MRF solutions for probabilistic optical flow formulations

Sebastien Roy
University of Montreal
Montreal, Quebec, CANADA

Venu Govindu
NEC Research Institute
Princeton, NJ 08540
USA

Abstract

In this paper we propose an efficient, non-iterative method for estimating optical flow. We develop a probabilistic framework that is appropriate for describing the inherent uncertainty in the brightness constraint due to errors in image derivative computation. We separate the flow into two one-dimensional representations and pose the problem of flow estimation as one of solving for the most probable configuration of one-dimensional labels in an Markov Random Fields (MRF) with linear clique potentials. The global optimum for this problem can be efficiently solved for using the maxflow computation in a graph. We develop this formulation and describe how the use of the probabilistic framework, the parametrisation and the MRF formulation together enables us to capture the desirable properties for flow estimation, especially preserving motion discontinuities. We demonstrate the performance of our algorithm and compare our results with that of other algorithms described in the performance evaluation paper of Barron et. al [2].

1 Introduction

Motion estimation is an important problem since it arises in different computer vision tasks. When the motion between images is small, it is described by the optical flow defined as the two-dimensional motion field between two different views. Under constant brightness assumptions (CBA), we can constrain the pixel motion along a single dimension. Since the flow at a pixel has two components, optical flow estimation is an inherently ill-posed problem. Most methods overcome this limitation by ”regularising” the flow field, ie. by enforcing some form of smoothness on the flow field [4, 8]. The CBA can also be cast as an energy minimisation, where the flow field is estimated by minimising the least squares difference between two images [1, 11]. Lucas [7] compute the flow by “intersecting” local brightness constraints over a small patch. The smoothness problem can also be addressed by fitting a parametric global motion model (eg. [12]). In contrast, in this paper we are interested in solving for the dense, non-parameteric flow while preserving discontinuities in the flow field.

Most estimation methods achieve a balance between the brightness constraint and smoothness by minimising a cost function. Since they depend on iterative, non-linear techniques, these methods are not guaranteed to converge to the global minimum and thus give unsatisfactory results when they converge to a local minimum. We overcome this limitation by formulating the problem of flow estimation as a labelling problem in a Markov Random Field (MRF) framework. For certain classes of MRF’s, the exact Maximum A Posteriori (MAP) estimate can be obtained efficiently by a maximum flow computation on a graph. Being guaranteed to be optimal, this computation avoids the problem of local minima. As will be clarified later, this is a huge advantage especially in preserving motion discontinuities. Some recent methods that use the MRF formulation and a graph-theoretic solution are [9, 5, 3].

Another significant problem in flow estimation is the computation of image derivatives. Since the image is discretized in the spatial, temporal and intensity dimensions, the accuracy of the discrete computation of spatio-temporal derivatives is limited. This problem is partially addressed by sophisticated derivative filters. In practice the derivatives are also corrupted due to deviations from the constant brightness assumption such as change of illumination, brightness scaling and specularities. Hence the brightness constraint should not be considered to be a ”true” rigid constraint. To capture this notion of uncertainty, we cast the brightness constraint in a probabilistic framework. A related example of a probabilistic interpretation of optical flow is that of Simoncelli *et al* [10] which we will discuss later.

In Section 2, we formulate the problem of optical flow estimation. In Section 3 we develop the probabilistic inter-

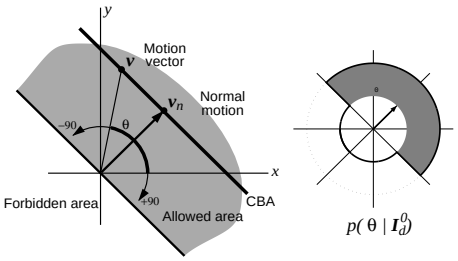

Figure 1. The Brightness Constraint and corresponding conditional distribution of flow orientations. All orientations in the half-circle centered on the normal flow vector are equally likely.

pretation of the brightness constraint. The MRF formulation and its solution are described in Section 4. In Section 5 we discuss the results obtained using our method and compare them with the results in Barron *et al* [2]. We also show an experiment which illustrates the performance of our algorithm in the presence of flow discontinuities.

