Mixture Models for Optical Flow Computation

Allan Jepson[1] and Michael Black[2]

[1] *Canadian Institute for Advanced Research, and Department of Computer Science, University of Toronto*

[2] *Xerox PARC*

Abstract

The computation of optical flow relies on merging information available over an image patch to form an estimate of 2D image velocity at a point. This merging process raises a host of issues, which include the treatment of outliers in component velocity measurements and the modeling of multiple motions within a patch which arise from occlusion boundaries or transparency. We present a new approach which allows us to deal with these issues within a common framework. Our approach is based on the use of a probabilistic *mixture model* to explicitly represent multiple motions within a patch. We use a simple extension of the *EM-algorithm* to compute a maximum likelihood estimate for the various motion parameters. Preliminary experiments indicate that this approach is computationally efficient and can provide robust estimates of the optical flow values in the presence of outliers and multiple motions. The basic approach can also be applied to other problems in computational vision, such as the computation of 3D relative motion, which require the integration of several partial constraints to obtain a desired quantity.

Category: Motion analysis.
Keywords: Optical flow, probabilistic mixture models, motion boundaries.

This paper is the University of Toronto, Department of Computer Science, Technical Report: RBCV-TR-93-44, April 1993. Correspondence should be sent to A. Jepson, Department of Computer Science, University of Toronto, 6 Kings College Road, Toronto, Ontario M5S 1A4, or to jepson@cs.toronto.edu

1 Introduction

The computation of optical flow relies on merging information available over an image patch to form an estimate of 2D image velocity at a point. The well known aperture problem for optical flow computations [11] states that, given information available from only a small spatial aperture, we can expect to derive only a partial constraint on the image motion. In order to fully constrain the optical flow we need to integrate several such constraints obtained over a larger spatial neighborhood. As the size of this neighborhood grows there is an increased likelihood that it will span an object boundary in the scene which will result in multiple motions within the region. Multiple motions can also be the result of transparency, highlights and shadows. In these situations, the assumption of a single motion within the region results in inaccurate estimates of the optical flow. We relax the single-motion assumption and, instead, assume that the motions within any particular region can be described by a probabilistic *mixture* of distributions.

We observe that, when multiple motions are present, the motion estimates within a region form distinct *clusters*. We employ a simple extension of the *EM-algorithm* [14] to isolate these clusters and estimate their likelihood. This approach has a number of benefits. Like robust regression techniques [3], the approach allows us to robustly estimate the dominant motion within a region. Morevoer, by assuming the motion is due to a mixture of distributions we are able to recover multiple coherent motions and identify *outlying* measurements which do not correspond to a coherent motion. The recovered information about the presence of multiple motions may prove useful for the early detection of surface boundaries.

In this paper we describe the problems caused by multiple motions and briefly review previous approaches for dealing with them. We then introduce the theory of *mixture models* and describe the EM-algorithm. This basic approach has more general applicability than motion estimation and can be applied to other problems in computational vision, such as the computation of 3D relative motion, which require the integration of several partial constraints to obtain a desired quantity. Here we illustrate the theory with a series of experiments with natural image sequences containing motion boundaries, noise, and transparency. These preliminary experiments indicate that this approach is computationally efficient and can provide robust estimates of the optical flow values in the presence of outliers and multiple motions.

2 Integration of Partial Constraints

We consider the problem of estimating the optical flow from constraints available within a particular image region. Let $S(\vec{x},t)$ denote an image sequence formed, possibly, by some preprocessing of the original sequence $I(\vec{x},t)$. For example, $S(\vec{x},t)$ might be obtained from a smoothed or filtered version of $I(\vec{x},t)$. To extract motion information we apply the *data conservation constraint*, which states that S is preserved locally in space and time in the direction of image motion. That is, upon differentiating the conservation constraint

$$S(\vec{x}(t),t) = constant,$$

we obtain the "motion constraint equation"

$$\vec{c}(\vec{x},t) \cdot \vec{v}(\vec{x},t) = 0. \tag{2.1}$$

Here the "motion constraint vector", $\vec{c}(\vec{x},t)$, is the spatiotemporal gradient of S, namely $\vec{\nabla}S(\vec{x},t)$, and $\vec{v}$ is a 3-vector representing the local image velocity. Usually $\vec{v}$ is taken to be $(v_1, v_2, 1)$, where v_1 and v_2 are the components of the 2-D image velocity in the image directions x_1 and x_2, respectively.

The motion constraint equation (2.1) provides a single (linear) constraint on the two unknowns v_1 and v_2, and is therefore insufficient to determine a unique 2D image velocity. This is commonly referred to as the *aperture problem* [11]. As a result, we are faced with collecting several such constraints, say from a spatio-temporal neighborhood of the point $(\vec{x},t)$, in an attempt to infer a particular 2D image velocity. Within this neighborhood one typically assumes that the motion can be described by a single parametric model which is commonly taken to be constant, affine, or quadratic. With this approach, the neighborhood must be taken to be sufficiently large to include several constraints having different orientations. There are other constraints, however, on the choice of the "aperture" size which require that the aperture be kept small. For example, our model of the motion typically will only provide a good approximation to the true image motion over small neighborhoods. Additionally, as the region size grows, it is more likely to contain multiple surfaces with different motions whose constraints will contaminate the single-motion estimate. We refer to this dilemma surrounding the choice of aperture as the *generalized aperture problem*.

A second issue that must be faced is that the constraint (2.1) arises from an assumption about the conservation of the image structure $S(\vec{x},t)$, and this assumption may sometimes be inappropriate. In the simplest example, with S just taken to be the original image I, we have the assumption that $I(\vec{x}(t),t)$ is (locally) constant for paths $(\vec{x}(t),t)$ moving with the correct image motion. This constraint is clearly violated in many natural situations, such as cases with shading variation, highlights, transparency, or occlusion boundaries. While more effort can be spent on developing more sophisticated structure assumptions to deal with some of these cases (as has been done, for example, by Fleet [6]), we cannot avoid the need to make some assumption in connecting image structure to the motion of points in the scene. When this assumption is violated we can expect our motion constraint vector, $\vec{c}(\vec{x},t)$, to be meaningless. Thus we should expect any method for the measurement of motion constraint vectors to produce at least the occasional outlier, and therefore the method used for the integration of these constraints must be robust to such outliers.

