# Motion Segmentation Using Inference in Dynamic Bayesian Networks

Marc Toussaint

TU Berlin
Franklinstr. 28/29
10587 Berlin, Germany
`mtoussai@cs.tu-berlin.de`

Volker Willert, Julian Eggert, Edgar Körner

Honda Research Institute Europe GmbH
Carl-Legien-Str. 30
D-63073 Offenbach/Main, Germany
`volker.willert@honda-ri.de`

### Abstract

Existing formulations for optical flow estimation and image segmentation have used Bayesian Networks and Markov Random Field (MRF) priors to impose smoothness of segmentation. These approaches typically focus on estimation in a single time slice based on two consecutive images. We develop a motion segmentation framework for a continuous stream of images using inference in a corresponding Dynamic Bayesian Network (DBN) formulation. It realises a spatio-temporal integration of optical flow and segmentation information using a transition prior that incorporates spatial and temporal coherence constraints on the flow field and segmentation evolution. The main contribution is the embedding of these particular assumptions into a DBN formulation and the derivation of a computationally efficient two-filter inference method based on factored belief propagation (BP) that allows for on- and offline parameter optimisation. The spatio-temporal coupling implemented in the transition priors ensures smooth flow field and segmentation estimates without using MRFs. The algorithm is tested on synthetic and real image sequences.

## 1   Introduction

Optical flow estimation is a fundamental problem in image processing. The analysis of movement in the image allows one to infer the motion of objects in the environment as well as the self-motion relative to the environment. As it is generally the case in information processing problems, the quality of the estimation can be greatly enhanced when information from different sources is integrated. In image sequences, an important source of information is prior knowledge about the structure of optical scenes. One may assume that images are composed of segments which refer to different physical objects. Each object induces a coherent flow field in its segment reflecting its 3D motion. An elegant way to integrate structural

segmentation and flow field estimation is to formulate a generative probabilistic model in terms of Bayesian Networks [13, 3, 11, 10]. These existing approaches focus on a single time slice flow field coupled to two consecutive images and usually implement smoothness of segmentation using Markov Random Field (MRF) priors [4, 3, 5, 9]. This can be applied on continuous image sequences by applying the technique on each time slice. However, this would neglect an additional source of information: the temporal (Markovian) coupling of the flow field and segmentation evolution. Further, when the application requires online motion and segmentation filtering, the computational cost of MRF inference in each time step may be too high [11]. Other approaches take the temporal coupling into account but do not consider the segmentation problem [1, 2, 12].

Our focus is on the temporal coupling of the flow field and segmentation in image sequences, and we avoid using intra-time slice MRF priors but instead achieve smoothness with a proper spatio-temporal coupling. A crucial aspect is formulating appropriate transition probabilities for the flow field and segmentation evolution. We formulate transition probabilities that allow us to propagate information across time slices but also imply a spatial smoothness of the flow field estimation and segmentation. Hence, instead of iterating BP (during MRF inference) within one time slice to impose spatial smoothness, we "unroll" this constraint and encode it in the spatio-temporal coupling such that BP along a temporal sequence yields spatial smoothness after a few time steps. This approach is particularly well-suited for online applications since the computational cost at each time step is limited to a small and fixed number of message passings.

After we have formulated the DBN model and the corresponding transition probabilities in section 2 we describe in section 3 our inference and parameter learning algorithm. During inference we make several factorisation assumptions and use a version of the Factored Frontier Algorithm [6]. For parameter training using a batch or online EM-algorithm we use either a two-filter or a forward filter approach to compute the flow field and segmentation posteriors. Based on the computed posteriors we update parameters in an M-step. Section 4 discusses experimental results and section 5 concludes the paper.

