

## BAYESIAN DENSE MOTION FIELD ESTIMATION WITH LANDMARK CONSTRAINT

Yi Chin and Chun-Jen Tsai<sup>1+</sup>

Department of Computer Science

National Chiao Tung University, Hsinchu, Taiwan, R.O.C.

<sup>+</sup>[cjtsai@cs.nctu.edu.tw](mailto:cjtsai@cs.nctu.edu.tw)

### ABSTRACT

*In this paper, a dense motion field estimation technique based on the Bayesian framework is proposed to estimate the true dense motion fields of video sequences. Previous stochastic techniques of dense motion field estimation adopts piecewise smooth motion model and use MAP estimation to find the motion field with joint minimization of motion compensation errors and maximization of motion smoothness. However, such random process does not guarantee to converge to the true motion field. In this paper, the motion of landmark points in the video sequence is introduced into the MAP estimation process to regularize the estimated motion field. Experimental results show that the proposed algorithm produces estimated motion fields which preserve piecewise smooth nature and are visually close to the true motion of the video sequences.*

### 1. INTRODUCTION

For the past decades, block-based motion estimation is one of the most popular topics for video coding [1]. However, the goal for motion estimation in video coding is to find a block-based motion fields that minimizes sum-of-absolute-differences (SAD) motion compensation errors. However, the motion field that minimizes SAD may not be the true motion field of a video sequence. With the arrivals of new applications, such as frame-rate upsampling for new LCD displays [2], intelligent content analysis or surveillance systems, and decoder-side motion compensation for distributed video coding [3], the estimation of dense true motion fields becomes more and more important. Unfortunately, it is also a well-known ill-posed high complexity problem in computational vision.

Without incorporation of high-level knowledge of the video sequences, e.g. understanding of the objects in the scene, true motion estimation from low-level features such as local intensity changes always lead to multiple possible solutions. In order to rule out possible yet wrong solutions, a general assumption that derived from true motion model is usually adopted to regularize the solution fields. One of the well-accepted characteristics is that a true motion field should be piecewise smooth [4][5].

In [6], a stochastic approach is proposed for dense motion estimation. In the paper, the dense motion field is modeled as a two-dimensional random field  $\mathbf{D}_t$  at time  $t$ , while the true motion field  $\mathbf{d}_t$  is a realization of  $\mathbf{D}_t$ . The motion fields are

segmented into piecewise smooth fields by a line process  $l_t$ . Two different estimation techniques, maximum a posteriori (MAP) probability estimation and minimum expected cost (MEC) estimation, are then used to jointly estimate both  $\mathbf{d}_t$  and  $l_t$ . Although the mathematical formulation in [6] is quite elegant, the proposed algorithm does not always produce true motion fields even within an image region with underlying piecewise smooth true motion fields. One major problem with the Bayesian framework in [6] is that the conditional probabilities of the motion fields given the observed image fields are computed using local information only. At image areas where there are few textures, smoothness constraint alone cannot regularize the ill-posed problem to produce the true motion field.

In this paper, the motion of image points with salient textures, i.e. landmark points, are incorporated into the MAP motion estimation process such that the estimate of the dense motion fields can be further regularized to close to true motion fields. Landmark points can provide semi-global information into the estimation process and produce better motion field estimates. This paper is organized as follows. In section 2, MAP estimation of dense motion fields is introduced. The landmark point selection and initial motion field estimation process is described in section 3. The proposed dense motion field estimation algorithm is presented in section 4. Experimental results are given in section 5 and some conclusions are made in section 6.

### 2. MAP ESTIMATION OF DENSE MOTION FIELD

The formulation of the two-dimensional random field model for Bayesian reconstruction of corrupted image was presented in Geman and Geman [7]. The concept can be generalized to estimate other image features, such as dense motion field. In this section, the formulation of the MAP approach used in this paper to estimate piecewise smooth motion fields based on the concept in [6] is summarized first. Later in section 3, we will propose adding a landmark-point constraint to the cost function so that the estimated motion field will resemble the true motion fields better.

