To appear in IEEE ICIP-94, Austin, Texas, November 13-16, 1994.

MOTION SEGMENTATION AND ESTIMATION

Tina Yu Tian and Mubarak Shah

Computer Vision Lab
Computer Science Department
University of Central Florida
Orlando, FL 32816
email: {tian, shah}@cs.ucf.edu

ABSTRACT

In this paper, we apply mean field technique and present a deterministic algorithm to determine the optical flow and motion boundaries. To deal with the problem of large motion, we present an adaptive multigrid approach, which also greatly reduces the computation time. This algorithm is fully parallelizable and iterative. The performance of the proposed method is compared against known algorithms using performance measures proposed by Barron et al [3]. Experimental results indicate that our approach provides good estimates of optical flow and motion boundaries.

1. INTRODUCTION

Motion segmentation and estimation is the essential first step to solve general structure-from-motion (SFM) problem. The conventional gradient-based motion estimation approaches impose a global smoothness constraint on the image, which results in inaccurate optical flow estimates near motion boundaries. Some researchers, Poggio et al [12], Konrad and Dubois [11], Heitz and Bouthemy [8], Harris and Koch et al [7], introduced line processes and a piecewise smoothness constraint into the regulation formulation to improve the motion estimates. The problem is reduced to minimizing a non-convex cost functional, for which optimization methods can be used, e.g. simulated annealing [11], iterated condition mode (ICM) [8], and graduated nonconvexity (GNC) algorithm. Simulated annealing converges very slowly to the global minimum, ICM gets stuck into a local minimum very easily unless a good initial estimate is provided, and GNC, which is a deterministic simulated annealing, still converges slowly. Mean field technique recently has been shown [5] to be more appealing for minimizing the non-convex cost functional, since it converges much faster to the global minimum.

In this paper, we apply mean field technique and present a deterministic algorithm to determine the optical flow and motion boundaries. Recently, Vlontzos and Geiger [13] proposed an algorithm for optical flow estimation using mean field technique. In their method, the sparse optical flow field is computed first by window matching scheme, then the dense flow field is computed through the formulation of interpolating and smoothing the sparse flow field except at the motion boundaries. Our formulation is directly in terms of intensity, avoiding the intermediate step to compute flow by correlation. Zhang and Hanauer [14] applied mean field theory to compute *displacement* field, while we compute *optical flow* field. Our formulation is different from theirs, and the computation cost is small. Abdelqader and Rajala [1] proposed a mean field annealing approach to motion estimation in which only displacement field is involved in the MRF formulation. Our coupled MRF model includes flow field, horizontal and vertical line fields, and motion estimation and segmentation are solved simultaneously. Irani and Peleg [10] proposed a method for motion detection and segmentation. The difference between their method and ours is that they did not compute optical flow or displacement field, but approximated the motion of objects by 2D parametric models, then compute the parameters of these models and segmentation. Harris and Koch et al. [7] used the similar formulation, but implemented motion discontinuities by analog circuits. Unlike their work, we present a deterministic solution for the MRF motion formulation. The performance of the proposed method is compared against known algorithms using performance measures proposed by Barron et al [3], and it is shown the proposed algorithm outperforms those algorithms.

2. DETERMINING OPTICAL FLOW AND MOTION BOUNDARIES BY MRFS

2.1. Problem Formulation based on MRFs

For rigid motion, since the optical flow field $\vec{\mathbf{U}}$, where $\vec{\mathbf{U}} = (u, v)^T$, consists of patches of vectors of similar orientation and length with potential discontinuities at motion boundaries, the probability distribution for $\vec{\mathbf{U}}$ at site i depends only upon the neighborhood N_i. Therefore, the optical flow field $\vec{\mathbf{U}}$ can be modeled as a 2D vector MRF. Moreover, Geman and Geman [6] introduced another MRF field in the dual lattice, a binary field (the line process), which is composed of the horizontal line process, $d_{h_{ij}}$, and the vertical line process, $d_{v_{ij}}$. The line process $d_{h_{ij}}$ connects the site (i, j) to the site $(i, j-1)$, while $d_{v_{ij}}$ connects the site (i,j) to the site $(i-1, j)$ in the lattice. One of the main attractions of MRF models is that they can deal directly with discontinuities represented by the line process. In our approach,

The research reported in this paper was partially supported by NSF grants CDA 9122006 and IRI-9220768.

the horizontal line process creates motion discontinuities between pixels along the vertical direction, and similarly the vertical line process creates motion discontinuities between pixels along the horizontal direction.