## 2  Dynamic Bayesian Network Model

We start by specifying a complete data likelihood of a sequence $I^{0:T}$ of $T + 1$ images. We do this by assuming the generative model for such an image sequence as given by the DBN in Fig. 1A. Here, $I^t$ is the grey value image at time slice $t$ with entries $I_x^t$ that are the grey values at all pixel locations $x \in X$ of the image. Similarly, $V^t$ is a flow field at time slice $t$ defined over the image range with entries $v_x^t \in W$ at each pixel location $x$ of the image. Throughout this paper we consider discrete flow fields ($W = \mathbb{Z}^2$). Further, we assume that each image is composed by a finite number of segments. Each segment has a shape and the discrete labelling $s_x^t \in \{1, .., K\}$ of each pixel specifies which image pixels stem from which of $K$ possible segments. Since every segment has a typical vector field describing the optical flow of its appearance, the segment labelling variable $S^t$ is coupled to the flow field variable $V^t$.

To define the model precisely we need to specify (i) the observation likelihood $P(I^{t+1} | V^t, I^t)$ of a pair of images $I^{t+1}$ and $I^t$, (ii) the transition probability
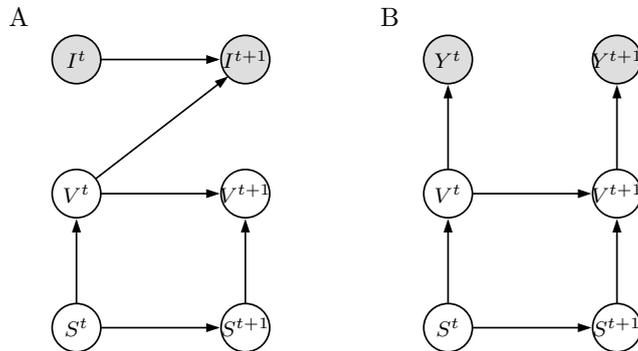
Figure 1: Dynamic Bayesian Network for motion estimation.

$P(V^{t+1} \mid S^{t+1}, V^t)$ of the flow field depending on the segmentation, and (iii) the transition probability $P(S^{t+1} \mid S^t)$ of the segmentation.

To simplify the notation we can introduce an alternative observation variable $Y^t = (I^{t+1}, I^t)$ that subsumes a pair of consecutive images. Since images are observed, the likelihood $P(I^t)$ in the term $P(I^{t+1} \mid V^t, I^t) \, P(I^t) = P(I^{t+1}, I^t \mid V^t)$ is only a constant factor we can neglect. This leads to the DBN shown in Fig. 1B with observation likelihoods $P(I^{t+1} \mid V^t, I^t) \propto P(I^{t+1}, I^t \mid V^t) = P(Y^t \mid V^t)$. For all transition probabilities we assume that they factorise over the image as follows,

$$P(Y^t \mid V^t) = \prod_x \ell(Y^t \mid v_x^t) \ , \tag{1}$$

$$P(V^{t+1} \mid S^{t+1}, V^t) = \prod_x P(v_x^{t+1} \mid S^{t+1}, V^t) \ , \tag{2}$$

$$P(S^{t+1} \mid S^t) = \prod_x P(s_x^{t+1} \mid S^t) \ . \tag{3}$$

## 2.1 Observation likelihood

We define the observation likelihood $P(Y^t \mid V^t)$ by assuming that the likelihood $\ell(Y^t \mid v_x^t)$ of a local velocity $v_x^t$ should be related to finding the same or similar image patch $\mathbf{I}_x^{t+1}$ centred around $x$ at time $t+1$ that was present at time $t$ but centred around $x - v_x^t \Delta t$. In the following, we neglect dimensions and set $\Delta t = 1$. Let $\mathcal{S}(x, \mu, \Sigma, \nu)$ be the Student's t-distribution and $\mathcal{N}(x, \mu, \Sigma) = \lim_{\nu \to \infty} \mathcal{S}(x, \mu, \Sigma, \nu)$ be the normal distribution of a variable $x$ with mean $\mu$, covariance matrix $\Sigma$ and the degrees of freedom $\nu$. We define

$$\ell(Y^t \mid v_x^t) = \mathcal{S}(\mathbf{I}_x^{t+1}, \mathbf{I}_{x - v_x^t}^t, \frac{\sigma_I^2}{\mathcal{N}(x, x', \varrho_I)}, \nu_I) \ . \tag{4}$$