The a posteriori probability of a motion field  $\mathbf{D}_t$  given the observation of luma images  $\mathbf{g}_t$  and  $\mathbf{g}_{t+1}$  at time  $t$  and  $t+1$  can be written as

$$P(\mathbf{D}_t = \hat{\mathbf{d}}_t | \mathbf{g}_t, \mathbf{g}_{t+1}) = \frac{1}{Z} e^{-U(\hat{\mathbf{d}}_t, \mathbf{g}_t, \mathbf{g}_{t+1})}, \quad (1)$$

where  $Z$  is a normalization constant for the probability distribution,  $\mathbf{D}_t$  is the random field model of the true motion,

$\hat{d}_t$  is an estimate of the true motion field, and  $U(\cdot)$  is a cost function that is derived from the likelihood model of the observed images  $\mathbf{g}_t$  and  $\mathbf{g}_{t+1}$ , and the prior knowledge that the true motion field should be piecewise continuous. Note that we do not adopt the line process field that models motion discontinuity as in [6] because it increase the parameter space and makes the optimization problem much more complex to solve.

In (1), the conditional likelihood function is modeled by motion-compensated pixel differences while the prior term is modeled by the smoothness of the motion field estimate. To be more precise, the cost function can be expressed as follows,

$$U(\cdot) = \lambda_g \cdot \sum_{i \in \text{image frame}} [\tilde{r}(d(x_i), x_i)]^2 + \lambda_d \cdot \sum_{\{x_i, x_j\} \in C_d} \|d(x_i) - d(x_j)\|^2, \quad (2)$$

where  $\lambda_g$  is the weight of the reliability of observations,  $\lambda_d$  is the weight of the smoothness of motion fields,  $\tilde{r}(\cdot)$  is the motion-compensated pixel difference function,  $C_d$  is the set of all neighborhood pairs in a Markov Random Field (MRF), and  $x_i$  is the 2D position of the image pixel  $i$ . Since the parameter space of Eq (2) is very large, it is nearly impossible to find globally optimal solution of the cost function. Therefore, simulated annealing technique with Gibbs sampler as proposed in [6] is adopted in our implementation.

The initial dense motion field for simulated annealing is computed using 16×16 block-based motion estimator at every pixel positions. The block matching window for SAD computation is weighted using a Gaussian window so that the pixels near the center have higher weights.

### 3. LANDMARK POINT SELECTION

In order to further regularize the dense motion field to the true motion fields, landmark point information is proposed in this paper to modify the cost function of the MAP estimation problem. A landmark point is defined to be the image point that has two properties. First of all, an image landmark point should have salient textures around it such that block-based motion estimation can be used to estimation the true motion at the landmark point reliably. Since block-based motion matching is used to find the true motion for landmark points, a second criterion for a landmark point is that the motions of pixels in a small (block) area around the landmark point should have the same translational motion such that the block-based motion model is not violated.

In this section, the two-step algorithm for landmark point selection is presented. The algorithm begins with the initial dense motion field obtain from block-based motion estimation as described in section 2. Typically speaking, the initial motion field has many outliers around motion discontinuities and is over-smoothed around areas with very small motions. In the first step of landmark point selection algorithm, the image gradient of the luma channel is computed using the Sobel operator. If the sum of the luma gradient within a 16×16 block centered on an initial motion vector is larger than a threshold  $\gamma$ , the motion is considered to be estimated from texture-rich area.

Such motion vectors and the associated positions are collected into a set  $L_1$ .

However, a motion vector in  $L_1$  may not be a true motion vector. The second step of the selection process filters out the motion vectors in  $L_1$  that violates the block-based translational motion model. This is achieved by performing block-based motion compensation on both gradient and luma images. Assume that  $v$  is a motion vector in  $L_1$  and its associated pixel position is  $p$ . If the pixels in a 16×16 block area centered on  $p$  follows translational motion model, the majority of the pixels should have small motion-compensated differences between the luma image pair and the gradient image pair using the vector  $v$ . The ratio  $r_1$  of pixels 16×16 block area produce small motion-compensated difference in gradient images and the ratio  $r_2$  of pixels produce small motion-compensated difference in luma images are used to estimate the reliability of the landmark motion. If both  $r_1 > 0.7$  (i.e. 70%) and  $r_2 > 0.7$ , the landmark point is considered reliable. The set of reliable landmark points is denoted as  $L_2$ .

