

# A COMPARISON OF STOCHASTIC AND DETERMINISTIC SOLUTION METHODS IN BAYESIAN ESTIMATION OF 2-D MOTION<sup>†</sup>

Janusz Konrad<sup>‡</sup> and Eric Dubois  
INRS-Télécommunications, 3 Place du Commerce  
Verdun, Québec, Canada, H3E 1H6

## Abstract

A new stochastic motion estimation method based on the Maximum *A Posteriori* Probability (MAP) criterion is developed. Deterministic algorithms approximating the MAP estimation over discrete and continuous state spaces are proposed. These approximations result in known motion estimation algorithms. The theoretical superiority of the stochastic algorithms over deterministic approximations in locating the global optimum is confirmed experimentally.

## 1. INTRODUCTION

Substantial work has been carried out recently in the application of stochastic models to the estimation of two-dimensional motion. Based on Markov random field (MRF) models, the problem has been formulated using the MAP criterion and solved by stochastic [1],[2],[3],[4] and deterministic [5],[6] methods. Although both approaches have proved successful, no experimental comparison has been carried out. In this paper we extend our previous work by developing stochastic MAP estimation over a continuous state space of solutions. Also, by instantaneously “freezing” a Markov chain produced by a stochastic relaxation algorithm, we propose two deterministic estimation methods. These approximations result in known motion estimation algorithms.

## 2. FORMULATION

### 2.1 Terminology

Let  $u$  and  $g$  denote the true underlying and the observed time-varying images, respectively. Let  $g$  be a sample from a random field (RF)  $G$ , and be quantized in amplitude and sampled on a lattice  $\Lambda_g$  in  $R^3$ . Let  $(\mathbf{x}, t)$  be a site in  $\Lambda_g$ , where  $\mathbf{x}$  and  $t$  denote spatial and temporal positions, respectively. Let also  $\mathbf{d}$  be the true (unknown) displacement field associated with  $u$ . Since it is not feasible to estimate  $\mathbf{d}$  on a continuum of spatial positions, it will be estimated on a lattice  $\Lambda_d$  in  $R^3$ , which may be different than  $\Lambda_g$  as in the case of temporal interpolation.

It is assumed that  $\Lambda_g, \Lambda_d$  are rectangular lattices with horizontal, vertical and temporal sampling periods  $(T_g^h, T_g^v, T_g)$  and  $(T_d^h, T_d^v, T_d)$ , respectively. Each field of the image sequence contains  $M_g$  picture elements, and each motion field consists of  $M_d$  vectors.

The true displacement field  $\mathbf{d}$  is assumed to be a sample (realization) from random field  $\mathbf{D}$ . Let  $\mathbf{d}$  denote any sample field from  $\mathbf{D}$  and allow  $\hat{\mathbf{d}}$  be an estimate of the true displacement field  $\mathbf{d}$ . Assuming a linear motion trajectory between two images we define a displacement field as follows:

---

<sup>†</sup> Work supported by the Natural Sciences and Engineering Research Council of Canada under Strategic Grant G-1357

<sup>‡</sup> This research was conducted when the author was also with the Department of Electrical Engineering, McGill University, Montreal, Canada, on leave from the Technical University of Szczecin, Poland

The displacement field  $\mathbf{d}$  defined over  $\Lambda_{\mathbf{d}}$  is a set of 2-D vectors such that for all  $(\mathbf{x}_i, t) \in \Lambda_{\mathbf{d}}$  the *preceding* image point  $(\mathbf{x}_i - \Delta t \cdot \mathbf{d}(\mathbf{x}_i, t), t_-)$  has moved to the *following* point  $(\mathbf{x}_i + (1.0 - \Delta t) \cdot \mathbf{d}(\mathbf{x}_i, t), t_+)$ , where  $\Delta t = \frac{t}{T_g} - \left\lfloor \frac{t}{T_g} \right\rfloor$ ,  $t_- = t - \Delta t \cdot T_g$  and  $t_+ = t + (1.0 - \Delta t) \cdot T_g$ .

To model abrupt changes in displacement vector length and/or orientation we use the concept of motion discontinuity. The true motion discontinuities  $l$  are defined over continuous coordinates  $(\mathbf{x}, t)$ , and are unobservable like the true motion fields. They can be understood as indicator functions for each  $(\mathbf{x}, t)$ . We assume that  $l$  is a sample from a RF  $L$ . Let  $\hat{l}$  be an estimate of  $l$ . The RF  $L$  will be called a *line process*, its sample  $l$  will be called a *line field* while individual discontinuities from  $l$  will be named *line elements*. We will estimate  $l$  on a union of shifted lattices  $\Psi_l = \psi_h \cup \psi_v$ , where  $\psi_h = \Lambda_{\mathbf{d}} + [0, T_{\mathbf{d}}^v/2, 0]^T$  and  $\psi_v = \Lambda_{\mathbf{d}} + [T_{\mathbf{d}}^h/2, 0, 0]^T$  are orthogonal cosets of horizontal and vertical line elements, respectively.

We assume that the random field  $\mathbf{D}_t$  is defined over the state space  $\mathcal{S}_{\mathbf{d}} = (\mathcal{S}'_{\mathbf{d}})^{M_{\mathbf{d}}}$ , where  $\mathcal{S}'_{\mathbf{d}}$  is the single vector state space. Two cases of  $\mathcal{S}'_{\mathbf{d}}$  are considered: a discrete state space (square 2-D grid) and a continuous state space  $R^2$ . It is also assumed that the random field  $L$  is defined over the discrete state space  $\mathcal{S}_l = (\mathcal{S}'_l)^{M_l}$ , where  $\mathcal{S}'_l$  is the single line element state space. Finally, let the subscript  $t$  denote the restriction of a random field (RF) or of its realization to time  $t$ .

