

Robust motion field estimation and segmentation via appearance model-free, aperture-free coherence propagation

Steven J. Rennie

PSI-TR-2005-012, February 2005 *

Abstract

In this paper, we present a aperture-free, appearance model free, algorithm for the estimation and segmentation of the motion field of arbitrary input sequences, that is applicable even when the appearance of the underlying components of the scene are changing rapidly, are partially or totally occluded at times, or whose deformation patterns are highly irregular. The algorithm is built upon a low-level, generative probability model of motion that captures our prior intuitions about motion fields are structured, while avoiding the representation of notions that break down, or are unfeasible or intractable to model under general conditions. Estimation of the motion field and segmentation under the defined model is achieved by applying Loopy Belief Propagation (LBP) to make the inference step (E-Step) of the Expectation Maximization (EM) algorithm used to learn the model parameters tractable. Results for the case of the application of a constant motion model formulation of the algorithm to the analysis of a natural image sequence containing two independently moving persons (the image sequence Jojic and Frey used to demonstrate their early work on learning flexible sprites in video layers [5]) are presented, and demonstrate very good motion field estimation and segmentation performance.

1 Introduction

The recovery and segmentation of the motion field of a video sequence is a fundamental problem in computer vision that has received much attention. Although much progress has been made in motion analysis and scene analysis in general in recent years [9, 3, 2, 13, 12, 11, 1, 5, 8], a robust solution to the motion estimation and segmentation problem under the most general conditions still alludes us today.

It is not immediately clear what is so challenging about the motion estimation and segmentation problem. A multitude of techniques for the estimation of optical flow exist [6, 7], and can provide, in regions of

*based upon a course project completed in April 2004

adequate texture, good local estimates of the motion field. There are several fundamental issues with these local, low-level approaches to motion estimation however. Most of these approaches require that a decision about the aperture to use when computing local flow estimates be made, and show high sensitivity to this decision in the resulting flow estimates. The spatial resolution of the flow field estimated using these approaches is generally limited by the choice of aperture, making the motion field estimate inaccurate around motion boundaries. For gradient based approaches, spatial smoothing must be done before estimating flow, also damaging the spatial resolution of the resulting motion field estimate. In addition, in general there will be large regions of the scene under analysis with little or no texture, that no reasonable choice of aperture can overcome. In this case the optical flow is computed as zero regardless of the motion field.



Figure 1: Local flow computation followed by clustering is an ineffective, unreliable way to obtain motion field and segmentation estimates. At left, the first of an input source image pair. At middle, the result obtained by applying the local optical flow estimation method [6] with an aperture of ± 15 . At right, the result obtained by applying the local optical flow estimation method [6] with an aperture of ± 5 . In both cases the smoothing parameter $\sigma = 2.75$. Note that an aperture of ± 15 was the smallest value that yielded reasonable segmentation results in our experiments. Clearly in this (the optimal) case, the achieved segmentation is of very low spatial resolution, with large regions of segmentation error.

Figure 1 depicts one frame of an natural image sequence where two people are walking towards each other at roughly constant velocities of approximately 3 and -3 pixels/second, respectively, along with the motion segmentation results obtained via the application of Mixture of Gaussians (MOG) clustering on locally computed flow estimates obtained via Horn and Schunck's optical flow estimation technique on an image pair [6]. In the first case, the aperture during flow estimation is ± 15 pixels with (isotropic) gaussian smoothing parameter $\sigma = 2.75$. Here the motion is (at least roughly) successfully segmented, but the resulting segmentation is of very poor spatial resolution, with large regions of segmentation error. In the second case, the aperture has been turned down to ± 5 pixels. Here we can see that the spatial resolution of the segmentation is much improved (the background segmentation is reasonably accurate) but the resulting optical flow estimates are actually very poor, resulting in negligible segmentation of the underlying motions. It should be noted that here an aperture of ± 15 was the smallest value that yielded reasonable segmentation results in our experiments. Clearly simple local optical flow estimation followed by clustering will not in general be an accurate, robust

way to do motion estimation and segmentation. More generally, we expect that *any* approach for motion segmentation based local flow estimates will not be able to recover a high resolution motion segmentation of the scene, since the observation data will generally either be of low spatial resolution, or highly inaccurate.

