An Efficient Two-pass MAP-MRF Algorithm for Motion Estimation Based on Mean Field Theory

Jie Wei, Ze-Nian Li
School of Computing Science
Simon Fraser University
Burnaby, B.C.
CANADA V5A 1S6
jiew@cs.sfu.ca, li@cs.sfu.ca

Abstract

This paper presents a two-pass algorithm for estimating motion vectors from image sequences. In the proposed algorithm, the motion estimation is formulated as a problem of obtaining the *Maximum A Posteriori in the Markov Random Field* (MAP-MRF). An optimization method based on the *Mean Field Theory* (MFT) is opted to conduct the MAP search. The estimation of motion vectors is modeled by only two MRF's, namely, the motion vector field and unpredictable field. Instead of utilizing the line field, a *truncation function* is introduced to take care of the discontinuity between the motion vectors on neighboring sites. In this algorithm, a "double threshold" preprocessing pass is first employed to partition the sites into three regions, whereby the ensuing MFT-based pass for each MRF is conducted on one or two of the three regions. With this algorithm, no significant difference exists between the block-based and pixel-based MAP searches any more. Consequently, a good compromise between precision and efficiency can be struck with ease. In order to render our algorithm more resilient against noises, the *Mean Absolute Difference* (MAD) instead of *Mean Square Error* (MSE) is selected as the measure of difference which is more reliable according to the knowledge of robust statistics. This is supported by our experimental results from both synthetic and real-world image sequences. The proposed two-pass algorithm is much faster than any other MAP-MRF motion estimation methods reported in the literature so far.

Keywords: Motion estimation, Markov Random Field, Mean Field Theory, Video processing

1 Introduction

The importance of motion estimation can hardly be overestimated in the processing of image sequences. The motion estimation is essential for video compression, object segmentation, object recognition, and a plethora of computer vision applications. Indeed, even stereo images can be viewed as a special case of image sequences [33]

In the video compression community, the Block Matching Algorithm (BMA) is generally employed [17]. With the BMA, each $k \times k$ block of the current video frame is compared to a block of the same size in the previous (reference) frame in the vicinity of its corresponding position, the one with the least Mean Square Error (MSE) or Mean Absolute Difference (MAD) is considered as a match and the difference of their positions is saved in the corresponding position on the *motion map* as the motion vector of the block in the current frame. It is evident that the *brute-force* search for the least MSE/MAD match is computationally intensive. In order to cut the computational overhead of the BMA, several effective enhancements have been developed, such as the logarithmic search [17], three-step search [19] and conjugate direction search [29], which render the BMA practical in real-time video compression.

Nevertheless, the BMA can only work efficiently at the price of the precision of the estimation. In order to obtain more refined motion vectors, more complex scheme must be taken than that used in the BMA. The refined motion vectors should reflect some natural interactions existing between contiguous vectors, *i.e.*, the motion vectors generally manifest themselves with smooth change except on the boundaries of those objects undergoing motion. With these refined motion vectors, in case of video compression, the entropies of the motion map will be reduced and a better compression result can be accomplished. In case of object segmentation, an improved object contour can be obtained. Apparently, the block-based least MSE criteria, giving no consideration to contextual constraints in the spatial and temporal domain, cannot generate satisfactory results due to the so-called *"aperture problem"*[1] and

[1]The component of the velocity of a moving edge in the direction of the edge cannot be determined uniquely

noises existed in those frames, such as object segmentation based on motion in computer vision.

In order to take the contextual constraints into account, the Markov Random Field (MRF) has been widely employed. Generally speaking, in the MRF the impact posed to the value of the random variable on one site, the pixel in case of images, by those on other sites is restricted to its close neighbors, where the "closeness" is determined by the definition of the *neighborhood system*. A description of the MRF will be given in the next section. The *Maximum a Posteriori* (MAP) probability is utilized most commonly as a statistical criterion for optimality and thus often chosen in conjunction with the MRF in vision modeling [22]. The resulting framework, referred to as MAP-MRF, is obtained. In the settings of motion estimation, with the MRF as the model, the estimation of the motion vectors is translated into a problem of optimization, *i.e.*, to locate the configuration of the MAP of the MRF, which is the MAP-MRF problem.

There are mainly two schools of techniques in conducting the optimization process, one is of deterministic nature, or referred to as *local methods*, the other is of the stochastic nature, or referred to as *global methods* [22]. The local methods include such methods as: 1) Iterated Conditional Modes (ICM) [3]: a "greedy" strategy in the iterative local maximization. 2) Dynamic Programming (DP) [2]: an optimization technique for problems where only a part of all variables has interactions simultaneously. 3) Neural Networks (NN) [34]: an alternative to Bayesian classifiers in deriving the MAP. The NN method does not always need to know the prior probabilities. Instead, it can learn the stochastic properties of the specific problem from the training set. On the other hand, Simulated Annealing (SA) and Graduated Non-Convexity (GNC) belong to the domain of global methods. The brief descriptions for them are: 1) SA [18]: a simulation of the physical annealing procedure. In order to avoid the local minima, instead of a simple gradient descent method, a stochastic search, such as Metropolis algorithm [25] which is of the nature of random fluctuations, is employed to find the next configuration. 2) GNC [5]: a method that approximates the global minima through locating the minimum of a convex approximation of the non-convex function. It is a successful effort to ease the computational burden of SA. The common feature shared by all local methods is that the computation converges quickly, but they are susceptible to getting stuck to local minima. For global methods, the global minima can be attained, however, often with extremely high computational cost.

For the purpose of estimating motion vectors in the framework of MAP-MRF, many methods have been developed with various degrees of success [20, 14, 30, 1, 33]. In these existing methods, either global methods or local methods are used to search for the MAP configuration of the MRF based on their respective prior beliefs. Thereby the strong and weak points of the two different schools of MAP-search schemes are all inherited. In order to strike a compromise between efficiency and performance in the estimation of motion vectors, a method based on Mean Field Theory (MFT) was proposed in [35, 36]. There the MFT, originally an approximation scheme in dealing with phase transitions in Statistical Mechanics [7], is utilized to approach the global minima with rapid convergence rate. It is claimed that results nearly as good as SA and convergence rate comparable to the ICM can be achieved with this method [36].

In this paper, after careful investigation of the motion estimation in the framework of MAP-MRF, we developed an efficient two-pass algorithm (TPA) using the MAP-MRF paradigm where fewer MRF's in formulating the problem are involved and smaller number of sites are needed in the MAP-search procedure. As a result, a substantially reduced cost in the MFT-based optimization procedure is achieved. Within our formulation, we will argue that the line field (LF), which is first introduced in [10] and employed extensively in the literature ever since, can be discarded in the process of estimating motion vectors. Instead, the discontinuity problem is taken care of by a *truncation function*. As such, only two MRF's are involved in our MAP-search process, namely, the motion vector field and unpredictable field. With the TPA, after a "double threshold" preprocessing pass, the sites of the current frame are partitioned into three regions of different characteristics. The second pass, which is the MAP-search procedure, is only conducted on some portions of the whole sites, thereby further improving the efficiency in the MFT procedure. In order to render our algorithm less sensitive to the gross errors, or "outliers" in the terminology of robust statistics [13], the Mean Absolute Difference (MAD) instead of Mean Square Error (MSE) is selected as the energy function where the contributions of the outliers are much less than those in the MSE.

This paper is organized as follows: Section 2 begins with a brief introduction to the concepts of the MRF and MFT and their application to the motion estimation; the proposed MAP-MRF model of the motion estimation is discussed in section 3; the two-pass MAP-MRF algorithm based on MFT is delineated in section 4; experiments conducted based on the proposed algorithm are presented in section 5. We conclude in section 6 with some remarks and discussions about the proposed scheme.

2 Markov Random Field and Mean Field Theory for Motion Estimation

In this section, fundamental concepts of the MRF and its applications are briefly introduced. We then proceed to present the MFT which plays a crucial role in the MAP search procedure of the proposed algorithm. Next, the motion estimation in the paradigm of MAP-MRF is presented. The choice of energy functions to make its value more robust against the gross errors is discussed in the last subsection.

2.1 Markov Random Field and its application in image processing

In what follows the basic concepts of the MRF are reviewed. For rigorous expositions one may refer to [28, 22]. Let $S_{m,n} = \{(i,j) : 1 \leq m,\ 1 \leq n\}$ be an $m \times n$ integer lattice; then $F = \{F_{i,j},\ (i,j) \in S_{m,n}\}$ denotes a family of random variables, *i.e.*, a random field, defined on $S_{m,n}$. Evidently each image f can be viewed as a discrete sample realization of F, with $f_{i,j}$ assuming the intensity value on each pixel site. $f = \{f_{1,1}, f_{1,2}, \cdots, f_{m,n}\}$ is referred to as a *configuration* of F, the complete set of all the configurations is denoted as $\mathbf{F}$.