## 2.2 MAP estimation criterion

To estimate the pair  $(\mathbf{d}_t, l_t)$  of true displacement and line fields corresponding to image  $u$  on the basis of the observations  $g$ , a pair  $(\hat{\mathbf{d}}_t, \hat{l}_t) \in \mathcal{S}_{\mathbf{d}} \times \mathcal{S}_l$  which maximizes the *a posteriori* probability  $P(\mathbf{D}_t = \hat{\mathbf{d}}_t, L_t = \hat{l}_t | g_{t_-}, g_{t_+})$  must be found. Applying Bayes rule this probability can be factored as follows [2]

$$P(\mathbf{D}_t = \mathbf{d}_t, L_t = l_t | g_{t_-}, g_{t_+}) = \frac{P(G_{t_+} = g_{t_+} | \mathbf{d}_t, l_t, g_{t_-}) \cdot P(\mathbf{D}_t = \mathbf{d}_t | l_t, g_{t_-}) \cdot P(L_t = l_t | g_{t_-})}{P(G_{t_+} = g_{t_+} | g_{t_-})}. \quad (1)$$

Note that since the probability in the denominator of (1) is not a function of  $\mathbf{d}_t$ , it can be ignored when maximizing (1) with respect to  $(\hat{\mathbf{d}}_t, \hat{l}_t)$ . If displacement vectors are defined over a continuous state space  $\mathcal{S}'_{\mathbf{d}} = R^2$ , then Bayes rule for mixed random variables results in a similar probability distribution where *a priori* probability  $P(\mathbf{D}_t = \hat{\mathbf{d}}_t | \hat{l}_t, g_{t_-})$  is replaced by the probability density  $p(\hat{\mathbf{d}}_t | \hat{l}_t, g_{t_-})$ .

### 2.2.1 Displaced pel difference model

To estimate motion from images a *structural* model relating motion vectors and image intensity values is needed. Disregarding illumination and occlusion effects we assume that over the time interval  $[t_-, t_+]$  the intensity of image  $u$  along  $\mathbf{d}$  is constant i.e.,  $u(\mathbf{x} + (1.0 - \Delta t) \cdot \mathbf{d}(\mathbf{x}, t), t_+) - u(\mathbf{x} - \Delta t \cdot \mathbf{d}(\mathbf{x}, t), t_-) = 0$ . Extrapolating this relationship to the observed image  $g$ , which is a transformed and noise-corrupted version of  $u$ , we model the displaced pel differences (DPDs)

$$\tilde{r}(\mathbf{d}(\mathbf{x}_i, t), \mathbf{x}_i, t, \Delta t) = \tilde{g}(\mathbf{x}_i + (1.0 - \Delta t) \cdot \mathbf{d}(\mathbf{x}_i, t), t_+) - \tilde{g}(\mathbf{x}_i - \Delta t \cdot \mathbf{d}(\mathbf{x}_i, t), t_-)$$

by independent Gaussian random variables ( $\tilde{g}(\mathbf{x}, t)$  denotes an intensity value at  $(\mathbf{x}, t) \notin \Lambda_g$  obtained by interpolation). Consequently, the likelihood  $P(G_{t_+} = g_{t_+} | \mathbf{d}_t, l_t, g_{t_-})$  from (1) can be expressed as the following Gaussian distribution<sup>†</sup>

$$P(G_{t_+} = g_{t_+} | \hat{\mathbf{d}}_t, g_{t_-}) = (2\pi\sigma^2)^{-M_{\mathbf{d}}/2} \cdot e^{-U_g(g_{t_+} | \hat{\mathbf{d}}_t, g_{t_-})/2\sigma^2}, \quad (2)$$

<sup>†</sup> Note that  $\mathbf{d}_t$  constitutes a complete description of motion and a line field  $l_t$  is only an aid in estimation of  $\mathbf{d}_t$ . Hence, the conditioning on  $L_t$  in  $P(G_{t_+} = g_{t_+} | \hat{\mathbf{d}}_t, l_t, g_{t_-})$  can be dropped.

with energy  $U_g$  defined as follows

$$U_g(g_{t_+}|\hat{\mathbf{d}}_t, g_{t_-}) = \sum_{i=1}^{M_d} [\tilde{\tau}(\hat{\mathbf{d}}(\mathbf{x}_i, t), \mathbf{x}_i, t, \Delta t)]^2. \quad (3)$$

### 2.2.2 Displacement field model

Since motion fields are smooth functions of spatial position  $\mathbf{x}$  (fixed  $t$ ) except for occasional abrupt changes in vector length and/or orientation, we will model displacement fields  $\mathbf{d}_t$  and displacement discontinuities  $l_t$  by vector and binary MRFs  $(\mathbf{D}_t, L_t)$  [2],[4],[6].

Recall that in (1) the *a priori* displacement model is expressed by the probability (density)  $P(\mathbf{D}_t = \mathbf{d}_t | l_t, g_{t_-})$ . Since the discontinuity model expressed by  $P(L_t = l_t | g_{t_-})$  depends on the data  $g_{t_-}$ , we assume that  $\mathbf{D}_t$  can be described by the Gibbs distribution:

$$P(\mathbf{D}_t = \mathbf{d}_t | l_t, g_{t_-}) = P(\mathbf{D}_t = \mathbf{d}_t | l_t) = \frac{1}{Z_d} e^{-U_d(\mathbf{d}_t | l_t) / \beta_d}, \quad (4)$$

where  $Z_d, \beta_d$  are constants and  $U_d(\mathbf{d}_t | l_t)$  is an energy function defined as:

$$U_d(\mathbf{d}_t | l_t) = \sum_{c_d = \{\mathbf{x}_i, \mathbf{x}_j\} \in \mathcal{C}_d} V_d(\mathbf{d}_t, c_d) \cdot [1 - l(\langle \mathbf{x}_i, \mathbf{x}_j \rangle, t)]. \quad (5)$$

$c_d$  is a clique of vectors, while  $\mathcal{C}_d$  is a set of all such cliques defined over lattice  $\Lambda_d$ .  $(\langle \mathbf{x}_i, \mathbf{x}_j \rangle, t) \in \Psi_l$  denotes a site of line element located between vector sites  $\mathbf{x}_i$  and  $\mathbf{x}_j$  which belong to  $\Lambda_d$ .  $V_d$  is a *potential function* crucial to characterization of the properties of displacement field  $\mathbf{d}_t$ .

We specify the *a priori* displacement model by using  $\|\mathbf{d}(\mathbf{x}_i, t) - \mathbf{d}(\mathbf{x}_j, t)\|^2$  as the potential function  $V_d$  for each clique  $c_d = \{\mathbf{x}_i, \mathbf{x}_j\}$ , as well as the first-order neighbourhood system  $\mathcal{N}_d^1$  with 2-element horizontal and vertical vector cliques [2].

