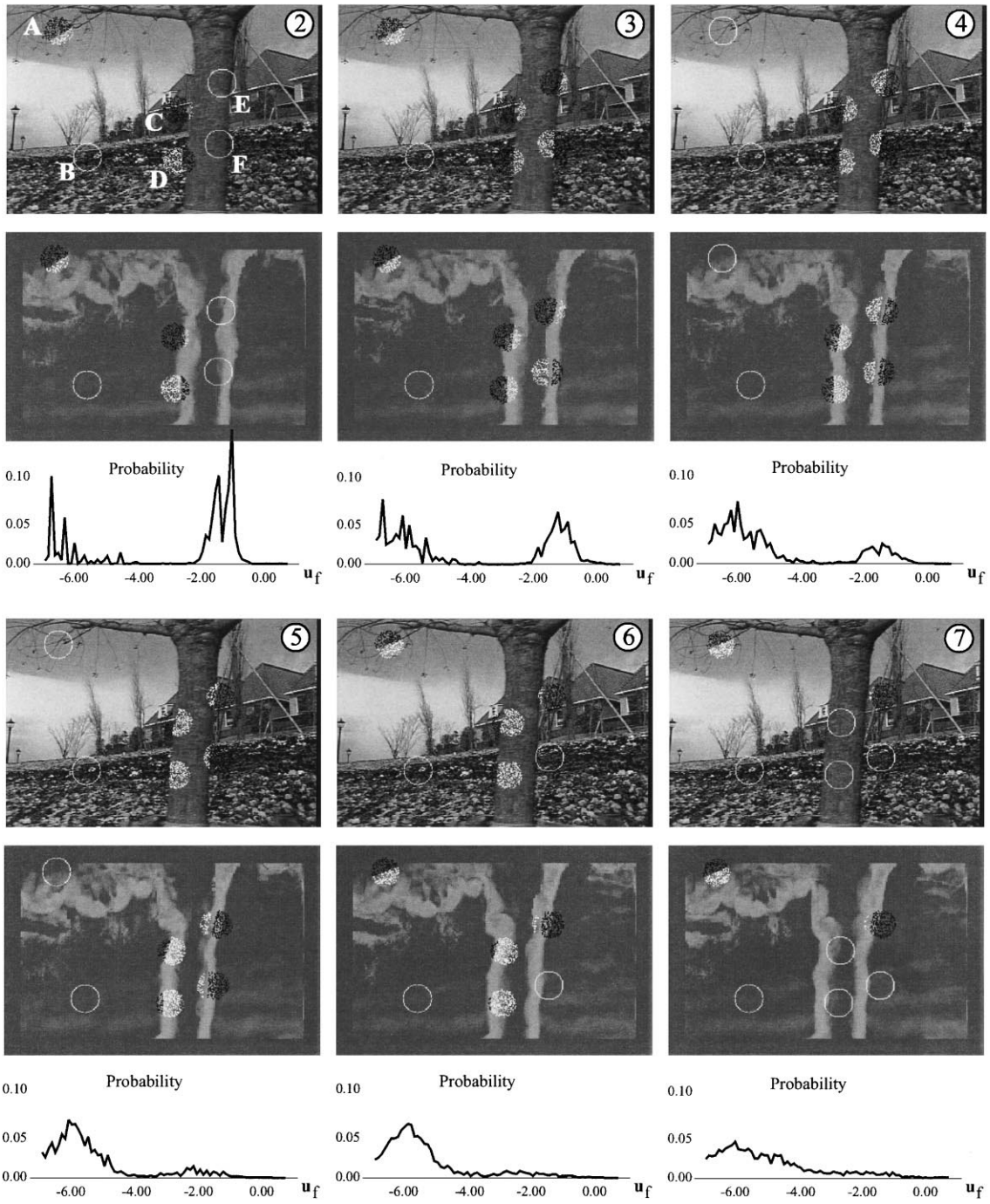

Figure 6. Flower Garden results (frames 2–7). Most likely models are drawn on images and low-level confidence maps. Marginal distributions over foreground velocity in region D also shown.

model the tree boundary (both occlusion and disocclusion) and, after the first three frames, correctly assign the tree trunk as the foreground side. Initially, in frame 2, regions C and D detect a motion boundary, but region D has incorrectly assigned the foreground to the flower garden rather than the tree. As discussed above, this is not surprising because we expect the correct foreground assignment to require more than two frames. By the third frame, the most likely mode of the posterior corresponds to the correct assignment of the foreground. Regions E and F are initially labeled with the smooth motion model since the tree boundary is just touching the right-most edge of the regions. These regions switch to boundary models in the next frame as the tree edge enters the regions. Motion boundary models then remain in all four regions along the tree trunk boundary until the last frame when the edge of the tree leaves the regions.