2 Problem Formulation

The brightness constraint is derived assuming the constancy of image brightness of a pixel. Therefore, we have

$$\begin{aligned} & I(x,y,t) = I(x+v_x, y+v_y, t+1) \\ \Rightarrow \quad & (\frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial t})I = 0 \\ \Rightarrow \quad & v_x I_x + v_y I_y + I_t = 0 \end{aligned} \tag{1}$$

where I_x, I_y and I_t are the spatio-temporal image derivatives and v_x, v_y are the components of flow along the x and y directions. This constraint describes the equation of a line (See Fig. 1). As previously mentioned, the brightness constraint has to be relaxed due to the inherent uncertainty in image derivatives. As will be shown later, using simple and intuitive priors in a Bayesian framework can yield useful models of the CBA.

For convenience, we shall use the following notation throughout the rest of the paper. The spatial derivatives $[I_x, I_y]$ are denoted as ∇I and the spatio-temporal derivatives $[I_x, I_y, I_t]$ are denoted as I_d. The flow at a given pixel is denoted as v, ie. $v = [v_x, v_y, 1]$.

The true spatio-temporal derivatives, denoted ${I_d}^0$, constrains the flow vector v to lie on the line described by ${I_d}^0 \cdot v = 0$, as shown in Fig. 1. In a probabilistic manner, we define $P(v|I_d)$ as the probability of flow conditioned on the noisy image derivatives I_d. We define the error model for image derivatives as

$$I_d = {I_d}^0 + n \qquad n \sim N(0, \Sigma) \tag{2}$$

where n is the error in our observation and is assumed to be Gaussian distributed with zero mean and some covariance Σ . To derive $P(v|I_d)$, we use Bayes rule to obtain

$$P(v|I_d) = \int P(v|{I_d}^0) P({I_d}^0|I_d) d{I_d}^0 \tag{3}$$

Since we have an additive noise model, the conditional probability $P({I_d}^0|I_d)$ is a Gaussian distribution with mean I_d and covariance Σ . Hence given a prior distribution of flow conditioned on the true image derivatives $P(v|{I_d}^0)$ we can describe the desired conditional probability $P(v|I_d)$. We would like to point out that Simoncelli *et al* [10] use a similar probabilistic formulation. In fact, they describe the same conditional probability as Eqn. 3. However, their formulation differs from ours in two ways. Firstly, their noise model places the source of errors on the flow vectors instead of the image derivatives which are known to be erroneous in the first place ! Secondly, their formulation requires them to choose a prior distribution on the flow vectors $P(v)$. This prior is very hard to describe and changes with the type of motion, distribution of depths in the scene etc. Moreover, analytical tractability imposes a choice of zero-mean, Gaussian distribution for $P(v)$ which is seldom realised in practice.

In our formulation we only need to choose the conditional distribution $P(v|{I_d}^0)$, the flow probability given the true image derivatives of a pixel. This prior is easier to motivate and does not require knowledge of the global motion patterns $P(v)$. The selection of this prior and its impact on the solutions is discussed in the following section.

3 Probability models for the Brightness Constraint

As can be observed from Fig. 1 , the unknown component of the flow vector v lies on the CBA line and can be parametrised by an angle θ. This separates the space of possible values of θ into an acceptable (shaded) and an unacceptable zone. The acceptable zone is the half-circle centered on the normal flow vector v_n associated with ${I_d}^0$. The requisite prior $P(v|{I_d}^0)$ can now be described as the conditional prior of θ. In its weakest form the prior on θ simply assumes that flow orientations in the acceptable zone are

equally likely (See $P(\theta|{I_d}^0)$ in Fig. 1). The prior "kernel" is

$$P(\theta|{I_d}^0) = \begin{array}{l} \frac{1}{\pi}, \quad |\theta - \theta_n| \leq \frac{\pi}{2} \\ 0 \quad otherwise \end{array} \tag{4}$$

where θ_n is the orientation of the normal flow vector v_n. If desired any specific knowledge of flow can be used to change the conditional distribution of flow orientations. As an example, for strictly bounded flow magnitude the range of acceptable angular deviations from θ_n can be reduced.