### 4. BAYESIAN FRAMEWORK FOR TRUE DENSE MOTION ESTIMATION

In this section, we describe the method the landmark point constraint is incorporated into the cost function Eq. (2). For each landmark point  $p$  in  $L_2$ , its motion  $v$  is assigned a reliability level  $\sigma^2 = r_1 \times r_2$ . The cost term for landmark point conformance is computed as follows:

$$\left( \lambda_m \cdot e^{-\frac{|x_i - p|^2}{2\sigma^2}} \right) \cdot [d(x_i) - v]^2. \quad (3)$$

Note that the exponential term in Eq. (3) is a weighting function of the landmark cost based on the distance to the landmark point. The closer the pixel  $x_i$  to a landmark point  $p$ , the higher the weight of landmark cost of Eq. (3) is added to Eq. (2) to form total cost function. The overall framework of the proposed algorithm can be decomposed into the following three steps:

*Step 1: Create initial dense motion fields.*

*Step 2: Pick out landmark motions.*

*Step 3: Energy minimization process.*

### 5. EXPERIMENTAL RESULTS

Firstly in this section, two image sequences are visually exhibited to demonstrate the performance of the proposed algorithm. Secondly, eight standard test sequences [8] are tested and compared in a table.

Fig. 1 and Fig. 4 show the original images of RACE HORSES and TEMPETE, respectively. In each figure, the two marked areas are the target areas for illustrating the performance of the proposed algorithm. The initial motion fields (Fig. 2a, 3a, 5a, 6a) are estimated using block-based dense motion estimator as described in section 2. The  $\lambda$ -parameters of the cost function are set to  $\lambda_g = 0.1$ ,  $\lambda_d = 1$ ,  $\lambda_m = 1$  for this two sequences.

In RACE HORSES, one can easily see that block-based motion estimation typically does not perform well across motion discontinuities along object boundaries. Furthermore, for areas with low texture information,

block-based method may also produce erroneous motions. On the other hand, MAP estimation without the proposed landmark constraint tends to produce curvy smooth motion for low texture areas. In Fig. 2c and Fig. 3c, the over-smooth of motion field across object boundaries can be reduced by introducing line process model as described in [6]. However, line process will not improve the curvy behavior of MAP motion estimates within a low-texture object with consistent smooth motion.

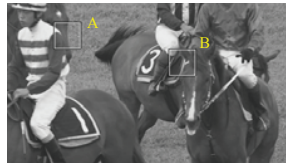


Fig. 1. RACE HORSES.

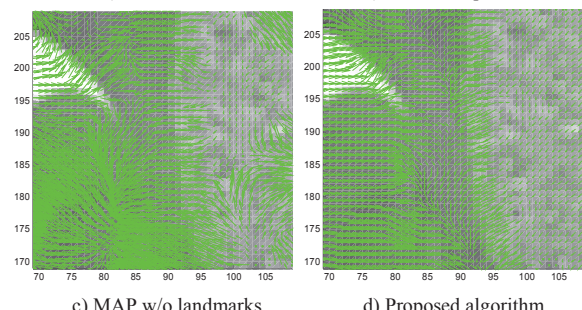
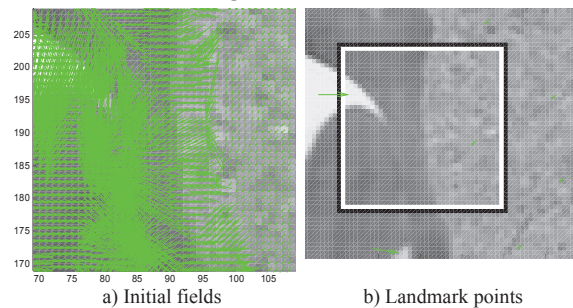


Fig. 2. Motion fields in area A of RACE HORSES.

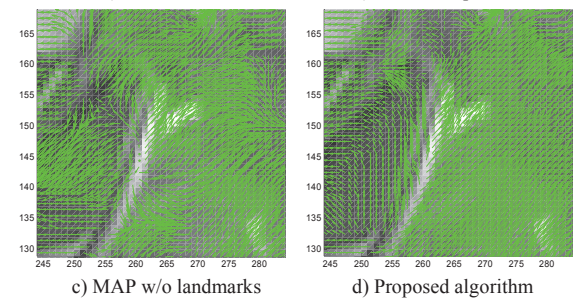
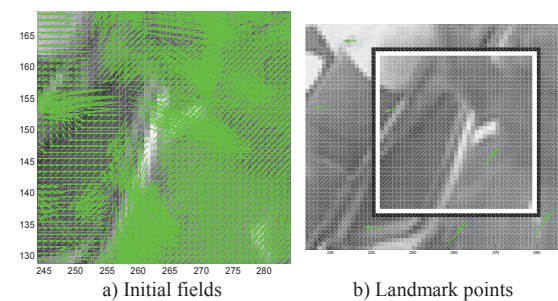


Fig. 3. Motion fields in area B of RACE HORSES.