Beneath each of the images in Fig. 6 are plots that show the marginal posterior distributions for the horizontal component of the foreground velocity for region D. Initially, at frame 2, there are two clear modes in the distribution. One mode corresponds to a fast speed, approximately equal to the image speed of the tree trunk. The other mode corresponds to the slower speed of the flower garden. These two modes reflect the foreground ambiguity, where there is evidence for assigning the foreground to both sides. In frame 2 it is the case that the probability of assigning the foreground to the flower garden is higher. However, with the accumulation of evidence through time, and because this foreground assignment is not consistent with the motion of the boundary, the probability of assigning the foreground to the flower garden decreases, while the probability of assigning the foreground to the tree trunk increases. In frame 3 the probability of assigning the foreground to the tree trunk is slightly larger, and hence the foreground assignment in region D switches between frame 2 to frame 3. As time continues the probability associated with this correct foreground assignment increases to become the dominant interpretation.

Region B corresponds to translation and is correctly modeled as such. While translation can be equally well accounted for by the motion boundary model, the low-level detectors do not respond in this region and hence the distribution is initialized with more samples corresponding to the translational model. Region A is more interesting; if the sky were completely uniform, this region would also be modeled as translation. Note, however, that there are significant low-level detector responses in this area (Fig. 5) due to the fact that the sky is not uniform. The probabilities of the translation and motion boundary models are roughly equal here and the displayed model flips back and forth between them. For the motion boundary model, the orientation corresponds to the orientation of the tree branches in the region.

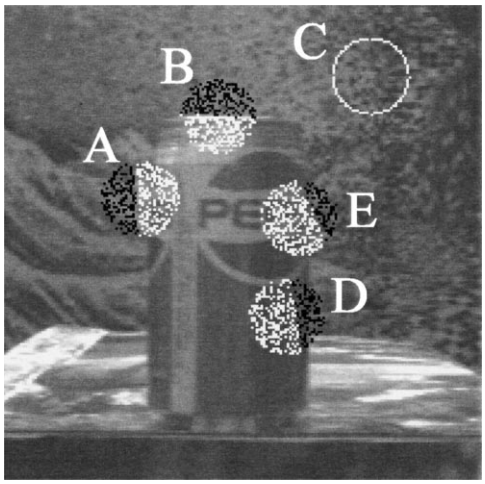

Figure 7. Pepsi Sequence. Discontinuity (filled) and translational (empty) models shown superimposed on frame 1 with labeled regions.

6.2. Pepsi Sequence

Figure 7 shows the results of applying the particle filter to the Pepsi can sequence over two frames. The regions in the figure highlight various issues raised by the approach. Note that each region has the correct model assignment (translation or boundary) and that the boundary models have the correct foreground assignment (the surface of the can). Also note that after only two frames the estimate of boundary position in Region A is not very accurate and that the boundary orientations in Regions E and D are incorrect.

To understand the complexity of the posterior distribution in this case it is useful to examine some of the marginal distributions. Figure 8 shows the marginal probability distribution for the horizontal component of the foreground velocity in region A. The top-left plot shows the marginal distribution at frame 1, which is, in fact, composed of two significant modes, associated with different foreground assignments. The numbered graphs show the marginal distributions of the

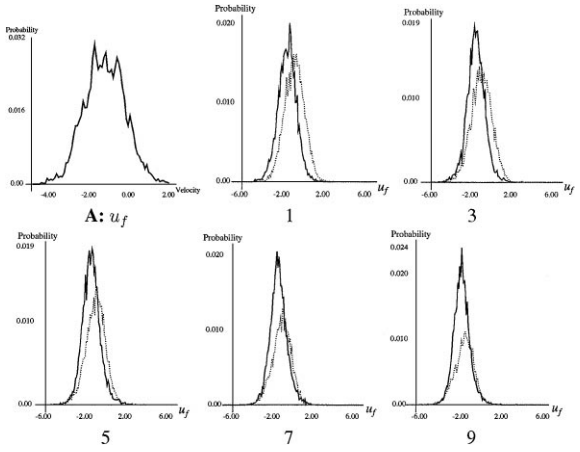

Figure 8. Pepsi Sequence. The marginal distribution of the horizontal component of the foreground velocity in region A is a mixture of two distributions. Its evolution over time is shown every other frame (1–9).