In order to introduce the MRF and its utility, a neighborhood system N on $S_{m,n}$ and the corresponding concept of *cliques* C should also be introduced. N is defined as:

$$N = \{N_{i,j}, (i,j) \in S_{m,n}\ \}, \tag{1}$$

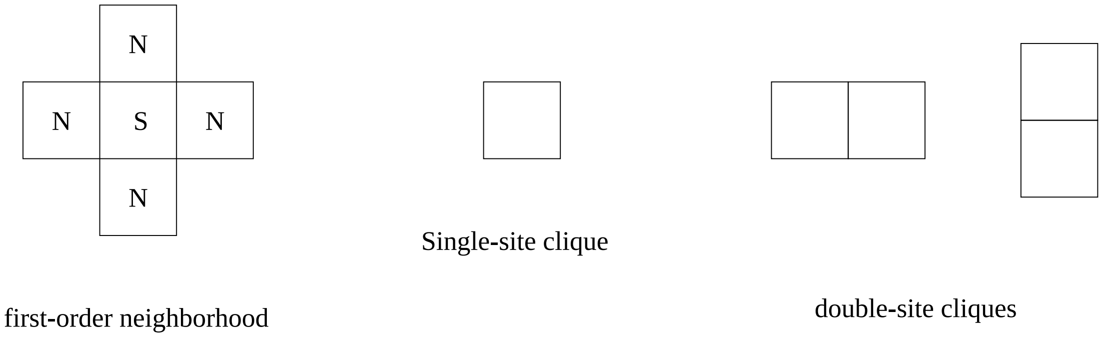


Figure 1: The first-order neighborhood system and the corresponding cliques.

where $N_{i,j}$ is the set of sites on the neighborhood of (i,j), whose formal definition is as follows:

$$N_{i,j} = \{(i',j') \ \mid \ (i',j') \in S_{m,n},\ (i',j') \neq (i,j), (i-i')^2 + (j-j')^2 \ \leq \ d\ \}, \tag{2}$$

where d is a positive integer. An N is called nth order neighborhood system if d assumes the value of n. For instance, the first-order neighborhood system is the 4-connection system in the glossary of computer vision, while the second-order neighborhood system is the 8-connection system.

Based on $(S_{m,n}, N)$, a clique c is defined as a subset of sites in $S_{m,n}$, each site is a neighbor of the other sites in c in the sense of N. As an illustration, for the first-order neighborhood system, c can be of single site or double sites; whereas for the second-order neighborhood system c can be of single, double and triple sites. The possible cliques in the first-order neighborhood system are depicted in Figure 1. If the collection of single site cliques is denoted as C_1 while that of double sites is denoted as C_2, *etc.*, then the collection of all the cliques for kth-order neighborhood system on $S_{m,n}$ is:

$$C = C_1 \cup C_2 \cup \cdots C_k \tag{3}$$

Now we are ready to define the MRF. For ease of exposition, let $P(f)$ denote $P(F = f) = P(F_{1,1} = f_{1,1}, F_{1,2} = f_{1,2}, \cdots, F_{m,n} = f_{m,n})$, that is, the probability density functions (pdf's) of the event of the configuration f. Next denote the pdf's of one random variable $F_{i,j}$: $P(F_{i,j} = f_{i,j})$ as $P(f_{i,j})$. F is called an MRF if the following properties are satisfied:

$$P(f) > 0,\ \forall f \in \mathbf{F} \tag{4}$$

$$P(f_{i,j}|f_{S_{m,n}-(i,j)}) = P(f_{i,j}|f_{N_{i,j}}) \tag{5}$$

Eq. (5) is the so-called *Markovianity*, which indicates the behavior of the random variable on site (i, j) is only affected by those sites in $N_{i,j}$. In other words, the Markovianity means a local interaction on the random field, which can be substantiated by most real-world images.

The MRF lends us a powerful tool in modeling the local property for images. However, for the purpose of image processing, the computation based on the pdf's for each site is prohibitively intensive. Due to the equivalence of the MRF to Gibbs Random Field (GRF) [10, 12], a method to characterize the global feature of a random field from the local features is provided. With this equivalence, the pdf $P(f)$ for a configuration f, which is a joint event, has the following close-form expression:

$$P(f) = exp[-\beta U(f)]/Z, \tag{6}$$

where

$$U(f) = \sum_{c \in C} V_c(f) \tag{7}$$

is the *energy function* for f. $V_c(\cdot)$ is the *clique potential* for a clique c, its value depends on the local configuration of the clique c.

$$Z = \sum_{f} exp[-\beta U(f)] \tag{8}$$

is the *partition function* for f, which acts as a normalization factor in eq. (6).

Eq. (6) has a crucial role to play in all the applications of the MRF. Its importance lies in the fact that the joint probability of the configuration f is fully determined by the nature of its local characteristics, *i.e.*, the choice of the neighborhood system and the values assigned to those corresponding clique potentials. Thereby, out of the prior belief, one can define a neighborhood system and assign values to those corresponding clique potentials to reflect the contextual constraints. The most possible result obtained from this prior is the MAP, which, according to eq. (6), turns out to be the one that has the minimal energy $U(f)$. Thereby a rich arsenal of optimization methods developed in various fields such as statistics, statistical mechanics can be employed in the search of the minimal energy.

2.2 Mean Field Theory

As mentioned earlier, the MFT has the promise of cutting a nice tradeoff between efficiency and effectiveness, *i.e.*, approximating the global minima at a cost comparable to the local methods. The MFT was originally developed in statistical mechanics [7, 27, 26] for systems undergoing phase transitions. As pointed out in [21], problems in statistical mechanics and image processing bear much resemblance to each other: the overall property, the macro state for statistical mechanics and image for image processing, is determined by the local interactions of local properties, *i.e.*, micro states in statistical mechanics and pixels in image processing respectively. Henceforth, many methods developed in statistical mechanics, such as SA [18, 10], MFT, and the renormalization group theory [11], have been exploited extensively in dealing with image processing applications.

The basic idea of MFT is, as described in [7], to assume that the effect to one particle p imposed by its neighboring particles through interactions is approximated by an average magnetic *field* formed by these particles. Therefore, instead of computing the interactions with all its neighboring particles, one can obtain the mean field generated by its neighboring particles and then compute the impact posed to p by this field. The MFT reduces the many-body statistical mechanics problem into a one-body problem. The resultant procedures can be very accurate and useful [27, 7].

By analogy of MFT in statistical mechanics, for image processing tasks, in order to search for the minimal energy of eq. (6), one should first calculate the mean field for each pixel. Denote the mean field for $f_{i,j}$ as $\langle f_{i,j} \rangle$. By definition, one has the following:

$$\langle f_{i,j} \rangle \approx \sum_{f_{i,j}} f_{i,j} P(f) \tag{9}$$

Substitute $P(f)$ by use of eq. (6), eq. (9) becomes:

$$\langle f_{i,j} \rangle = \frac{1}{Z} \sum_{f_{i,j}} f_{i,j} exp[-\beta U(f)] \tag{10}$$

Provided that the first-order neighborhood system is adopted for the MRF F, the following approximation of the mean field is obtained:

$$\langle f_{i,j} \rangle = \frac{1}{Z_{i,j}^{mf}} \sum_{f_{i,j}} f_{i,j} exp[-\beta U_{i,j}^{mf}(f_{i,j})], \tag{11}$$

where the new partition function is defined as:

$$Z_{i,j}^{mf} = \sum_{f_{i,j}} exp[-\beta U_{i,j}^{mf}(f_{i,j})] \tag{12}$$

And the new energy function is defined as:

$$U_{i,j}^{mf}(f_{i,j}) = V_c(f_{i,j}) + \sum_{(i',j') \in N_{i,j}} V_c(f_{i,j}, \langle f_{i',j'} \rangle). \tag{13}$$

For rigorous expositions of this approximation formula, one may refer to [36] [2].

Henceforth, the marginal pdf $P(f_{i,j})$ has the following approximation formula:

$$P(f_{i,j}) \approx \frac{1}{Z_{i,j}^{mf}} exp[-\beta U_{i,j}^{mf}(f_{i,j})] \tag{14}$$

From eq. (14), the joint pdf for a configuration f under the MFT is approximated as follows:

$$P(f) \approx \prod_{i,j} \frac{1}{Z_{i,j}^{mf}} exp[-\beta U_{i,j}^{mf}(f_{i,j})] \tag{15}$$

Based on the MFT approximation of the pdf for a configuration f as defined in eq. (15), the estimation of the MAP of f amounts to minimizing the energy functions $U_{i,j}^{mf}(f_{i,j})$. With the MFT procedure, one need only to compute the joint fixed points of those $\langle f_{i,j} \rangle$'s to achieve the approximated MAP. Suppose the mean field of $f_{i,j}$ computed in iteration k is denoted as $\langle f_{i,j} \rangle^k$, a prescribed small number is ϵ, the joint fixed points of $\langle f_{i,j} \rangle$'s is said to be achieved if the following condition is satisfied:

$$\sum_{i,j} \parallel \langle f_{i,j} \rangle^{k+1} - \langle f_{i,j} \rangle^k \parallel^2 \leq \epsilon \tag{16}$$

the resulting $\langle f_{i,j} \rangle^{k+1}$'s are the estimation of the MFT process.