Global Bayesian estimation associated with MRF models provides a mathematical framework to recover motion and detect motion boundaries from the observed image sequence. The problem we are considering is to estimate 2D optical flow field $\vec{\mathbf{U}}$ and line process fields, d_h and d_v, from the image gradient ∇I, where $\nabla I = (I_x, I_y, I_t)^t$. Therefore, we are concerned with a conditional probability for the fields $\vec{\mathbf{U}}$, d_h, and d_v given the image gradient ∇I: $P(\vec{\mathbf{U}}, \mathrm{d_h}, \mathrm{d_v} \mid \nabla \mathrm{I})$.

There are two possible types of errors in the estimation process. One type of error is in the derivative computations due to camera and quantization noise, aliasing, imprecision in the derivative filters, etc. If we describe this type of uncertainty by white Gaussian noise n_1 with standard deviation σ_1, then we have

$$P(\nabla I \mid u, v) = \frac{1}{C_1} e^{-\sum_{i,j} \frac{(u_{i,j} I_{x_{i,j}} + v_{i,j} I_{y_{i,j}} + I_{t_{i,j}})^2}{2\sigma_1^2}}, \quad (1)$$

where u_{ij} and v_{ij} are the optical flow field defined at site (i, j) of the lattice and C_1 is a normalization constant. The term in the exponential in (1) is based on the brightness constancy equation of Horn and Schunck [9].

Another type of error is caused by changes in lighting or reflectance, or the presence of multiple motions, in which case the brightness constancy constraint is not satisfied. This type of error can similarly be modeled independently of n_1 by white Gaussian noise n_2 with standard deviation σ_2. Within the Bayesian decision approach, a prior probability, $P(u, v, d_h, d_v)$, can be defined to impose piecewise smoothness constraint on the optical flow field:

$$P(u, v, d_h, d_v) = \frac{1}{C_2} e^{-V_1(u, v, d_h, d_v)}, \quad (2)$$

where C_2 is normalization constant, and

$$\begin{aligned} V_1(u, v, d_h, d_v) = \sum_{i,j} \quad \{ \quad & \frac{1}{2\sigma_2^2} [\|\vec{\mathbf{U}}_{\mathrm{i,j}} - \vec{\mathbf{U}}_{\mathrm{i-1,j}}\|^2 (1 - \mathrm{d}_{\mathrm{h_{i,j}}}) \\ & + \|\vec{\mathbf{U}}_{\mathrm{i,j}} - \vec{\mathbf{U}}_{\mathrm{i,j-1}}\|^2 (1 - \mathrm{d}_{\mathrm{v_{i,j}}})] \\ & + \quad \gamma^h d_{h_{i,j}} + \gamma^v d_{v_{i,j}} \}. \end{aligned} \quad (3)$$

$$\begin{aligned} \|\vec{\mathbf{U}}_{\mathrm{i,j}} - \vec{\mathbf{U}}_{\mathrm{i-1,j}}\|^2 &= (\mathrm{u_{i,j}} - \mathrm{u_{i-1,j}})^2 + (\mathrm{v_{i,j}} - \mathrm{v_{i-1,j}})^2, \\ \|\vec{\mathbf{U}}_{\mathrm{i,j}} - \vec{\mathbf{U}}_{\mathrm{i,j-1}}\|^2 &= (\mathrm{u_{i,j}} - \mathrm{u_{i,j-1}})^2 + (\mathrm{v_{i,j}} - \mathrm{v_{i,j-1}})^2. \end{aligned}$$

The first two terms in (3) contain the sum of local interactions of the flow field and line processes between the nearest neighbors in terms of smoothness constraint to solve the well-known aperture problem. If the magnitude of u-flow or v-flow in the horizontal or vertical direction is very high at site (i, j), the corresponding line process will be likely to signal a discontinuity ($d_{v_{ij}} = 1$ or $d_{h_{ij}} = 1$) to increase $P(u, v, d_h, d_v)$. The third term in (3) takes into account the penalty paid each time a discontinuity is created, and is needed to prevent the creation of discontinuities everywhere.

