Estimation of Motion Vector Fields

Rasmus Larsen

Institute of Mathematical Statistics and Operations Research
Technical University of Denmark

Abstract

This paper presents an approach to the estimation of 2-D motion vector fields from time varying image sequences. We use a piecewise smooth model based on coupled vector/binary Markov random fields. We find the maximum a posteriori solution by simulated annealing. The algorithm generate sample fields by means of stochastic relaxation implemented via the Gibbs sampler.

1 Introduction

The computation of a displacement vector field, that links a pixel in one frame to the corresponding pixel position in another frame, is of great importance for the interpretation of image sequences with temporal variation. The 2-D displacement vector field can be used to infer 3-D motion or to compute structure from motion. It can however also be used directly for interpolation and noise reduction or compression of image sequences. There is also a analogy to the extraction of spatial information from stereo-pairs.

Existing approaches motion detections algorithms either rely on low level vision techniques such as block matching, computation of optical flow based on spatio-temporal gradients and Fourier methods, or high level techniques, that use image analysis to extract key object features, such as edges, boundaries or complete objects, and use these to solve the correspondence problem.

This problem of motion computation is difficult due to its ill-posedness and its complexity. It is ill-posed since many different vector fields can describe the data, and it is complex due to its high dimensionality. In this paper we will apply the stochastic optimisation approach of simulated annealing to determine a displacement field based on a coupled Markov random field model.

Optical flow was first defined by [1] as the apparent motion of the brightness patterns contained in two frames. Horn and Schunk used a motion constraint obtained from an assumption of constancy of image brightness, together with a motion smoothness constraint in an iterative scheme to obtain a solution. The coupled vector/line Markov Random Field (MRF) framework has been used by [2, 3].

2 A Markov Random Field Description of the Displacement

The formulation involves specification of a deterministic structural model, and stochastic observation and motion field models. We will use the proposed piecewise smooth model derived from coupled vector-line(binary) Markov random fields. The maximum *a posteriori* (MAP) estimation is performed using simulated annealing, in which sample fields are generated by means of stochastic relaxation implemented via the Gibbs sampler.

The observed image, g, which is related to the true underlying image, I, by some random transformation is considered to be a sample of a random field, G.

Disregarding occlusions and newly exposed areas, for every point in the preceding image, $t = t_1$, there exist a corresponding point in the following image, $t = t_2$. Let the 2-D projection of the straight lines connecting these pairs of points be referred to as the displacement field, $\boldsymbol{u}$, associated with the underlying image I.

The true displacement field, $\tilde{\boldsymbol{u}}$, is a set of 2-D vectors such that for all $(\boldsymbol{x}_i, t)$, the preceding point $(\boldsymbol{x}_i, t_1)$ has moved to the following point $(\boldsymbol{x}_i, t_2)$. $\tilde{\boldsymbol{u}}$ is assumed to be a sample from a random field $\boldsymbol{U}$. Let $\hat{\boldsymbol{u}}$ be an estimate of $\tilde{\boldsymbol{u}}$ and $\boldsymbol{u}$ denote any sample field from $\boldsymbol{U}$.

It is obvious that we will encounter motion discontinuities along boundaries of objects with different motion. We will describe these discontinuities with a binary field l. Let the true field be denoted $\tilde{l}$. $\tilde{l}$ will be represented by a discrete discontinuity field. The line elements are located midway between points and take on the value 1 if there exists a motion boundary between the neighbouring displacement sites, and the value 0 otherwise. In analogy with the displacement field we assume $\tilde{\boldsymbol{l}}$ to be a sample from a random field $\boldsymbol{L}$, $\hat{\boldsymbol{l}}$ be an estimate of $\tilde{\boldsymbol{l}}$ and $\boldsymbol{l}$ denote any sample field from $\boldsymbol{L}$.