Therefore, the optimization procedure under the MFT strategy is to compute the mean field $\langle f_{i,j} \rangle$'s based on eq. (11) iteratively until the difference between the results of two contiguous iterations is less than a certain small number, which indicates that the equilibrium is attained.

[2]Chandler [7] provided another formula under the Ising model.

2.3 Motion estimation as an MAP-MRF problem

As in [15], in the application of motion estimations, the *intensity constancy* constraint is generally assumed:

$$f^{t+1}_{(i+u,j+v)} \approx f^t_{(i,j)} \tag{17}$$

where $f^t_{(i,j)}$ is the intensity value of site (i,j) in the frame at time t, $u(i,j)$ and $v(i,j)$ are the projections of the motion vector on (i,j) along the horizontal and vertical direction respectively. For simplicity, denote $(u(i,j),\ v(i,j))$ as $\vec{d}^t_{(i,j)}$, and the set of the $\vec{d}^t_{(i,j)}$'s on all sites as $\vec{d}^t$. The motion estimation in the settings of MAP-MRF can be stated as the process of maximizing the conditional pdf $P(\vec{d}^t|f^{t+1}, f^t)$, *i.e.*,

$$\vec{d}^t = argmax_{\vec{d}^t} P(\vec{d}^t|f^{t+1}, f^t), \tag{18}$$

which reads: in the presence of f^{t+1} and f^t locate the $\vec{d}^t$ that maximizes the conditional pdf.

As described in [36, 20], since $P(f^{t+1}|f^t)$ is not a function of $\vec{d}^t$, it can be ignored in maximizing $P(\vec{d}^t|f^{t+1}, f^t)$ with respect to $\vec{d}^t$. Therefore, after the application of Bayesian theorem, eq. (18) turns out to be:

$$\vec{d}^t = argmax_{\vec{d}^t}[P(f^t|\vec{d}^t, f^{t+1}) \cdot P(\vec{d}^t|f^{t+1})] \tag{19}$$

In the RHS of eq. (19), the first term assumes the role of measuring the "likelihood" between f^{t+1} and f^t undergoing a motion represented by $\vec{d}^t$, while the second term addressing the contextual constraints between neighboring motion vectors, *e.g.*, the smoothness of them in case they are of the same object.

Under the MRF model, the first-order neighborhood system is chosen throughout the following presentation. As discussed before, eq. (19) amounts to minimizing the corresponding energy functions. For instance, for $P(f^t|\vec{d}^t, f^{t+1})$, the function for the likelihood energy, which can be viewed as the collection of the clique potentials for single sites, is usually formulated as:

$$U(f^t|\vec{d}^t, f^{t+1}) = \sum_{i,j}(f^{t+1}_{(i+u,j+v)} - f^t_{(i,j)})^2 \tag{20}$$

As far as the robustness is concerned, as described in the next subsection, the proposed two-pass algorithm will employ an improved energy function in order to be more resilient against gross errors.

2.4 MAD energy function to measure the likelihood

As described in [13], the least-squares function for eq. (20) is extremely sensitive to outliers. The problem is that, for those outliers, their squares contribute "too much" to the overall value of the MSE. Various functions have been proposed in [16, 13] in order to discount the contribution of those outliers, Mean Absolute Difference (MAD) is one of the them which is more robust than the MSE. In this paper, due to its simplicity and robust feature, the MAD is opted as the energy function. Therefore, the function measuring the likelihood of neighboring frames becomes:

$$U(f^t|\vec{d^t}, f^{t+1}) = \sum_{i,j} |f^{t+1}_{(i+u,j+v)} - f^t_{(i,j)}| \tag{21}$$

For $P(\vec{d^t}|f^{t+1})$, one can define the following function, which can be treated as the collection of the clique potentials for double sites:

$$U(\vec{d^t}|f^{t+1}) = \sum_{i,j} \sum_{(i',j') \in N_{i,j}} |\vec{d^t}_{i,j} - \vec{d^t}_{i',j'}| \tag{22}$$

Due to the option of the first-order neighborhood system, the foregoing two energy functions totally determine the prior belief for the MAP-MRF problem. In the proposed Two-Pass Algorithm (TPA), the MAD instead of the MSE is used as a measure of the likelihood. We call this MRF model the *"simple model for motion estimation"*. In reality the definition formulated in this way, similar to the case when using the MSE, imposes a universal smooth factor to the resulting motion vectors on all sites since the motion vector on each site has the same amount of interaction to its first-order neighbors, as such the resulting estimations of the motion vectors from the MAP-MRF will exhibit the artifact of over-smoothness. Efforts to attack it will be explicated in the next section to render the estimation more naturally.

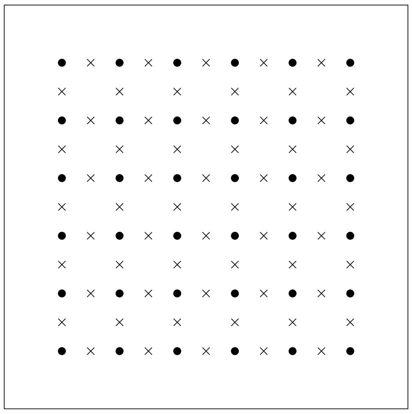


Figure 2: The line field (cross) is the dual of the original field (dot).

3 The Proposed MAP-MRF Model Using MFT for Motion Estimation

In this section, first the modeling of the interactions of neighboring motion vectors is discussed, then a model for the MFT estimation of motion vectors which reflects the prior belief is proposed.

As mentioned briefly before, without introducing other techniques, the simple model for motion estimation introduced in the last section imposes a universal smooth factor to the random variables on every site, the resultant motion estimations will be over-smooth or often a total failure. In a pioneering work by Geman and Geman [10], the *line field* (LF) is introduced to enforce the discontinuities for some sites, thereby the effect of over-smoothness can be avoided elegantly.

In essence, as described in [10], the sites of LF are the *"dual"* of those of the original lattice $S_{m,n}$, namely, the sites of the LF are in the middle of each horizontal or vertical pair of sites. Therefore, between every two neighboring sites in $S_{m,n}$, there exists exactly one site on the LF correspondingly, as depicted in Figure 2. The random variables $l_{i,j}$ on each site (i,j) of the LF is of boolean nature, *i.e.*, of the value 0 or 1 which relies on the difference of the corresponding values of f on $S_{m,n}$. Suppose the two neighbors of (i,j) on the LF are (i_1,j_1) and (i_2,j_2) on $S_{m,n}$, $l_{i,j}$ takes its value according to the following rules:

$$l_{i,j} = \begin{cases} 1 & \text{if } |f^t_{i_1,j_1} - f^t_{i_2,j_2}| \geq \gamma \\ 0 & \text{otherwise,} \end{cases} \tag{23}$$

where γ is a prescribed threshold. It is evident that the value of $l_{i,j}$ indicates the presence or absence of an "edge element". Thus it is set to 1 when a discontinuity exists. With the introduction of the LF, the prior models for the MRF are so formed that, for any two neighboring sites on $S_{m,n}$, only if the corresponding $l_{i,j}$ on the LF is of the value 0, will the interaction between them be counted in the energy formulation. In this way no over-smoothing artifacts will arise in the resulting MAP. As a result, the LF is employed extensively in the applications of the MRF in image processing and computer vision [8].

In order to attack the over-smoothness effect in motion estimation, based on eq. (11), Zhang *et al.* [36] employed the line fields and *unpredictable fields*, wherein an impressive result was accomplished. The unpredictable fields are exploited to account for the situation where no correspondence of some sites in f^{t+1} can be made in f^t because of the occlusions or some uncovered areas, which is employed in most literature addressing the MRF-based motion estimation, *e.g.*, [20, 33].

The LF is of interest in its power to enforce the interactions based on the contents of the F, it finds its usefulness in a wide spectra of applications. However, we will argue below that it is not appropriate to exploit the LF literally in formulating the interactions of motion vectors on neighboring sites or blocks (an ensemble of sites). Instead, a function whose value depends on the difference of the magnitude of motion vectors of neighboring sites or blocks will be employed to enforce discontinuity in an appropriate manner.

There are two observations justifying the removal of the LF in the process of motion estimation:

- For image sequences, if one employs the LF in the formulation of the prior models to estimate motion vectors, the continuities of motion vectors are only enforced for neighboring sites of similar intensity values, which indicates that objects of interest are of smooth intensity changes. Unfortunately, on the surface of an object, the intensity can undergo dramatic changes. Moreover, although the boundaries of objects often cause intensity discontinuities, intensity discontinuities themselves do not necessarily indicate object boundaries. Thereby, in the process of motion estimation, discontinuities of neighboring motion vectors incurred solely out of the LF will be inclined to generate "over-scattered" motion vectors as the estimation results.