By Bayes's theorem, we have a posterior probability for variables u, v, d_h, and d_v, given the image gradient ∇I:

$$P(u, v, d_h, d_v \mid \nabla I) = \frac{1}{Z} e^{-V(u, v, d_h, d_v)}, \quad (4)$$

where Z is the normalization constant called partition function ([6]). Z can be computed by

$$Z = \sum_{\{u, v, d_h, d_v\}} e^{-V(u, v, d_h, d_v)},$$

where $\sum_{\{u, v, d_h, d_v\}}$ means the sum over all the possible configurations $\{u, v, d_h, d_v\}$ of the system. V is a cost function of the following form:

$$\begin{aligned} V(u, v, d_h, d_v) = \sum_{i,j} \quad \{ \quad & \frac{1}{2\sigma_1^2} (u_{i,j} I_{x_{i,j}} + v_{i,j} I_{y_{i,j}} + I_{t_{i,j}})^2 \\ & + \quad \frac{1}{2\sigma_2^2} [\|\vec{\mathbf{U}}_{\mathrm{i,j}} - \vec{\mathbf{U}}_{\mathrm{i-1,j}}\|^2 (1 - \mathrm{d}_{\mathrm{h_{i,j}}}) \\ & \qquad + \|\vec{\mathbf{U}}_{\mathrm{i,j}} - \vec{\mathbf{U}}_{\mathrm{i,j-1}}\|^2 (1 - \mathrm{d}_{\mathrm{v_{i,j}}})] \\ & + \quad \gamma^h d_{h_{i,j}} + \gamma^v d_{v_{i,j}} \}. \end{aligned} \quad (5)$$

Finding the maximum a posteriori estimate of fields u, v, d_h, and d_v is reduced to the problem of minimizing the energy function V.

2.2. Energy Minimization Using Mean Field Techniques

In this section, we apply the mean field technique to the minimization problem of the energy function associated with line processes. Since the mean field essentially represents the minimum variance Bayesian estimator, it can be used as a measure of field value.

A well-known result [5] from statistical mechanics and probability theory indicates that all mean values of the system can be obtained from the partition function Z, so we must compute the function Z. By introducing β, we have

$$Z = \sum_{\{u, v, d_h, d_v\}} e^{-\beta V(u, v, d_h, d_v)}.$$

The introduction of β will allow us to apply continuation methods similar to simulated annealing. From (5), the partition function becomes

$$\begin{aligned} Z \quad = \quad & \sum_{\{u, v\}} e^{-\beta \sum_{i,j} [\frac{(u_{i,j} I_{x_{i,j}} + v_{i,j} I_{y_{i,j}} + I_{t_{i,j}})^2}{2\sigma^2} + \gamma^h + \gamma^v]} \\ \times \quad & \sum_{\{d_h, d_v\}} e^{-\beta \sum_{i,j} [(1 - d_{h_{i,j}}) G^h_{i,j} + (1 - d_{v_{i,j}}) G^v_{i,j}]}, \end{aligned} \quad (6)$$

where

$$G^h_{ij} = \frac{1}{2\sigma_2^2} \|\vec{\mathbf{U}}_{\mathrm{i,j}} - \vec{\mathbf{U}}_{\mathrm{i-1,j}}\|^2 - \gamma^{\mathrm{h}}, \qquad \mathrm{G^v_{ij}} = \frac{1}{2\sigma_2^2} \|\vec{\mathbf{U}}_{\mathrm{i,j}} - \vec{\mathbf{U}}_{\mathrm{i,j-1}}\|^2 - \gamma^{\mathrm{v}}.$$

The line process term in (6) is the partition function of two independent spin systems, d_h and d_v, in an external field, G^h or G^v, with no interaction between neighboring

sites. Each spin contributes to the partition function independently of the others. Therefore, the line process can be successfully averaged out to yield an effective cost function depending only upon the fileds u and v. The partition function can then be rewritten as

$$Z = \sum_{\{u,v\}} e^{-\beta V_{eff}(u,v)},$$

where

$$V_{eff}(u,v) = \sum_{i,j}\{\frac{1}{2\sigma_1^2}(u_{i,j}I_{x_{i,j}} + v_{i,j}I_{y_{i,j}} + I_{t_{i,j}})^2 + \gamma^h + \gamma^v - \frac{1}{\beta}\ln[(1+e^{-\beta G^h_{i,j}})(1+e^{-\beta G^v_{i,j}})]\}.$$