Since the true flow is fully determined by θ the choice of the conditional prior $P(\theta|{I_d}^0)$ automatically fixes the conditional prior $P(v|{I_d}^0)$. It can be shown that

$$v = [\cos(\theta), \sin(\theta)]\|v_n\| \sec(\theta - \theta_n) \tag{5}$$

where v_n has a magnitude equal to $\frac{\|I_t\|}{\|\nabla I\|}$.

By combining Eqn. 3, Eqn. 4 and Eqn. 5 we can express $P(v|I_d)$ as a function of $P({I_d}^0|I_d)$. However this function does not have a simple analytic form. In practice it will be evaluated numerically.

Each pixel yields an image derivatives I_d. Subsequently, we generate a series of realisations drawn from the distribution $P({I_d}^0|I_d)$. For each realisation, the prior kernel is accumulated on the desired distribution, $P(v|I_d)$ [1]. Intuitively the conditional distribution of true flow $P(v|I_d)$ is the weighted average of different orientation kernels where the weights are determined by the conditional distribution $P({I_d}^0|I_d)$.

To illustrate the probability distributions described above, Fig. 2 shows the normal flow and conditional flow distributions $P(v|I_d)$ for three image derivatives, $[20, 20, 10]$, $[10, 10, 5]$ and $[4, 4, 2]$. These derivatives correspond to the same normal flow vector of $[-0.35, -0.35]$ observed in areas featuring different amounts of image texture. The error in image derivatives is assumed to be uniform, Gaussian distribution with a standard deviation of 1 in each of the spatio-temporal dimensions. For high levels of texture ($I_d = [20, 20, 10]$), the brightness constraint and the normal flow vector are reliable. Hence the resulting normal flow distribution is very compact and the full flow distribution is uncertain only along the brightness constraint line. In the case of medium texture ($I_d = [10, 10, 5]$), the uncertainty in both the position of the normal flow vector and the full flow increases. When the amount of image texture is low ($I_d = [4, 4, 2]$), the degree of uncertainty in both the normal and full flow values increases significantly. This

[1] For our experiments, we have used the kernel shown described by Eqn. 4 as the conditional distribution, $P(\theta|{I_d}^0)$

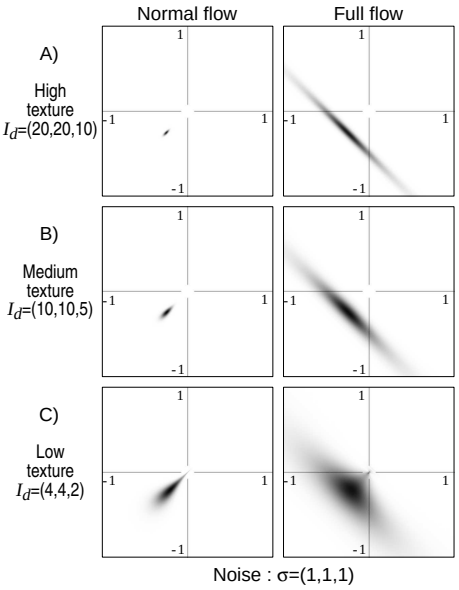

Figure 2. Probability distributions of normal and true flow for different amounts of texture.

corresponds to the intuition that the reliability of the normal flow and the brightness constraint depends on the amount of image texture present in a local patch. In low texture areas this model does not penalise large deviations from the brightness constraint line. In Fig. 3, we show the distributions obtained for the flow orientation and magnitude using the same values of I_d as in Fig. 2. As can be observed, the distribution of flow orientations is essentially invariant to the amount of texture available. However, the amount of texture significantly affects the probabilites of flow magnitudes. In the case of high texture, the normal flow is reliable, hence the full flow magnitude has to be greater than the normal flow magnitude (indicated by the vertical line). As the amount of texture decreases, the normal flow magnitude is less reliable and we have an increase in the probability of flow magnitudes that are less than the normal flow magnitude. This confirms the intuition that unreliable normal flows should not overly restrict the range of full flow values. An extreme case would be when there is no discernable motion, ie. $I_d \approx [0, 0, 0]$. In this case the simulated orientation distribution is uniform in the range $[-\pi, \pi]$. As a result, the orientation of such a pixel will be completely determined by the orientation of its neighbours, due to the imposed smoothness.