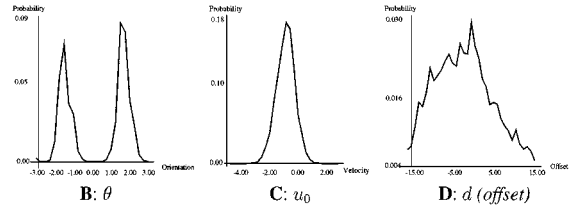

Figure 9. Pepsi Sequence. Marginal posterior probability distributions for specific state parameters in the regions labeled according to Fig. 7.

individual modes at several different times. Note that the two models are approximately centered at speeds of -1.7 and -0.8 pixels per frame, which correspond to the image velocities of the can and of the surface behind the can. The closeness of the two solutions helps to show how individual modes that correspond to the different interpretations can be difficult to resolve in the marginal distribution. It is also interesting to note that, as time evolves, the difference between the two modes becomes more pronounced as the distribution becomes dominated by the true interpretation that the foreground corresponds to the Pepsi can surface.

Other marginal distributions are shown in Fig. 9. The foreground/background ambiguity is pronounced in region B. The image motion is parallel to the motion boundary and hence we cannot disambiguate the foreground and background locally. The result is a bi-modal marginal distribution for the edge orientation. In region C the particle filter detects the smooth motion model and there no ambiguity evident in this marginal distribution over the horizontal component of the translation. The right-most plot in Fig. 9 shows the marginal probability distribution for edge location, d, in region D. In this case, the distribution is also non-Gaussian and is skewed to one side of the boundary.

Figure 10 shows the tracking behavior of the method. Note that in region A, the edge is tracked correctly and, in the detail, we see that the accuracy of the edge boundary location improves over time. Similarly in

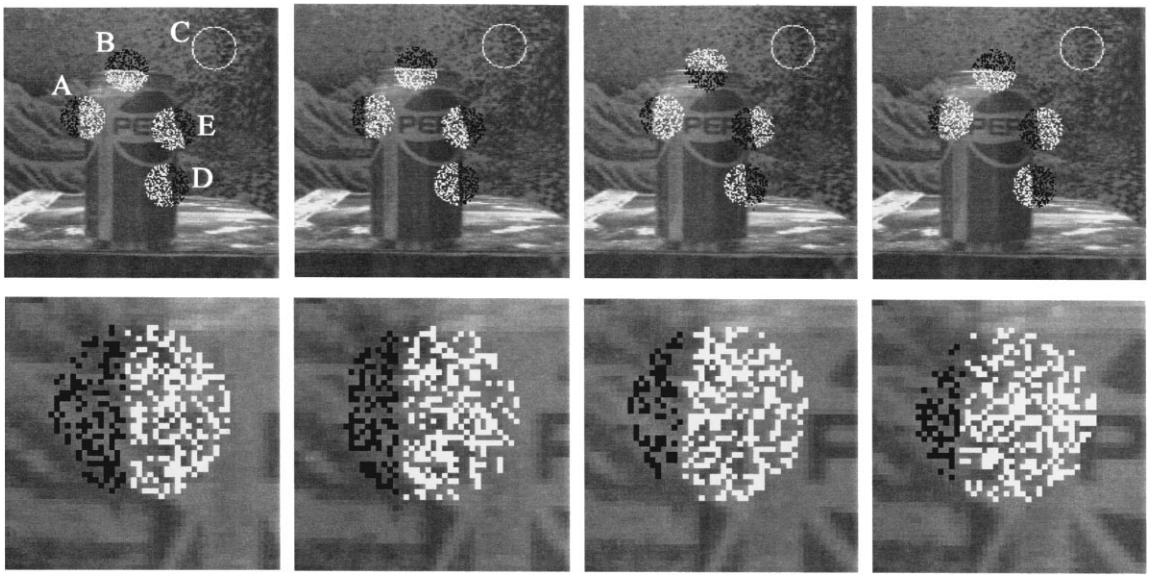

Figure 10. Pepsi Sequence. (top) Mean states for most likely models at frames 2, 4, 6, and 8. (bottom) Enlarged image of region A.

region D, the correct foreground assignment is made to the soda-can surface and the edge is accurately tracked. Region C is correctly classifed as translational motion throughout the sequence.