- When using the LF, difficulties are witnessed if motion vectors of macro-sites (*i.e.*, blocks) are to be estimated, which is of more interest as far as efficiency is concerned. One solution is to resort to the hierarchical method where a pyramidal representation of the original frames should first be given as mentioned in [36]. This "top-down" strategy can be expected to generate some enhancement to the original method as far as the efficiency is concerned. However, the LF is not well-preserved on the sites of subimages of lower frequencies, since the low-pass filtering would blur the image across the boundaries of the blocks.

The efficiency of the removal of the LF has been discussed in [5], where the minimization is first conducted over the LF by use of a truncation function. In [4], the LF and the robust statistical techniques are elegantly unified, in which the LF is formulated in terms of the outlier process, as such a general framework is provided by using the existing robust statistical methods. In this paper the LF is removed to reduce the computation involved in the MAP search through the introduction of a so-called *truncation function* $g(\cdot,\cdot)$. Suppose γ_d is a threshold whose value can change dynamically in the MFT process, $\vec{d}_1$ and $\vec{d}_2$ are the motion vectors of two neighboring sites, the definition of g is as follows:

$$g(\vec{d}_1,\vec{d}_2)=\begin{cases} \parallel \vec{d}_1-\vec{d}_2 \parallel & \text{if } \parallel \vec{d}_1-\vec{d}_2 \parallel \leq \gamma_d \\ \eta \quad (\eta<\gamma_d) & \text{otherwise} \end{cases} \tag{24}$$

The difference of the truncation function defined here from the one in [5] is that, a smaller penalty, *e.g.*, $\frac{\gamma_d}{2}$, will be levied if the value of $\parallel \vec{v}_1-\vec{v}_2 \parallel$ exceeds the threshold in our truncation function to encourage a disconnection. In their definition [5], a penalty equal to γ_d is applied in case the prescribed threshold is surpassed. Our truncation function is defined on the same track as the scheme proposed by B.K.P. Horn in [24] to cope with the discontinuities in optical flow. In our implementation, γ_d can alter its value over the iterations. By trial and error, γ_d is found to be $\max\{8e^{-i/8},4\}$, where i is the number of the current iteration. The rationale behind this definition is: if the magnitude of the difference of the motion vectors on two neighboring sites is too large, the two sites will not be considered in the same object and a small penalty is levied for the appearance of this discontinuity.

By substituting the LF with the truncation function, which is a check on the difference of neighboring motion vectors, the line field is totally removed in the process of the MAP-search. Therefore, the computational load is eased greatly, which is one of the reasons for the efficiency of the proposed algorithm. Plus, through using the proposed truncation function, it does not matter whether the estimation of motion vectors is based on each individual pixel or a block of pixels, a trade-off between efficiency and effectiveness can be cut with ease by changing the block sizes and controlling parameters.

It is not unusual that for some sites or blocks in f^{t+1} their correspondence cannot be found in f^t due to the fact that they belong to the unpredictable area. Mathematically, for a site (i,j), the MAD is defined as[3]:

$$U^t_{d_{(i,j)}} = |f^{t+1}_{(i,j)} - f^t_{(i_1,j_1)}| \tag{25}$$

where $(i,j)+d_{(i,j)}$, denoted as (i_1,j_1), are the sites determined by all possible motion vectors. If the maximum value of the magnitude of motion vectors is k, then the set of (i_1,j_1)'s, denoted as $C_{i,j}$, is the square of size $2k+1$ centered on (i,j) in f^t. Denote

$$\hat{E}^t_{d_{(i,j)}} = min_{(i_1,j_1)\in C_{i,j}}|U^t_{d_{(i,j)}}| \tag{26}$$

If $\hat{E}^t_{d_{(i,j)}}$ is very large, (i,j) is considered as unpredictable. For instance, if we are concerned with compression, the cost of transmitting/saving the site or macro-site (i,j) will consume less bandwidth/space than that of the estimated motion vectors and the corresponding displaced residual site or block. For segmentation-by-motion tasks, evidently singling this out will facilitate a more meaningful object segmentation results.

In the literature, *e.g.*, [36], an unpredictable field O, whose sites are the same as $S_{m,n}$ and the value on each site is of the boolean nature, is proposed to account for the existence of

[3] In case of blocks, one can use the coordinate of its left upper corner to indicate one block, then the corresponding MAD can be defined in the similar manner to the pixel case.

some unpredictable sites. For a site or macro-site, if the value of the corresponding random variable $o_{i,j}$ of the unpredictable field O is 1, a constant value instead of a certain $U^t_{d(i,j)}$ will be the penalty. In [36] O is also estimated for every site in the MAP-search procedure. From our observation, in the process of MAP-MRF motion estimation, in order to address the unpredictable cases efficiently, we propose a preprocessing pass, called "double-threshold" preprocessing, which assumes the role of partitioning those sites or blocks into three groups based on two predefined thresholds γ_{p1} and γ_{p2}, where $\gamma_{p1} > \gamma_{p2}$, *i.e.*,

$$S_{unpredict} = \{(i,j) | (i,j) \in S_{m,n},\ \hat{E}^t_{d_{(i,j)}} \geq \gamma_{p1}\} \tag{27}$$

$$S_{uncertain} = \{(i,j) | (i,j) \in S_{m,n}, \gamma_{p2} \leq\ \hat{E}^t_{d_{(i,j)}} < \gamma_{p1}\} \tag{28}$$

$$S_{predict} = \{(i,j) | (i,j) \in S_{m,n},\ \hat{E}^t_{d_{(i,j)}} < \gamma_{p2}\} \tag{29}$$

Based on this partition, there are three cases correspondingly:

- the sites in $S_{unpredict}$ are excluded from the MAP search procedure;
- for sites in $S_{predict}$, no unpredictable random variables are defined;
- for sites in $S_{uncertain}$, both the motion vector field and the unpredictable field are defined.

Thereby, in the proposed model, there are two MRF's, namely, motion vector field $\vec{d}$ and unpredictable field O. The sites for $\vec{d}$ entailed to compute are $S_{predict} \cup S_{uncertain}$, while on $S_{unpredict}$ the corresponding $d_{i,j}$'s are always set to 0. The sites for O to be estimated consist only of $S_{uncertain}$, for $S_{unpredict}$ and $S_{predict}$ the corresponding $o_{i,j}$'s are always set to 1 and 0 respectively. In the next section the corresponding energy function for either MRF will be given and the resulting algorithm will be proposed.

4 The Proposed Two-pass Algorithm for Motion Estimation

In the last section the MAP-MRF model for the estimation of motion vectors has been discussed, where two MRF's are employed in the estimation: the motion vector and unpredictable fields. For the motion vector field on $S_{predict} \cup S_{uncertain}$, the energy function is:

$$U^{mf}_{i,j}(\vec{d}_{i,j}) = (1 - o_{i,j})U_{d_{(i,j)}} + \lambda_d \sum_{(i',j') \in N_{(i,j)}} g(\vec{d}_{(i,j)}, \langle \vec{d}_{(i',j')} \rangle) \tag{30}$$

where λ_m and λ_d are two control parameters generally of small positive values. Notice here the truncation function g is a bit different from the one defined in the last section with one more tenor:

$$g(\vec{d}_{(i,j)}, \langle \vec{d}_{(i',j')} \rangle) = 0 \text{ if } \langle o_{i,j} \rangle = 1 \text{ or } \langle o_{i',j'} \rangle = 1 \tag{31}$$

The rationale behind this is: if either of the two sites is unpredictable based on the value of the corresponding value on the unpredictable random field O, no penalty should apply. With this definition, only if the two neighboring sites are both predictable and the difference of the corresponding motion vectors in the current iteration is sufficiently small, will the interactions between them be counted in.

For the unpredictable field on $S_{uncertain}$, the energy function is:

$$U^{mf}_{i,j}(o_{i,j}) = o_{i,j}[C_o - \lambda_p U_{\langle d_{(i,j)} \rangle}] + \lambda_q \sum_{(i',j') \in N_{(i,j)}} h(o_{i,j}, \langle o_{i',j'} \rangle) \tag{32}$$

where C_o is a constant, *i.e.*, the penalty is always levied when an unpredictable site makes its appearance. λ_p and λ_q are two control parameters. The function $h(\cdot, \cdot)$ is formally defined as follows:

$$h(o_{i,j}, \langle o_{i',j'} \rangle) = \begin{cases} |o_{i,j} - \langle o_{i',j'} \rangle|, & \text{if } \ (i',j') \notin S_{uncertain} \\ (1 - sgn(\| \vec{d}_{i,j} - \vec{d}_{i',j'} \| - \gamma_d))(1 - 2\delta(o_{i,j} - \langle o_{i',j'} \rangle)), & \text{otherwise,} \end{cases} \tag{33}$$

where the function $\delta(\cdot)$ is the *Kronecker function*, *i.e.*, it is of the value 1 if the input is 0, and 0 otherwise, and the function $sgn(\cdot)$ is the sign function, *i.e.*,

$$sgn(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{34}$$

Essentially, in eq. (32), the first term is of the nature of the clique energy for the single-site clique; whereas the second term is for the clique of double sites, which takes care of the interaction of neighboring sites on the unpredictable field.

