Robust Motion-Based Image Segmentation Using Fusion

Michael E. Farmer[§], Xiaoguang Lu*, Hong Chen*, and Anil K. Jain*
[§] Eaton Corporation
* Department of Computer Science and Engineering, Michigan State University
Email: {farmerm3, lvxiaogu, jain}@ cse.msu.edu, chenhon2@egr.msu.edu

ABSTRACT

To support real-time tracking of objects in video sequences, there has been considerable effort directed at developing optical flow and general motion-based image segmentation algorithms. The goal is to segment multiple moving objects in the image based on their relative motion. This task can be complicated by the presence of lighting variations. Furthermore, a combination of multiple motions and complex lighting effects can lead to dramatic image variations that may not be adequately accounted for by a single motion-based segmentation algorithm. We propose to fuse the results of multiple motion segmentation algorithms to improve the system robustness. Our approach uses the Expectation Maximization (EM) algorithm as a fusion engine. It also uses Principal Components Analysis (PCA) to perform dimensionality reduction to improve the performance of the EM algorithm and reduce the processing burden. The performance of the proposed fusion algorithm has been demonstrated in the "smart airbag" application of monitoring occupants in a moving automobile to determine if they are too close to the instrument panel (airbag). Through fusion of the outputs of multiple algorithms we are able to reduce the percentage of pixels missed on the target by 35 %.

1. INTRODUCTION

Object motion in a sequence of images is a powerful cue for performing segmentation of the object from the background [3,5,6,7]. Considerable work has been performed in the areas of both motion-based segmentation and motion estimation [1,2,3,4,5,6,7,10]. Motion segmentation is concerned with isolating and extracting all of the objects within an image while not necessarily requiring an accurate estimate of the motion. On the other hand, motion estimation requires the motion parameters be accurately calculated. This paper addresses the problem of motion segmentation.

Image segmentation based on motion can be performed in a number of ways, including [3], (i) dense flow field estimation, (ii) feature point correspondence tracking, and (iii) high level contour or region template -based tracking. We focus on computing the dense flow field of the input images sequences. The specific approaches to computing a dense flow field that we will investigate include: (i) gradient methods, (ii) phase methods, and (iii) correlation methods [2,3,6,7,10,11,12]. Each of these methods has some advantages for a particular type of motion environment but none of them has been shown to be consistently more robust than the others for all tracking environments. In the presence of temporally and spatially varying illumination, it becomes even less clear if there is a single best approach.

Temporal lighting effects can include global image illumination changes as well as narrow bands of shadow or highlights moving across the image [4,5,6]. The illumination effects in an image have been shown to be well-modeled by two contributions [5]: (i) a multiplicative contribution corresponding to the motion of diffuse shadows (this is equivalent to modeling the reflectivity of the materials imaged), and (ii) an additive contribution corresponding to the motion of illumination (this is equivalent to additional illumination falling on the object).

The lack of a universally best approach to motion segmentation implies that benefits can be derived if multiple approaches are combined or fused together. This is similar to the approach of classifier combination in pattern recognition, where combining the outputs from multiple classifiers can improve performance [9]. When applying fusion algorithms to a problem, there are three key problems that must be addressed, (i) at what level of abstraction the data should be combined, (ii) what mechanism will be used to combine the information, and (iii) how to manage the processing and memory burdens.

For motion segmentation, the possible data abstractions for fusion, include: (i) pixel level fusion with the velocity information, (ii) fusion of regions of common velocity, and (iii) object level fusion where the motion estimates are abstracted to objects and then combined.

Since the motion segmentation algorithms we are using provide an estimate of the dense flow field, it is natural for us to concentrate on combining the information at the pixel level of the directional component images. Combining at a higher level of abstraction would require us to interpret the dense flow fields into possible objects and then perform the fusion, which may result in the introduction of errors.

For motion segmentation we propose addressing the choice of a data combining mechanism by using the Expectation Maximization (EM) algorithm. EM has been successfully applied to the motion estimation problem for resolving multiple motions in a single image [7]. Jepson and Black in [7] applied EM at the feature level in a constraint space, so we will extend its usage and apply it at the pixel level.