In summary, optical flow computation relies on a somewhat tenuous link to the motion of the scene provided by assumptions about data conservation and, moreover, these assumptions are insufficient on their own to determine a unique image velocity. However, on the positive side, in many situations of interest the desired flow field is spatially coherent over relatively large regions of the image. For example, it has been shown that an affine flow model is a reasonable approximation in many cases [1], such as for a smooth surface having a sufficiently small variation in relative depth. If such a surface is textured then it can be expected to give rise to a large number of motion constraint vectors, and thus the equations for the six parameters of the affine flow can be massively redundant. In this sort of situation we should expect to be able to compute an accurate representation of the flow field for the region.

Consider, for example, the situation that occurs when the spatial neighborhood is centered at a motion boundary. In this case, approximately half of the constraints will correspond to one side of the boundary and half to the other. Such an example is shown in Figure 2.1 where the bottom edge of the figure is the v_1 axis, the left edge is the v_2 axis, and the dark lines are the constraints lines, $\vec{c}(\vec{x},t)$. In the figure two "clusters" of constraint

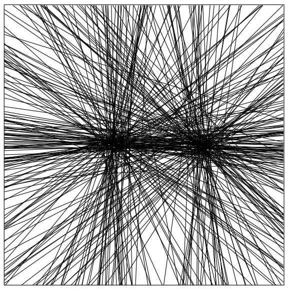


Figure 2.1: Constraint lines from differently moving surfaces within an aperture give rise to distinct "clusters" of constraint line intersections.

intersections can be observed. If we assume a single translational motion over the region the "optimal" motion estimate will lie somewhere between the two actual motions and will result in a *blurring* of the flow field at the motion boundary. A similar situation results in cases of multiple transparent motions and fragmented occlusion where no clear surface boundary exists. These situations result in exactly the same type of clusters of constraints.

To cope with situations such as this we relax the single motion assumption in two ways. First, we assume that a region may contain multiple coherent motions. We can think of these multiple motions as corresponding *layers* [15] whose spatial extent may be the entire region. Each layer contains a single consistent motion and each layer may be described by different parametric motion models. This is an important point since, in the case of transparency, there is more than one motion at each point in the image corresponding to different surfaces in the scene. Second, we assume that multiple motions and noise will occasionally result in constraint vectors which are outliers and, these should be identified and rejected.

2.1 Previous Approaches

A host of techniques have been developed to deal with the generalized aperture problem and outliers.[1] Example techniques include the use of "adaptive windows" which adjust their size and shape in an attempt to capture constraints from a single smooth surface patch [16]. While such an approach cannot cope with transparency and fragmented occlusion, the area-based regression approach of Bergen et al. [2] provides an iterative method for recovering two transparent motions from three frames. For coping with motion boundaries, an alternative approach is provided by regularization, in which the effective integration domain for constraints on a single smooth surface is dictated implicitly through a smoothness model and iteration. Smoothing over motion discontinuities is avoided either by explicitly introducing boundaries (as line processes) [9] or by using weak continuity constraints [4]. To cope with cases of fragmented occlusion, Darrell and Pentland [5] have proposed a method for segmenting the motions into distinct layers. The use of multiple layers within a robust regularization framework is discussed further in the chapter by Madarasmi and Kersten in this volume [13].

A host of other techniques fall under the category of *robust estimation* [10] in which the goal is to recover the dominant motion while treating the inconsistent constraints as outliers

[1] See [3] for a review.

and reducing their influence on the solution. The robust estimation framework introduced by Black [3] has been applied to area-based regression techniques as well as correlation. The framework also generalizes the regularization techniques by applying outlier rejection to both the data conservation and spatial smoothness assumptions.

One problem with the robust statistical techniques is that, in cases of multiple motions, they treat "secondary" structure as noise. In the previous figure which contained multiple motions, the robust techniques would accurately recover one of the two motions. An additional procedure would be needed to recognize that the "outliers" have a coherent structure and that two motions are present. Our approach, as opposed to making the single-motion assumption, explicitly models multiple motions and outliers, and hence is able to capture the more complex structure present in the data.

A somewhat different approach, referred to as "constraint line clustering", has been proposed by Schunck [17]. The basic idea is that the redundancy in the motion constraint vectors $\vec{c}(\vec{x}, t)$ arising from a smooth surface patch should be recognizable from the constraint vectors themselves. Indeed, for a patch moving with a nearly constant velocity, the "constraint lines" will all nearly pass through the same point in the (v_1, v_2)-plane. Therefore, given the constraints from an image patch, the idea is to seek such clusters of constraint lines. If the cluster detection process could be designed to be insensitive to outliers, and if the location of each identified cluster could be made insensitive to other clusters in the data set, then the approach should be able to provide accurate flow estimates without the need for a detailed knowledge of the appropriate integration regions. We follow this general strategy here, although our cluster detection process is quite different than the one-dimensional technique proposed by Schunck. In particular, Schunck does not model multiple motions within a patch and therefore cannot detect and exploit information about multiple motions when it exists.

3 Mixture Models of Flow

For a given image region we attempt to model the flow in terms of a handful of smoothly varying layers. For example, $\vec{v}(\vec{x}; \vec{a})$ may represent a constant velocity field for one layer, or it could denote an affine flow where the components v_1 and v_2 are given by linear functions of the image position $\vec{x}$. In the first case the parameter vector $\vec{a}$ is 2-dimensional, while it is 6-dimensional in the affine case.[2] Multiple motions within a particular patch are represented by selecting more than one set of parameters $\vec{a}$. However, note that at this stage of analysis we have not modeled *where* in the image patch each of the various models are appropriate. Thus transparent motion, with two different velocity fields realized over the whole patch, will be initially modeled in the same way as an occlusion boundary. A subsequent level of analysis is needed to determine which of these two interpretations is appropriate for a particular patch.