### 2.2.3 Line field model

Let the line field model be based on a binary MRF  $L_t$  with the Gibbs probability distribution

$$P(L_t = l_t | g_{t_-}) = \frac{1}{Z_l} e^{-U_l(l_t | g_{t_-}) / \beta_l}, \quad U_l(l_t | g_{t_-}) = \sum_{c_l \in \mathcal{C}_l} V_l(l_t, g_{t_-}, c_l), \quad (6)$$

where  $Z_l, \beta_l$  are the usual constants,  $c_l$  is a line clique and  $\mathcal{C}_l$  is a set of all line cliques defined over  $\Psi_l$ . The line potential function  $V_l$  provides a penalty associated with introduction of a line element. Separate neighbourhood systems are associated with cosets  $\psi_h$  and  $\psi_v$  [4]. To model the smoothness and continuity of motion boundaries as well as to disallow formation of isolated displacement vectors inconsistent with their neighbours we chose the potential  $V_{l_4}$  defined over four-element cliques [4]. We also used potential  $V_{l_2}$  for two-element cliques to prevent formation of double contours.

Since the *a priori* probability of the line process (6) is conditioned on the observations, the image information  $g_{t_-}$  should be considered when computing the line samples  $l_t$ . Similarly to Hutchison *et al.* [7] we assume that an introduction of a motion boundary coincides with an intensity edge. We use the following potential function for one-element cliques:

$$V_{l_1}(l_t, g_{t_-}, c_l) = \begin{cases} \frac{\alpha}{(\nabla_v g_{t_-})^2} \cdot l_h(\langle \mathbf{x}_i, \mathbf{x}_j \rangle, t) & \text{for horizontal } c_l = \{\mathbf{x}_i, \mathbf{x}_j\} \\ \frac{\alpha}{(\nabla_h g_{t_-})^2} \cdot l_v(\langle \mathbf{x}_i, \mathbf{x}_j \rangle, t) & \text{for vertical } c_l = \{\mathbf{x}_i, \mathbf{x}_j\}, \end{cases} \quad (7)$$

where  $l_h, l_v$  are horizontal and vertical line elements,  $\nabla_h, \nabla_v$  are horizontal and vertical components of the spatial gradient at  $(\langle \mathbf{x}_i, \mathbf{x}_j \rangle, t)$ , and  $\alpha$  is a non-negative constant. The total line potential function  $V_l(l_t, g_{t_-}, c_l)$  is simply a sum of  $V_{l_4}$ ,  $V_{l_2}$  and  $V_{l_1}$ .

### 2.3 A posteriori probability

Combining (2), (4) and (6) it follows that probability (1) is Gibbsian with energy function:

$$U(\hat{\mathbf{d}}_t, \hat{l}_t, g_{t-}, g_{t+}) = \lambda_g \cdot U_g(g_{t+} | \hat{\mathbf{d}}_t, g_{t-}) + \lambda_d \cdot U_d(\hat{\mathbf{d}}_t | \hat{l}_t) + \lambda_l \cdot U_l(\hat{l}_t | g_{t-}). \quad (8)$$

The conditional energies are defined in (3), (5) and (6) respectively, and  $\lambda_g = 1/(2\sigma^2)$ ,  $\lambda_d = 1/\beta_d$ ,  $\lambda_l = 1/\beta_l$ . The MAP estimation can be achieved by minimization of energy (8) with respect to  $(\hat{\mathbf{d}}_t, \hat{l}_t)$ . Note that the minimized energy consists of three terms and can be viewed as regularization:  $U_g$  describes the ill-posed matching problem of the data  $g_{t-}, g_{t+}$  by the motion field  $\hat{\mathbf{d}}$ , while  $U_d$  and  $U_l$  are responsible for conforming to the properties of the *a priori* displacement and line models.

## 3. SOLUTION TO MAP ESTIMATION

The minimization of energy (8) is very complex because of the number of unknowns involved and because of multimodality of the objective function (dependence on  $\hat{\mathbf{d}}_t$  via  $g$ ). We will carry out the MAP estimation using *simulated annealing* and appropriate deterministic approximations.

### 3.1 Stochastic optimization via simulated annealing

To implement the MAP estimation using simulated annealing [8], samples from MRFs  $\mathbf{D}_t$  and  $L_t$  are needed as well as an annealing schedule to control temperature  $\mathbf{T}$ . We will generate such samples using the *Gibbs sampler* [9] which, like any stochastic relaxation algorithm, produces states according to probabilities of their occurrence i.e., the unlikely states are also generated (however less frequently). This property, incorporated into simulated annealing, allows the algorithm to escape local minima unlike the case of standard methods.

We will use the Gibbs sampler based on the *a posteriori* probability (1) with energy (8). The displacement Gibbs sampler at location  $(\mathbf{x}_i, t)$  is driven by a (Gibbs) marginal conditional probability characterized by the following energy function [4]

$$U_d^i(\hat{\mathbf{d}}(\mathbf{x}_i, t) | \hat{\mathbf{d}}_t^c, \hat{l}_t, g_{t-}, g_{t+}) = \lambda_g \cdot [\tilde{r}(\hat{\mathbf{d}}(\mathbf{x}_i, t), \mathbf{x}_i, t, \Delta t)]^2 + \lambda_d \cdot \sum_{j: \mathbf{x}_j \in \eta_d(\mathbf{x}_i)} \|\hat{\mathbf{d}}(\mathbf{x}_i, t) - \hat{\mathbf{d}}(\mathbf{x}_j, t)\|^2 \cdot [1 - \hat{l}(\langle \mathbf{x}_i, \mathbf{x}_j \rangle, t)], \quad (9)$$

where  $\hat{\mathbf{d}}_t^c = \{\hat{\mathbf{d}}(\mathbf{x}_j, t) : j \neq i\}$  and  $\eta_d(\mathbf{x}_i)$  is a spatial neighbourhood of displacement vector at  $\mathbf{x}_i$ . Corresponding local energy function  $U_l^i$  driving the Gibbs sampler for  $l_t$  can be found in [4].

#### 3.1.1 Discrete state space Gibbs sampler

For each candidate vector  $\hat{\mathbf{d}}(\mathbf{x}_i, t) \in \mathcal{S}_d^i$ , the marginal probability distribution is computed from the local energy (9). Then, two vector coordinates are sampled from this distribution. The necessity to obtain the complete probability distribution of a displacement vector at each  $\mathbf{x}_i$  is decisive in the computational complexity of the discrete state space Gibbs sampler. A similar procedure applies to line elements, except that the state space  $\mathcal{S}_l^i$  is binary.