Note that the foreground/background ambiguity remains for region B even over many frames. In frame 6 we see the most likely mode switch to the incorrect foreground assignment and then switch back to the correct assignment in frame 8. As was illustrated in Fig. 9 the distribution has two modes of similar probability corresponding to the two interpretations. In general, propagation of information from neighboring distributions would be needed to resolve such ambiguities.

Finally, it is important to note that the particle filter does not detect and track motion boundaries in all cases as desired. For example, consider region E in Fig. 10. A motion boundary is detected in this region, but in the first frame, the most likely mode does not place the edge in the correct location at the correct orientation. Over time, the estimate of the edge orientation improves but the region switches to the incorrect foreground assignment. This behavior may be the result of low image contrast in this region and the similarity of the image velocities of the two surfaces. In any case, improving these results by incorporating additional sources of information in the likelihood computation, or by introducing some degree of spatial propagation between neighboring regions, remains a topic for future research.

7. Conclusions

Research on image motion estimation has typically exploited limited models of spatial smoothness. Our goal is to move towards a richer description of image motion using a vocabulary of motion primitives. Here we describe a step in that direction with the introduction of an explicit non-linear model of motion boundaries and a Bayesian framework for representing a posterior probability distribution over models and model parameters. Unlike previous work that attempts to find a maximum-likelihood estimate of image motion, we represent the probability distribution over the parameter space using discrete samples. This facilitates the correct Bayesian propagation of information over time when ambiguities make the distribution non-Gaussian.

The applicability of discrete sampling methods to high dimensional spaces, as explored here, remains an open issue. We find that an appropriate initialization prior is needed to direct samples to the portions of the state space where the solution is likely. We have proposed and demonstrated such a prior here but the more general problem of formulating such priors and incorporating them into a Bayesian framework remains open.

This work represents an early effort in what we hope will be a rich area of inquiry. In particular, we can now begin to think about the spatial interaction of these local motion models. For this we might formulate a probabilistic spatial "grammar" of motion features and how they relate to their neighbors in space and time. This requires incorporating the spatial propagation of probabilities in our Bayesian framework. This also raises the question of what is the right vocabulary for describing image motion and what role learning may play in formulating local models and in determining spatial interactions between them (see Freeman and Pasztor, 1999). In summary, the techniques described here (generative models, Bayesian propagation, and sampling methods) permit us to explore problems within motion estimation that were previously inaccessible.

Acknowledgments

We thank Allan Jepson for many discussions about motion discontinuities, generative models, sampling methods, and probability theory.

Note

1. Physical situations that violate this assumption include rotating objects, such as a baseball where the edge of the ball moves in one direction, but, due to the spin on the ball, the surface texture of the ball moves in another direction. However, assuming that the edge moves with the foreground velocity, as we do in this paper, allows one to handle most cases of interest.

References

Ayer, S. and Sawhney, H. 1995. Layered representation of motion video using robust maximum-likelihood estimation of mixture models and MDL encoding. In *Proc. IEEE International Conference on Computer Vision*, Boston, MA, IEEE, pp. 777–784.

Bab-Hadiashar, A. and Suter, D. 1997. Optic flow calculation using robust statistics. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, Puerto Rico, pp. 988–993.

Barron, J.L., Fleet, D.J., and Beauchemin, S.S. 1994. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77.

Beauchemin, S.S. and Barron, J.L. 2000. The local frequency structure of 1d occluding image signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(2):200–206.

Bergen, J.R., Anandan, P., Hanna, K., and Hingorani, R. 1992. Hierarchical model-based motion estimation. In *Proc. European Conference on Computer Vision*, Springer-Verlag, pp. 237–252.

Black, M.J. 1999. Explaining optical flow events with parameterized spatiotemporal models. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Fort Collins, pp. 326–332.

Black, M.J. and Anandan, P. 1990. Constraints for the early detection of discontinuity from motion. In *Proc. National Conference on Artificial Intelligence, AAAI-90*, Boston, MA, pp. 1060–1066.

Black, M.J. and Anandan, P. 1996. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104.

Chou, G.T. 1995. A model of figure-ground segregation from kinetic occlusion. In *IEEE International Conference on Computer Vision*, Boston, MA, pp. 1050–1057.