In this notation, the image patches can be regarded vectors and the covariance matrix is a diagonal with entries $\sigma_I^2 / \mathcal{N}(x, x', \varrho_I)$ that depend on the position $x'$ relative to the centre (i.e., a heteroskedastic variance). In effect, the term $\mathcal{N}(x, x', \varrho_I)$ implements a Gaussian weighting of locality centred around $x$ for the patch $\mathbf{I}_x^{t+1}$ and around $x - v_x^t$ for the patch $\mathbf{I}_{x - v_x^t}^t$. The parameter $\varrho_I$ defines the spatial range

of the image patches and $\sigma_I$ the grey value variance. The univariate Student's t-distribution realises a robust behaviour against large grey-value differences within image patches, which means the euclidean distance between the two patches is treated as an outlier if it is too large.

## 2.2 Flow field transition probability

The definition of the flow field transition probability $P(V^{t+1} \mid S^{t+1}, V^t)$ includes two factors. First, we assume that the flow field transforms according to itself. Second, the segmentation imposes an additional factor on the transition probability acting similarly to a prior over $V^{t+1}$ depending on the segmentation.

Let us first discuss the first factor: We assume that the origin of a local flow vector $v_x^{t+1}$ at position $x$ was a previous flow vector $v_{x'}^t$ at some corresponding position $x'$,

$$v_x^{t+1} \sim \mathcal{S}(v_x^{t+1}, v_{x'}^t, \sigma_V, \nu_V) \ . \tag{5}$$

So, we assume robust spatio-temporal coherence because evaluations on first derivative optical flow statistics [7] and on prior distributions that allow to imitate human speed discrimination tasks [8] provide strong indication that they resemble heavy tailed distributions. Now, asking what the corresponding position $x'$ in the previous image was, we assume that we can infer it from the flow field itself via

$$x' \sim \mathcal{N}(x', x - v_x^{t+1}, \varrho_V) \ . \tag{6}$$

Note that here we use $v_x^{t+1}$ to retrieve the previous corresponding point. Combining both factors and integrating $x'$ we would get (still neglecting the coupling to the segmentation)

$$v_x^{t+1} \mid V^t \sim \sum_{x'} \mathcal{N}(x', x - v_x^{t+1}, \varrho_V) \ \mathcal{S}(v_x^{t+1}, v_{x'}^t, \sigma_V, \nu_V) \ . \tag{7}$$

The parameter $\varrho_V$ defines the spatial range of a flow-field patch, so we compare velocity vectors within flow-field patches at different times $t$ and $t + 1$.

We introduced new parameters $\varrho_V$ and $\sigma_V$ for the uncertainty in spatial identification between two images and the transition noise between $V^t$ and $V^{t+1}$, respectively. The robustness against outliers is controlled by $\nu_V$, with smaller/larger $\nu_V$ decreasing/increasing the influence of incoherently moving pixels within the observed spatial range $\varrho_V$.

So far we have not discussed the coupling of the segmentation to the flow field transition. Each segment corresponds to a typical flow pattern $q_s(V^t)$. On its own (i.e., within one time slice), a flow pattern $q_s$ corresponds to a prior over the flow field,

$$q_s(V^t) = \prod_x \mathcal{N}(v_x, A_s x + t_s, \sigma_s) \ . \tag{8}$$

This is equivalent to assuming the world is approximately a set of planar objects such that each of their movements are completely described by an affine parameterisation $A_s$ and $t_s$, with $A_s$ being a $2 \times 2$ matrix describing any combination of rotation, divergence and shear, and $t_s$ being a $2 \times 1$ translational vector.

The segmentation field $S^t$, which contains for every pixel a label $s_x^t \in \{1, .., K\}$, specifies the correspondence of each pixel to each flow pattern. This segmentation couples to the flow field transition probability as an additional factor. Combining this with (7) we finally define the flow field transition probability as

$$P(v_x^{t+1} \mid S^{t+1}, V^t) \propto q_{s_x^{t+1}}(v_x^{t+1}) \sum_{x'} \mathcal{N}(x', x - v_x^{t+1}, \varrho_V) \, \mathcal{S}(v_x^{t+1}, v_{x'}^t, \sigma_V, \nu_V) \, .$$

(9)