#### 3.1.2 Continuous state space Gibbs sampler for $\mathbf{D}_t$

We avoid a very fine quantization of  $\mathcal{S}_d^i$  (to obtain the continuous state space) by approximating the local energy (9) by a quadratic form in  $\hat{\mathbf{d}}_t$  so that the Gibbs sampler is driven by a Gaussian probability distribution.

Assume that an approximate estimate  $\hat{\mathbf{d}}_t$  of the true displacement field is known, and that the image intensity is locally approximately linear. Then, using the first-order terms of the Taylor expansion the DPD  $\tilde{\mathbf{r}}$  can be expressed as follows:

$$\tilde{\mathbf{r}}(\hat{\mathbf{d}}(\mathbf{x}_i, t), \mathbf{x}_i, t, \Delta t) \approx \tilde{\mathbf{r}}(\dot{\mathbf{d}}(\mathbf{x}_i, t), \mathbf{x}_i, t, \Delta t) + (\hat{\mathbf{d}}(\mathbf{x}_i, t) - \dot{\mathbf{d}}(\mathbf{x}_i, t)) \cdot \nabla_{\mathbf{d}} \tilde{\mathbf{r}}(\dot{\mathbf{d}}(\mathbf{x}_i, t), \mathbf{x}_i, t, \Delta t),$$

where the spatial gradient of  $\tilde{\mathbf{r}}$  is defined as

$$\nabla_{\mathbf{d}} \tilde{\mathbf{r}}(\dot{\mathbf{d}}(\mathbf{x}_i, t), \mathbf{x}_i, t, \Delta t) = \begin{bmatrix} \tilde{\mathbf{r}}^x(\dot{\mathbf{d}}(\mathbf{x}_i, t), \mathbf{x}_i, t, \Delta t) \\ \tilde{\mathbf{r}}^y(\dot{\mathbf{d}}(\mathbf{x}_i, t), \mathbf{x}_i, t, \Delta t) \end{bmatrix}. \quad (10)$$

$\tilde{\mathbf{r}}^x$  and  $\tilde{\mathbf{r}}^y$  are computed as an average of appropriate derivatives at the end points of vector  $\dot{\mathbf{d}}(\mathbf{x}_i, t)$  [10]. Including the temperature  $\mathbf{T}$  the local energy  $U_{\mathbf{d}}^i$  can be written as follows

$$U_{\mathbf{d}}^i(\hat{\mathbf{d}}(\mathbf{x}_i, t) | \hat{\mathbf{d}}_t^c, \hat{t}_t, g_{t-}, g_{t+}) \approx \frac{\lambda_g}{\mathbf{T}} \cdot [\tilde{\mathbf{r}}(\dot{\mathbf{d}}(\mathbf{x}_i, t), \mathbf{x}_i, t, \Delta t) + (\hat{\mathbf{d}}(\mathbf{x}_i, t) - \dot{\mathbf{d}}(\mathbf{x}_i, t)) \cdot \nabla_{\mathbf{d}} \tilde{\mathbf{r}}(\dot{\mathbf{d}}(\mathbf{x}_i, t), \mathbf{x}_i, t, \Delta t)]^2 + \frac{\lambda_d}{\mathbf{T}} \cdot \sum_{j: \mathbf{x}_j \in \eta_d(\mathbf{x}_i)} \|\hat{\mathbf{d}}(\mathbf{x}_j, t) - \hat{\mathbf{d}}(\mathbf{x}_i, t)\|^2 \cdot [1 - \hat{l}(\langle \mathbf{x}_i, \mathbf{x}_j \rangle, t)],$$

where  $\dot{\mathbf{d}}$  is fixed. It can be shown that the conditional probability density with the above energy is a 2-D Gaussian with the following mean vector  $\mathbf{m}$  at location  $(\mathbf{x}_i, t)$  [10]:

$$\mathbf{m} = \bar{\mathbf{d}}(\mathbf{x}_i, t) - \frac{\varepsilon_i}{\mu_i} \nabla_{\mathbf{d}}^T \tilde{\mathbf{r}}(\dot{\mathbf{d}}(\mathbf{x}_i, t), \mathbf{x}_i, t, \Delta t),$$

where the scalars  $\varepsilon_i$  and  $\mu_i$  are defined as follows

$$\begin{aligned} \varepsilon_i &= \tilde{\mathbf{r}}(\dot{\mathbf{d}}(\mathbf{x}_i, t), \mathbf{x}_i, t, \Delta t) + (\bar{\mathbf{d}}(\mathbf{x}_i, t) - \dot{\mathbf{d}}(\mathbf{x}_i, t)) \cdot \nabla_{\mathbf{d}} \tilde{\mathbf{r}}(\dot{\mathbf{d}}(\mathbf{x}_i, t), \mathbf{x}_i, t, \Delta t) \\ \mu_i &= \xi_i \frac{\lambda_d}{\lambda_g} + \|\nabla_{\mathbf{d}} \tilde{\mathbf{r}}(\dot{\mathbf{d}}(\mathbf{x}_i, t), \mathbf{x}_i, t, \Delta t)\|^2, \end{aligned} \quad (11)$$

and  $\bar{\mathbf{d}}(\mathbf{x}_i, t)$  is an average vector

$$\bar{\mathbf{d}}(\mathbf{x}_i, t) = \frac{1}{\xi_i} \sum_{j: \mathbf{x}_j \in \eta_d(\mathbf{x}_i)} \hat{\mathbf{d}}(\mathbf{x}_j, t) \cdot [1 - \hat{l}(\langle \mathbf{x}_i, \mathbf{x}_j \rangle, t)], \quad (12)$$

with  $\xi_i = \sum_{j: \mathbf{x}_j \in \eta_d(\mathbf{x}_i)} [1 - \hat{l}(\langle \mathbf{x}_i, \mathbf{x}_j \rangle, t)]$ . Note that averaging is disallowed across a motion boundary, which is a desirable property. The horizontal and vertical component variances  $\sigma_x^2$ ,  $\sigma_y^2$ , as well as the correlation coefficient  $\rho$ , which comprise the covariance matrix, have the following form

$$\begin{aligned} \begin{bmatrix} \sigma_x^2 \\ \sigma_y^2 \end{bmatrix} &= \frac{\mathbf{T}}{2\xi_i \lambda_d \mu_i} \begin{bmatrix} \xi_i \frac{\lambda_d}{\lambda_g} + [\tilde{\mathbf{r}}^y(\dot{\mathbf{d}}(\mathbf{x}_i, t), \mathbf{x}_i, t, \Delta t)]^2 \\ \xi_i \frac{\lambda_d}{\lambda_g} + [\tilde{\mathbf{r}}^x(\dot{\mathbf{d}}(\mathbf{x}_i, t), \mathbf{x}_i, t, \Delta t)]^2 \end{bmatrix} \\ \rho_{\sigma_x \sigma_y} &= \frac{-\tilde{\mathbf{r}}^x(\dot{\mathbf{d}}(\mathbf{x}_i, t), \mathbf{x}_i, t, \Delta t) \tilde{\mathbf{r}}^y(\dot{\mathbf{d}}(\mathbf{x}_i, t), \mathbf{x}_i, t, \Delta t)}{2\xi_i \lambda_d \mu_i}. \end{aligned}$$