4 Solving for optical flow

Most methods estimate optical flow by minimising a cost function that trades off fidelity to the brightness constraint with a local smoothness assumption on the flow field. Due to depth discontinuities the flow field is typically piecewise smooth, ie. it contains smooth motion patches separated by large discontinuities. In most algorithms, the enforcement of smoothness causes the flow estimate to smooth across these boundaries resulting in incorrect estimates of flow. Generally, the resulting cost functions are minimised using iterative schemes for non-linear optimisations and are not guaranteed to converge to the global minimum. By formulating the flow estimation as a labelling problem for a restricted class of MRF models, we can avoid iterative methods and guarantee a globally optimal solution. The exact Maximum A Posteriori (MAP) estimate of this labelling problem can be obtained by a transformation into a maximum-flow problem on a graph. This global minimum tends to preserves large discontinuities.

The maximum-flow solution to the MAP estimate of the MRF requires the labels to be one-dimensional. Therefore, for our implementation we chose to parametrise the two dimensional flow field $[v_x, v_y]$ into a corresponding anglemagnitude representation $[\theta; m]$. We are unable to discuss this choice of parametrisation here due to space constraints.

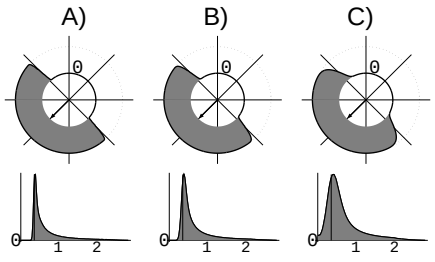

Figure 3. Probability distributions of flow orientation and magnitude for different amounts of texture. Each column shows the conditional distribution of flow orientation and magnitude for the different values of I_d used in Fig. 2.

For a detailed discussion on MRF's, the reader is referred to [6]. In our case, we have a problem of assigning a unique label (orientation or magnitude) at every site (pixel). One assignment of labels constitutes a configuration F drawn from a distribution. The Markovian property of an MRF is defined to be such that the probability of a site taking on a certain label f_i is dependent only on its neighbourhood. The Hammersley-Clifford theorem establishes that this probability can be related to a "clique potential" $V_c(f)$ through the Gibbs distribution [6], ie. $P(F = f) \propto e^{-U(f)}$ where $U(f) = \sum_{c \epsilon \mathcal{C}} V_c(f)$, ie. the clique potentials summed over all cliques. We consider cliques over a local neighbourhood $\mathcal{N}$ (in our case, the 4-neighbourhood of a pixel). In a Bayes formulation, we are interested in maximising the posterior probability $P(F = f|X = x)$ where x is the observed data. Using Bayes rule, we can observe that

$$P(F = f|X = x) \propto P(X = x|F = f)P(F = f). \quad (6)$$

Assuming that the noise is iid, we can define the likelihood term to be

$$P(X = x|F = f) = \prod_{i \epsilon \mathcal{S}} (X_i = x_i|F_i = f_i) \quad (7)$$

where the product is over all sites, ie. all pixels. In summary, the MAP estimate can be rewritten as an energy minimisation problem where the energy is

$$E(f) = \sum_{i \epsilon \mathcal{S}} \sum_{i' \epsilon \mathcal{N}_i} V(f_i, f_{i'}) - \sum_{i \epsilon \mathcal{S}} \ln(P(X_i = x_i|F_i = f_i)) \quad (8)$$

which contains a contribution from the label configuration and a contribution from the resulting clique potentials. Typically, the clique potentials reflect our prior knowledge of the problem and in the case of optical flow they are used to impose smoothness on the estimated flow field.