2.1 Estimation Criteria

We seek to estimate the pair $(\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{l}})$ of displacement and line field at a time t corresponding to an underlying image I on the basis of the observation g. In the maximum a posteriori sense the best displacement estimate $(\hat{\boldsymbol{u}}_t^*, \hat{\boldsymbol{l}}_t^*)$ of must satisfy

$$p(\hat{\boldsymbol{u}}_t^*, L_t = \hat{l}_t^* \mid g_{t_1}, g_{t_2}) \geq p(\hat{\boldsymbol{u}}_t, L_t = \hat{l}_t \mid g_{t_1}, g_{t_2}), \tag{1}$$

$\forall \hat{\boldsymbol{u}}_t, \hat{l}_t$, where p is the conditional probability distribution of the displacement and line field given the observation. By using the Bayes rule for random variables, we get

$$p(\hat{\boldsymbol{u}}_t, L_t = \hat{l}_t \mid g_{t_1}, g_{t_2}) = \frac{p(\boldsymbol{u}_t, L_t = l_t \mid g_{t_1})}{p(g_{t_2} \mid g_{t_1})} p(g_{t_2} \mid \boldsymbol{u}_t, l_t, g_{t_1}) \tag{2}$$

Note that as the denominator is not a function of $(\boldsymbol{u}_t, L_t)$, it can be omitted when maximising the posterior probability with respect to $(\boldsymbol{u}_t, l_t)$.

2.2 Models

We now formulate models for the probabilities in equation (2). We link displacement vectors and intensity values by assuming constant image intensity along motion trajectories. This relationship is extrapolated to the observed image, g. The displaced pixel differences

$$r(\tilde{\boldsymbol{u}}(\boldsymbol{x}_i, t), \boldsymbol{x}_i) = g_{t_2}(\boldsymbol{x}_i) - g_{t_1}(\boldsymbol{x}_i + \tilde{\boldsymbol{u}}(\boldsymbol{x}_i, ti_1)) \tag{3}$$

are modelled by independent Gaussian random variables. Given these assumptions, we have

$$p(g_{t_2} \mid \boldsymbol{u}_t, l_t, g_{t_1}) = (2\pi\sigma^2)^{-M\boldsymbol{u}/2} \cdot \exp\left(-\frac{H_g(g_{t_2} \mid \boldsymbol{u}_t, g_{t_1})}{2\sigma^2}\right) \tag{4}$$

where the energy function H_g is defined as $H_g(g_{t_2} \mid \boldsymbol{u}_t, g_{t_1}) \approx \sum [r(\tilde{\boldsymbol{u}}(\boldsymbol{x}_i, t), \boldsymbol{x}_i)]^2$.

As the motion in most scenes is the result of change of position of rigid or near rigid bodies, the motion field of such images will consist of patches of similar vectors with possible discontinuities at motion boundaries. Therefore we will assume that motion fields are smooth functions of spatial position. We will model this by a pair $(\boldsymbol{U}_t, L_t)$ of coupled vector and binary Markov random fields. Remember that a MRF with respect to a neighbourhood is uniquely characterised by a Gibbs distribution with respect to the same neighbourhood system. The properties of the motion model are described by $p(\boldsymbol{u}_t, L_t = l_t \mid g_{t_1})$ from equation (2), which can be factored using the Bayes rule

$$p(\boldsymbol{u}_t, L_t = l_t \mid g_{t_1}) = p(\boldsymbol{u}_t \mid l_t, g_{t_1}) \cdot P(L_t = l_t \mid g_{t_1}) \tag{5}$$

If these two factors are Gibbsian, so is he product. Assuming that a single image field contribute little to the motion vector field, we omit conditioning on g_{t_1} in $p(\boldsymbol{u}_t, L_t = l_t \mid g_{t_1})$ as an approximation. We will, however, keep the conditioning on g_{t_1} in $P(L_t = l_t \mid g_{t_1})$, as motion discontinuities are likely to coincide with positions of intensity discontinuities at object boundaries. Under the assumed Markov properties, that every configuration can be attained with a non-zero probability and that the probability of a site having a specific value is dependent only on the values of the sites in a pre-defined neighbourhood, $\boldsymbol{U}_t$ can be expressed by the Gibbs distribution

$$p(\boldsymbol{u}_t \mid l_t) = \frac{1}{Z_{\boldsymbol{u}}} \exp\left(-\frac{H_{\boldsymbol{u}}(\boldsymbol{u}_t \mid l_t)}{\beta_{\boldsymbol{u}}}\right), \tag{6}$$