Using the mean field technique [5], we can derive the following mean field equations for the line processes:

$$d_{h_{ij}} = \frac{1}{1+e^{\beta(\gamma^h - \frac{1}{2\sigma_2^2}\|\vec{\mathbf{U}}_{\mathrm{i,j}} - \vec{\mathbf{U}}_{\mathrm{i-1,j}}\|^2)}}, \tag{7}$$

$$d_{v_{ij}} = \frac{1}{1+e^{\beta(\gamma^v - \frac{1}{2\sigma_2^2}\|\vec{\mathbf{U}}_{\mathrm{i,j}} - \vec{\mathbf{U}}_{\mathrm{i,j-1}}\|^2)}}. \tag{8}$$

For simplicity we assume $\gamma^h = \gamma^v = \gamma$, where γ is one of the predefined parameters. Notice that the mean values of the line processes can vary continuously from 0 to 1.

We pointed out earlier that the estimate of u and v obtained by minimizing the energy function V coincides with the maximum a posteriori estimate of (4) after summing out the line process fields. This corresponds to minimizing V_{eff} with respect to u and v, and yields the following differential equations:

$$\frac{\partial}{\partial u_{ij}}V_{eff}(u,v) = 0, \qquad \frac{\partial}{\partial v_{ij}}V_{eff}(u,v) = 0.$$

We define the local average of discontinuity at pixel (i,j) as

$$\bar{d}_{i,j} = \frac{1}{4}[4 - d_{h_{i,j}} - d_{h_{i+1,j}} - d_{v_{i,j}} - d_{v_{i,j+1}}], \tag{9}$$

and the local averages of u and v at pixel (i,j) as

$$\bar{u}_{i,j} = \frac{1}{4}[(1-d_{h_{i,j}})u_{i-1,j} + (1-d_{v_{i,j}})u_{i,j-1} + (1-d_{h_{i+1,j}})u_{i+1,j} + (1-d_{v_{i,j+1}})u_{i,j+1}], \tag{10}$$

$$\bar{v}_{i,j} = \frac{1}{4}[(1-d_{h_{i,j}})v_{i-1,j} + (1-d_{v_{i,j}})v_{i,j-1} + (1-d_{h_{i+1,j}})v_{i+1,j} + (1-d_{v_{i,j+1}})v_{i,j+1}]. \tag{11}$$

Assume $\lambda = 4\frac{\sigma_1^2}{\sigma_2^2}$.

The iterative formulae are obtained as follows:

$$u_{i,j}^{(n+1)} = \frac{1}{\bar{d}_{i,j}^{(n)}}\{\bar{u}_{i,j}^{(n)} - \frac{I_{x_{i,j}}\bar{u}_{i,j}^{(n)} + I_{y_{i,j}}\bar{v}_{i,j}^{(n)} + \bar{d}_{i,j}^{(n)}I_{t_{i,j}}}{\lambda\bar{d}_{i,j}^{(n)} + I_{x_{i,j}}^2 + I_{y_{i,j}}^2}I_{x_{i,j}}\}, \tag{12}$$

$$v_{i,j}^{(n+1)} = \frac{1}{\bar{d}_{i,j}^{(n)}}\{\bar{v}_{i,j}^{(n)} - \frac{I_{x_{i,j}}\bar{u}_{i,j}^{(n)} + I_{y_{i,j}}\bar{v}_{i,j}^{(n)} + \bar{d}_{i,j}^{(n)}I_{t_{i,j}}}{\lambda\bar{d}_{i,j}^{(n)} + I_{x_{i,j}}^2 + I_{y_{i,j}}^2}I_{y_{i,j}}\}, \tag{13}$$

where the superscript denotes the iteration number.

During an updating cycle, the new approximation of optical flow, $u^{(n+1)}$ and $v^{(n+1)}$, can be determined from the estimated brightness derivatives, I_x, I_y, and I_t, and from the local average of the previous flow, $\bar{u}^{(n)}$ and $\bar{v}^{(n)}$, and line process estimates, $\bar{d}^{(n)}$, respectively. The line process can be updated from the previously estimated mean values of $u^{(n)}$ and $v^{(n)}$. Note that the forms of (7), (8), (12), and (13) are well suitable for a fast, parallel, and iterative scheme for a solution.