4.1 Maxflow solution for optical flow

As mentioned earlier we want to solve the labelling problem using a non-iterative, global minimisation method. We can achieve this by expressing $E(f)$ as a flow graph on which a maximum-flow computation is performed. In this context we require the clique potential $V(.,.)$ to be linear, yielding a smoothness term of the form $V(f_i, f_{i'}) = \beta|f_i - f_{i'}|$, where β is a proportionality constant that controls the amount of smoothness desired in the solution. Thus, the cost function to be minimized using maximum-flow computation is

$$E(f) = \sum_{i \epsilon \mathcal{S}} \sum_{i' \epsilon \mathcal{N}_i} \beta|f_i - f_{i'}| - \sum_{i \epsilon \mathcal{S}} \ln(P(X_i = x_i|F_i = f_i)) \quad (9)$$

The interested reader is referred to [9, 5], for details of the maxflow formulation and the MRF interpretation. The

proof of the global optimality of the minimum cost cut associated to the MAP estimate is a simple extension of the proof of Boykov *et al* in [3] and is beyond the scope of this paper. As mentioned earlier, our parametrisation of the flow field is the (θ, m) representation. To solve for the flow, we compute the conditional probabilities $P(\theta|I_d)$ as described in Section 3 by simply factorising the flow velocity distribution $P(v|I_d)$ into its angle and magnitude components $P(\theta|I_d)$ and $P(m|I_d)$ respectively. In order to solve for the orientation flow field Θ (denoting the configuration of orientation for all pixels), Eqn. 9 now takes the form

$$\min_{\Theta} \sum_{i \epsilon \mathcal{S}} \sum_{i' \epsilon \mathcal{N}_i} \beta |\theta_i - \theta_{i'}| - \sum_{i \epsilon \mathcal{S}} \ln(P(\theta|I_{di})) \qquad (10)$$

It may be noted that since the MRF scheme uses a finite number of labels, we need to discretise the range of values of $\theta = [-\pi, \pi]$ into a finite number of steps. Having solved for the flow orientation, we need to solve for the magnitude m for each pixel. Since computing the magnitude is much more difficult than the orientation of the flow, we chose to modify the conditional distribution $P(m|I_d)$ by using the information provided by the computed orientation estimate, ie. to $P(m|\theta_s, I_d)$ where θ_s is our solution for the orientation of a pixel. This modification is found to dramatically improve performance and can be explained by the fact that the orientation estimate restricts the full flow to a line thereby reducing the uncertainty of the magnitude distribution. By combining the two estimates ie. θ and m we obtain the optical flow between the two images.

5 Results and Discussion

In this section, we evaluate the performance of our algorithm by using synthetic and real datasets from Barron *et al.* [2], as well as comparing our results with those of various methods also mentioned in [2]. Run time for most experiments range between a few seconds for small images and upto 10 minutes on large images on a fast workstation. These run times can be easily reduced by using a coarser discretization of the motion parameters without significantly affecting the solutions. All results presented in this section are the raw flow fields obtained by our algorithms without applying any post-processing.

5.1 Synthetic Images

Our algorithm was run on 5 synthetic sequences of [2], for which ground truth was provided. We compared our results with those of the 5 algorithms in [2] that yield 100% flow field density. The results are summarized in Fig. 4.

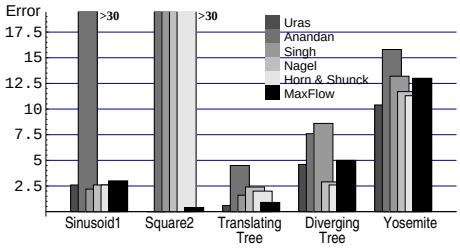

Figure 4. Results of various algorithms from [2] for various synthetic datasets, compared to ours (maximum-flow).

Our performance on these datasets is consistently good. However, these datasets all feature very smooth motion fields which do not reveal the behaviour of algorithms near motion discontinuities. This is an an important aspect of optical flow computation which is handled especially well by our method.

The most striking result is for Square2, where we outperform all other methods by orders of magnitude. This is a case where very sparse derivative information is available and therefore demonstrates the advantage of enforcing smoothness globally rather than locally. We also notice that we perform consistently better than the correlation-based algorithms (Anandan, Singh).