We wish to consider fitting a layered flow model to the set of motion constraint vectors measured within an image patch. In particular, we seek the parameter values $\vec{a}_n$, $n = 1, \ldots, N$ for N possibly distinct smooth fields, one for each layer. For the n^{th} layer, the probability of observing a constraint vector $\vec{c}_k$, given that the observation is at the spatial location $\vec{x}_k$, is modeled by the "component probability" distribution $p_n(\vec{c}_k | \vec{x}_k, \vec{a}_n)$. In addition we also

[2] In practice, it is often useful to add parameters representing the uncertainty of $\vec{a}$.

have a model for outlier processes given by $p_0(\vec{c}_k)$. Finally, the probability of selecting layer n is given by the "mixture probabilities" m_n, which are treated as further parameters we need to fit. Together these pieces provide the overall probability of observing the constraint $\vec{c}_k$, namely

$$p(\vec{c}_k|\vec{x}_k, \vec{m}, \vec{a}_1, \ldots, \vec{a}_N) = \sum_{n=0}^{N} m_n p_n(\vec{c}_k|\vec{x}_k, \vec{a}_n). \tag{3.1}$$

Here the mixture probabilities m_n, for $n = 0, 1, \ldots, N$ must sum to one.

Given a set of motion constraint vectors obtained within a patch at time $t = t_0$, say $\{\vec{c}_k(\vec{x}_k, t_0)\}_{k=1}^{K}$, we seek parameter values $\{\vec{a}_n\}_{n=1}^{N}$ and mixture probabilities $\{m_n\}_{n=0}^{N}$ which provide a *maximum likelihood* fit to the data set. In particular, the log likelihood of generating this set of observations from a specific model is

$$\log L(\vec{m}, \vec{a}_1, \ldots, \vec{a}_N) = \sum_{k=1}^{K} \log p(\vec{c}_k|\vec{x}_k, \vec{m}, \vec{a}_1, \ldots, \vec{a}_N). \tag{3.2}$$

At a local extrema, it can be shown that the parameters $\vec{m}$ and $\vec{a}_n$ for $n = 0, \ldots, N$ must satisfy

$$\sum_{k=1}^{K} q_{nk} = \lambda m_n, \tag{3.3a}$$

$$\sum_{k=1}^{K} q_{nk} \frac{\partial}{\partial \vec{a}_n} \log p_n(\vec{c}_k|\vec{x}_k, \vec{a}_n) = 0. \tag{3.3b}$$

Here the quantities q_{nk} represent the "ownership probabilities", that is, the probability that the k^{th} constraint belongs to the n^{th} layer. These ownership probabilities are defined by

$$q_{nk} = \frac{m_n p_n(\vec{c}_k|\vec{x}_k, \vec{a}_n)}{\sum_{j=0}^{N} m_j p_j(\vec{c}_k|\vec{x}_k, \vec{a}_j)}. \tag{3.4}$$

These equations (3.3) for a maximum likelihood fit have been derived by a number of authors; for further details see [14]. The first equation (3.3a) comes from the condition that the partial derivative of $\log L$ with respect to the mixture proportion m_n must be equal to the Lagrange multiplier λ. This Lagrange multiplier arises by imposing the constraint that the mixture proportions must sum to one. The second equation is obtained simply by requiring that the partial derivative of $\log L$ with respect to the parameters $\vec{a}_n$ must vanish.

These equations suggests an iterative algorithm, known as the EM-algorithm [14], for obtaining a maximum likelihood fit for the parameters m_n and $\vec{a}_n$, for $n = 0, \ldots, N$. Given an initial guess for these parameters we first estimate the ownership probabilities q_{nk} for each constraint belonging to each component. This is the expectation, or "E"-step, and simply involves the evaluation of the right hand side of (3.4). Given these ownership probabilities q_{nk}, we need to find parameter values $\vec{a}_n$ which satisfy (3.3b). This is equivalent to performing a maximization step, that is the "M"-step, on the expected value of the log probabilities $\log p_n(\vec{c}_k)$. As we see below, for Gaussian distributions this maximization step can be easily solved. The result is a simple iterative algorithm which is guaranteed to increase the log likelihood of its fit each iteration.

3.1 Mixtures of Constant Velocity Models

Our purpose in this paper is simply to demonstrate the utility of considering mixture models of optical flow. As such we restrict our attention to the simplest case, in which the flow field is decomposed into patches with the flow in each patch treated as constant velocity plus noise. To deal with simple occlusion boundaries and transparency, we allow two different constant velocity layers to be extracted for each patch. As a result, we take N to be 2, and take the parameters $\vec{a}_n$ to be simply $\vec{v}_n$, the 3-vector representing the constant velocity of the model (recall that an image velocity is represented by the 3-vector $(v_1, v_2, 1)$). In addition, we also attempt to identify outliers, corresponding to the 0^{th} component of the mixture. This component is, roughly speaking, modeled by a uniform distribution and does not require any parameters to be fit other than its mixing proportion m_0. Details of the outlier model are given further below.

Given the choice of using noisy uniform flow in each patch, the next step is to define the component densities p_n of the mixture distribution. For $n > 0$ (i.e. other than the outlier process), $p_n(\vec{c}_k|\vec{v}_n)$ is meant to represent the likelihood of measuring the motion constraint vector $\vec{c}_k$ within a patch which has mean image velocity $\vec{v}_n$.