The initial vector  $\mathbf{d}$  can be assumed zero throughout the estimation process, but then with increasing displacement vector estimates the error due to intensity non-linearity would significantly increase. Hence, it is better to "track" an intensity pattern by modifying  $\dot{\mathbf{d}}$  accordingly. An interesting result can be obtained when it is assumed that at every iteration of the Gibbs sampler  $\dot{\mathbf{d}} = \bar{\mathbf{d}}$  i.e., the initial (approximate) displacement field is equal to the average from the previous iteration. Then, the estimation process can be described by the following iterative equation:

$$\hat{\mathbf{d}}^{n+1}(\mathbf{x}_i, t) = \bar{\mathbf{d}}^n(\mathbf{x}_i, t) - \frac{\varepsilon_i}{\mu_i} \nabla_{\mathbf{d}}^T \tilde{\mathbf{r}}(\bar{\mathbf{d}}^n(\mathbf{x}_i, t), \mathbf{x}_i, t, \Delta t) + \mathbf{n}_i, \quad (13)$$

where  $n$  is the iteration number, and  $\varepsilon_i$ ,  $\mu_i$  and the covariance matrix are defined as before except for  $\dot{\mathbf{d}} = \bar{\mathbf{d}}$ . At the beginning, when the temperature is high, the random term  $\mathbf{n}_i$  has a large variance and

the estimates assume quite random values. As the temperature  $T \rightarrow 0$ ,  $\sigma_x^2, \sigma_y^2, \rho\sigma_x\sigma_y$  get smaller, thus reducing  $n_i$ . In the limit the algorithm performs a deterministic update. Note that the variance  $\sigma_x^2$  of the horizontal vector component for fixed values of  $\lambda_g, \lambda_d, \lambda_l$  and  $\tilde{r}$  decreases with growing  $\tilde{r}^x$ . It means that when there is a significant horizontal gradient (detail) in the image structure the uncertainty of the estimate in horizontal direction is small. The same applies to  $\sigma_y^2$ . Hence, the algorithm takes into account the image structure in determining the amount of randomness allowed at a given temperature.

Note the similarity of the iterative update equation (13) to the update equation of the Horn-Schunck algorithm [11]. Except for  $\varepsilon_i$  equal to DPD instead of the motion constraint equation and except for different image model used, they are identical for  $T=0$ . It is interesting that similar update equations result from two different approaches: Horn and Schunck establish necessary conditions for optimality and solve them by deterministic relaxation, while here a 2-D Gaussian distribution is fitted into the conditional probability driving the Gibbs sampler.

## 3.2 Deterministic optimization using steepest descent method

### 3.2.1 Discrete state space

Note that for  $T=0$  (deterministic update) the discrete state space Gibbs sampler generates only states with minimum local energy (9). Hence, the final result is only an approximation to the MAP estimate.

Besag [12] proposed a similar approach called *iterated conditional modes* (ICM). He argued that since it is difficult to maximize the joint *a posteriori* probability over the complete field, it should be divided into a minimal number of disjoint sets (or colours) such that any two random variables from a given set are conditionally independent given the states of the other sets. Using this approach displacement vectors or line elements can be computed individually (e.g., exhaustive search) for each location  $(\mathbf{x}_i, t)$  one colour at a time. Note that also this technique does not result in maximization of probability (1), but provides separate MAP estimates for joint probabilities defined over corresponding colours. The difference between the ICM method and the Gibbs sampler with  $T=0$  is only the update order of variables [10]. Both techniques can be classified as a (pel) matching algorithm with smoothness constraint.

### 3.2.2 Continuous state space

Let the displacement vector state space be continuous ( $\mathcal{S}'_d = R^2$ ). The energy function under minimization (8) is non-quadratic in  $\hat{\mathbf{d}}_t$  as well as in  $\hat{l}_t$ . We will perform interleaved optimization with respect to  $\hat{\mathbf{d}}_t$  and  $\hat{l}_t$ . If  $\hat{l}_t$  is known, then  $U_l$  in (8) is constant and only minimization of  $\lambda_g U_g + \lambda_d U_d$  must be performed. Using the linearization of the DPD  $\tilde{r}$  and establishing necessary conditions for optimality at each location  $\mathbf{x}_i$  as well as assuming that  $\dot{\mathbf{d}} = \bar{\mathbf{d}}$ , it follows that the iterative update for this deterministic method is [10]:

$$\hat{\mathbf{d}}^{n+1}(\mathbf{x}_i, t) = \bar{\mathbf{d}}^n(\mathbf{x}_i, t) - \frac{\varepsilon_i}{\mu_i} \nabla_{\mathbf{d}}^T \tilde{r}(\bar{\mathbf{d}}^n(\mathbf{x}_i, t), \mathbf{x}_i, t, \Delta t), \quad (14)$$

where  $\varepsilon_i$  and  $\mu_i$  are defined in (11) with  $\dot{\mathbf{d}} = \bar{\mathbf{d}}$ . To resemble the Gibbs sampler as close by as possible, the Gauss-Seidel relaxation will be used in (14) rather than the Jacobi relaxation. Once an estimate  $\hat{\mathbf{d}}_t$  is known, an improved estimate  $\hat{l}_t$  should be obtained. For a fixed  $\hat{\mathbf{d}}_t$  the minimized energy  $\lambda_d U_d + \lambda_l U_l$  is non-linear in  $\hat{l}_t$ . Since  $l(\mathbf{x}_i, t)$  is binary for each  $i$ , the ICM method reported above can be used.