Fig. 5 and Fig. 6 are the results from the TEMPETE sequence. In area-A, the dark area causes issues to both block-based motion estimator and MAP estimation without landmark constraint. In both cases, noises in the image sequence dominate the matching (cost) function and results in wild outliers in block-matching method or small, curvy perturbations in MAP technique. On the other hand, when landmark point constraint is introduced into the optimization process, reliable motions from nearby objects such as the fence by area-A will play the role of regulator to overcome the aperture problem of low texture area.

For area-B in the TEMPETE sequence, it seems that block-based motion estimation does a reasonable job. However, although there seems to be no large outliers from block-based motion estimation, the estimated motion field is probably not correct. Through visual inspection, the video sequence appears to have both zoom and rotational motions, while the motion fields obtained from block-based techniques produces only translational motion. One way to judge the overall quality of a dense motion estimation algorithm is to compute both the motion-compensated errors (SAD) between two frames, as well as the entropy of the motion fields, as shown in Table I.



Fig. 4. TEMPETE.

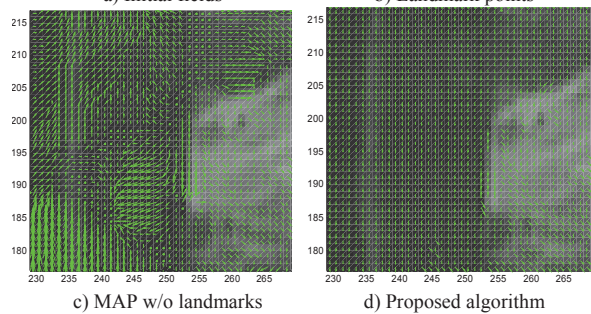
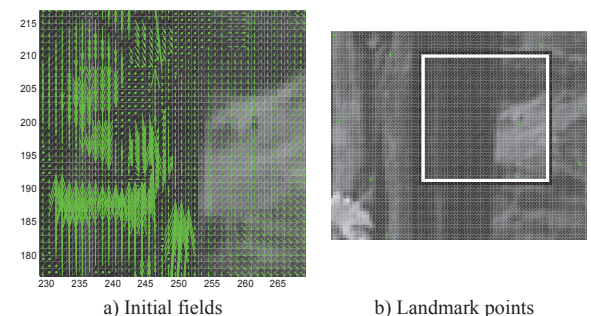


Fig. 5. Motion fields in area A of TEMPETE.

Note that the entropy is estimated by the file size of lossless compression of the dense motion fields using BZip2 (the column BZ2 in the table). As shown in

TABLE I, the proposed algorithm produces smallest SAD with low motion field entropy.

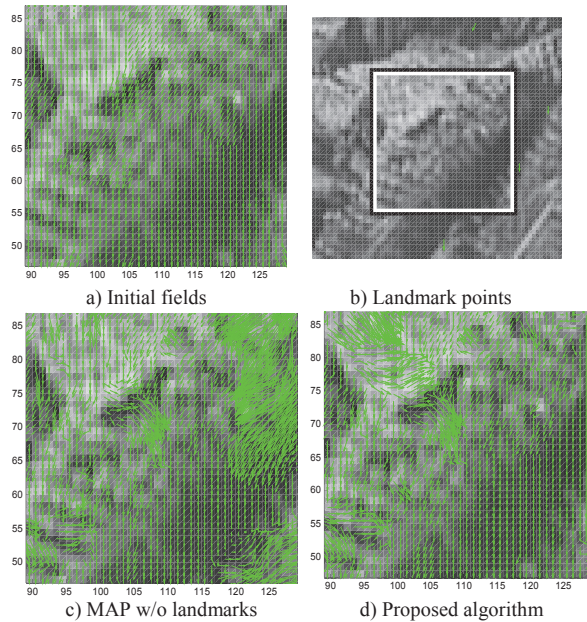


Fig. 6. Motion fields in area B of TEMPETE.