## 2.3  Segmentation transition probability

For the transition of the segmentation field itself we assume, following exactly the same reasoning as previously for $v_x^{t+1} \mid V^t$ in equation (7),

$$P(s_x^{t+1} \mid S^t) \propto \sum_{x'} \mathcal{N}(x', x - \bar{q}_{s,x}^{t+1}, \varrho_S) \, Q(s_x^{t+1}, s_{x'}^t, \nu_S) \, ,$$

(10)

where $\bar{q}_{s,x}^{t+1}$ is the mean of $q_{s_x^{t+1}}$ at $x$, and $Q(s, s', \nu_S)$ adds uniform noise on a discrete random variable,

$$Q(s, s', \nu_S) = \begin{cases} 1 - \nu_S(K-1) & \text{for s'=s} \\ \nu_S & \text{otherwise} \end{cases} .$$

(11)

That is, we assume that the segmentation field transforms according to the flow field priors $q_s$ as specified by the segmentation itself; when a pixel is labelled $s$ we expect that it transforms according to the flow field $\bar{q}_s$.

# 3  Factored frontier inference

The finite set of flow patterns $q_s$ is a basic mechanism to introduce global (w.r.t. the image range, not the time horizon) structure to the prior over flow fields, similar to a mixture of factored models. As a consequence, the exact belief over the flow field (e.g., during filtering) and the segmentation will not decouple. Dealing with a full joint distribution over the flow field $V^t$ and the segmentation $S^t$ is infeasible. Hence, we will use an approximate inference technique based on factored belief propagation, which can be regarded as a factored frontier algorithm [6]. The factored observation likelihoods and transition probabilities we introduced ensure that the forward propagated messages will remain factored as well. We will describe the algorithm here in the jargon of loopy BP. Below we briefly restate the factored frontier interpretation of the algorithm.

We start by assuming the belief over $V^t$ and $S^t$ at time $t$ to be factored,

$$P(V^t, S^t \mid Y^{1:t}) = P(V^t \mid Y^{1:t}) \, P(S^t \mid Y^{1:t}) =: \alpha(V^t) \, \alpha(S^t) = \prod_x \alpha(v_x^t) \, \alpha(s_x^t) \, .$$

(12)

We first propagate a message forward from $S^t$ to $S^{t+1}$, resulting in

$$\widetilde{\alpha}(S^{t+1}) = \prod_x \widetilde{\alpha}(s_x^{t+1}) \, , \quad \widetilde{\alpha}(s_x^{t+1}) \propto \mu_{s \to s'}(s_x^{t+1})$$

(13)

$$\mu_{s \to s'}(s_x^{t+1}) \propto \sum_{S^t \in \{1,..,K\}^X} P(s_x^{t+1} \,|\, S^t)\, \alpha(S^t)$$

$$= \sum_{S^t} \sum_{x'} \mathcal{N}(x', x - \bar{q}_{s,x}^{t+1}, \varrho_S)\, Q(s_x^{t+1}, s_{x'}^t, \nu_S) \prod_z \alpha(s_z^t) \qquad (14)$$

$$= \sum_{x'} \mathcal{N}(x', x - \bar{q}_{s,x}^{t+1}, \varrho_S) \sum_{s_{x'}^t} Q(s_x^{t+1}, s_{x'}^t, \nu_S)\, \alpha(s_{x'}^t) \underbrace{\sum_{S^t \backslash s_{x'}^t} \prod_{z \neq x'} \alpha(s_z^t)}_{=1} \ .$$

Note that the summation $\sum_{S^t \in \{1,..,K\}^X}$ is summing over all possible segmentation fields ($X$ is the pixel range), i.e. it represents $|X|$ summations $\sum_{s_1^t} \sum_{s_2^t} \sum_{s_3^t} \cdots$ over each local segmentation label. We separated these into a summation $\sum_{s_{x'}^t}$ over the label at $x'$ and a summation $\sum_{S^t \backslash s_{x'}^t}$ over all other labels at $x \neq x'$. Hence we can use $\sum_{S^t \backslash s_{x'}^t} \prod_{z \neq x'} \alpha(s_z^t) = \prod_{z \neq x'} \sum_{s_z^t} \alpha(s_z^t) = 1$.