For the moment assume that the actual velocity is given by $\vec{v}_n$, then an exact constraint vector $\vec{c}_k$ would lie on the plane perpendicular to $\vec{v}_n$. In [6] it is shown that a reasonable approximation for the distribution of errors in component velocity measurements is given by a roughly Gaussian distribution for the *angular* error between $\vec{c}_k$ and the plane perpendicular to $\vec{v}_n$. The appearance of this angle should not be too surprising since, after all, we are simply measuring the orientation of a surface in space and time. We will make an additional assumption that this angular error distribution is independent of the actual image velocity $\vec{v}_n$ and is roughly isotropic.[3] Given these assumptions, the probability of observing $\vec{c}_k$, given that the actual image velocity is $\vec{v}_n$, is modeled by a Gaussian distribution in $d(\vec{c}_k, \vec{v}_n)$ defined by

$$d(\vec{c}_k, \vec{v}_n) = \frac{\vec{c}_k \cdot \vec{v}_n}{||\vec{c}_k||\,||\vec{v}_n||}.$$

Here $d(\vec{c}_k, \vec{v}_n)$ is simply the sine of the angular error which, for the small angles we are concerned with here, is roughly equal to the angular error itself. An important point about this error distribution is that it is only meant to model the measurements that are reasonably accurate. While the actual error distributions have longer tails than one might expect from such a Gaussian model (see [6]), this is not critical for our current situation since we also incorporate a model for outliers as a separate component in the mixture model. In effect, the longer tails are modeled by this outlier process, rather than by the above Gaussian distribution.

In addition to measurement noise within each patch, we also wish to accommodate the deviations of the actual flow from our constant velocity approximation. This additional variability is also modeled using angular errors from the mean velocity $\vec{v}_n$. This is chosen simply for convenience, since the variance of this error can simply be added to the variance of the measurement error, to obtain the following Gaussian model for the n^{th} component

[3] The measurement scheme we use in Section 4 uses only two frames while the spatial support of the component velocity measurements is considerably larger. Thus we would expect the spatial orientation error to be smaller than the speed error, that is, the noise should not be isotropic. This could be taken into account in a more detailed model.

distribution

$$p_n(\vec{c}|\vec{v}) = \frac{1}{\sqrt{2\pi}\sigma_v} \exp(-\frac{d^2(\vec{c},\vec{v})}{2\sigma_v^2}). \tag{3.5}$$

Here σ_v^2 is an estimate for the combined variance of the component velocity measurement errors and the modeling error in assuming a uniform velocity within the patch.

Given this specification of p_n we can solve the maximization step in closed form. Recall that this involves finding a solution of (3.3b), for a fixed set of mixture probabilities m_n and ownership probabilities q_{nk}. We omit the derivation, and simply state that the solution is given by choosing the new approximation for $\vec{v}_n$ as the eigenvector corresponding to the minimum eigenvalue of the 3×3 matrix

$$D_n \equiv \sum_{k=1}^{K} \frac{q_{nk}}{\sigma_v^2 ||\vec{c}_k||^2} \vec{c}_k \, \vec{c}_k^{\,T}. \tag{3.6}$$

This result is easy to justify intuitively. Consider the quadratic form

$$\begin{aligned} \vec{v}_n^{\,T} D_n \vec{v}_n &= \sum_{k=1}^{K} \frac{q_{nk}}{2\sigma_v^2 ||\vec{c}_k||^2} (\vec{c}_k \cdot \vec{v}_n)^2, \\ &= \sum_{k=1}^{K} \frac{q_{nk}}{2\sigma_v^2} d^2(\vec{c}_k, \vec{v}_n), \end{aligned}$$

which we recognize as minus one times the expected value of the exponent in the probability distribution p_n. By choosing the eigenvector $\vec{v}_n$ associated with the minimum eigenvalue of D_n we are simply maximizing the expected value of this exponent, as is standard in maximum likelihood estimates.

3.2 Modeling Outlier Processes

The distribution $p_0(\vec{c}_k)$ in the mixture is meant to model outliers. One common approach is to choose p_0 to be a Gaussian distribution with a large variance. The large variance attempts to model the long tails in the typical error distributions for component velocity measurements. Unfortunately, a single Gaussian outlier model is often insufficient. For example, an important case for motion estimation is the situation in which a patch overlaps an occlusion boundary. In such a case we may obtain two "signal" distributions, with one peak at the velocity of the foreground and a second peak at the velocity of the background. Each of these peaks has the typical long tails, and moreover, measurements straddling the boundary can provide constraints far from either peak. In order to model this behaviour we would need at least two broad Gaussians to approximate the tails around the peaks, and a third distribution to capture the yet more widely distributed responses scattered by the occlusion boundary. The main point being that it is very unlikely that we can find a reasonable model of the outlier processes in terms of a single Gaussian distribution.

It is worthwhile to consider the opposite extreme, where we use many Gaussian distributions to model the outlier distribution. As we show below, this leads to a computationally simple implementation. Imagine that we covered the sphere of possible image velocities (represented by unit vectors) using Gaussians having a significantly larger standard deviation than σ_v, the standard deviation of the signals we are seeking. At any point on the sphere,

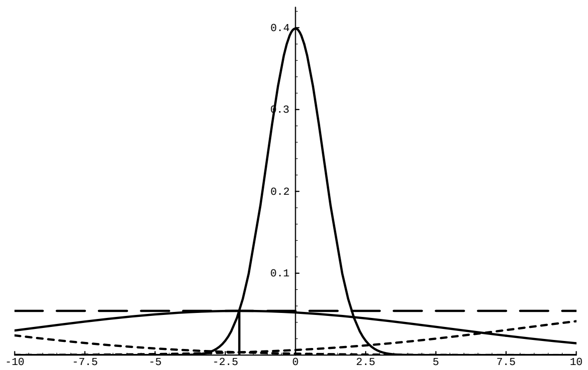


Figure 3.1: A coherent motion is represented by a Gaussian with small variance while outliers are represented by a tiling of broad Gaussian distributions. The dashed line represents an outlier threshold which is an upper bound on the sum of the outlier distributions; in this case, the intersection with the coherent motion distribution is set to be two standard deviations.

the sum of the values of all the Gaussians in this cover should be roughly given by a constant, as is depicted in Figure 3.1. By using a larger standard deviation for this cover we are constraining any fit of an outlier distribution to be a smoothly varying function over the sphere. Appropriate mixture coefficients for the various Gaussians, holding the means and variances fixed, is provided by the ownership probabilities, q_{nk}, for each broad Gaussian. As we saw in the discussion of the E-step above, these quantities can easily be computed from an equation of the form (3.4).