The above approximation to the continuous state space MAP estimation is a spatio-temporal gradient technique which can be viewed as a modified version of the Horn and Schunck algorithm, except that:

1. the modified algorithm (14) allows computation of displacement vectors for arbitrary  $\Lambda_d$  unlike the original Horn and Schunck algorithm in which  $\Lambda_d = \Lambda_g + [0.5, 0.5, 0.5]^T$ ,
2. the scalar  $\varepsilon_i$  is a displaced pel difference in the modified version rather than a motion constraint equation: no temporal derivative is needed,
3. the spatial intensity derivatives are computed from a separable polynomial model in both images and appropriately weighted (10), instead of the finite difference approximation over a cube as proposed in [11].

The ability to estimate motion for arbitrary  $\Lambda_d$  is crucial for motion-compensated interpolation of sequences (original Horn-Schunck algorithm would require 3-D interpolation of motion fields).

The use of  $\tilde{r}$  instead of the motion constraint equation in  $\varepsilon_i$  is important because it allows intensity pattern tracking thus permitting more accurate intensity derivative computation, and also does not require the computation of the purely temporal derivative (actually,  $\tilde{r}$  is an approximation to the directional derivative). The purely temporal derivative used in the Horn-Schunck algorithm is a reliable measure of temporal intensity change due to motion as long as small displacements are applied to linearly varying intensity pattern. Otherwise, significant errors may result, for example an overestimation at moving edges of high contrast.

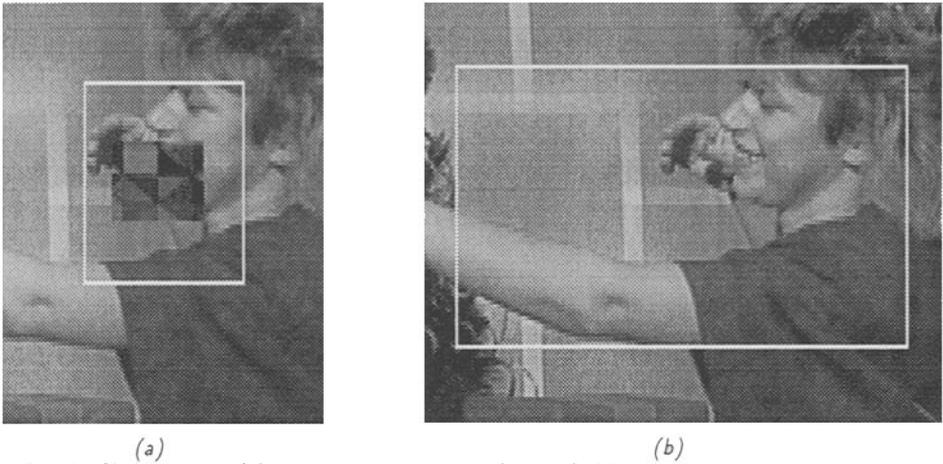
The deterministic algorithm (14) together with the ICM method for  $l_t$  is related to the algorithm proposed in [7]. The major differences are those reported above for the Horn-Schunck algorithm as well as the line potentials: the potential  $V_{l_1}$  for single-element cliques is binary (0 or  $\infty$ ) in [7] while here it varies continuously according to the local intensity gradient.

#### 4. EXPERIMENTAL RESULTS

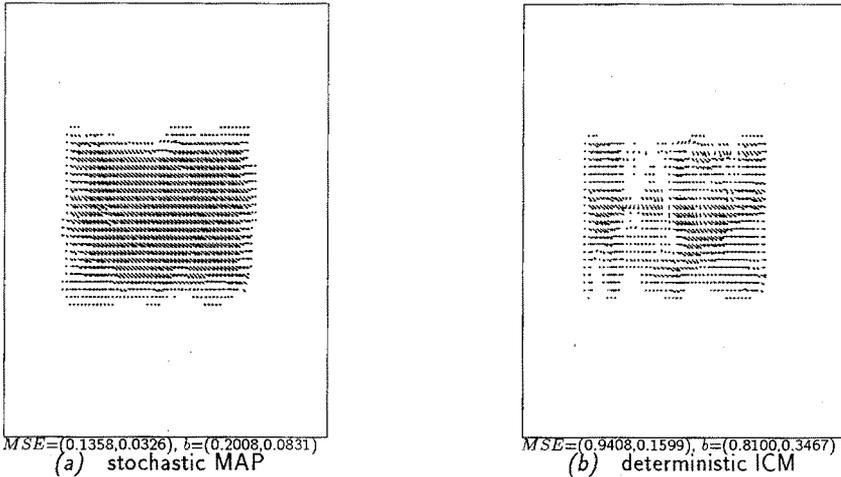
The algorithms described above have been tested on a number of images with synthetic and natural motion. Results for two of images with natural data and inter-field distance  $\tau_{60}=1/60$  sec. are presented below.

To provide a quantitative test we generated test image 1 (Fig. 1(a)) with stationary background provided by the test image from Fig. 1(b) and a moving rectangle  $\mathcal{R}$  obtained from another image through low-pass filtering, subsampling and pixel shifting. This test pattern permits non-integer displacements so that there is no perfect data matching. Fig. 1(b) shows the test image 2 containing natural motion, acquired by a video camera.

The stochastic relaxation used was based either on the discrete state space  $S_d^l$  with maximum displacement  $\pm 2.0$  pixels and 17 quantization levels in each direction or on the continuous state space  $R^2$ . The first-order displacement neighbourhood system and, if applicable, the line neighbourhood with four-, two- and one-element line cliques, as proposed in [4], have been used. The ratio  $\lambda_d/\lambda_g=20.0$  has been chosen experimentally, however, as pointed out in [10], even a change of 2 orders magnitude did not have an excessively severe impact on the estimate quality. The motion estimates presented in the sequel have been obtained from pairs of images (fields) separated by  $T_g = 2\tau_{60}$ . All estimates have been obtained with Keys bicubic interpolator [10] except for the discrete state space estimation applied to test image 2, when bilinear interpolation was used.