Next we propagate forward from $V^t$, $S^{t+1}$, and $Y^{t+1}$ to $V^{t+1}$ resulting in

$$\alpha(V^{t+1}) = \prod_x \alpha(v_x^{t+1})\ , \quad \alpha(v_x^{t+1}) \propto \ell(Y^{t+1} \,|\, v_x^{t+1})\, \mu_{v \to v'}(v_x^{t+1})\, \mu_{s \to v}(v_x^{t+1})\ ,$$

$$\mu_{s \to v}(v_x^{t+1}) \propto \sum_{s_x^{t+1}} q_{s_x^{t+1}}(v_x^{t+1})\, \widetilde{\alpha}(s_x^{t+1})\ , \qquad (15)$$

$$\mu_{v \to v'}(v_x^{t+1}) \propto \sum_{V^t \in W^X} \sum_{x'} \mathcal{N}(x', x - v_x^{t+1}, \varrho_V)\, \mathcal{S}(v_x^{t+1}, v_{x'}^t, \sigma_V, \nu_V) \prod_z \alpha(v_z^t)$$

$$= \sum_{x'} \mathcal{N}(x', x - v_x^{t+1}, \varrho_V) \sum_{v_{x'}^t} \mathcal{S}(v_x^{t+1}, v_{x'}^t, \sigma_V, \nu_V)\, \alpha(v_{x'}^t)\ , \qquad (16)$$

analogous to (14). Finally we pass back a message from $V^{t+1}$ to $S^{t+1}$,

$$\alpha(S^{t+1}) = \prod_x \alpha(s_x^{t+1})\ , \quad \alpha(s_x^{t+1}) \propto \widetilde{\alpha}(s_x^{t+1})\, \mu_{v \to s}(s_x^t) \qquad (17)$$

$$\mu_{v \to s}(s_x^t) \propto \sum_{v_x^{t+1}} q_{s_x^{t+1}}(v_x^{t+1})\, \frac{\alpha(v_x^{t+1})}{\mu_{s \to v}(v_x^{t+1})}$$

$$\propto \sum_{v_x^{t+1}} q_{s_x^{t+1}}(v_x^{t+1})\, \ell(Y^{t+1} \,|\, v_x^{t+1})\, \mu_{v \to v'}(v_x^{t+1})\ . \qquad (18)$$

This inference procedure can be interpreted as a Factor Frontier Algorithm (FFA) as follows. We start with a factored frontier belief (12) over $V^t$ and $S^t$. The FFA first adds the node $S^{t+1}$ to formulate the joint $P(V^t, S^t, S^{t+1})$. For this joint the marginals over $V^t$ and $S^{t+1}$ are computed and these define the new factored frontier. In this step it turns out that the marginal over $V^t$ remains unchanged whereas the marginal over $S^{t+1}$ is our $\widetilde{\alpha}(S^{t+1})$ we computed in equation (13). To the new frontier, the node $V^{t+1}$ is added to formulate the joint over $P(V^t, S^{t+1}, V^{t+1})$ (including the new observation likelihood factor $P(Y^{t+1} \,|\, V^{t+1})$). For this joint the marginals over $S^{t+1}$ and $V^{t+1}$ are computed to yield the new factored frontier. It turns out that these marginals are $\alpha(S^{t+1})$ and $\alpha(V^{t+1})$ as we computed them in (17) and (15).

## 3.1 Two-filter inference

If we have access to a batch of data (or a recent window of data) we can compute smoothed posteriors as a basis for an EM-algorithm and train the free parameters. In our two-filter approach we derive the backward filter as a mirrored version of the forward filter, using

$$P(v_x^t \mid S^t, V^{t+1}) \propto q_{s_x^t}(v_x^t) \sum_{x'} \mathcal{N}(x', x + v_x^t, \varrho_V) \, \mathcal{S}(v_x^t, v_{x'}^{t+1}, \sigma_V, \nu_V) \qquad (19)$$

$$P(s_x^t \mid S^{t+1}) \propto \sum_{x'} \mathcal{N}(x', x + \bar{q}_{s,x}^t, \varrho_S) \, Q(s_x^t, s_{x'}^{t+1}, \nu_S) \ . \qquad (20)$$