where $Z_{\boldsymbol{u}}$ is a partition function, $\beta_{\boldsymbol{u}}$ is a constant controlling characteristic properties of $\boldsymbol{U}_t$, and the energy function is defined as

$$H_{\boldsymbol{u}}(\boldsymbol{u}_t \mid l_t) = \sum_{c_{\boldsymbol{u}} = \{\boldsymbol{x}_i, \boldsymbol{x}_j\} \in C_{\boldsymbol{u}}} V_{\boldsymbol{u}}(\boldsymbol{u}_t, c_{\boldsymbol{u}}) \left[1 - l(\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle, t)\right] \tag{7}$$

where $c_{\boldsymbol{u}}$ is a clique of vectors, and $C_{\boldsymbol{u}}$ is a set containing all such cliques derived from a neighbourhood system. $(\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle, t)$ is the site of the line element located between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. $V_{\boldsymbol{u}}$ is a potential characterising the displacement process $\boldsymbol{U}_t$. This potential is the cost associated with each vector clique. The second term makes sure that there is no penalty for introducing a abrupt changes in displacement. Later we will penalise the insertion of a line element. In order to model the above mentioned smoothness assumption we define the potential function to be

$$V_{\boldsymbol{u}}(\boldsymbol{u}_t, c_{\boldsymbol{u}}) = \|\boldsymbol{u}(\boldsymbol{x}_i, t) - \boldsymbol{u}(\boldsymbol{x}_j, t)\|^2 \tag{8}$$

We will use the first order neighbourhood system depicted in Figure 1(a), which consists of two-element horizontal and vertical cliques, cf. Figures 1(c) and 1(b).

As for the line field model, this is described by the Gibbs probability distribution

$$P(L_t = l_t \mid g_{t_1}) = \frac{1}{Z_l} \exp\left(-\frac{H_l(l_t|g_{t_1})}{\beta_l}\right), \tag{9}$$

where the energy function is defined like this $H_l(l_t|g_{t_1}) = \sum_{c_l \in C_l} V_l(l_t, g_{t_1}, c_l$, where c_l is a line clique and C_l is the set of all line cliques derived from the neighbourhood system N_l defined over S_l. The potential function V_l penalises introduction of line elements. The second-order neighbourhood system N_l^2 for the lattice S_l is shown in Figure 2. Note that we, due to having both horizontal and vertical line elements, have two neighbourhood systems, see Figures 2(a) and 2(b). There are two types of four element line cliques. The cross-shaped cliques from Figure 2(c) are used to model the shape of motion boundaries, whereas the

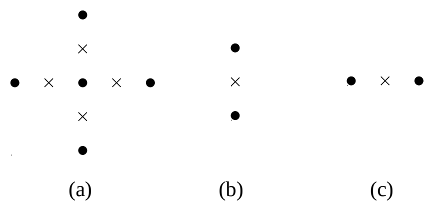

Figure 1. (a) First-order neighbourhood system $N_{\boldsymbol{u}}^1$ for vector field $\boldsymbol{u}_t$ defined over $S_{\boldsymbol{u}}$ with discontinuities l_t defined over S_l; (b) vertical cliques; (c) horizontal cliques (•–vector site, ×–line site).

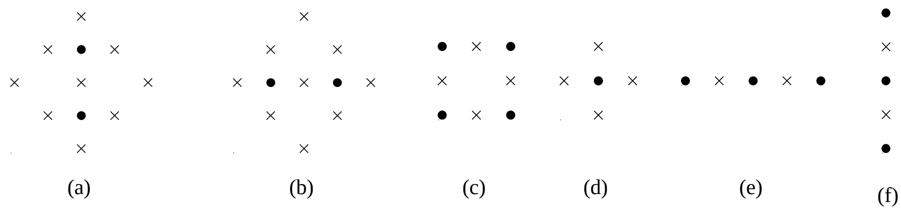

Figure 2. Second-order neighbourhood system N_l^2 for line field defined over S_l. (a) Horizontal line element; (b) Vertical line element; (c),(d) Four-element cliques; (e),(f) Two-element cliques; (•–vector site, ×–line site).

square-shaped cliques from Figure 2(d) are used to inhibit isolated vectors. The two-element horizontal and vertical cliques in Figures 2(e) and 2(f) are used to prevent double edges.