In summary the TPA is as follows:

1. The first pass—"double threshold" preprocessing: compute the $\hat{E}$ as defined in eq. (26) for each site, whereby the three partitions $S_{predict}, S_{uncertain}$ and $S_{unpredict}$ of S are acquired.

2. The second pass—MAP search:

   (a) Use eq. (30) and eq. (11) to compute the mean field of $\vec{d}$;

   (b) Use eq. (32) and eq. (11) to compute the mean field of O;

   (c) Similar to [36], the normalized difference e_k for the two mean fields on the current iteration k can be formally defined as follows:

   $$e_k = [\ \| \langle \vec{d}^{(k)} \rangle - \langle \vec{d}^{(k-1)} \rangle \|^2 + \| O^{(k)} - O^{(k-1)} \|^2\ ]^{1/2} / \| S \| \tag{35}$$

   If e_k is greater than a prescribed threshold ϵ, then exit; otherwise, goto step (a).

It can be observed that in the proposed algorithm, due to the existence of the two passes, not only the number of mean fields but also the number of sites to be computed on is reduced, as such the efficiency is improved greatly.

5 Experimental Results

In this section, experiments using the proposed TPA are reported. All the image sequences consist of 256 gray level black and white images. With synthetic image sequences the displacements and moving objects can be controlled with ease, hence our first group of experiments are conducted on two synthetic image sequences, one consists of one moving object, while the other contains two objects undergoing certain motions. Next, experimental results conducted on two real-world image sequences are presented.

The controlling parameters adopted throughout the experiments are listed in Table 1. The following parameters are used:

1. β: it assumes the same role as that of the temperature in simulated annealing. As discussed in [36], to ease the computation burden, instead of making it change dynamically, it is chosen to be held fixed, *i.e.*, 1.0.

Table 1: The values assigned to each parameter in the experiment.

parameter	**opted value**	**parameter**	**opted value**	**parameter**	**opted value**
β	1.0	λ_q	5	λ_d	12.8
λ_p	1.0	γ_{p1}	40	γ_{p2}	10
γ_d	$\max\{8e^{-i/8}, 4\}$	η	$\gamma_d/2$	C_o	16
ϵ	0.01				

2. λ_d: used to enforce the smoothness of the motion vectors of neighboring blocks.

3. γ_{p1}, γ_{p2}: meant to determine the range of uncertain blocks. Whether or not those blocks with average difference in between γ_{p1} and γ_{p2} are unpredictable is to be determined by the following MAP search pass.

4. $C_o, \lambda_p, \lambda_q$: these are the control parameters for the uncertain blocks. The first two are the singleton clique energy, while the last one is used to enforce the "smoothness" of neighboring blocks with regard to their unpredictabilities.

Since there are many parameters and equalities to deal with, it is very difficult, if not impossible, to estimate them automatically. The parameters given here are thus obtained by trial and error. Except γ_{p1}, which is a hard threshold labeling all those blocks with MAD larger than it to be unpredictable, minor alterations to all other parameters will not induce drastically different estimation results.

The searching window is always set to 5×5.

As mentioned before, no significant difference exists between block-based motion estimation and single-pixel based estimation with the proposed TPA, thus the proposed algorithm is conducted on 4×4, 2×2, and 1×1 blocks, the last one is actually the pixel-based estimation. They are denoted as 4×4 TPA, 2×2 TPA and pixel TPA, respectively.

The needlegram, a visualization of the estimated motion vectors for each sites or blocks as defined in [36] is used to show the performance of different schemes.

In order to evaluate the proposed algorithm, several criteria are employed:

- *Displacement field error,* it is of the following form [36]:

$$\epsilon_{\vec{d}} = \| \vec{d}_0 - \vec{d} \|_{S_1}^2 \; / \; \| S_1 \| \tag{36}$$

where S_1 is the image lattice with the unpredictable sites removed, $\| S_1 \|$ is the number of sites in S_1. $\vec{d}_0$ is the known motion vectors. Evidently this criteria is only applicable to the synthetic image cases where the motion map is known beforehand.

- *Entropy of motion field,* it is defined as follows [31]:

$$H = -\sum_u P(u) \log_2 P(u) - \sum_v P(v) \log_2 P(v) \tag{37}$$

where $P(u)$ and $P(v)$ denote the relative frequency of occurrence of the horizontal and vertical components of the motion vector $\vec{d}$. H is used to reflect the randomness of the results of the motion estimation process. Usually, a lower H indicates a more consistent motion estimation. For the same image sequence, a lower H also means less space/bandwidth is required to store/transmit the motion map. Together with the needlegram, it offers an indication as to the performance of the evaluated methods.

- *Number of iterations,* it is to indicate how many iterations the proposed algorithm takes before the converging point is attained. The first pass of TPA, a brute-force BMA to compute $\hat{E}$, is counted as one iteration.

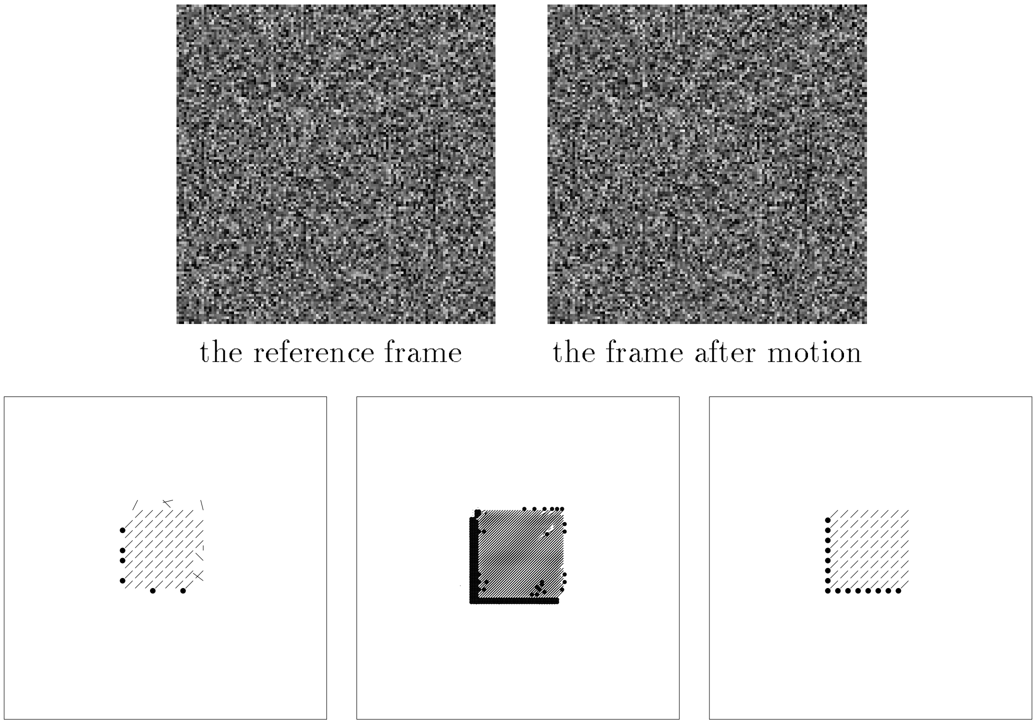
Detailed comparisons are presented with the block matching algorithm with the MAD as the energy function and the MRF based method proposed in [36], which is denoted as BMA and MRF later.

5.1 Synthetic image sequences

In this part, experiments conducted on two types of synthetic image sequences are reported. One contains a single moving object while the other has two moving objects.

- Single moving object:

The sequence of two 128×128 synthetic images are obtained in the following way: an image I is generated through a realization of an identically independent distributed Gaussian random process with mean value $\mu = 100$, and standard deviation $\sigma = 60$. Then the 32×32 block in the center of I is moved with a motion vector $(3, 3)$. The uncovered area was filled with samples from the same Gaussian random process as that generating I. The resulting image is denoted as I'. Finally, a white Gaussian noise with $\sigma = 5$ is added to both I and I'. They are viewed as the reference and current frame respectively in the ensuing motion estimation process. The two frames and the corresponding needlegrams for the 4×4 BMA, the MRF, and the 4×4 TPA are depicted in Figure 3.



the reference frame　　the frame after motion

4×4 BMA　　MRF　　4×4 TPA

Figure 3: The synthetic image sequence with one moving object and the corresponding needlegrams. Legend: 1) Dot: the block estimated as unpredictable; 2) line: the block with a non-zero motion vector; 3) blank: the block whose motion vector is of the magnitude of zero.