Several more advanced approaches have been taken to the motion analysis and segmentation problem [9, 3, 2, 12, 11, 1, 5, 8]. One major theme has to enforce spatial coherence during both motion and segmentation estimation to improve performance. In [12] for example, Weiss presents a motion segmentation algorithm where the notion of spatial coherence among neighboring pixels is enforced via a Markov Random Field (MRF) over the postulated generating motion clusters, and a mean-field algorithm to facilitate computationally tractable estimation of both the number of clusters (when the variance of the underlying clusters is assumed to be common and known), and their cluster parameters is derived. In [3], Cheung, Maclean et. al. present a coupled motion estimation and segmentation algorithm, developed for the purpose of initializing their contour tracking algorithm, that incorporates spatial coherence by doing 'post-clustering' image warping to recover a 'corrected posterior' of the motion segmentation, based on the agreement of the warped input image intensity and output image intensity under each identified cluster motion.

A multitude of 'higher-level' approaches to scene analysis that incorporate or achieve motion estimation and segmentation as a by-product have been developed [9, 1, 5], and have demonstrated excellent results under certain conditions. A common theme of many of these approaches is to learn a representation of the appearance of the underlying components of the scene to achieve enhanced motion estimation and segmentation. In [9, 5], for example, Frey and Jovic present algorithms for automatically learning flexible sprites in video layers, and demonstrate robust, high resolution motion estimation and segmentation performance under the conditions of occlusion and small appearance deformations. In cases where the underlying appearances are very similar, rapidly changing, very complex, or irregularly changing, however, it is not clear at this point that such approaches will be feasible or of fidelity. Under such conditions, can high fidelity estimation and segmentation of the motion field be achieved?

In this paper, we present a aperture-free, appearance model free, algorithm for the estimation and segmentation of the motion field of arbitrary input sequences, that is applicable even when the appearance of the underlying components of the scene are changing rapidly, are partially or totally occluded at times, or whose deformation patterns are highly irregular. The algorithm is built upon a low-level, generative probability model of motion that captures our prior intuitions about how motion fields are structured, while avoiding the representation of notions that break down, or are unfeasible or intractable to model under general conditions. Estimation of the motion field and segmentation under the defined model is achieved by applying Loopy Belief Propagation (LBP) to make the inference step (E-Step) of the Expectation Maximization (EM) algorithm used to learn the model parameters tractable. Results for the case of the application of a constant motion-mixture model formulation of the algorithm to the analysis of a natural image sequence contain-

ing two independently moving persons (the image sequence Jojic and Frey used to demonstrate their early work on learning flexible sprites in video layers [5]) are presented, and demonstrate very good motion field estimation and segmentation performance.

2 A Low-Level Generative Probability Model of Motion

We assume that we have an image pair $\{I_1, I_2\}$ consisting of successive frames of an image sequence, taken at a frame rate sufficient such that brightness constancy is a reasonable assumption, and that the spatial and temporal derivatives of the image pair $\{I_x, I_t\}$ are (approximately) related by the brightness constancy constraint:

$$\mathbf{I}_x(\mathbf{x})^T \mathbf{u}(\mathbf{x}) + I_t(\mathbf{x}) = 0 \quad (1)$$

at each image location, where $\mathbf{u}(\mathbf{x})$ is the motion field at image location \mathbf{x} . In the spirit of the formulation presented in [3], but in a slightly different form, we suppose the existence of underlying motion cluster variables $c(\mathbf{x})$, which together with the measured spatial derivative $\mathbf{I}_x(\mathbf{x})$ define a conditional distribution on $I_t(\mathbf{x})$ that is Gaussian:

$$p(I_t(\mathbf{x})|c(\mathbf{x}), \mathbf{I}_x(\mathbf{x})) = N(I_t(\mathbf{x}); \mathbf{I}_x(\mathbf{x})^T \mathbf{u}_c(\phi), \sigma_c(\phi)) \quad (2)$$

Where the parameters of \mathbf{u}_c and σ_c , ϕ , will depend on the assumed motion model. Note that in general \mathbf{u}_c and σ_c will be a function of \mathbf{x} .