For the one-element cliques the following potential function is used.

$$V_{l_1}(l_t, g_{t_1}, c_l) = \begin{cases} \frac{\alpha}{(\frac{\partial g_{t_1}}{\partial x})^2} l_h(\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle, t) & \text{for horizontal } c_l \\ \frac{\alpha}{(\frac{\partial g_{t_1}}{\partial y})^2} l_v(\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle, t) & \text{for vertical } c_l, \end{cases} \tag{10}$$

where l_h and l_v are horizontal and vertical line elements and α is a constant. V_{l_1} represents a penalty only if the line element is on and the appropriate gradient is small. The total potential function for the line field can be expressed as $V_l(l_t, g_{t_1}, c_l) = V_{l_1}(l_t, g_{t_1}, c_l) + V_{l_2}(l_t, c_l) + V_{l_4}(l_t, c_l)$, where the potentials for the various two- and four-element clique configurations are tabulated in Figure 3. Note from Figures 3(a)–3(f), that we apply small penalties for straight lines and high penalties for intersections. The square shaped clique configuration potential in Figure 3(g) prohibit isolated points. Figures 3(h)–3(k) show potentials to penalise double edges.

2.3 A Posteriori Probability

When we combine the observation model (4), the a priori probability of the displacement (6) and the a priori probability of the line field (9), using (2), we get the following Gibbs form of the a posteriori probability

$$P(\boldsymbol{U}_t = \hat{\boldsymbol{u}}_t, L_t = \hat{l}_t \mid g_{t_1}, g_{t_2}) = \frac{1}{Z} \exp\left(-H_{\boldsymbol{u}}(\boldsymbol{u}_t, \hat{l}_t, g_{t_1}, g_{t_2})\right), \tag{11}$$

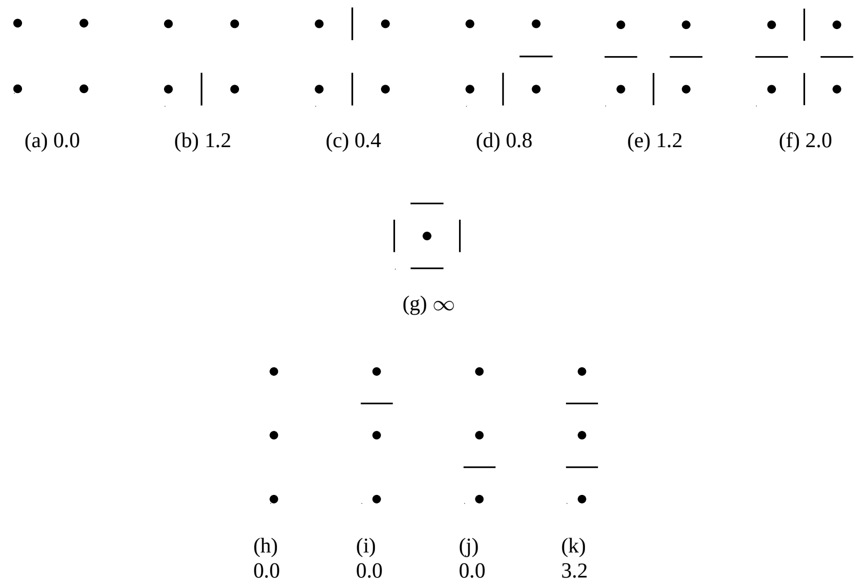

Figure 3. Potentials associated with various configurations (up to a rotation): (a)-(f) four-element cross shaped cliques; (g) four element square shaped clique; (h)-(k) two-element cliques (•–vector site, — –line element ”on”).

where Z is a new normalising constant and the new energy function is defined as

$$H_{\boldsymbol{u}}(\boldsymbol{u}_t, \hat{l}_t, g_{t_1}, g_{t_2}) = \lambda_g H_g(\tilde{g}_{t_2} \mid \hat{\boldsymbol{u}}_t, g_{t_1}) + \lambda_{\boldsymbol{u}} H_{\boldsymbol{u}}(\hat{\boldsymbol{u}}_t \mid \hat{l}_t) + \lambda_l H_l(\hat{l}_t | g_{t_1}) \tag{12}$$