instead of (9) and (10). These equations are motivated in exactly the same way as we motivated (7): e.g., we assume that the $v_x^t \sim \mathcal{S}(v_x^t, v_{x'}^{t+1}, \sigma_V, \nu_V)$ for a corresponding position $x'$ in the subsequent image, and that $x' \sim \mathcal{N}(x', x - v_x^t, \varrho_V)$ is itself defined by $v_x^t$. However, note that using this symmetry of argumentation is actually an approximation to our model because applying Bayes rule on (9) or (10) would lead to a different, non-factored $P(V^t \mid S^t, V^{t+1})$. The backward filter equations are exact mirrors of the forward equations. To derive the smoothed posterior we need to combine the forward and backward filters. In the two-filter approach this reads

$$
\begin{aligned}
\gamma(v_x^t) &:= P(v_x^t \mid Y^{1:T}) = \frac{P(Y^{t+1:T} \mid v_x^t) \, P(v_x^t \mid Y^{1:t})}{P(Y^{1:T})} \\
&= \frac{P(v_x^t \mid Y^{t+1:T}) P(Y^{t+1:T}) P(v_x^t \mid Y^{1:t})}{P(v_x^t) P(Y^{1:T})} \propto \alpha(v_x^t) \, \beta(v_x^t) \, \frac{1}{P(v_x^t)} \quad (21)
\end{aligned}
$$

with $P(Y^{t+1:T})$ and $P(Y^{1:T})$ being constant. If both the forward and backward filters are initialised with $\alpha(v_x^0) = \beta(v_x^T) = P(v_x)$ we can identify the normalisation $P(v_x^t)$ with the prior $P(v_x)$. The same holds for the smoothed segmentation estimate $\gamma(s_x^t)$.

## 3.2 Parameter adaptation

This paper focuses on adapting the parameters that define each segment. These are the 6 parameters $A_s$ and $t_s$ and the variance $\sigma_s$ associated with each of the $K$ segment models $q_s$. Following the EM-algorithm we use inference on an image sequence to derive posteriors $\gamma(v_x^t)$ and $\gamma(s_x^t)$ over the latent variables $V^{1:T}$ and $S^{1:T}$. The exact M-step would then compute parameters that maximise the expected data log-likelihood, expectations taken by integrating over all latent variables. However, given the high-dimensionality of our latent variables a full integration is computationally too expensive. Thus, we use approximate M-steps based on the MAP estimates of the latent variables.

More precisely, the estimation of $A_s$ and $t_s$ are based on the MAP flow field estimates $\hat{v}_x^t = MAP(\gamma(v_x^t))$. Since $q_s$ is assumed to be Gaussian, parameter estimates can be found by weighted linear regression of the MAP flow field. Further, let $w_{s,x}^t = 1$ if the pixel $x$ at time $t$ corresponds most likely to segment $s$, i.e. iff $\hat{s}_x^t = MAP(\gamma(s_x^t)) = s$, and zero otherwise. Then we update the variance $\sigma_s$ as

$$\sigma_s = \frac{\sum_{x,t} \gamma(\hat{v}_x^t) \, w_{s,x}^t \, (\hat{v}_x^t - \bar{q}_{s,x}^t)^2}{\sum_{x,t} \gamma(\hat{v}_x^t) \, w_{s,x}^t} \ . \qquad (22)$$
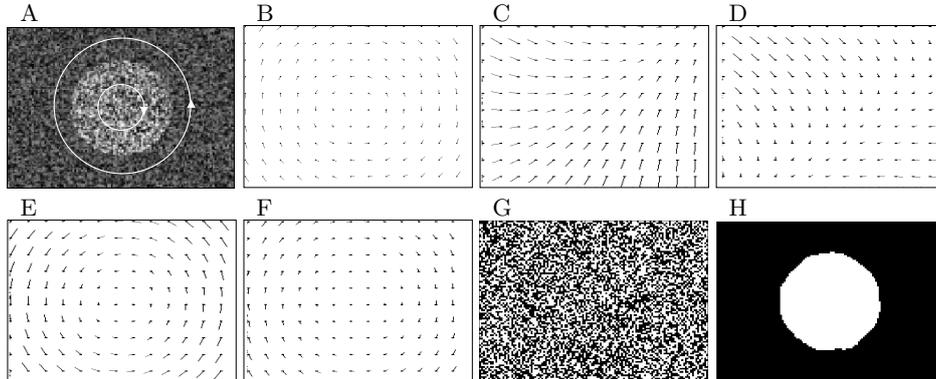
Figure 2: Rotating disc example.