To manage the processing burden in motion segmentation, we apply dimensionality reduction. In fusion for motion segmentation, as more and more component motion field images are included, the number of mixture components that must be estimated by the EM algorithm increases quadratically, since EM is estimating covariance matrices [7]. Additionally, some of the component motion field estimates may provide only a minimal amount of information and, therefore, unnecessarily increase the processing burden and possibly add noise. We want to find the set of salient features that capture the maximal amount of information available in the incoming video stream. We use Principal Components Analysis (PCA) to determine the combination of motion fields that provide the optimal representation [8].

We have applied the above techniques to the difficult problem of tracking a human occupant inside a moving automobile [13]. This problem has a unique combination of the following attributes: (i) a non-rigid body to track, (ii) complex background motions, and (iii) complex illumination variations due to moving shadows and highlights. Some examples of these conditions are shown in Figure 1. We will show the robustness of the proposed system for segmenting the occupant in this complex environment.

(a) (b)

Figure 1. Difficulty of occupant segmentation, (a) limb motion plus illumination changes, (b) body, limb and outside vehicle motions, and illumination changes.

2. MOTION SEGMENTATION AND CLUSTERING

Motion clustering provides regions of common motion that may correspond to objects of interest. In most motion segmentation problems, there are often multiple moving objects within the image. This is clearly the case in the automotive occupant tracking application where there is often motion outside the window, as well as passenger motion inside the vehicle. Additionally, the occupant motion can be composed of multiple motions corresponding to motion of limbs, turning of the head, and motion of the torso. Multiple motions cause discontinuities in the motion field as multiple objects move past each other in 3-D space and generate distinct motions in the image [7].

There are three methods of motion segmentation that we propose to use for the fusion inputs: (i) phase, (ii) gradient, and (iii) correlation methods. For the phase-based methods, the local phase gradient is measured from the output of the individual directional Gabor filters to estimate the component velocity v_n [1,2]:

$$v_n = -\phi_t(x,t)\frac{\nabla\phi(x,t)}{\|\nabla\phi(x,t)\|^2}, \qquad (1)$$

where $\phi_t(x,t)$ and $\nabla\phi(x,t)$ are the temporal and spatial derivatives of phase. The U and V component velocities are then generated by combining these finer angular estimates.

For correlation processing, we compute the U and V components and the illumination changes by modeling the correlation between two sequential images denoted f(x,y,t) and f(x,y,t-1) as [3,4,11]:

$$m_3 f(x,y,t) + m_4 = f(x+m_1, y+m_2, t-1), \qquad (2)$$

where m_1 and m_2 denote the translations in x and y, m_3 represents the variation in contrast, and m_4 represents the variation in brightness.

For the gradient-based methods, we use the following model for optical flow [3,5]:

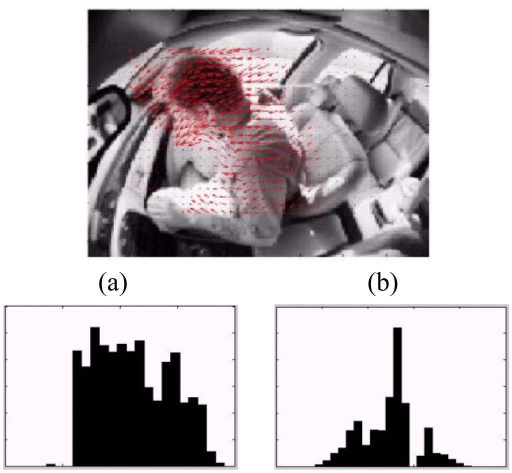
$$\frac{\partial f}{\partial t} = -\vec{v}\cdot grad(f) + f\cdot\frac{\partial m}{\partial t} + \frac{\partial c}{\partial t}, \qquad (3)$$

where f is the brightness distribution of an image, v is the velocity vector at each point in the image, the $\partial m/\partial t$ is the multiplicative illumination term, and $\partial c/\partial t$ is the additive illumination term [5,6].

It has been shown that the probability distributions of each of the motions in the image can be modeled by a Gaussian distribution [7]. Thus, the combination of multiple motions can be defined as a mixture of Gaussians. To compute the parameters for these individual Gaussians, we employ the EM algorithm [7]. We input the images corresponding to the U and V component of the motion field for each of the above motion segmentation algorithms.