Additionally, conditioned on $c(\mathbf{x})$, we model the relationship between the warped version of $I_1(\mathbf{x})$ and $I_2(\mathbf{x})$ as zero mean and Gaussian:

$$p(I_1(\mathbf{x} + \mathbf{u}_c) - I_2(\mathbf{x})|c(\mathbf{x})) = p(\Delta I(\mathbf{x})|c(\mathbf{x})) = N(\Delta I(\mathbf{x}); 0, \psi) \quad (3)$$

to enforce coherence of the obtained motion estimates with the known image pair in the model, as was done as a post-processing step in [3]. To enforce the notion that neighboring pixels will generally have related motions, we impose a MRF prior on $\{c(\mathbf{x})\}$:

$$p(\{c(\mathbf{x})\}) = \frac{1}{Z} \prod_{\mathbf{x}} \pi_{c(\mathbf{x})} \prod_{\|\mathbf{x}_2 - \mathbf{x}_1\|=1} \Psi(c(\mathbf{x}_1), c(\mathbf{x}_2), \boldsymbol{\lambda}); \quad (4)$$

whose compatibility functions $\Psi(c(\mathbf{x}_1), c(\mathbf{x}_2), \boldsymbol{\lambda})$ may be defined to represent our belief about how the estimation of the motion of neighboring should be related.

The overall generative probability model for the observed image pair, their spatial and temporal derivatives, and the postulated hidden motion cluster variables $\{c(\mathbf{x})\}$ is then given by:

$$\begin{aligned}
p(\{I_1\}, \{I_2\}, \{\mathbf{I}_x(\mathbf{x})\}, \{I_t(\mathbf{x})\}, \{c(\mathbf{x})\}) &= p(\{c(\mathbf{x})\}) \prod_{\mathbf{x}} p(I_t(\mathbf{x})|c(\mathbf{x}), \mathbf{I}_x(\mathbf{x})) p(\Delta I(\mathbf{x})|c(\mathbf{x})) \\
&= \frac{1}{Z} \prod_{\mathbf{x}} \pi_{c(\mathbf{x})} \prod_{\|\mathbf{x}_2 - \mathbf{x}_1\|=1} \Psi(c(\mathbf{x}_1), c(\mathbf{x}_2), \boldsymbol{\lambda}) \prod_{\mathbf{x}} N(I_t(\mathbf{x}); \mathbf{I}_x(\mathbf{x})^T \mathbf{u}_c(\boldsymbol{\phi}), \sigma_c(\boldsymbol{\phi})) N(\Delta I(\mathbf{x}); 0, \psi) \quad (5)
\end{aligned}$$

In the case of the assumption of a constant motion model for each motion cluster the model simplifies to:

$$\begin{aligned}
p(\{I_1\}, \{I_2\}, \{\mathbf{I}_x(\mathbf{x})\}, \{I_t(\mathbf{x})\}, \{c(\mathbf{x})\}) \\
= \frac{1}{Z} \prod_{\mathbf{x}} \pi_{c(\mathbf{x})} \prod_{\|\mathbf{x}_2 - \mathbf{x}_1\|=1} \Psi(c(\mathbf{x}_1), c(\mathbf{x}_2), \boldsymbol{\lambda}) \prod_{\mathbf{x}} N(I_t(\mathbf{x}); \mathbf{I}_x(\mathbf{x})^T \mathbf{u}_c, \sigma_c) N(\Delta I(\mathbf{x}); 0, \psi) \quad (6)
\end{aligned}$$

The application of this model to problems in motion estimation and segmentation will be the focus of the remainder of this paper.

3 A Loopy Belief Propagation Algorithm for Motion Estimation and Segmentation

Given the low-level generative probability model of motion defined in the previous section, the task of motion estimation and segmentation maps to the estimation of the motion cluster velocities \mathbf{u}_c and the posterior probability of each motion class at each pixel location given the observed image data, $\mathbf{D} = \{\Delta I(\mathbf{x}), I_t(\mathbf{x}), \mathbf{I}_x(\mathbf{x})\}$. We achieve this by applying an EM algorithm to simultaneously learn the parameters of the model and estimate the motion cluster posterior in iterative fashion [10].

Because the motion cluster prior $p(\{c(\mathbf{x})\})$ is locally coupled, the exact computation (inference) of $p(c(\mathbf{x})|\mathbf{D})$, which is required in the Expectation-Step (E-Step) of EM algorithm will scale as $O(N^4)$, where N is the number of pixels in the source images, and so will generally not be computationally feasible. To achieve computationally feasible approximate inference during the E-Step we apply Loopy Belief Propagation [13], passing local messages between the random variables of the model to communicate the influence coupled random variables have on one another.