5.2 Real Images

To demonstrate the performance of our algorithm under realistic conditions, we computed the flow for 3 real images. These are the familiar Rubic Cube, Hamburg Taxi, and NASA Sequence also discussed in Barron *et al.* [2]. Since no ground truth is available, we can only display qualitative results. For a comparative evaluation with the other algorithms evaluated by Barron *et al*, we recommend that the reader consult [2].

The estimated flow field for the Rubic Cube is displayed in Fig. 5. This data set features a cube rotating on a turntable. We observe that the flow closely follows the motion of the turntable and the cube in both orientation and magnitude. The flow is well propagated over textureless regions like the top of the turn table. Moreover, motion discontinuities are well preserved.

The NASA Sequence features a divergent flow field induced by a camera zoom. The magnitudes of the motion are very small, typically much less than 1 pixel. As illustrated in Fig. 6, the divergence of the flow is well

recovered. Notice errors in the middle of the Coke can mostly induced by specularities coupled with low motion.

The Hamburg Taxi sequence is an example of multiple independent motions. Three vehicles are moving independently across the sequence. The resulting flow is shown in Fig. 7. The motions of the vehicles are well recovered and well localised, making it possible to segment the motions by using a simple thresholding of the motion magnitudes. This is an example where accurate recovery of motion discontinuities is critical.

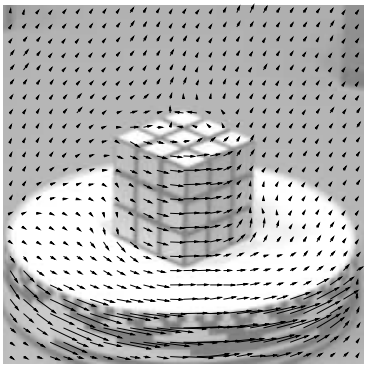

Figure 5. Estimated flow field for the Rubic's cube. The motion estimate is accurate and the discontinuities are preserved.

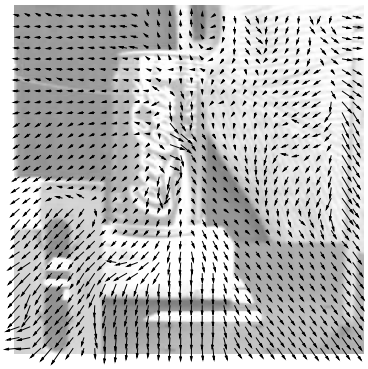

Figure 6. Estimated flow field for the NASA image pair. The true motion is a divergence.

5.3 Recovering Flow Discontinuities

The data sets with ground truth in Barron ([2]) are smooth flow fields. However, an important issue in optical

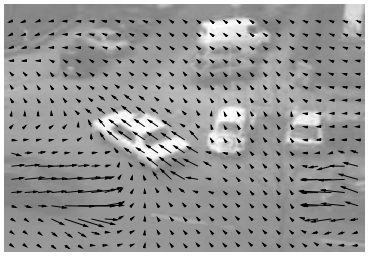

Figure 7. Hamburg Taxi sequence. An example of multiple, independent motions

flow estimation is the accurate recovery of flow field discontinuities since such motion discontinuities occur very frequently. Unlike most of the non-parametric methods for flow estimation, our algorithm is particularly suited for handling motion discontinuities in a natural manner. In general, since we can compute the exact global minimum of our cost function, we can accurately recover motion discontinuities that may be present.

To illustrate this property, we show a data set with explicit motion discontinuities. A textured pattern was used and this pattern has an associated flow field which is shown in Fig. 8. The flow field has a magnitude of 0.15 pixels every where and the orientation is either 0° or 90° as indicated. Thus in the case of infinite (or very high smoothness) the resultant flow field would have an orientation of 45° throughout. This in turn implies that the maximum possible error is 45°. In our experiment we are interested in comparing the behaviour of the Horn-Schunck algorithm and our method under high smoothness conditions. Since the interpretation of the smoothness parameter is different in each case, we chose to establish a baseline for comparison by applying smoothness values in each case till both of the flow estimates have the same average flow error. In our case, we chose a high error level of 25°, ie. we run both algorithms with higher levels of smoothness till we attain the requisite error level. This is a very high level of smoothing and helps illustrate the behaviour of each method under such circumstances.