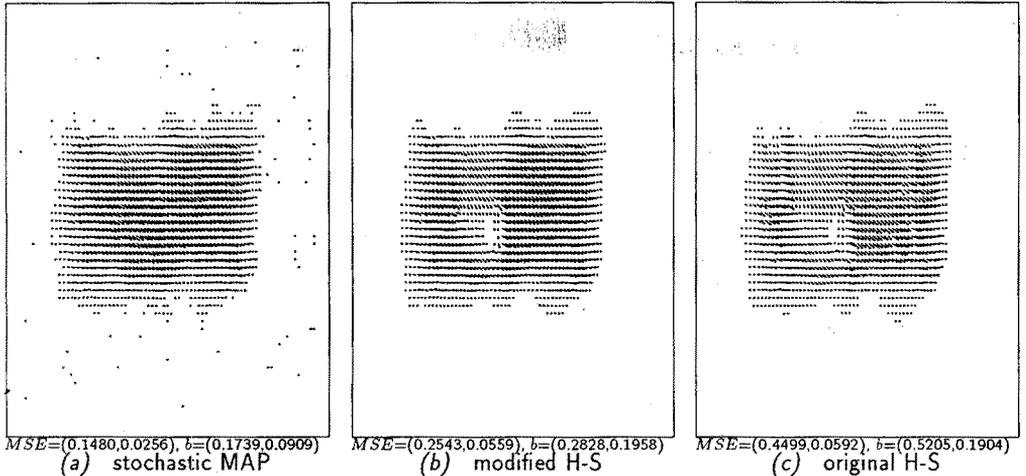
**Fig. 1** Test images: (a) synthetic motion,  $d_s=[1.5,0.5]$ ,  $|\mathcal{R}| = 45 \times 20$ , (b) natural motion (white frame outlines the area used in estimation: (a)  $77 \times 49$ , (b)  $221 \times 69$ ).



**Fig. 2** Discrete state space MAP and ICM estimates for the test image 1:  $\Lambda_d = \Lambda_g$ .

Since the true motion field is known for the test image with synthetic motion (except for the occlusion and newly exposed areas), it is possible to assess the quality of motion field estimates. The Mean Squared Error ( $MSE$ ) and the bias ( $b$ ) measuring the departure of estimate  $\hat{d}$  from the known motion field  $d_s$ , are computed within the rectangle and shown below appropriate estimates.

Fig. 2 shows the discrete state space MAP and ICM displacement estimates from the test image 1. The stochastic MAP estimate is superior to the ICM estimate both subjectively and objectively ( $MSE, b$ ). In both cases the zero displacement field has been used as an initial state. In other experiments ML estimates ( $\lambda_d/\lambda_g=0.0$ ) have been computed and used as a starting point (as suggested by Besag for ICM estimation). The ML estimates were characterized by substantial randomness in vector lengths and orientations, which can be explained by the lack of a prior model. As expected the initial state had no impact on the stochastic MAP estimate, but the final ICM estimate was inferior to the ICM estimate presented above both subjectively and in terms of  $MSE$ .

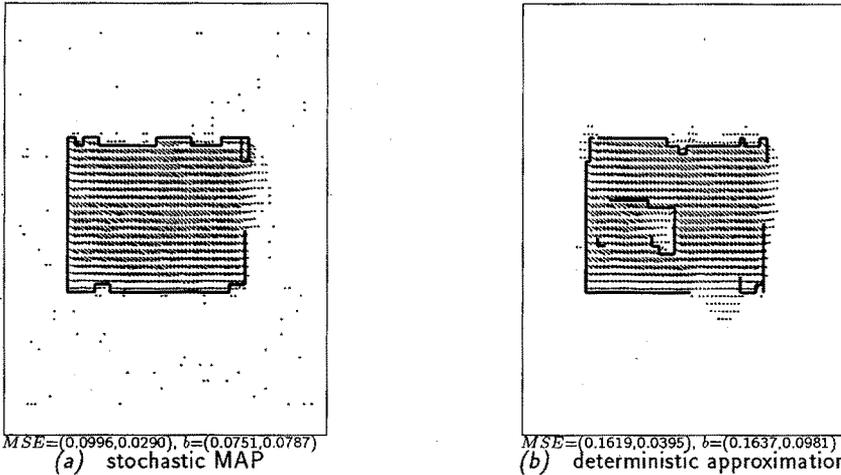


**Fig. 3** Continuous state space MAP, modified and original Horn-Schunck estimates for test image 1:  $\Lambda_d = \Lambda_g + [0.5, 0.5, 0.5]^T$ .

To compare the stochastic MAP estimate and its deterministic approximation (modified Horn-Schunck method) with the original Horn-Schunck algorithm (Fig. 3), the condition that  $\Lambda_d = \Lambda_g + [0.5, 0.5, 0.5]^T$  was imposed. Note that the Horn-Schunck algorithm produces the worst result, both subjectively and in terms of *MSE*. The motion tends to be overestimated at strong edges (due to the purely temporal gradient), while it is underestimated in uniform areas. The deterministic approximation has produced a significantly lower *MSE*, and also subjectively the estimate is more uniform. Except for the visible triangle of underestimated displacements, the motion has been quite well computed. Superiority of the stochastic approach is clear from Fig. 3.a. Subjectively this estimate is closest to the true motion, *MSE* is the lowest of the three estimates and also the total energy is lower than for the deterministic approximation (original Horn-Schunck algorithm cannot be compared in terms of energy since it assumes different intensity model).

Fig. 4 shows the stochastic and deterministic estimates for the piecewise smooth motion model. The parameters used are the same as before. During experimentation we have observed that the ratio  $\lambda_I/\lambda_d$  had to be substantially lower for the deterministic algorithm in order to obtain results comparable with the stochastic MAP estimation. This may be explained by explicit averaging used in the deterministic algorithm. The continuous state space MAP estimation uses similar averaging, but it also involves a randomness factor thus allowing switching line elements off and on, even if motion discontinuity does not quite allow it. Note that both subjectively and in terms of *MSE* the deterministic estimate is clearly inferior.

Fig. 5 shows the discrete state space MAP and ICM displacement estimates for the test image 2. The ICM estimate is again subjectively poorer than the stochastic MAP estimate. The ICM algorithm failed to compute correctly the motion of the forearm and of the arm, except for the displacement vectors along the edge of the shirt sleeve. Also the vectors on the neck and parts of the face suggest that there is no motion, which is incorrect.



**Fig. 4** Continuous state space stochastic and deterministic MAP estimates with piecewise smooth motion model for test image 1:  $\Lambda_d = \Lambda_g$ ,  $\lambda_l/\lambda_d = 0.8$  (a) and 0.15 (b),  $\alpha = 10.0$ .