## 4 Examples

We first test the algorithm on a scene composed of a rotating disc on a rotating background. Fig. 2A shows an image with a circular random pattern in the centre of the image rotating with constant angular velocity of $\omega = -\pi/30\Delta t$ and a random background pattern rotating with $\omega = \pi/60\Delta t$. The sequence contains 5 images. Fig. 2B displays the MAP flow field after 5 EM iterations using two-filter inference on the sequence batch. The pair of Figures 2C&D shows the mean of the random initialisation of $q_s$ for both segments, while the pair of Figures 2E&F displays $q_s$ for both segments after the 5 EM iterations. The two flow patterns $q_s$ nicely specialised to model either the left or right rotating fields (the right rotation with double velocity). Finally, 2G displays the initial random segmentation of the image while 2H shows the final MAP segmentation after 5 EM iterations, which is close to the best possible pixel-based accuracy.

Next we tested the algorithm on the Flower Garden sequence. Fig. 3A displays the 15$^{\text{th}}$ image of the original sequence. In a first experiment we allowed for 4 different segments ($K = 4$) and initialised all $(A_s, t_s)$ randomly and $\sigma_s$ large. We first used only 2 EM-iterations on the whole sequence to train the parameters. Fig. 3B shows the result in terms of the MAP segmentation. The $\sigma_s$'s became smaller and two flow field segments are dominating, which correspond to the tree and the background. Then, after 8 iterations of the EM-algorithm, the $\sigma_s$'s further decreased and three segments are detected (Fig. 3C), which correspond to the tree, the flower garden and back branches, and the house and background. Note the smoothness of segmentation achieved based on the spatio-temporal coupling. Fig. 3D also displays the probabilistic segmentation: The forward filtered belief $\alpha(S^t)$ (in the 8$^{\text{th}}$ EM iteration) for the four different segments $s = 1, .., 4$ are displayed. Compared to the MAP segmentation the probabilistic segmentation is smoother and captures more detailed information.

The last experiment tests a pure online segmentation using only forward filtering and an online EM which adapts the parameters at each time step. This targets towards real-time segmentation that does not rely on batch processing. Since in the flower garden sequence the local flow field measurement is rather precise we pre-specified a low variance $\sigma_s = 0.25$ for all $s$. We allowed for 3 segments ($K = 3$)
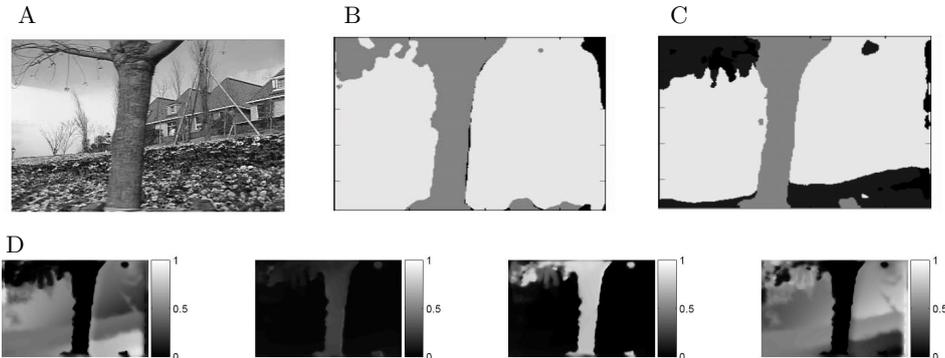
Figure 3: Flower Garden test using two-filter inference.