Perhaps the most straightforward to describe form of Belief Propagation is the application of the Sum-Product algorithm [4] to the factor graph representing the global function on which marginals are to be computed. The factor graph representing our generative probability model of motion is depicted in figure 2. Here the square nodes represent 'function nodes' that are connected to dependent variables, and the circular nodes are 'variable nodes' representing the variables of the model.

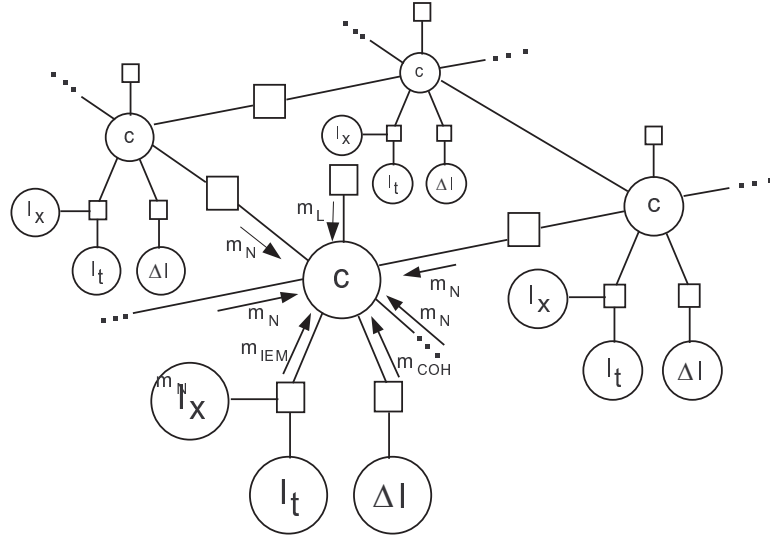


Figure 2: The Factor Graph representation of our generative probability model of motion.

At each iteration of the sum-product algorithm, outgoing messages are computed as follows:

- At variable nodes, the outgoing message on each connected edge is computed as the product of all incoming messages on the other edges.
- At function nodes, the outgoing message on each connected edge is computed by taking the product of all incoming messages on the other edges, multiplying by the local function, and marginalizing over all variables less the destination variable (and then normalizing)

Note that all messages sent to and from function and variable nodes are a function of only the variable connected to that edge. The product of *all* incoming messages to a given function node represents the current estimate of the marginal of that variable. In the case that the relationship between all random variables forms a tree, allowing messages to propagate in both directions along all edges in parallel for a number of iterations proportional to depth of the tree results in the computation of the *exact* marginals for variables in the network. In the case that the variable network has loops, which is the case here, the computation is approximate (the network has loops but each variable treats the incoming beliefs as independent information). LBP has been applied to inference problems on graphs with loops, including MRFs, with much success in many areas [12, 13].

To compute an approximate E-Step at each iteration of EM under our probability model of motion then, we perform an iteration of the sum product algorithm on the graph describing the model, passing messages

between all function and variable nodes, and then simply take the product of all incoming messages as our current estimate of the marginal posterior of $c(\mathbf{x})$. At a given cluster node then, the marginal estimate at each E-Step iteration is computed (except on the image edges) as product of 7 incoming messages; 4 from it's local neighbors in the MRF, which we will denote as m_N , and three local beliefs, m_L , m_{IEM} and m_{COH} , which express the influence of local prior π_c , $\mathbf{I}_x(\mathbf{x})$ and $I_t(\mathbf{x})$, and $\Delta I(\mathbf{x})$ respectively on $c(\mathbf{x})$ given the current model parameters. m_L , m_{IEM} , and m_{COH} are given, respectively, as follows:

$$m_L(c(\mathbf{x})) = \pi_c \quad (7)$$

$$m_{IEM}(c(\mathbf{x})) = p(I_t(\mathbf{x})|c(\mathbf{x}), \mathbf{I}_x(\mathbf{x})) = N(I_t(\mathbf{x}); \mathbf{I}_x(\mathbf{x})^T \mathbf{u}_c(\phi), \sigma_c(\phi)) \quad (8)$$

$$m_{COH}(c(\mathbf{x})) = p(\Delta I(\mathbf{x}) = N(\Delta I(\mathbf{x}); 0, \psi) \quad (9)$$

Each message $m_N(c(\mathbf{x}))$ is defined by the marginalized product of the message sent by each neighbor and its associated local compatibility function, and cannot be expressed in terms of the model parameters in closed form.