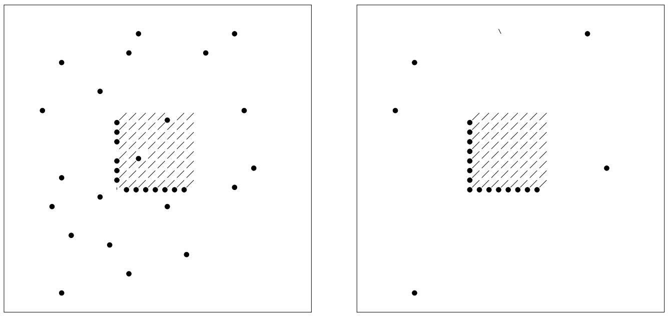


Figure 4: The needlegrams for 4×4 block after adding some pepper-like sparse noises into the reference frame of the single-object synthetic sequence.

It can be seen that the motion vectors and unpredictable sites estimated the BMA are the worst among all the three methods. As to the MRF, the estimation is quite acceptable, but not as good as the one generated by TPA; in addition, the time consumption is much more than the 4×4 TPA (16 vs. 4). The reason for the worse results of the MRF is twofold: 1) the existence of line field. In this synthetic image sequence, the intensity values of neighboring pixels, which are sampled from an i.i.d., are no longer smooth, according to the definition of the LF, many positions will be labeled as line elements, the "over-scattered" effects as discussed in the last section will have some negative impacts on the final estimation results. 2) the use of MSE energy function, which is more sensitive to the noises the the MAD. Detailed test results are listed in Table 2. It can be seen from Table 2 that much fewer iterations are needed for the TPA to generate comparable estimation results than the MRF. This efficiency manifests itself consistently in the other two experiments.

In order to show the difference between the MAD and the MSE, some high intensity (150 in our experiment) pepper-like sparse noises are added to the reference frame. The resulting needlegrams with the two versions are demonstrated in Figure 4. It can be observed that more robust estimation can be achieved with the MAD. This is further confirmed by all of our extensive experiments.

- Two moving objects:

Table 2: Test results for the first simulated image sequence.

Method	Displacement field error	Entropy of motion field	Number of iterations
4×4 BMA	0.355	10.63	
8×8 BMA	0.372	9.043	
MRF	0.180	9.691	16
4×4 TPA	0.025	8.211	4
2×2 TPA	0.029	9.701	6
pixel TPA	0.104	10.817	10

Table 3: Test results for the second simulated image sequence.

Method	Displacement field error	Entropy of motion field	Number of iterations
4×4 BMA	0.562	8.622	
8×8 BMA	0.642	8.921	
MRF	0.325	8.578	21
4×4 TPA	0.245	8.539	4
2×2 TPA	0.317	8.960	4
pixel TPA	0.382	9.775	9

The reference image in this test is generated in the same manner as in the last experiment. Instead of moving only one block, this time two neighboring 32×32 blocks around the central portion undergo motions, one with the vector of $(0, 3)$ while the other with $(0, -3)$. They are shown in Figure 5.

The corresponding needlegrams based on 4×4 blocks of BMA, the MRF, and the TPA are illustrated in Figure 5. Table 3 lists more test results conducted on this sequence. It can be observed that the smoothness and discontinuity are taken good care of by the TPA than the BMA and the MRF.

5.2 Real-world image sequences

Extensive experiments have been made on many real-world clips, such as Miss America, Susie, and Tennis, *etc.* In this part the test results of two different real-world image sequences

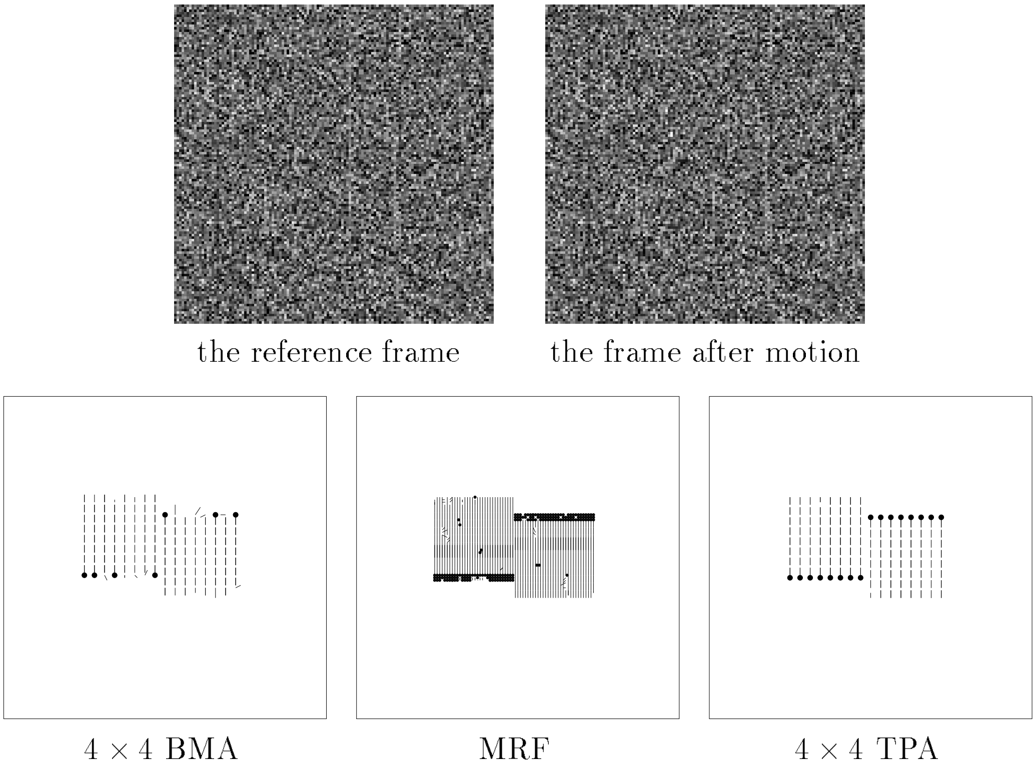


Figure 5: The synthetic image sequence with two moving objects and the corresponding needlegrams.

are presented to show the performance of the proposed algorithm. One is the “Toronto tourism commercial” clip while the other is the “Moon walking” clip, some of the image frames are shown in Figure 6.

- The “Toronto tourism commercial” clip:

  The images are acquired from the “Toronto tourism commercial”. By examining Figure 6 one can find that initially two persons are moving in the clip, then a third person steps in. The needlegrams of the BMA, the MRF[4], and the proposed scheme based on 4×4 block, *e.g.*, for Frame 24 and Frame 25, are depicted in Figure 7. It is observed that the result from the TPA reflects the scenario much better than that of the BMA. It is of interest to notice that based on the result of the TPA the three persons can be separated by use of the consistency of the motion vectors. The result generated from the 4×4 TPA can be employed to obtain a quick object segmentation, while from

[4]In the two real-world clips, the results of the MRF are always generated with 30 iterations.

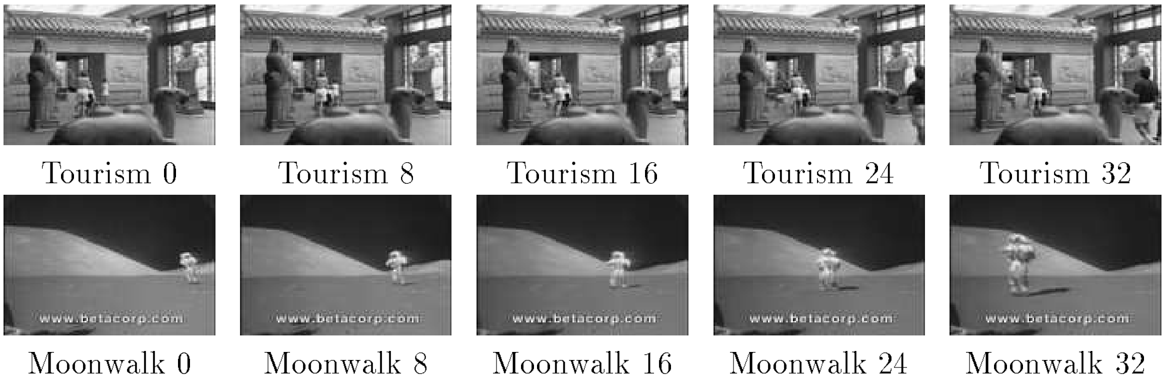


Figure 6: The real-world image sequences on which the experiments are conducted.

the result generated by BMA it is impossible to obtain any object segmentation. The time consumed is merely three times that of the BMA, which is superior to any other MAP-MRF method reported so far. For the MRF method, due to the use of the line field and MSE, some incorrect motion vectors make their appearances around those areas with noises or busy textures. The overall estimation result of the MRF is quite good, with a long processing time though.

The complete results conducted on this image sequence are given in Fig. 8 and Table 4. It can be seen that the entropy of the motion field increases in the later stage which is in accordance with the fact that a third person entered in the view.