3. RESULTS

The real image sequence used for our algorithm contains two toys, Dale and Scrooge (Disney World characters), moving in the opposite directions. The measured displacement is about 1 pixel/frame for each toy. The edge map from Canny's edge detector is used as initial value for line processes. The results for optical flow and motion boundaries are shown in Figure 1.

To deal with the problem of large motion, we apply an adaptive multigrid approach, which is tested on Hamburg Taxi sequence Figure 2. This sequence contains a taxi turning the corner and a car moving from left to right, with approximately 1 and 3 pixels/frame, respectively. A three-level Laplacian pyramid is used for this sequence. Since the motion of the taxi is small, the computation for taxi portion continues from the coarsest level to the finest level because the estimated error of optical flow is always above a predefined threshold, and we have to keep on refining optical flow to reduce the error. However, for the car portion, computation terminates at the intermediate level since the estimated error of optical flow due to derivative estimation error and/or quantization error is below the threshold at the intermediate level. If the computation were continued further for the car portion, the optical flow would be corrupted at the finer level because of poor derivative estimation at this level. Laplacian pyramid for Frame 1 and the resulting optical flow at three different levels are shown in Figure 2. It is clear that optical flow has been refined through multiscales. We can observe that there is no oversmoothing due to introduction of line process, and adaptive multigrid algorithm does give better and faster result even for image sequences with different ranges of optical flow velocity.

We also applied our algorithm to a sequence containing occlusion shown in Figure 3(a) and (b). This sequence contains a toy fire engine and a toy truck, moving in the opposite directions. The measured displacement is about 1-2 pixels/frame for each toy. The results for optical flow obtained from Horn and Schunck and our algorithm are shown in Figure 3(d) and (c), respectively.

3.1. Comparison

Barron, Fleet and Beauchemin [3] present a comprehensive evaluation and comparison of existing optical flow methods. Here we compare our method with Anandan's method [2], Horn and Schunck's method [9] and Fleet and Jepson's

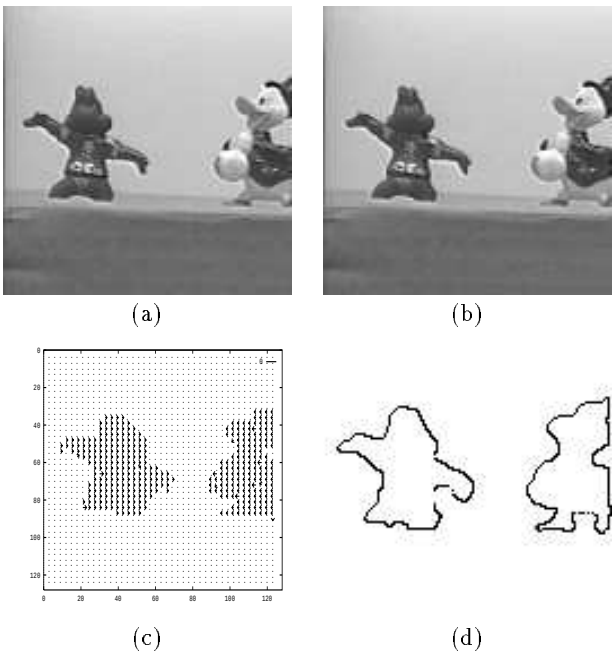

Figure 1: Results for toy images with $\sigma_1 = 1.73$, $\gamma = 5.0$, $\sigma_2 = 0.71$, and $\beta = 5.0$. (a) Frame 1. (b) Frame 2. (c) Optical flow. (d) Motion boundaries.