The Maximization Step (M-Step) is computed by taking the derivative of the expected complete likelihood or the hidden and observed data given the current estimates of $p(\{c(\mathbf{x})\}|\mathbf{D})$ with respect to the model parameters. Here another approximation is required to make the computation tractable. Since the coherence-based likelihood $p(\Delta I(\mathbf{x}))$ is a function of \mathbf{u}_c , an exact M-step would require this term be included when maximizing the likelihood of the data w.r.t. \mathbf{u}_c . This term however, is not tractable to store or compute at each iteration (it could be represented, for example, by a mixture of gaussians, one for each pixel, or as a discrete belief) even when the underlying motions are reasonably bounded. We proceed by ignoring the dependence of this term in the likelihood function when minimizing \mathbf{u}_c , which is certainly suboptimal. Note however, that the coherence term still has an influence on the estimation of \mathbf{u}_c at each iteration via the m_{COH} message passed during the E-Step, which discriminates against incoherent motions.

The Approximate M-Step updates are given by:

$$\mathbf{u}_c = (B^T A)^{-1} B^T b \quad (10)$$

$$B = [p(c(\mathbf{x}_1)|\mathbf{D})\mathbf{I}_x(\mathbf{x}_1), p(c(\mathbf{x}_2)|\mathbf{D})\mathbf{I}_x(\mathbf{x}_2), \dots, p(c(\mathbf{x}_N)|\mathbf{D})\mathbf{I}_x(\mathbf{x}_N)]^T$$

$$A = [\mathbf{I}_x(\mathbf{x}_1), \mathbf{I}_x(\mathbf{x}_2), \dots, \mathbf{I}_x(\mathbf{x}_N)]^T$$

$$b = -[I_t(\mathbf{x}_1), I_t(\mathbf{x}_2), \dots, I_t(\mathbf{x}_N)]^T$$

$$\sigma_c = \frac{\sum_{\mathbf{x}} p(c(\mathbf{x})|\mathbf{D})(\mathbf{I}_x(\mathbf{x})^T \mathbf{u}_c - I_t(\mathbf{x}))^2}{\sum_{\mathbf{x}} p(c(\mathbf{x})|\mathbf{D})} \quad (11)$$

Here we set ψ arbitrarily at unity, but ψ can be set according prior information (for example the average variance of the video camera output intensity for constant input intensity) if available. Looking at the M-Step updates we can see that the update for \mathbf{u}_c is effectively a weighted-least-squares estimate of the motion of cluster c , where the weights are given by the current posterior estimate for $c(\mathbf{x})$ over the image. The update for rule for σ_c can be similarly understood.

By iterating between the approximate E and M Steps described above, we can learn the parameters of the our probability model of motion, recovering estimates of the underlying motions \mathbf{u}_c , and simultaneously recover a motion segmentation of the image via the inferred posterior $\{p(c(\mathbf{x})|D)\}$.

4 Experimental Results

The motion segmentation algorithm described above was implemented in Matlab and tested on the Frey and Jojic’s video of two people walking towards and then past each other in parallel but non-intersecting planes, as was introduced in an earlier discussion in the introduction. Because the E-Step is implemented approximately via LBP, we can trivially enable and disable the propagation of classes of beliefs to investigate their influence on the motion estimation and segmentation results obtained.

Figure 3 depicts the motion segmentation results obtained for several variants of our motion estimation and segmentation algorithm for the image frame depicted in figure 1. In the top left image is the motion segmentation obtained when message passing across the MRF, and the coherence enforcing messages m_{COH} , are disabled. In this case the algorithm reduces to MOG clustering of $\{I_t(\mathbf{x})\}$ given $\{\mathbf{I}_x(\mathbf{x})\}$, as was done in [3] (in [3] the results were then tested for coherence and the posterior estimates of the segmentation ‘replaced’ by the coherence likelihood). In this case very poor motion estimation and segmentation is achieved. While in [3] excellent segmentation results were demonstrated for the case of one moving object on a static background, we can see that the algorithm breaks down when these conditions are relaxed.

In the top right image, we have turned on the coherence messages m_{COH} during the E-Step. Here a dramatic improvement in the segmentation (and motion estimates) is obtained over the previous result. The coherence belief steers the estimation towards estimates of motion that are coherent with the observed image. In contrast, the former version of the algorithm is grossly unconstrained and the optimization space has many local minima.