- The “Moon walk” clip:

  Our last test is conducted on the image sequence of NASA’s “Moon walk” clip as shown in Figure 6, where the astronaut is jumping and the background is static. It should be pointed out that here the quality of the images is far worse than the “Toronto tourism commercial” sequence—lots of noises make their appearance. Moreover, the background lacks textures and thereby the mere minimal MAD criterion employed by BMA can hardly find the correct match. As can be seen in Figure 7, the performance of the BMA method is so poor that hardly any useful information can be induced. For the MRF, with the use of line field and MSE, it is found that this pixel based

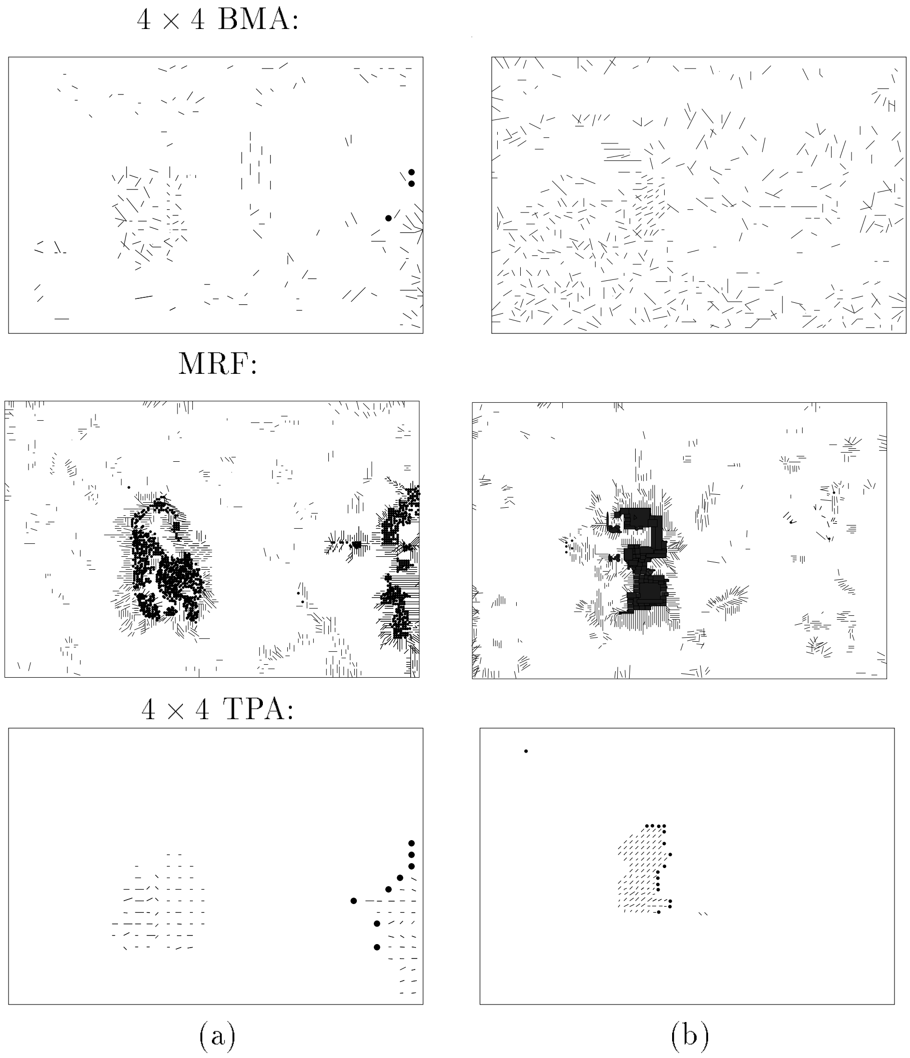
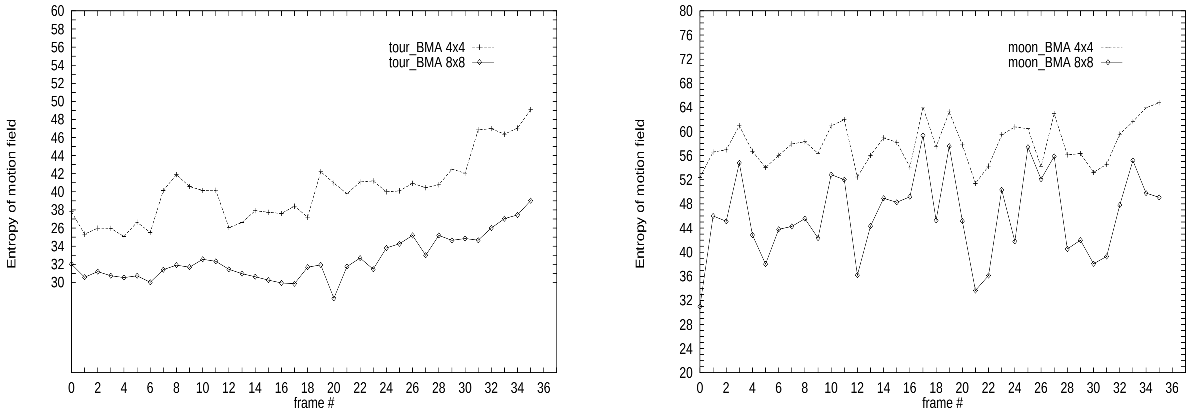


Figure 7: Column a: needlegrams of the clip "Toronto tourism" for Frame 24 and Frame 25. Column b: needlegrams of the clip "Moon walk" for Frame 23 and Frame 24.

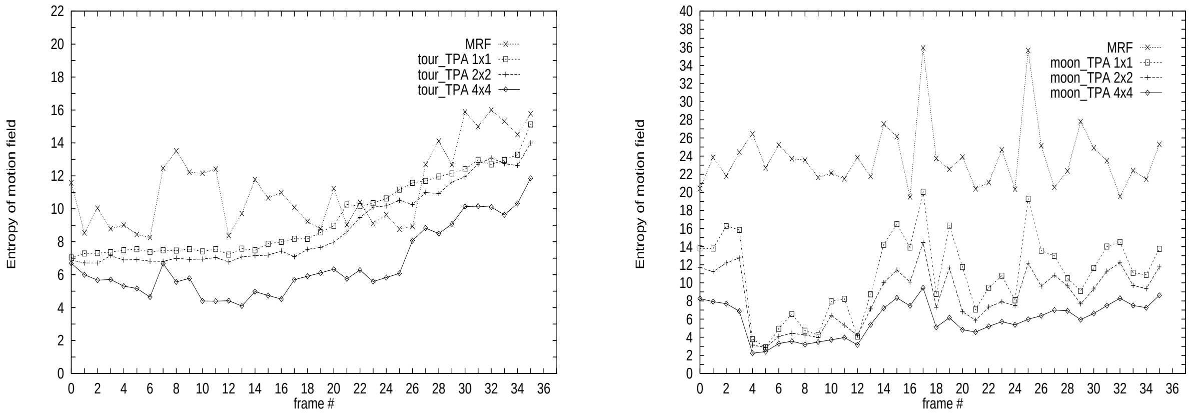
method is also relatively sensitive to noises, thus the resultant motion vectors are not that good, however, they are still much better than the BMA. By contrast, the TPA is quite robust in this scenario. After merely 4 iterations the obtained motion map as illustrated in Figure 7 demonstrates very satisfactory result where the astronaut can be located easily.

The complete test results on this clip are depicted in Fig. 8 and Table 4, from which it can be observed that the entropies of motion field generated by the TPA using different block sizes are consistently better than those done by the BMA and MRF.

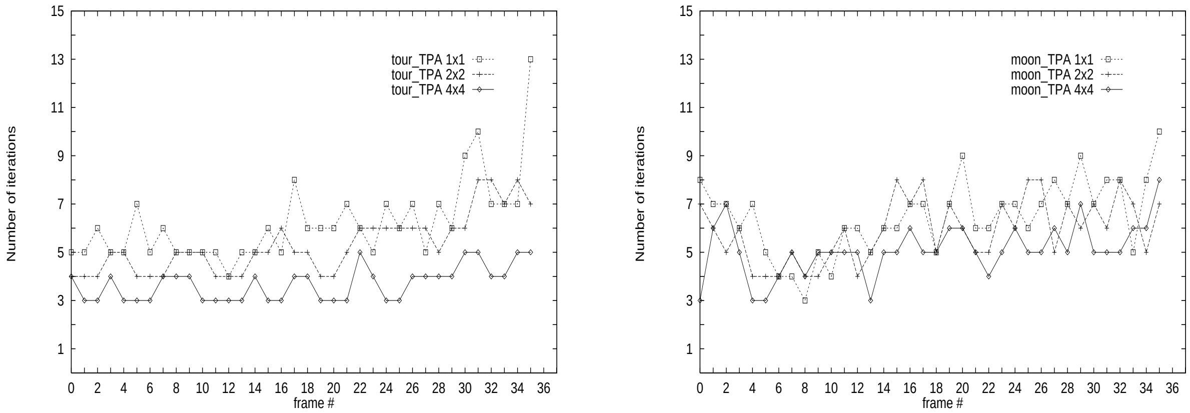
MV entropy using the BMA:



MV entropy using MRF and TPA:



Iterations using TPA:



(a) (b)

Figure 8: The statistics of the results of the two real-world clips: column (a) the Tourism clip; column (b) the Moonwalk clip

Table 4: Statistics of the MV entropy for the two real-world image sequence using the BMA and the TPA.