We have introduced the new parameters $\lambda_g = \frac{1}{2\sigma^2}$, $\lambda_{\boldsymbol{u}} = \frac{1}{\beta_{\boldsymbol{u}}}$ and $\lambda_l = \frac{1}{\beta_l}$. The neighbourhood system of this Gibbs distribution is a combination of $N_{\boldsymbol{u}}$ and N_l. As the posterior distribution is Gibbsian, the MAP estimate can be found by the following minimisation

$$\min_{\{\hat{\boldsymbol{u}}_t, \hat{l}_t\}} \lambda_g H_g(\tilde{g}_{t_2} \mid \hat{\boldsymbol{u}}_t, g_{t_1}) + \lambda_{\boldsymbol{u}} H_{\boldsymbol{u}}(\hat{\boldsymbol{u}}_t \mid \hat{l}_t) + \lambda_l H_l(\hat{l}_t | g_{t_1}) \tag{13}$$

3 Simulated Annealing Formulation

To solve this minimisation problem (13) we will employ simulated annealing. Introducing a control parameter, the temperature T, in the Gibbs distribution yields

$$p(\hat{\boldsymbol{u}}_t, L_t = \hat{l}_t \mid g_{t_1}, g_{t_2}) = \frac{1}{Z} \exp\left(-H_{\boldsymbol{u}}(\boldsymbol{u}_t, \hat{l}_t, g_{t_1}, g_{t_2}/T)\right) \tag{14}$$

We generate sample configurations from the Gibbs distribution using stochastic relaxation. The Gibbs sampler will be incorporated with the annealing scheme specified by the initial temperature T_0, the final temperature T_s and the temperature changing rule $T_k = \phi(T_0, k)$.

In the case of a discrete displacement process $\boldsymbol{u}_t$, the Gibbs sampler generates a new vector at every position $(\boldsymbol{x}_i, t) \in S_{\boldsymbol{u}}$ according to the marginal conditional probability distribution

$$p(\hat{\boldsymbol{u}}(\boldsymbol{x}_i,t) \mid \hat{\boldsymbol{u}}(\boldsymbol{x}_j,t), j \neq i, \hat{l}_t, g_{t_1}, g_{t_2}) = \frac{\exp\left(\frac{-H^i_{\boldsymbol{u}}(\hat{\boldsymbol{u}}_t,\hat{l}_t,g_{t_1},g_{t_2})}{T}\right)}{\sum_{\boldsymbol{z}\in S_{\boldsymbol{u}}} \exp\left(\frac{-H^i_{\boldsymbol{u}}(\hat{\boldsymbol{u}}^z_t,\hat{l}_t,g_{t_1},g_{t_2})}{T}\right)} \tag{15}$$

with the local displacement energy function $H^i_{\boldsymbol{u}}$ defined as

$$H^i_{\boldsymbol{u}}(\hat{\boldsymbol{u}}^z_t, \hat{l}_t, g_{t_1}, g_{t_2}) = \lambda_g \hat{r}(\boldsymbol{z}, \boldsymbol{x}_i, t)^2 + \lambda_{\boldsymbol{u}} \sum_{\boldsymbol{x}_j \in \aleph_{\boldsymbol{u}}(\boldsymbol{x}_i)} \|\boldsymbol{z} - \hat{\boldsymbol{u}}(\boldsymbol{x}_j,t)\|^2 (1 - \hat{l}(\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle, t)), \tag{16}$$

where $\aleph_{\boldsymbol{u}}(\boldsymbol{x}_i)$ is a spatial neighbourhood of the displacement vector at $\boldsymbol{x}_i$, and $\hat{\boldsymbol{u}}^z_t$ is a displacement field identical to the field $\hat{\boldsymbol{u}}_t$, except for the vector at the spatial location $\boldsymbol{x}_i$, which is $\boldsymbol{z}$. Similarly we express the conditional probability for the Gibbs sampler for the displacement discontinuities at the spatio-temporal location $(\boldsymbol{y}_i, t)$ as