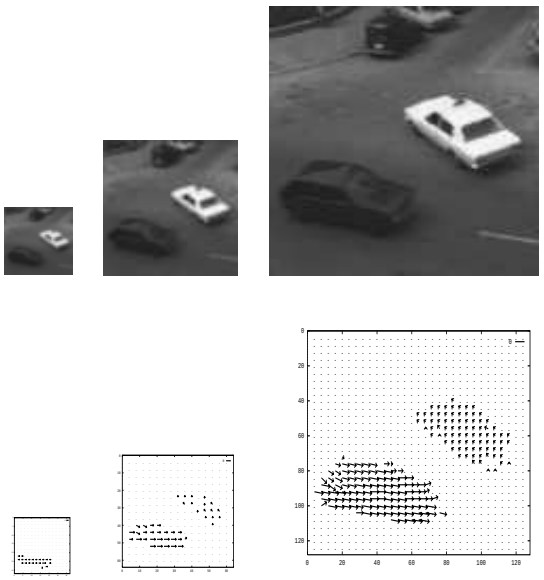

Figure 2: Laplacian pyramid for Frame 1 and the resulting optical flow at correspondent levels. $\sigma_1 = 3.75$, $\gamma = 7.0$, $\sigma_2 = 0.71$, and $\beta = 5.0$.

method [4] on the same inputs using the performance measures of Barron et al. The random-dot synthetic square image sequence is used for quantitative performance comparisons, since textured images are more realistic for the real scenes. Table 1 shows the average and standard deviation of error and density of the optical flow computed by these methods. The optical flow error is computed in terms of the angle between the correct velocity and the corresponding estimated velocity. It is surprising to see that Fleet and Jepson's method does not work well for the random-dot images which involve the texture and high frequency data, implying that their method may not work well with other real textured images. However, their method works well with uniform non-textured image data. In these experiments we have used Fleet and Jepson's program obtained by anonymous ftp (`ftp@csd.uwo.ca`). In order to be sure that we have used Fleet and Jepson's program properly, we also experimented with the translating uniform square image sequence, which is the same as the random-dot square sequence except that background and objects have constant gray levels, 0 and 255, respectively. 21 frames of the square images have been used as the input. Their program provided quite accurate results with average and standard deviation of error equal to 0.034^o and 0.028^o, respectively, which are very close to (0.07^o, 0.08^o), the error measurements of Fleet and Jepson's method for their uniform square sequence, as reported in [3]. The same program was then applied to the random-dot sequence, which provides the sparse optical flow field with average and standard deviation of error and density of the optical flow equal to 7.38^o, 11.79^o, and 13.2%, respectively.

Figure 4 illustrates the optical flow field obtained by

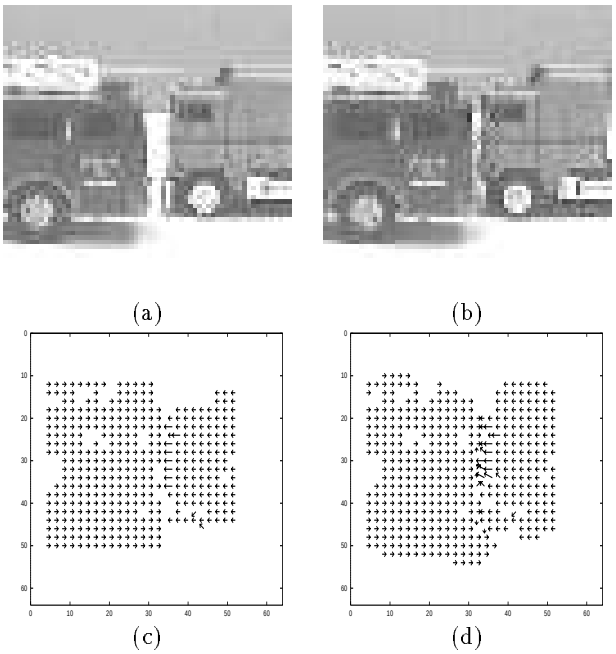

Figure 3: Occlusion and Multiple Motions: (a) Frame 1. (b) Frame 3. (c) Optical flow computed by our algorithm. (d) Optical flow computed by Horn and Schunck's algorithm.

Table 1: Performance comparison of optical flow methods on the random-dot square images.