Cornelius, N. and Kanade, T. 1983. Adapting optical flow to measure object motion in reflectance and X-ray image sequences. In *Proc. ACM Siggraph/Sigart Interdisciplinary Workshop on Motion: Representation and Perception*, Toronto, Ont., Canada, pp. 50–58.

Fleet, D.J. 1992. *Measurement of Image Velocity*. Kluwer: Boston.

Fleet, D.J. and Langley, K. 1994. Computational analysis of non-fourier motion. *Vision Research*, 22:3057–3079.

Fleet, D.J., Black, M.J., Yacoob, Y., and Jepson, A.D. 2000. Design and use of linear models for image motion analysis. *International Journal of Computer Vision*, 36(3):171–193.

Freeman, W. and Pasztor, E. 1999. Learning low-level vision. In *Proc. IEEE International Conference on Computer Vision*, Corfu, Greece, pp. 1182–1189.

Gordon, N.J., Salmond, D.J., and Smith, A.F.M. 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE Proceedings-F*, Vol. 140, No. 2, pp. 107–113.

Harris, J.G., Koch, C., Staats, E., and Luo, J. 1990. Analog hardware for detecting discontinuities in early vision. *International Journal of Computer Vision*, 4(3):211–223.

Heitz, F. and Bouthemy, P. 1993. Multimodal motion estimation of discontinuous optical flow using Markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(12):1217–1232.

Isard, M. and Blake, A. 1998a. Condensation-conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):2–28.

Isard, M. and Blake, A. 1998b. ICondensation: Unifying low-level and high-level tracking in a stochastic framework. In *Proc. European Conf. on Computer Vision, ECCV-98*, H. Burkhardt and B. Neumann (Eds.), Freiburg, Germany, Springer-Verlag, Vol. 1406, LNCS-Series, pp. 893–908.

Jepson, A. and Black, M.J. 1993. Mixture models for optical flow computation. In *Partitioning Data Sets: With Applications to Psychology, Vision and Target Tracking*, Ingmer Cox, Pierre Hansen, and Bela Julesz (Eds.), AMS Pub.: Providence, pp. 271–286. RI, DIMACS Workshop.

Konrad, J. and Dubois, E. 1988. Multigrid Bayesian estimation of image motion fields using stochastic relaxation. In *Proc. IEEE International Conference on Computer Vision*, Tampa, Florida, pp. 354–362.

Liu, J.S. and Chen, R. 1998. Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93(443):1032–1044.

Murray, D.W. and Buxton, B.F. 1987. Scene segmentation from visual motion using global optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(2):220–228.

Mutch, K. and Thompson, W. 1985. Analysis of accretion and deletion at boundaries in dynamic scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(2):133–138.

Nagel, H.H. and Enkelmann, W. 1986. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(5):565–593.

Niyogi, S.A. 1995. Detecting kinetic occlusion. In *Proc. IEEE International Conference on Computer Vision*, Boston, MA, pp. 1044–1049.

Otte, M. and Nagel, H.H. 1994. Optical flow estimation: Advances and comparisons. In *Proc. European Conference on Computer Vision*, Stockholm, Sweden, J. Eklundh (Ed.), Springer-Verlag, Vol. 800, LNCS-Series, pp. 51–60.

Potter, J.L. 1980. Scene segmentation using motion information. *IEEE Transactions S.M.C.*, 5:390–394.

Sawhney, H.S. and Ayer, S. 1996. Compact representations of videos through dominant and multiple motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):814–831.

Schunck, B.G. 1989. Image flow segmentation and estimation by constraint line clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(10):1010–1027.

Shulman, D. and Hervé, J. 1989. Regularization of discontinuous flow fields. In *Proc. IEEE Workshop on Visual Motion*, Irvine, CA, pp. 81–85.

Spoerri, A. and Ullman, S. 1987. The early detection of motion boundaries. In *Proc. IEEE International Conference on Computer Vision*, London, UK, pp. 209–218.

Thompson, W.B., Mutch, K.M., and Berzins, V.A. 1985. Dynamic occlusion analysis in optical flow fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7(4):374–383.

Weiss, Y. and Adelson, E.H. 1996. A unified mixture framework for motion segmentation: Incorporating spatial coherence and estimating the number of models. In *Proc. IEEE Computer Vision and Pattern Recognition*, San Francisco, pp. 321–326.