In fact, it can be shown that an *upper bound* for the outlier probability is provided by the case in which all of the mixture probability for the members of the cover is concentrated on a single element of the cover, and this element is such that the constraint agrees with its mean value. (We are free to play with the positioning of the elements in the cover, since we are just seeking the value of this upper bound.) This upper bound can also be obtained by simply using p_0 to be the constant value provided by the cover, and treating the mixture probability m_0 as the probability of getting an outlier, integrated over all possible positions. Using p_0 and m_0 in equation (3.4) gives an estimate q_{0k} which is an upper bound for the probability that constraint k belongs to the outlier distribution. The corresponding algorithm for identifying outliers is thus trivial and it is conservative with respect to which constraints are treated as signal rather than outliers.

All that remains is to discuss the choice of the constant p_0. We find it convenient to consider a situation in which the outlier probability should be about 1/2, and use this situation to set p_0. For example, consider a single Gaussian having a standard deviation of σ_v, with a mixture probability of m'_1, while outliers account for the remaining data. Moreover, assume that for this choice of mixture proportions, data a distance $\rho\sigma_v$ from the mean is to be assigned an outlier probability of 1/2. These parameters m'_1 and ρ are then used to set the value of p_0, which then remains fixed during the execution of the algorithm. In particular, using (3.4) we find

$$p_0 = \frac{m'_1}{(1-m'_1)\sqrt{2\pi}\sigma_v} \exp(-\rho^2/2). \tag{3.7}$$

In our experiments we use $m'_1 = 0.9$ and $\rho = 2.5$. In this case, when we have 10% outliers, the quantity ρ dictates some trust in constraints coming within 2.5 standard deviations of our estimated mean. However, as the mixture proportions change during the execution of the algorithm, so will this region of trust; the region decreases with an increase in the percentage of outliers even though p_0 remains fixed.

3.3 Summary of the EM-algorithm.

Given a choice of the constants p_0 and σ_v, the EM-algorithm proceeds as follows. First, using (3.4) the ownership probabilities are computed according to the current values of the mixture probabilities, $\vec{m}$, and the current estimates for the mean velocities $\vec{v}_n$, $n = 1, 2$. The mean velocity $\vec{v}_n$ can then be updated by constructing the 3×3 matrices D_n, and finding the eigenvector associated with the minimum eigenvalue. This needs to be done for $n = 1$ and 2, but there is no need to fit any such parameters for the outlier process. Also, a new set of mixture probabilities is obtained from (3.3a), followed by a renormalization to ensure the sum of the mixture probabilities is one. Note that here we also update the mixture probability m_0 for the outlier process. This entire EM-iteration is repeated until the change in the parameters is sufficiently small. For our experiments here we simply used 10 iterations, which turned out to be more than sufficient.

4 Computational Examples

In order to demonstrate the feasibility of the approach we consider two real image sequences, one of which involves a simple occlusion boundary, while the other involves transparency. We examine the behaviour of our approach in areas of the images which contain multiple motions as well as noisy regions and regions which do not conform to the simple uniform-motion assumption.

From the wide range of different measurement strategies for component velocities or, equivalently, for the motion constraint vectors $\vec{c}_k$, we chose a phase-based approach. In previous work a similar method has been shown to provide reasonably accurate component velocities, with a low outlier rate [6]. The particular approach we use is based on only two consecutive frames, and the actual component velocity measurement method is similar to the phase-based stereo disparity measurement scheme discussed in [12]. Briefly, the basic steps in the component velocity measurement are to first convolve each frame with the complex band-pass filter $G_2 + iH_2$, for each of four different spatial orientations of the filter kernel [8]. This kernel is chosen because it is compact (eg. the fine-scale version is 9×9), has a simple analytical form, and the real and imaginary parts nearly form a quadrature pair. Two different spatial scales were used, one tuned to a spatial wavelength of 4 pixels while the other is tuned to a wavelength of 8. The complex responses of these convolutions were sampled at the rate of 1/4 of the wavelength and quantized to 8 bits. The spatiotemporal phase derivatives were calculated using these subsampled and quantized results. The phase derivatives in the x_1, x_2 and t directions supplied the coefficients of the motion constraint vector $\vec{c}_k$. Points at which the complex convolution response was below 4% of the maximum possible value were discarded (since responses at this low level have been crudely quantized to only a few different gray levels), as were points where the phase and amplitude derivatives failed the test for a singularity neighborhood [7]. Finally, in order to bring the two frames into rough alignment, the frames were sometimes spatially shifted by an integral multiple of the filter spacing. The appropriate shift was easily obtained by applying the mixture model

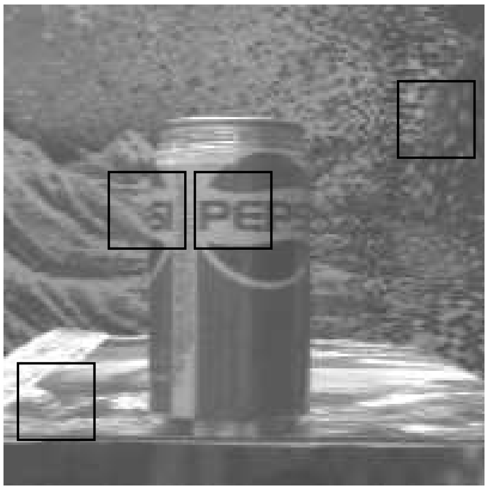


Figure 4.1: The **Pepsi Can** image sequence. The mixture approach is applied to the boxed regions of the scene (numbered 1 to 4 from left to right).

approach first to the wavelength 8 case, and using the results to choose the appropriate shift for the wavelength 4 case.