3. DIMENSIONALITY REDUCTION

Note that the combined set of motion segmentation methods generates a 6-dimensional space corresponding to the U and V component images for the phase, correlation and gradient algorithms. This number can grow if other image planes such as color or grayscale are also input. Executing the EM algorithm in a high dimensional space greatly increases the number of parameters which causes higher processing throughput and increases the likelihood of noise corruption since some of the image planes may have no useful information. Figure 2 shows the contrasting amounts of information available in the motion fields, where one can see there is considerably more information (dispersion) in the U component. Recall that the dispersion in the histogram is a measure of the amount of information available (entropy).



(a) (b)

(c) (d)

Figure 2. Example vector flow field, (a) vector image, (b) histogram of U component, and (c) histogram of V-component.

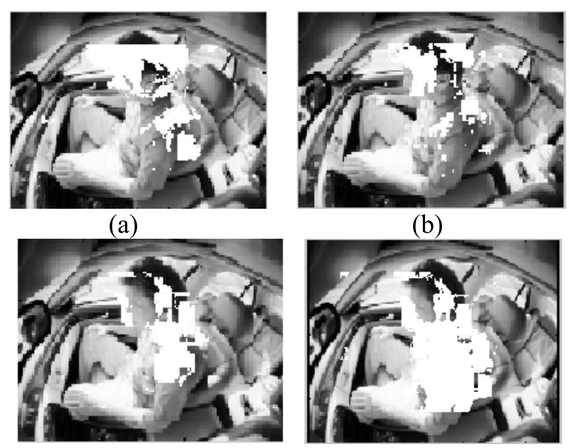
We can obtain a representational subspace that captures as much variability in the motion field as possible in the least number of image planes by using Principal Component Analysis (PCA) [8]. We will use PCA to convert our *MxNx*6 dimensional space into an *MxNxk* dimensional subspace, where *M* and *N* are the number of rows and columns in the input images respectively, and k<6.

4. EXPERIMENTAL RESULTS

The motion sequence used here is a subset of images taken from an outdoor driving sequence of over 10 minutes of streaming video at 30 frames per second. A small subset of this motion sequence is shown in Figure 3. To baseline the performance of the fusion algorithm, we initially performed the EM clustering on the U and V component outputs of the individual motion algorithms. We then applied the EM algorithm on the combined set of U and V component images from these motion algorithms. Finally, we apply the PCA on this combined set of U and V component images, and then we perform EM on the PCA representation for the final motion segmentation.

Figure 3. Image sub-sequence for motion processing.

The EM results for clustering the motions from the U and V component fields for each of the three motion segmentation algorithms are provided in Figures 4 (a), 4 (b), and 4 (c). Figure 4 (d) shows the EM results of the combined 6 component motion fields. We then perform the EM clustering on the image set corresponding to the k-highest principal components, k < 6. The ratio of the eigen-values will show us the relative information content in each dimension. We found that by retaining only the first two eigen-vectors we retain 82% of the information.



(a) (b)

(c) (d)

Figure 4. EM clustering of U and V fields derived from (a) phase method, (b) gradient method, (c) correlation method, and (d) EM clustering for combined motion fields.

Figure 5 (a) shows the segmentation results from the combined motion fields after PCA when we retain the top principal component and Figure 5 (b) shows the results for the top two principal components. Note the quality of the segmentation in Figure 5 is better than that for all of the individual methods and is also equivalent to the segmentation over the complete set of six motion fields. This is significant since the amount of processing for segmentation clustering is dramatically reduced.

Table 1 shows the quantitative results for the segmentation over the short image sequence. Note that maintaining the top two principal components results in the lowest percentage of missed pixels, and it is actually a factor of two better than the best individual segmentation algorithm.