In Fig. 9, we show the orientation for the flow field estimates for the Horn-Schunck method and our algorithm. As can be observed, the large amounts of smoothness has violated the motion discontinuity by simply smoothing across the discontinuities. In the case of our algorithm, the level of smoothness required to achieve the same amount of error is much higher. However, our globally optimal solution is such that the penalty for smoothing across boundaries is very high, hence the motion discontinuities

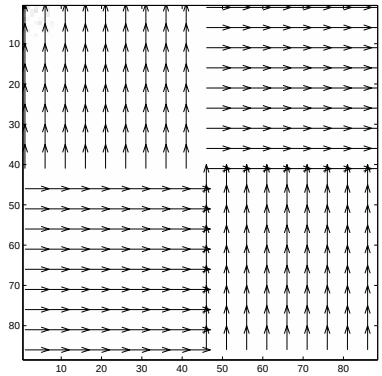

Figure 8. Flow pattern used for our experiments on preservation of motion discontinuities.

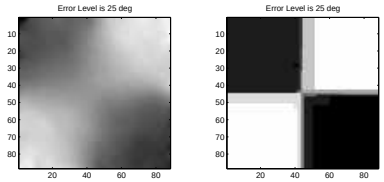

Figure 9. The figure on the left shows the flow orientation for the Horn-Schunck method and the one on the right shows the orientation for our method. In both cases, smoothness is applied till the average error is 25°. As is obvious, the motion discontinuities are well preserved in our method.

are accurately preserved. Instead, we observe that the flow field values in homogeneous regions are moved closer to each other so as to reduce the penalty paid at the boundaries. This implies that the algorithm trades off bias in the solution to preserving motion discontinuities. We would like to point out that such desirable preservation of flow field discontinuities will not be possible in sub-optimal estimates given by most minimisation methods, whereas it is inherent to our algorithm since we can solve for the global minimum of our cost function using efficient graph based formulations.

6 Conclusion

In this paper we have introduced a new method for estimating optical flow in a probabilistic framework. We explicitly account for the inherent inaccuracy of image derivatives using a simple noise model which in turn results in a probabilistic model of the full flow. By separating the flow into its angle-magnitude components, we compute the full flow in two steps, each based on a MAP estimate of an MRF with linear clique potentials. These estimates are optimal and obtained efficiently through a maximum-flow computation over a graph. The recovered flow fields are dense and retains sharp motion discontinuities. We believe that careful probabilistic modeling can achieve high levels of robustness to the significant errors inherent to the problem of optical flow estimation.

References

[1] P. Anandan. A computational framework and an algorithm for the measurement of structure from motion. 2:283–310, 1989.

[2] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical fbw techniques. 2(1):43–77, 1994.

[3] Y. Boykov, O. Veksler, and R. Zabih. Markov random fi elds with effi cient approximations. pages 648–655, 1998.

[4] K. Horn and B. Schunck. Determining optical fbw. *Artifi cial intelligence*, 17:185–203, 1981.

[5] H. Ishikawa and D. Geiger. Occlusions, discontinuities, and epipolar lines in stereo. Freiburg, Germany, 1998.

[6] S. Li. Markov random fi eld modeling in computer vision, springer. In *Tokyo 95*, page 1995, 1995.

[7] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *DARPA IU Workshop*, pages 121–130, 1981.

[8] H. Nagel. On the estimation of optical fbw: Relations between different approaches and some new results. *Artifi cial Intelligence*, 33:299–324, 1987.

[9] S. Roy and I. Cox. A maximum-fbw formulation of the n-camera stereo correspondence problem. In *International Conference on Computer Vision*, pages 492–499, Mumbai,India, 1998.

[10] E. Simoncelli, E. Adelson, and D. Heeger. Probability distributions of optical fbw. pages 310–315, 1991.

[11] A. Singh. *Optic Flow Computation: A Unifi ed Perspective*. IEEE Computer Society Press, 1992.

[12] S. Srinivasan and R. Chellappa. Optical fbw using overlapped basis functions for solving global motion problems. pages 288–304, 1998.