TABLE I. THE QUALITY OF THE DENSE MOTION FIELD.

	MAP w/o Landmarks		Proposed	
	SAD ( $\times 10^3$ )	BZ2 (KB)	SAD ( $\times 10^3$ )	BZ2 (KB)
RACE HORSES	389	46	340	24
TEMPETE	522	41	488	20

A comparison based on the ground-truth in [8] is presented in TABLE II. In addition to the SAD and BZip2 (BZ2) measures, we have added a new column, DIS, that shows the sum of absolute differences between the true motion field and the estimated motion field.

TABLE II. THE QUALITY OF THE DENSE MOTION FIELD.

	MAP w/o Landmarks			Proposed		
	SAD ( $\times 10^3$ )	DIS ( $\times 10^3$ )	BZ2 (KB)	SAD ( $\times 10^3$ )	DIS ( $\times 10^3$ )	BZ2 (KB)
DIMETRO DON	420	738	99.4	357	155	35.4
GROVE2	1,413	901	141	1,368	462	99
GROVE3	1,490	1,030	161	1,444	500	115
URBAN2	649	1,825	159	586	931	104
URBAN3	605	1,970	158	643	1,131	98
HYDRANG EA	530	800	107	439	187	43.4
RUBBER WHALE	404	701	89.6	322	74	22
VENUS	383	614	72	358	183	32.5

The computational increment of the proposed algorithm are mainly concentrated in the energy minimization process; meanwhile, the complexity of picking landmark motions can be considered as trivial. Among the ten sequences we examined, the computational increments are less than 17%, even less than 1% for some sequences. The  $\lambda$ -parameters for the following eight sequences in TABLE II are set to  $\lambda_g = 1$ ,  $\lambda_d = 1$ ,  $\lambda_m = 1$ .

## 6. CONCLUSIONS

In this paper, we propose a MAP dense motion algorithm with landmark point constraint. The proposed constraint incorporate semi-global information into the iterative MAP estimation process based on local cost updates. The experiments show that the proposed technique produces a dense motion field with very good quality.

## ACKNOWLEDGEMENT

This research is funded by National Science Council, Taiwan, ROC, under grant number NSC 98-2220-E-009-012.

## REFERENCES

- [1] Y.-W. Huang, C.-Y. Chen, C.-H. Tsai, C.-F. Shen, and L.G. Chen, "Survey on Block Matching Motion Estimation Algorithms and Architectures with New Results," *Journal of VLSI Signal Processing*, 42, pp. 297-320, 2006.
- [2] P. Didyk, E. Eisemann, T. Ritschel, K. Myszkowski, and H.-P. Seidel, "Perceptually-motivated real-time temporal upsampling of 3d content for high-refresh-rate displays," *Comput. Graph. Forum (Proc. Eurographics 2010, Norrkpping, Sweden)*, 2010.
- [3] Y.-C. Sun, S.-Y. Lian, and C.-J. Tsai, "Prioritized Side-Information Correction for Distributed Video Coding," *Picture Coding Symposium*, Chicago, 2009.
- [4] B. P. Horn and B. Schunck, "Determining Optical Flow," *Artificial Intelligence*, vol. 17, pp. 185-203, 1981.
- [5] H. H. Nagel and W. Enkelmann, "An Investigation of Smoothness Constraints for the Estimation of displacement Vector Fields from Image Sequences," *IEEE Trans. Patt. Anal. Machine Intelli.*, vol. 8, no. 9, pp. 565-593, 1986.
- [6] J. Konrad and E. Dubois, "Bayesian Estimation of Motion Vector Fields," *IEEE Trans. Patt. Anal. Machine Intelli.*, vol. 14, no. 9, pp. 910-927, 1992.
- [7] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Trans. Patt. Anal. Machine Intelli.*, vol. 6, no. 11, pp. 721-741, 1984.
- [8] S. Baker, D. Scharstein, J.P. Lewis, S. Roth, M.J. Black, and R. Szeliski, "A Database and Evaluation Methodology for Optical Flow", *MSR-TR-2009-179*, 2009