Method	Avg. Error	Std. Dev.	Dens.
Fleet and Jepson	7.38^o	11.79^o	13.2%
Horn and Schunck	2.89^o	9.01^o	100%
Anandan	2.16^o	8.47^o	100%
Proposed method	0.61^o	3.70^o	100%

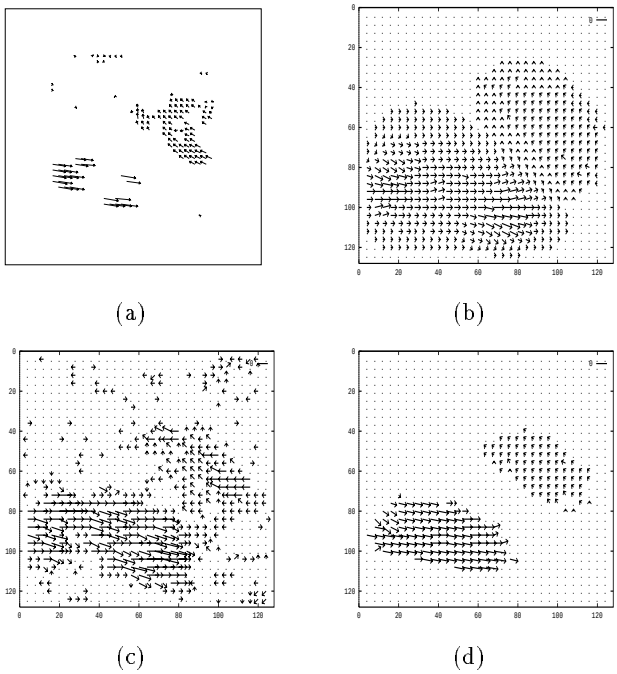

Figure 4: Comparison of optical flow for Hamburg Taxi sequence. (a) Fleet and Jepson's results. (b) Multigrid Horn and Schunck's results. (c) Anandan's results. (d) Our results.

multigrid Horn and Schunck's method, Anandan's method, Fleet and Jepson's method and our method for the Hamburg Taxi sequence shown in Figure 2. Fleet and Jepson's method provides a sparse optical flow field shown in Figure 4(d). In Horn and Schunck and Anandan's results shown in Figure 4(b)-(c), the optical flow field has been smoothed across the motion boundaries and some spurious optical flow vectors are present at the stationary regions. Neither Horn and Schunck's nor Ananadan's approach takes motion boundaries into account. Therefore, the results appear to be over-smoothed. However, our method explicitly takes motion boundaries into account, and provides better results as shown in Figure 4(a).

4. CONCLUSION

In this paper, we apply mean field technique and present a deterministic algorithm to compute the optical flow and motion boundaries. To deal with the problem of large motion, we present an adaptive multigrid approach, which also greatly reduces the computation time. Furthermore, the algorithm is fully parallelizable and iterative. Experimental results indicate that our approach provides good estimates of optical flow and motion boundaries. We can apply the SFM algorithm to each segmented moving object in the scene in order to obtain the motion parameters and depth maps after we get the optical flow and motion boundaries.

5. REFERENCES

[1] I. Abdelqader and S. Rajala. Motion estimation from noisy image data. In *ICASSP'93*, pages 209–212, 1993.

[2] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2:283–310, 1989.

[3] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of optical flow techniques. In *CVPR'92*, pages 236–242, 1992.

[4] D. Fleet and A. Jepson. Computation of component image velocity from local phase information. *International Journal of Computer Vision*, 5:77–104, 1990.

[5] D. Geiger and F. Girosi. Parallel and deterministic algorithms from mrf's: Surface reconstruction. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13:401–412, 1991.

[6] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

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

[8] F. Heitz and P. Bouthemy. Multimodal estimation of discontinous optical flow using markov random fields. In *IEEE Trans. on PAMI*, pages 1217–1232, 1993.

[9] B.K.P. Horn and G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[10] M. Irani, B. Rousso, and S. Peleg. Detecting and tracking multiple moving objects using temporal integration. In *ECCV'92*, pages 282–287, 1992.

[11] J. Konrad and E. Dubois. Bayesian estimation of discontinuous motion in images using simulated annealing. In *Conf. Vision Interface*, London, Ontario, June 1989.

[12] T. Poggio, E. Gamble, and J. Little. Parallel integration of vision modules. *Science*, 242:436–440, 1988.

[13] J. Vlontzos and D. Geiger. A mrf approach to optical flow estimation. In *CVPR'92*, pages 853–856, 1992.

[14] J. Zhang and J. Hanauer. The mean field theory for image motion estimation. In *ICASSP'93*, pages 197–200, 1993.