4.1 An Occlusion Boundary

Component velocities were obtained for the Pepsi can sequence in Figure 4.1. As suitable test cases for our mixture model approach we chose various 32×32 image patches, as depicted in Figure 4.1. The camera motion is purely translational and the image motion is to the left with speeds ranging roughly between 1.6 and 0.7 pixels/frame (for the can and the background, respectively). In all cases we take the standard deviation of the mixture model within a patch to be $\sigma_v = 0.2$ pixels/frame.

The second patch from the left in Figure 4.1 is roughly centered on an occlusion boundary, and we begin our discussion with this patch. A third of all the motion constraints obtained for this patch are depicted in Figures 4.2*d*. Note the presence of the two clusters associated with the motions on either side of the boundary which are made clearer in Figures 4.2*a* and *b* where the two clusters are shown separately. The two white "X"'s mark the peaks of the extracted mixture model for this example, while the convergence to these peak values is illustrated by black in "X"'s in Figure 4.2*d*. This rapid convergence behaviour was typical in all our tests and moreover the convergence appeared to be rather insensitive to the initial guess. The recovered velocities were $(-1.53, -0.02)$ for the portion of the can, and $(-0.70,$ $0.01)$ for the background. The mixture probabilities were $(m_0, m_1, m_2) = (0.03, 0.67, 0.30)$ for the outliers, the can, and the background, respectively. The method has clearly recovered the velocities of both sides of the occlusion boundary without difficulty.

A couple of additional points can be made using this same patch. Figure 4.2*a* shows all the constraints deemed to have an ownership probability for the first motion larger than 0.4 (i.e. $q_{1k} > 0.4$), with the darkness of the constraint lines increasing with the probability that the constraint belongs to the first motion. A similar plot is given in Figure 4.2*b* for the constraints having an ownership probability larger than 0.4 for the second motion. Clearly the mixture components have picked out appropriate clusters of constraints. Note that

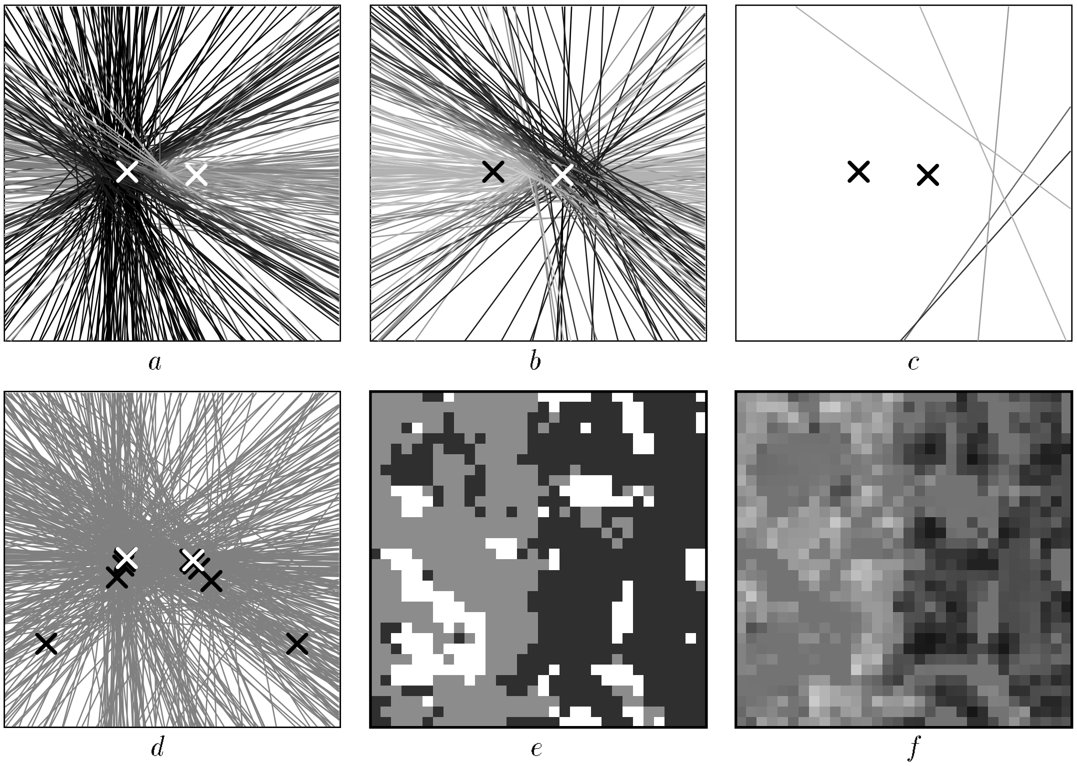


Figure 4.2: Region 2 (see text).

constraint lines that are roughly horizontal, and pass close to both peaks, have low ownership probabilities. This is appropriate since these constraints are roughly equally likely to arise from either motion, and thus their probabilities for any particular motion is close to 1/2. This illustrates the competition between the various components in the mixture model for the ownership of each constraint. There are only a handful of constraint lines that are outliers and are depicted in Figure 4.2*c*. Finally, in Figures 4.2*e* and 4.2*f* we show the spatial distribution of responses for the horizontal velocity $v_1(\vec{x})$ and the ownership probabilities $q_{2k}(\vec{x})$, respectively, for $\vec{x}$ varying over the patch. The general spatial distribution of the ownership probabilities reflects the structure and location of the occlusion boundary within the patch. To show the area of support for each motion, we have depicted the maximally probable horizontal velocity in Figure 4.2*e* (the white areas are regions where there were no component velocities, due to low amplitude or removal by the singularity neighborhood test). The majority of incorrectly classified pixels arise in areas where the ownership probabilities are near 1/2 (seen as neutral gray areas in Figure 4.2*f*).