In the lower left hand corner the results obtained when both the coherence and MRF messages are turned on, and the compatibility between neighboring pixels is set to be uniform (all neighboring pixels have equal

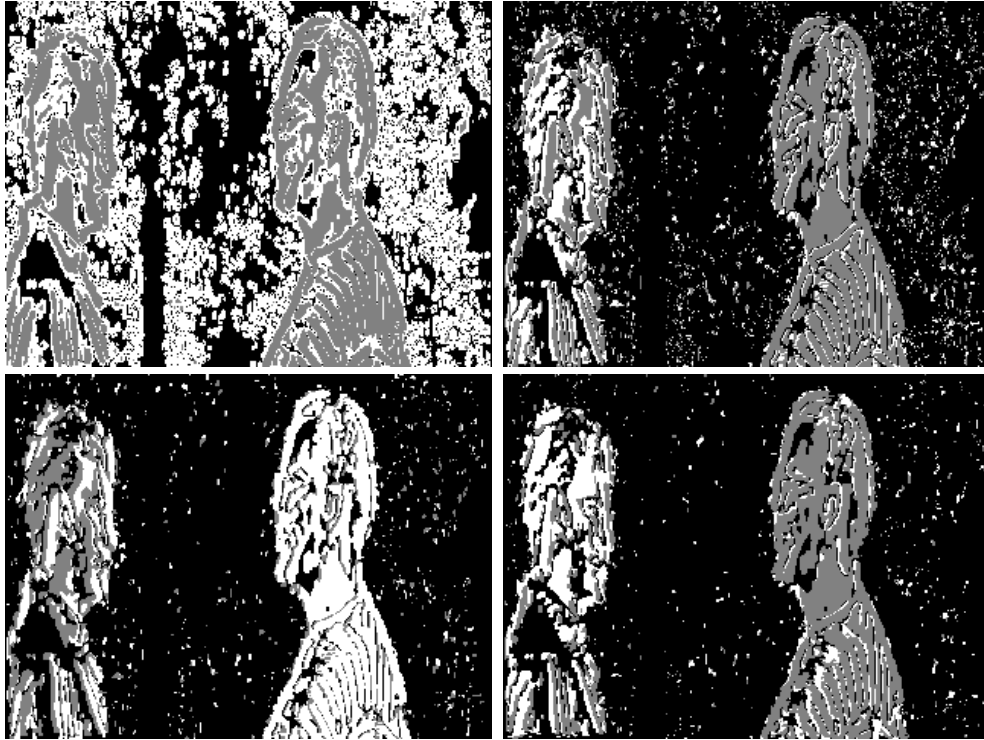


Figure 3: Motion segmentation results achieved by our algorithm when: (top left) message passing over the MRF and from the coherence constraints is disabled, (top right) message passing over the MRF is disabled and message passing from the coherence constraints is enabled, (bottom left) all message passing is enabled, with all MRF compatibilities defined uniformly (bottom right) all message passing is enabled, with the MRF compatibilities specified according to (12) with $K = 40$ to define our fully-enabled motion estimation and segmentation algorithm.

influence on each other). Looking at the segmentation result we can see that the incorporation of the notion that neighboring pixels will generally have similar motions, in addition to forcing coherence via the m_{COH} messages, improves the obtained results minimally, if at all.

In the lower right hand corner of figure 3 the results obtained when both the coherence and MRF messages are turned on, and the compatibility between neighboring pixels is set to be:

$$\psi(\mathbf{x}_1, \mathbf{x}_2) = \exp\left(\frac{|I_a(\mathbf{x}_2) - I_a(\mathbf{x}_1)|}{K}\right) - 1 \quad (12)$$

with 'coupling factor' K set to 40 are depicted. Here the coupling is weighted based on how similar the intensities of neighboring pixels are. When the coupling is low, the belief communicated from one cluster node to another becomes neutral (uniform). When the coupling is high, the belief of one cluster node about it's neighboring cluster node is communicated unaltered.

Looking at the result, we can see a large improvement over the segmentation result achieved by the more

primitive versions of the algorithm, and a high resolution, informative segmentation of the underlying motion of the scene is obtained. Note that the coupling factor was not tuned specifically for this image, but set based an adhoc estimate of how we cluster intensity information when doing motion segmentation. Note also that simple image processing such as low-pass and morphological operations could be trivially applied to improve the aesthetics of the result, if necessary.