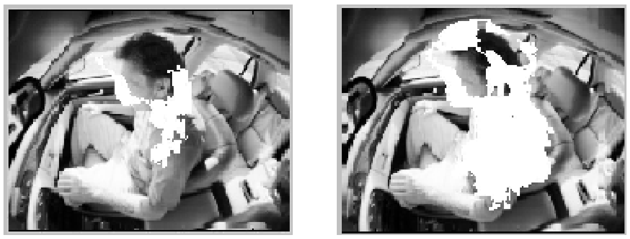


Figure 5. Segmentation results after PCA, (a) segmentation for 1st PC, (b) segmentation for 1st and 2nd principal components.

Table 1. Relative errors in segmentation.

Method	% Missing on Occupant	% Background Included
phase	62.8	2.58
gradient	75.7	0.41
correl	68.2	0.30
em_all	**36.4**	3.22
pca_1	77.4	0.31
pca_2	**28.4**	4.59

5. SUMMARY AND CONCLUSIONS

We have shown that the segmentation results of the three most popular optical flow algorithms need improvement when applied to the motion segmentation of complex semi-rigid objects undergoing compound motions in complex lighting conditions. We have shown an alternative way to use the EM algorithm for motion segmentation, where we treat the U and V component images as image planes and apply the EM to them.

We have also shown that the traditional individual motion estimation methods can be combined in the form of motion segmentation fusion using the EM Maximization algorithm. The result of the fusion is greatly improved by applying the Principal Component Analysis as a means of performing dimensionality reduction on the data. This not only allows the EM fusion algorithm to run more quickly but it also provided superior segmentation results. In our application we reduced the incoming data from six image components to two images for a segmentation that was more complete compared to any of the individual algorithms.

Future work will be directed at integrating the illumination fields into the fusion to see if the segmentation can be even further improved. We will also integrate the raw incoming grayscale values into the fusion processing and test against multiple moving objects in addition to temporally varying illumination. Finally, we will investigate methods to reuse the PCA projections over multiple image frames and to automatically determine when a new set of PCA transformations must be computed to make the algorithm feasible for real-time implementation.

References

[1] D. Fleet and A. Jepson, "Computation of component image velocity from local phase information," *International Journal of Computer Vision*, vol. 5, no. 1, pp. 77-104, 1990.

[2] T. Gautama and M. Hulle, "A phase-based approach to the estimation of the optical flow field using spatial filtering," *IEEE Trans. Neural Networks*, vol. 13, no. 5, pp.1127-1136, 2002.

[3] S. Beauchemin and J. Barron, "The computation of optical flow," *ACM Computing Surveys*, vol. 27, no. 3, pp. 433-467, 1996.

[4] S. Periaswamy and H. Farid, "Elastic registration in the presence of intensity variations, " *IEEE Trans. Medical Imaging*, in press, 2003

[5] S. Negahdaripour, "Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 9, 1988.

[6] L. Zhang, T. Sakurai, H. Miike, "Detection of Motion Fields Under Spatio-temporal non-uniform Illumination", *Image and Vision Computing*, vol. 17, 1999.

[7] A. Jepson and M. Black, "Mixture Models for Optical Flow Computation", Tech Rep RBCV-TR-93-44, University of Toronto, Dept of Computer Science, 1993.

[8] R.O. Duda, P.E. Hart and D.G. Stork, *Pattern Classification*, 2nd edition, Wiley, New York, 2000.

[9] R.P.W. Duin, "A note on comparing classifiers," *Pattern Recognition Letters*, vol. 1, pp. 529-536, 1996

[10] C.L. Fennema and W.B. Thompson, "Velocity determination in scenes containing several moving images," *Comput. Graph. Image Proc.*, vol. 9, pp. 301-315, 1979.

[11] M. .A. Sutton, W.J. Walters, W.H. Peters, W.F.Ranson, and S.R. McNeil, "Determination of displacement using an improved digital correlation method," *Image and Vision Computing.*, vol. 1, no. 3, pp. 133-139, 1983.

[12] D.S. Kalivas and A.A. Sawchuk,, "A region matching motion estimation algorithm," *Computer Vision, Graphics, and Image Processing*, vol. 52, no. 2, pp. 275-288, 1991.

[13] M. Farmer, R.L. Hsu, and A. Jain, "Interactive multiple model (IMM) Kalman filter for robust high speed human motion tracking", *Proc. IEEE International Conference on Pattern Recognition*, pp. 20-23, 2002.