The results for the other patches in Figure 4.1 are shown in Figure 4.3. Consider the results for Region 3 (the third patch from the left in the image) which contains only the single motion of the can. In this case, the mixture model collapses both peaks onto the same point, given by the velocity $(-1.61, -0.006)$ pixels/frame. This value is in excellent agreement with the velocity obtained for the can in the Region 2, indicating that the presence of the second motion there did not significantly perturb the responses.

In general, given a single motion within a patch, the situation is not this simple with both mixture components collapsing to a point. For example, in the patches on the extreme left and right (Regions 1 and 4) which both contain a single motion, the mixture model

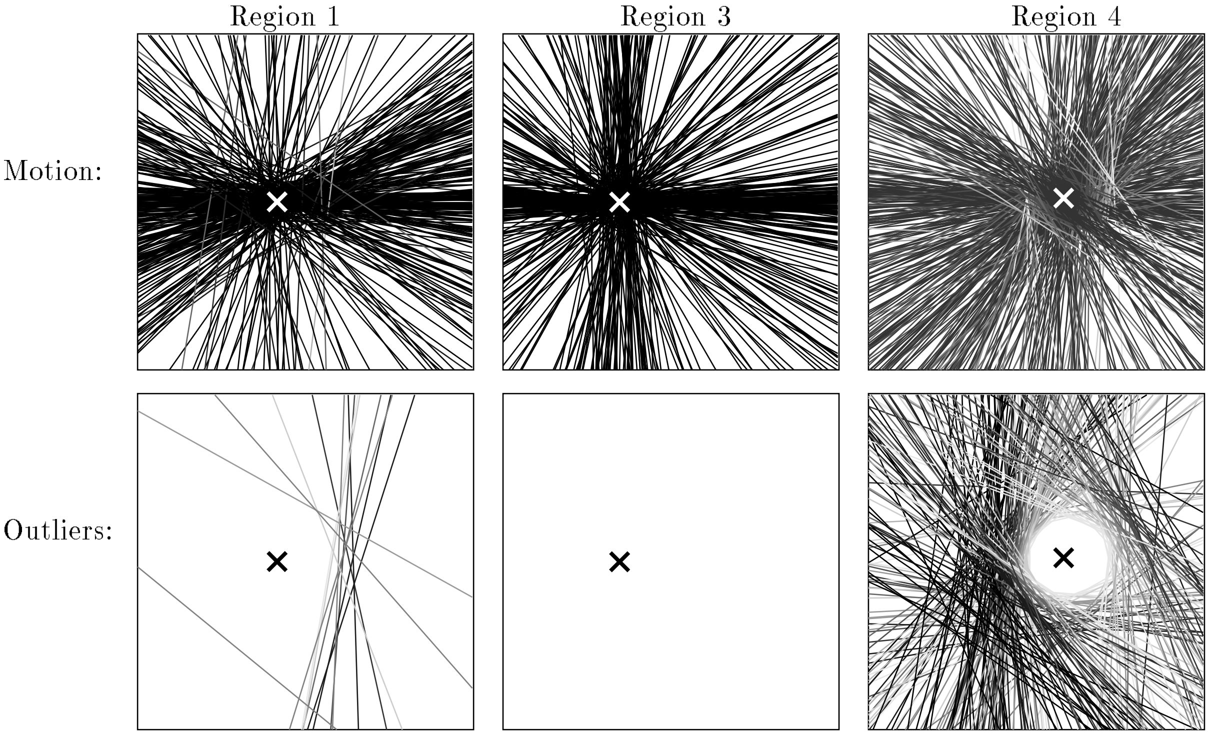
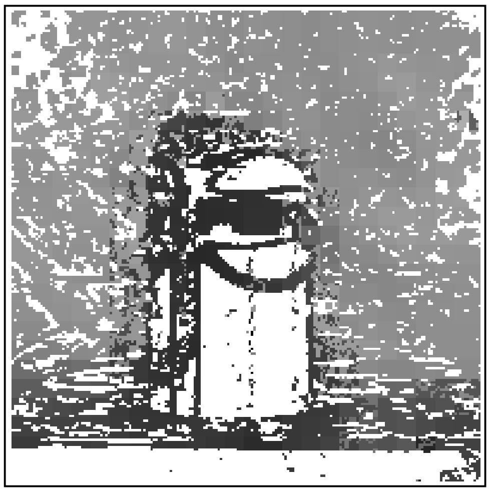


Figure 4.3: Constraint lines for Regions 1, 3, and 4. The top row shows the constraints corresponding to the dominant motion and the "X" marks the recovered motion (note that there is only a single motion in these regions). The bottom row shows the outliers (none for Region 3 and many in the noisy Region 4).

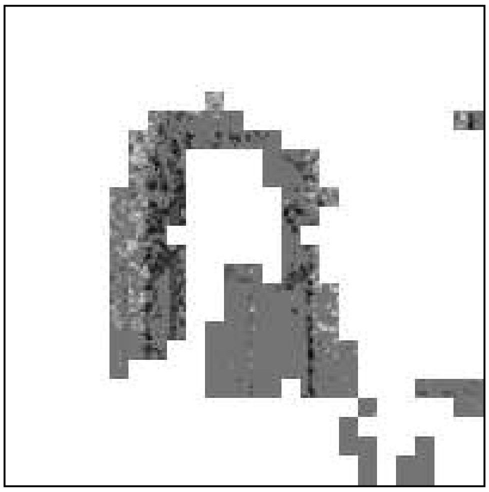
initially converges to descriptions having two different peak locations. Region 1 contains a planar surface slanted away from the camera. In this case, our simple assumption of uniform motion over the region is not a good approximation to the true motion and the resulting constraint lines do not form a tight cluster. When we assume that two uniform motions are present the method recovers horizontal component velocities of -1.4 and -1.0 pixels/frame, while the vertical components are essentially zero. The distance between these two motions, however, is only $2\sigma_v$, which is not a sufficient spread in order for the sum of the two Gaussian distributions to have more than a single peak. A similar situation occurs in Region 4 on the far right of the image. In this region there is a considerable amount of noise in the measurements, with 17% of the constraints labeled as outliers. In this noisy situation the mixture model also chooses a pair of velocities roughly separated by $2\sigma_v$ and, thus, a unimodal distribution.