$$P(L(\boldsymbol{y}_i,t) = \hat{l}(\boldsymbol{y}_i,t) \mid \hat{l}(\boldsymbol{y}_j,t), j \neq i, \hat{\boldsymbol{u}}_t, g_{t_1}) = \frac{\exp\left(\frac{-H^i_l(\hat{l}_t,\hat{\boldsymbol{u}}_t,g_{t_1})}{T}\right)}{\sum_{\boldsymbol{z}\in S_l} \exp\left(\frac{-H^i_l(\hat{l}^z_t,\hat{\boldsymbol{u}}_t,g_{t_1})}{T}\right)} \tag{17}$$

where the local line energy function H^i_l is defined as

$$H^i_l(\hat{l}^z_t, \hat{\boldsymbol{u}}_t, g_{t_1}) = \lambda_l \sum_{c_l \mid \boldsymbol{y}_i \in c_l} V_l(\hat{l}^z_t, g_{t_1}, c_l) + \lambda_{\boldsymbol{u}} \sum_{\substack{c_{\boldsymbol{u}} = \{\boldsymbol{x}_m, \boldsymbol{x}_n\} \mid \\ \langle \boldsymbol{x}_m, \boldsymbol{x}_n \rangle = \boldsymbol{y}_i}} \|\hat{\boldsymbol{u}}(\boldsymbol{x}_m,t) - \hat{\boldsymbol{u}}(\boldsymbol{x}_n,t)\|^2 (1 - z) \tag{18}$$

and $\hat{l}^z_t$ is a line field identical to the field $\hat{l}_t$, except for the line element at the spatial location $\boldsymbol{y}_i = \langle \boldsymbol{x}_m, \boldsymbol{x}_n \rangle$, where the line element is z.

The first term in the local energy of the displacement field is quadratic in $\hat{r}$, whereas the second term is quadratic in $\hat{\boldsymbol{u}}_t$. If the first term could be approximated with a quadratic form in $\hat{\boldsymbol{u}}_t$, the density function would be Gaussian, and generation of samples thus very efficient.

If we in the local displacement energy function (16) replace the displaced pixel difference $\hat{r}(\boldsymbol{z}, \boldsymbol{x}_i, t)$ with the Taylor expansion to the first order

$$\tilde{r}(\hat{\boldsymbol{u}}(\boldsymbol{x}_i,t), \boldsymbol{x}_i, t) \approx \hat{\boldsymbol{u}}(\boldsymbol{x}_i,t)^T \nabla g + \frac{\partial g}{\partial t} \tag{19}$$

the energy function becomes quadratic in $\hat{\boldsymbol{u}}_t$, and thus Gaussian. If we fit the conditional probability function $p(\hat{\boldsymbol{u}}(\boldsymbol{x}_i,t) \mid \hat{\boldsymbol{u}}(\boldsymbol{x}_j,t), j \neq i, \hat{l}_t, g_{t_1}, g_{t_2})$ with the local energy function resulting from the above approximation into a 2-D Gaussian distribution with mean, $\boldsymbol{m}$, and covariance matrix, $\boldsymbol{M}$.

We find, that

$$\begin{aligned} \boldsymbol{M} &= \frac{T}{2\lambda_{\boldsymbol{u}}\xi_i\mu_i}\left(\mu_i \boldsymbol{I} - \nabla g \nabla^T g\right) \\ \boldsymbol{m} &= \overline{\boldsymbol{u}}_i - \frac{1}{\mu_i}\nabla g\left(\frac{\partial g}{\partial t} + \overline{\boldsymbol{u}}_i^T \nabla g\right), \end{aligned}$$

which is very similar to the update equation of Horn and Schunks method. For $T = 0$ they are identical. Horn and Schunks algorithm thus corresponds to an instantaneously frozen simulated annealing.

References

[1] Berthold K. P. Horn and Brian G. Schunk. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[2] Janusz Konrad and Eric Dubois. Bayesian estimation of motion vector fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(9):910–927, September 1992.

[3] David W. Murray and Bernard F. Buxton. Scene segmentation from visual motion using global optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(2):220–228, March 1987.