Figure 4: Flower Garden test using an online forward filter and learning

and initialised one of them to zero $A_s = 0, t_s = 0$ (a prior that there are regions of low velocity) and the other two randomly. Fig. 4 displays the MAP segmentation for the online EM-filter after $t = 1, 2, 3, 4, 10, 20, 29$ time steps. In the first iteration the segmentation already distinguishes between the tree stem, the flower garden and branches, and the house and background. Initially this segmentation is rather noisy. During the online EM filter the segmentation becomes refined and also spatially more smooth due to the spatio-temporal priors that are build up during filtering.

## 5   Conclusion

Our approach focuses on exploiting spatio-temporal coherence in the flow field and segmentation evolution by formulating a Dynamic Bayesian Network framework as a basis for online filtering and inference over an image sequence. The core ingredients are the particular assumptions implied by our transition priors of the flow field (9) and the segmentation (10), and the efficient inference technique based on propagating factored beliefs over the flow field and segmentation forward and backward. Further, using the Student's t-distributions (4, 7) increases the robustness against outliers.

Both experiments have shown how the algorithm can extract a smooth segmentation of a sequence of images starting with a random initialisation of the segments' flow field and variance parameters. The smoothness is a result of the assumed transition prior rather than an additional MRF prior. This reduces the computational costs *per time step* considerably and is particularly interesting in view of online filtering problems as demonstrated by the last example. Here, an online parameter adaptation (online EM) can be based only on forward propagated beliefs with rather small computational cost per time step.

The second example also demonstrated the interesting effect of adapting the

variance associated to each segment. Large initialised $\sigma_S$'s have the effect that all segments $q_s$ try to learn similar global flow patterns. The resulting flat segmentation posterior has only minor influence on the flow field estimation. During the EM the $\sigma_S$'s decrease (to increase the overall data likelihood), the segments start to specialise on certain spatial regions and the segmentation becomes more and more detailed.

## Acknowledgements

## References

[1] M.J. Black and P. Anandan. Robust dynamic motion estimation over time. In *CVPR*, pages 296–302, 1991.

[2] P. Burgi, A.L.Yuille, and N.M. Grzywacz. Probabilistic motion estimation based on temporal coherence. *Neural Computation*, 12(8):1839–1867, 2000.

[3] R. Dupont, O. Juan, and R. Keriven. Robust segmentation of hidden layers in video sequences. *IEEE ICPR*, 3:75–78, 2006.

[4] F. Heitz and P. Bouthemy. Multimodal estimation of discontinuous optical flow using markov random fields. *PAMI*, 15(12):1217–1232, 1993.

[5] K.P. Lim, A. Das, and M.N. Chong. Estimation of occlusion and dense motion fields in a bidirectional bayesian framework. *PAMI*, 24(5):712–718, 2002.

[6] Kevin Murphy and Yair Weiss. The factored frontier algorithm for approximate inference in DBNs. In *Proc. of the 17th Conf. on Uncertainty in Artificial Intelligence (UAI 2001)*, pages 378–385, 2001.

[7] S. Roth and M.J. Black. On the spatial statistics of optical flow. In *ICCV*, pages 42–49, 2005.

[8] A. A. Stocker and E. P. Simoncelli. Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, 9(4):578–585, 2006.

[9] N. Vasconcelos and A. Lippmann. Empirical bayesian motion segmentation. *PAMI*, 23:217–221, 2001.

[10] Y. Wang, K.-F. Loe, T. Tan, and J.-K. Wu. Spatiotemporal video segmentation based on graphical models. *IEEE Trans. on Image Processing*, 14:937–947, 2005.

[11] Y. Weiss and E. H. Adelson. A unified mixture framework for motion segmentation: incorporating spatial coherence and estimating the numbers of models. *CVPR*, pages 321–326, 1996.

[12] V. Willert, J. Eggert, J. Adamy, and E. Körner. Non-gaussian velocity distributions integrated over space, time, and scales. *IEEE Transactions on Systems, Man and Cybernetics - Part B*, 36(3):482–493, 2006.

[13] K. Y. Wong, L. Ye, and M. E. Spetsakis. EM clustering of incomplete data applied to motion segmentation. *BMVC*, pages 237–246, 2004.