These observations suggest a simple decision criterion for whether the velocities within a patch belong to a single layer. In particular, we consider the criteria that the minimum probability on the line connecting the two peaks must be at least half the height of the peaks in order to be merged. This criterion can also be based on the Mahalanobis-distance between the two peaks. When we recognize that the motions within a region should be merged we rerun the EM-algorithm assuming a single motion and outliers. The results for Regions 1 and 4 are shown in Figure 4.3.

Figure 4.4 shows the result of applying the mixture model approach over the entire image in 32×32 patches which are separated from each other by 8 pixels in both directions. We



(a) Horizontal Flow



(b) Mulitple Motions

Figure 4.4: Mixture models applied to the entire Pepsi-can image (see text).

use a simple uniform motion model within the patches and begin by assuming two coherent motions and outliers. After convergence, we apply the above decision criteria to determine if there is one motion or two within each patch. The horizontal component of the flow is shown in Figure 4.4*a* (in the same format as Figure 4.2*e* where white indicates no information, and the magnitude of the dominant motion is displayed in shades of gray). The regions containing two well-separated motions are shown in Figure 4.4*b* (in the same format as Figure 4.2*f*). Note that most of the boundary of the can has been identified, along with a few incorrectly classified regions in which the flow is quite noisy.

4.2 Transparency

We next consider a case of additive transparent motion in which a face is reflected in the glass covering an Escher print (Figure 4.5*a*). The entire image was treated as a single region, two uniform translational motions were assumed, and a noise estimate of $\sigma_v = 0.1$ was used. The initial motion constraints were computed with wavelength 8.[4] The recovered motion parameters were $(-3.31, 0.02)$ for the Escher print and $(-0.79, 0.01)$ for the reflection of the face.

The accuracy of these motion parameters can be evaluated by performing a simple computation. We compute the difference between the second image and the first image shifted by one set of motion parameters. This has the effect of canceling the intensity structure which is consistent with the motion and revealing the structure of the other surface. Figures 4.5*b* and *c* show the results obtained by canceling the effect of the moving print and reflection respectively. These results obtained from two frames compare favorably with those of [2] which required three frames to recover the two motions.

[4] A large wavelength was necessary to simultaneously compute constraints for both motions since they differ by approximately 2.5 pixels.

a *b* *c*

Figure 4.5: **Where's Jim?** Transparency sequence containing the reflection of a mystery vision researcher (see text).

5 Conclusion

We have examined the problems posed by multiple motions and outliers in the intergration of motion information over a spatial neighborhood. We relax the assumption of a single motion and, instead, view image regions as containing multiple layers corresponding to surfaces with different image motions. We also cope with outliers which can decease the accuracy of the recovered flow. To achieve this we introduced the idea of using mixture models for integrating noisy constraints when there are multiple interpretations and we provided details of the EM-algorithm for computing the maximum-likelihood estimate of the motion parameters. Our experiments demonstrate the feasibility of the approach and indicate that it is computationally efficient and can provide robust estimates of the optical flow values in the presence of outliers and multiple motions.

Acknowledgements

We thank J. Heel for providing the Pepsi-can image sequence and J. Bergen for providing the transparency sequence. Funding for this work was provided, in part, by the ARK (Autonomous Robot for a Known environment) Project, which receives its funding from PRECARN Associates Inc., Industry Canada, the National Research Council of Canada, Technology Ontario, Ontario Hydro Technologies, and Atomic Energy of Canada Limited. The authors also gratefully acknowledge the financial support of NSERC Canada.

References

[1] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In G. Sandini, editor, *Proc. of Second European Conference on*

Computer Vision, ECCV-92, volume 588 of *LNCS-Series*, pages 237–252. Springer-Verlag, May 1992.

[2] J. R. Bergen, P. J. Burt, R. Hingorani, and S. Peleg. A three-frame algorithm for estimating two-component image motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(9):886–896, September 1992.

[3] M. J. Black. *Robust Incremental Optical Flow.* PhD thesis, Yale Univeristy, New Haven, CT, 1992. Research Report YALEU/DCS/RR-923.

[4] A. Blake and A. Zisserman. *Visual Reconstruction.* The MIT Press, Cambridge, Massachusetts, 1987.

[5] T. Darrell and A. Pentland. Robust estimation of a multi-layer motion representation. In *Proc. IEEE Workshop on Visual Motion*, pages 173–178, Princeton, NJ, October 1991.

[6] D.J. Fleet. *Measurement of Image Velocity.* Kluwer, Boston, 1992.

[7] D.J. Fleet, A.D. Jepson, and M. Jenkin. Phase-based disparity measurement. *CVGIP: Image Understanding*, 53(2):198–210, March 1991.

[8] W.T. Freeman and E.H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):891–906, 1991.

[9] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, November 1984.

[10] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions.* John Wiley and Sons, New York, NY, 1986.

[11] B. K. P. Horn. *Robot Vision.* The MIT Press, Cambridge, Massachusetts, 1986.

[12] A.D. Jepson and M. Jenkin. Fast computation of disparity from phase differences. In *Proc. Computer Vision and Pattern Recognition, CVPR-89*, pages 398–403, San Diego, 1989.

[13] S. Madarasmi and D. Kersten. The visual perception of surfaces, their properties, and relationships. this volume.

[14] G.J. McLachlan and K.E. Basford. *Mixture Models: Inference and Applications to Clustering.* Marcel Dekker Inc., N.Y., 1988.

[15] M. Nitzberg and D. Mumford. The 2.1-D sketch. In *Proc. Int. Conf. on Computer Vision, ICCV-90*, pages 138–144, Osaka, Japan, December 1990.

[16] M. Okutomi and T. Kanade. A locally adaptive window for signal matching. *International Journal of Computer Vision*, 7(2):143–162, January 1992.

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