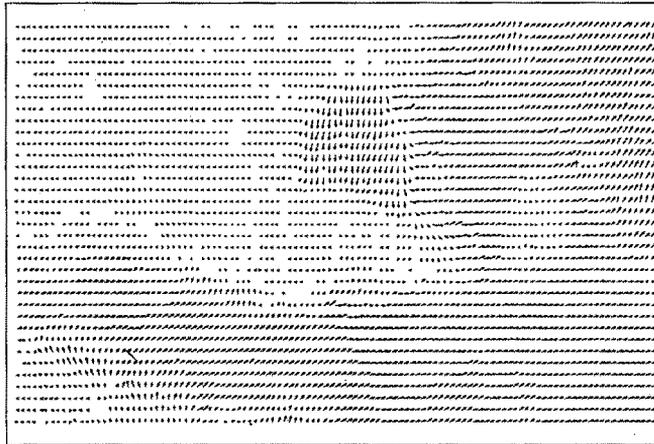
Similarly, the three continuous state space methods have been applied to the test image 2 (Fig. 6). The original Horn-Schunck estimate shows some overestimated vectors (edge of shirt sleeve) and numerous underestimated ones (uniform area to the right). The deterministic approximation performs better: it is more uniform and has smaller edge effects. The stochastic estimate, however, is superior in terms of the total energy  $U$  as well as motion field smoothness and lack of edge effects.

## 5. CONCLUSION

In this paper two types of solution methods to the problem of 2-D motion estimation via the MAP criterion have been presented and compared: stochastic and deterministic. It has been demonstrated that, as an example of instantaneous freezing, the deterministic methods may be incapable of localizing the global minimum not only theoretically but also in practice. Higher value of the energy function for the deterministic solutions was confirmed by inferior subjective and objective (synthetic motion) quality. Such an improvement in estimate quality comes at a cost of increased computational effort, however. The computational overhead (per iteration) of the continuous state space stochastic estimation compared to its deterministic approximation is small (less than 25%) because it includes only the computation of the random update term. The number of iterations required to provide sufficiently slow annealing schedule, however, makes the stochastic method more involved computationally by about an order of magnitude.

## REFERENCES

- [1] D.W. Murray and B.F. Buxton, "Scene segmentation from visual motion using global optimization," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. PAMI-9, pp. 220-228, March 1987.
- [2] J. Konrad and E. Dubois, "Estimation of image motion fields: Bayesian formulation and stochastic solution," in *Proc. IEEE Int. Conf. on Acoust., Speech, and Signal Process. ICASSP'88*, 1988, pp. 1072-1074.
- [3] J. Konrad and E. Dubois, "Multigrid Bayesian estimation of image motion fields using stochastic relaxation," in *Proc. IEEE Int. Conf. Computer Vision ICCV'88*, 1988, pp. 354-362.
- [4] J. Konrad and E. Dubois, "Bayesian estimation of discontinuous motion in images using simulated annealing," in *Proc. Conf. Vision Interface VI'89*, 1989, pp. 51-60.



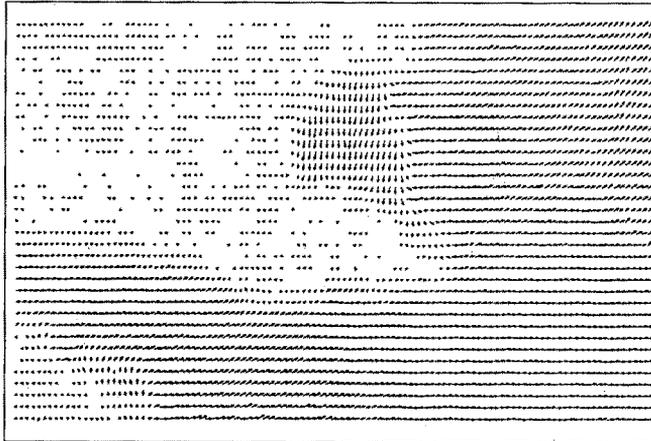
(a) stochastic MAP



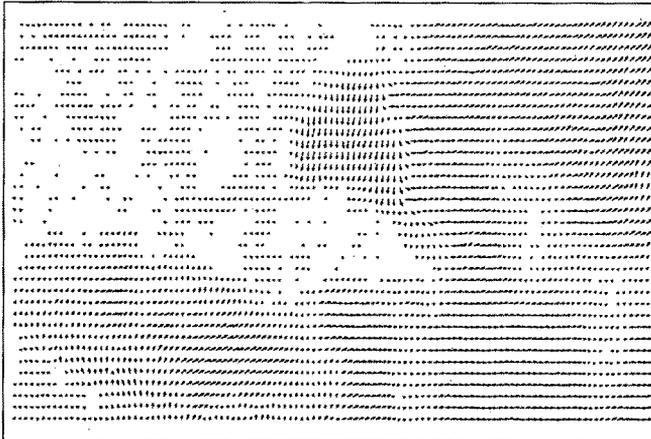
(b) deterministic ICM

**Fig. 5** Discrete state space MAP and ICM estimates for the test image 2:  $\Lambda_d = \Lambda_g$ .

- [5] P. Boutheymy and P. Lalande, "Motion detection in an image sequence using Gibbs distributions," in *Proc. IEEE Int. Conf. on Acoust., Speech, and Signal Process. ICASSP'89*, 1989, pp. 1651-1654.
- [6] F. Heitz and P. Boutheymy, "Estimation et segmentation du mouvement: approche bayésienne et modelisation markovienne des occlusions," in *Proc. of 7-th Congress AFCET - RFIA*, 1989.
- [7] J. Hutchinson, Ch. Koch, J. Luo and C. Mead, "Computing motion using analog and binary resistive networks," *Computer*, vol. 21, pp. 52-63, March 1988.
- [8] S. Kirkpatrick, C.D. Gelatt, Jr. and M.P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671-680, May 1983.
- [9] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. PAMI-6, pp. 721-741, November 1984.
- [10] J. Konrad, "Bayesian estimation of motion fields from image sequences," Ph.D. Thesis, McGill University, Dept. of Electr. Eng., 1989.
- [11] B.K.P. Horn and B.G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185-203, 1981.
- [12] J. Besag, "On the statistical analysis of dirty pictures," *J. R. Statist. Soc.*, vol. 48, B, pp. 259-279, 1986.



(a) stochastic MAP



(b) modified H-S



(c) original H-S

**Fig. 6** Continuous state space MAP, modified and original Horn-Schunck estimates for test image 2:  $\Lambda_d = \Lambda_g + [0.5, 0.5, 0.5]^T$ .