Figure 4: (left) source image. (middle) motion segmentation result achieved by our fully-enabled motion estimation and segmentation algorithm. (right) result achieved by local flow computation followed by MOG clustering.

Figure 4 (middle) depicts the segmentation result based on m_{IEM} obtained by the full version of our image segmentation algorithm, for a point in the test image sequence where one person is partially occluding the other and their associated motion boundaries meet. Here again we can see that a high resolution segmentation of the motion field, even along some sections of where the motion boundaries of the people meet, is obtained. In the lower portion of the image, we can see segmentation leakage on the striped shirt of the second person. In this case, because the shirt is striped and the displacement of both subjects over the image pair is similar to the distance between the stripes, and the motions are spatially local, the segmentation is ambiguous to the algorithm. Similarly, around the motion boundary in the facial area of the subjects, there is limited texture and the pixel intensity on both sides of the motion boundary is very similar, leading to segmentation leakage. Nevertheless, the obtained result is impressive. Figure 4 (right) depicts the result obtained via local computation of optical flow followed by MOG clustering. Here the obtained segmentation is unintelligible.

5 Conclusion

In this paper, an aperture-free, appearance model free, algorithm for the estimation and segmentation of the motion field of arbitrary input sequences, that is applicable even when the appearance of the underlying components are not tractably modelable, was presented. The algorithm is built upon a low-level, generative probability model of motion that captures our prior intuitions about motion fields are structured, and avoids the representation of notions that break down, or are unfeasible or intractable to model under general conditions. Estimation of the motion field and segmentation under the defined model is achieved by applying

Loopy Belief Propagation (LBP) to make the inference step (E-Step) of the Expectation Maximization (EM) algorithm used to learn the model parameters tractable. Results for the case of the application of a constant motion-mixture model formulation of the algorithm to the analysis of a natural image sequence containing two independently moving persons (the image sequence Jojic and Frey used to demonstrate their early work on learning flexible sprites in video layers [5]) are presented and demonstrate that high-resolution motion segmentation and estimation is possible even without appearance models. Directions of future research include the definition of a LBP algorithm under the assumption of more complex underlying motion models, and the development and analysis of an equivalent patch-feature as opposed to pixel-feature based models, with the aim of reducing segmentation leakage and obtaining sharper estimates of the posterior probability of the underlying motion clusters, at the expense of some spatial resolution.

References

- [1] Michael J. Black and Allan D. Jepson. A probabilistic framework for matching temporal trajectories: Condensation-based recognition of gestures and expressions. In *Proc. European Conference on Computer Vision*, 1998.
- [2] Micheal J. Black and David J. Fleet. Probabilistic detection and tracking of motion boundaries. In *ICCV*, 1999.
- [3] Desmond Cheung and James Maclean. Integrated region and boundary information for improved spatial coherence in object tracking. In *CVPR*, 2004.
- [4] B. J. Frey F. R. Kschischang and H.-A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 47:2:498–519, 2001.
- [5] B. Frey and N. Jojic. Learning graphical models of images, videos and their spatial transformations. In *In Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, 2000.
- [6] Berthod K. P. Horn (HP) and Brian G. Schunck (HP). Determining optical flow. In *Technical Report, MIT Artificial Intelligence Laboratory, Number AIM-572*, 1980.
- [7] W. F. Clocksin (HP). A new method for estimating optical flow. In *Technical Report, University of Cambridge, Computer Laboratory, Number UCAM-CL-TR-436*, 1997.
- [8] A. D. Jepson and M. J. Black. Mixture models for image representation, 1996.
- [9] Nebojsa Jojic and Brendan J. Frey. Learning flexible sprites in video layers. In *ICCV*, December 2001.
- [10] M. Jordon and C. Bishop. *An Introduction to Graphical Models*. (to appear).

- [11] P. Bouthemy M. Gelgon. A region-level motion-based graph representation and labeling for tracking a spatial image partition. *Pattern Recognition*, 33(4):725–740, 2000.
- [12] Weiss Y. and Adelson E.H. A unified mixture framework for motion segmentation: incorporating spatial coherence and estimating the number of models. In *CVPR*, 1996.
- [13] Freeman W.T. Yedidia J.S. and Weiss Y. Understanding belief propagation and its generalizations. *IJCAI 2001 Distinguished Lecture track*, 2001.