Using the BMA								
	Entropy of MV for Tourism				Entropy of MV for Moonwalk			
	Mean		std		Mean		std	
8×8	32.537		2.387		46.162		6.940	
4×4	40.146		3.575		57.922		3.561	
Using the MRF								
	Entropy of MV for Tourism				Entropy of MV for Moonwalk			
	Mean		std		Mean		std	
	11.278		2.425		23.817		3.584	
Using the TPA								
	Tourism				Moonwalk			
	Entropy of MV		Iterations		Entropy of MV		Iterations	
Block size	Mean	std	Mean	std	Mean	std	Mean	std
4×4	6.637	2.053	3.694	0.700	5.907	1.910	5.108	1.085
2×2	8.845	2.305	5.305	1.198	8.551	3.114	5.917	1.299
1×1	9.453	2.309	6.222	1.685	10.952	4.410	6.472	1.518

6 Conclusion

In this paper, a two-pass algorithm (TPA) employing the Mean Field Theory as the optimization method in the framework of MAP-MRF in a reliable manner is proposed for the purpose of motion estimation. In our algorithm, a preprocessing pass is initially applied to partition the set of sites into three different regions of different characteristics. In the second pass, the motion estimation is conducted, where two MRF's, namely, the motion vector field and the unpredictable field, are utilized. In this algorithm, the discontinuity is taken care of by a simple truncation function instead of introducing another MRF line field. Based on the partitions, the corresponding computation of the mean field is only incurred on the respective sites. Substantial amount of computations is saved with this algorithm. Because of the reduction of the number of MRF's, the partition of sites, and the inherited power from the MFT, compared to those existing scheme, a better balance is struck between the efficiency and effectiveness. In addition, due to the choice of the MAD which is more robust against outliers than the MSE, a more reliable estimation can be accomplished. Our experimental results substantiated this claim. Since no line field is induced in the estimation process, there exists little difference whether a site is a single pixel or a block. As such the proposed algorithm can be more flexible in achieving different tradeoffs. Through extensive experiments, it is observed that a satisfactory result can be achieved by simply applying the algorithm based on a 4×4 block, where the time consumed is only several times more than the BMA. This is superior to any other MAP-MRF motion estimation methods reported in the literature so far.

An efficient and effective method to obtain a representation of image sequences in terms of objects is of vital importance in the upcoming MPEG-4 and MPEG-7 standards [38]. The algorithm proposed in this paper, with its high efficiency, can be tailored to suit a different demand in practice due to the power of the MRF to model the contextual constraints and the MFT in approximating the global minima with reduced computational load. For instance, one extension of the algorithm is to obtain the object segmentation by use of the estimated motion field, where initial successes have been witnessed. Another extension is to accommodate the illumination alternations within image sequences by use of the techniques developed in our previous work [32, 9], where a generalized linear equality instead of the intensity constancy constraint is assumed between neighboring frames.

To further improve the efficiency of our algorithm, aside from decreasing the block size gradually, the hierarchical strategy can also be utilized. One can apply this algorithm to such pyramid representations of each frame as *Laplacian* [6] or wavelet [23] in a hierarchical manner [37].

References

[1] S. Barnard. Stereo matching. In R. Chellappa and A. K. Jain, editors, *Markov Random Fields*, pages 245–271. Academic Press, 1993.

[2] R. E. Bellman and S. E. Dreyfus. *Applied dynamic programming*. Princeton University Press, 1962.

[3] J. Besag. On the statistical analysis of dirty pictures (with discussions). *Journal of the Royal Statistical Society*, Series B 48:259–302, 1986.

[4] M.J. Black and A. Rangarajan. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *Int. J. of Computer Vision*, 19(1):57 – 91, 1996.

[5] A. Blake and A. Zisserman. *Visual Reconstruction.* The MIT Press, 1987.

[6] D. J. Burt and E. H. Andelson. The laplacian pyramid as a compact code. *IEEE Trans. on Commu.*, 31(4):532–540, 1983.

[7] D. Chandler. *Introductin to modern statistical mechanics.* Oxford University Press, 1987.

[8] R. Chellappa and A. K. Jain. *Markov Random Fields.* Academic Press, 1993.

[9] M. S. Drew, J. Wei, and Z.-N. Li. Illumination–invariant color object recognition via compressed chromaticity histograms of normalized images. In *Proc. Intl. Conf. on Computer Vision (ICCV'98)*, pages 533–540, 1998.

[10] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.

[11] B. Gidas. A renoumalization group approach to image processing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(2):164–180, 1989.

[12] J. M. Hammersley and P. Clifford. Markov field on finite graphs and lattices. *unpublished*, 1971.

[13] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics – the approach based on influence functions.* John Wiley and Sons, 1986.

[14] F. Heitz, P. Perez, and P. Bouthemy. Parallel visual motion analysis using multiscale markov random fields. In *Proc. IEE Workshop on Visual Motion*, pages 30–35, 1991.

[15] B. K. P. Horn. *Robot Vision.* MIT Press, 1986.

[16] P. J. Huber. *Robust Statistics.* John Wiley and Sons, New York, NY, 1981.

[17] J. Jain and A. Jain. Displacement measurement and its applications in interframe image coding. *IEEE Trans. on Commu.*, 29(12):1799–1808, 1981.

[18] S. Kirkpatrick, C. D. Gellatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

[19] T. Koga, *et al.* Motion compensated interframe coding for video conferencing. In *Proc. Nat. Telecommun. Conf.*, pages 5.3.1–5.3.5, 1993.

[20] J. Konrad and E. Dubois. Bayesian estimation of motion vector fields. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14:910–927, 1992.

[21] S. Krishnamachari and R. Chellappa. Multiresolution Gauss-Markov Random Field Models for texture segmentation. *IEEE Trans. on Image Processing*, 6(2):251–267, 1997.

[22] S. Z. Li. *Markov random field modeling in computer vision.* Springer-Verlag, 1995.

[23] S. Mallat. Multifrequency channel decompositions of images and wavelet models. *IEEE Trans. Acoust. Speech Signal Processing*, 17(12):2091–2110, 1989.

[24] D. Marr. *Vision.* W. H. Freeman, 1982.

[25] N. Metropolis. Equations of state calculations by fast computational machine. *Journal of Chemical Physics*, 21:1087–1091, 1953.

[26] G. Parisi. *Statistical Field Theory.* Addison-Wesley, 1988.

[27] C. Peterson and J. R. Anderson. A mean field theory learning algorithm for neural networks. *Complex Systems*, 1:995–1019, 1987.

[28] I. A. Rozanov. *Markov random fields.* Springer-Verlag, 1982.

[29] R. Srinivasan and K. R. Rao. Predictive coding based on efficient motion estimation. *IEEE Transactions on Commu.*, 33:888–896, 1985.

[30] C. Stiller. Object-based estimation of dense motion fields. *IEEE Trans. on Image Processing*, 6(2):234–250, 1997.

[31] A. M. Tekalp. *Digital Video Processing*. Prentice Hall, 1995.

[32] J. Wei and Z. N. Li. Motion compensation in color video with illumination variations. In *Proc. IEEE International Conference on Image Processing (ICIP'97)*, volume 3, pages 614–617, Santa Barbara, CA, 1997.

[33] W. Woo and A. Ortega. Stereo image compression with disparity compensation using the MRF model. In *Proc. visual Commu. and image processing(VCIP'96)*, 1996.

[34] E. Yair and A. Gersho. Maximum a posteriori decision and evaluation of class probabilities by Boltzman perceptron classifiers. *Proc. of IEEE*, 78(10):1620–1628, 1990.

[35] J. Zhang. Mean field theory in EM procedures for MRF's. *IEEE Trans. Signal Processing*, 40(10):2570–2583, 1992.

[36] J. Zhang and G. G. Hanauer. The application of mean field theory to image motion estimation. *IEEE Trans. on Image Processing*, 4(1):19–32, 1995.

[37] Y.-Q. Zhang and S. Zafar. Motion-compensated wavelet transform coding for color video compression. *IEEE Trans. on Circ. and Sys. for Video Tech.*, 2(3):285–296, 1992.

[38] Y.-Q. Zhang, *et al.* (ed). Special issues on MPEG-4. *IEEE Trans. on Circ. and Sys. for Video Tech.*, 7(1), 1997.