MEAN FIELD APPROXIMATION TO MULTIMODAL MOTION ESTIMATION PROBLEM

Thanh Dang Nguyen †, *Kalman Fazekas* ‡

† Department of Electrical Engineering, 200 Broun Hall
Auburn University, AL 36849, USA
e-mail: nguyet1@eng.auburn.edu

‡ Department of Microwave Telecommunications, Signal Processing Laboratory
Technical University of Budapest, Goldman ter 3, Budapest, H-1111, Hungary
Tel: +36 1 4631559; Fax: +36 1 4633289
e-mail: t-fazekas@nov.mht.bme.hu

ABSTRACT

The 2D Markov Random Field (MRF) model, combined with the Bayesian estimation framework, has proved to be an efficient and reliable computing tool to the optical flow estimation problem. Specifically, we are investigating the multimodal approach, where complementary constraints are imposed on the optical flow model. However, this approach suffers from expensive computational requirements, which is the direct consequence of the large dimensions of the optimization problem. Recently, a deterministic optimization technique, namely the mean field approximation has been proposed, which not only provides satisfactory estimation result, but also reduces the computational cost drastically. Here we apply this new technique to the above mentioned multimodal motion estimation problem.

1 INTRODUCTION

The application of MRF model-based technique in low-level visual motion analysis has been reported as an efficient and promising statistical approach due to its inherent ability to incorporate various sources of *a priori* information imposed by multiple constraints into the estimation process [6]. This model can be conveniently quantified using the Gibbs distribution and it is usually involved within the framework of the Bayesian estimation where the best estimate is determined from the *a posteriori* probability distribution based on a given error criterion.

Here a rather complex, but comprehensive motion field model is applied in order to entirely address the most troublesome difficulties that other local methods (e.g. block matching) usually suffer, e.g. proper detection of occlusion areas or motion boundaries. This multimodal approach (proposed by Heitz and Bouthemy in [3]) is essentially based on complementary constraints, where the validity of each contraint is also tested and included in the model. The leading idea here is that optical flow estimation should be achieved by estimating the motion information in those regions when reliable data and valid constraints are available, and then propagating this result to the ambiguous areas in the motion field.

In the Bayesian statistical framework, our task is equivalently formulated as global optimization of an energy function that contains the models of different local interactions [2][5]. Recently, the mean field approximation has been proposed as a very attractive deterministic optimization method, showing significant improvements in terms of both computational complexity (compared to the simulated annealing technique) and convergence characteristic (compared to the well-known ICM algorithm) [7]. In our work, we follow the same technique and extend it to the multimodal motion estimation problem. Experiments have been carried out on real-world image sequences in order to demonstrate the superior performance of the proposed technique.

2 2D MARKOV FIELD MODEL

Let $\mathbf{u}$ be the collection of random variables that are defined on the 2D lattice S to which a neighborhood system is defined. In the MRF approach, the joint probability distribution of $\mathbf{u}$ is usually formulated using the well-known Gibbs distribution as follows:

$$p(\mathbf{u}) = Z^{-1} e^{-\beta U(\mathbf{u})} \tag{1}$$

where Z is the normalization term usually referred to as the partition function, and $U(\mathbf{u})$ designates the global energy function. It can be defined as the sum of local potentials associated with all cliques ($c \in C$) defined on S according to the chosen neighborhood system:

$$U(\mathbf{u}) = \sum_c V_c(\mathbf{u}) \tag{2}$$

We can see that MRF model can be quantified using the clique potentials that describe the *local* interactions between different variables of the cliques. The lower the global energy, the higher the probability of its corresponding realizations. The main advantage provided by this model is that the global energy function $U(\mathbf{u})$ can be easily decomposed into the sum of different energy terms each of which representing different *a priori* local constraint that we claim on the field to be estimated.

3 MULTIMODAL MOTION ESTIMATION

In motion estimation, MRF models allow to jointly handle problems of optical flow estimation and issues of motion discontinuity and occlusion processing. The motion field model and observation model represented here are incorporated in the multimodal motion estimation scheme. In the following subsections we summarize the most important aspects of this model.

3.1 Motion Field Model

The motion field model is a coupled MRF field which consists of a velocity field $\mathbf{v}_{ij}$ assigning a local velocity vector to every site in the image plane and a set of discrete labels ν_{ij} representing motion boundaries [3]. These labels are given to edge sites which are located between the pixel sites, and can take one of the three values: 0 means that no motion discontinuity is present at the given site, where +1 or -1 value not only represents a motion discontinuity, but also helps to identify the occluded (and corresponding occluding) region. These labels can be easily interpreted as a superset of the well-known horizontal and vertical line fields introduced by Geman and Geman [2], except that here a given site can take three values as described above. Using these labels, we are able to quantify our *a priori* assumption on the motion field (segment-wise smoothness with motion discontinuities) as follows:

$$U(\mathbf{v}, \nu) = \sum_c \alpha(\|\mathbf{v}_s - \mathbf{v}_t\|^2)(1 - |\nu_{st}|) \tag{3}$$

where $\mathbf{v}_s$ and $\mathbf{v}_t$ designate the motion velocities of two neighbor sites in a given clique c, ν_{st} represents the label of the corresponding edge site, and α is a regularization parameter. However, in order to prevent discontinuities everywhere in the motion field, another energy term must be included that penalizes each time an edge site is created. We could also constraint the edge geometry in order to eliminate undesirable edge configurations [2][7].

3.2 Observation Model

Former observation models usually make use of the well-known optical flow constraint introduced by Horn and Schunck [4]:

$$\nabla I \mathbf{v} + I_t = 0 \tag{4}$$

where I denotes the image intensity, while ∇I and I_t are the spatial and temporal gradients, respectively.

However, these models usually ignore the question of the validity of this constraint in such areas as intensity or motion edges, occlusion regions and uniform areas. In fact, the optical flow equation assumes the local spatiotemporal linearity of the intensity function and that the first-order derivatives are not zero and well-defined. This is no longer true in the aforementioned ambiguous areas where additional constraint should be imposed in order to properly recover the motion field.

Heitz and Bouthemy [3] proposed a moving edge constraint that is very reasonable although it introduces some more complexity into the model including the need for an additional module that is responsible for detecting the moving edges and their corresponding velocities (only the component perpendicular to the edge can be detected due to the *aperture problem*) [1]. We should note that occlusion areas can be identified with more reliability using the moving edge constraint than other heuristic methods (e.g. assumption based on prediction error as in [7]).

The observation model involved here integrates these two complementary motion constraints and uses validation factors to appropriately apply these constraints where needed and disregard them in ambiguous areas, thus assuring reliable estimation and propagation. These validation factors are part of our *a priori* information and must be estimated and then fed as inputs to the Bayesian estimator.

Finally, as mentioned above, an energy term representing the cost of introducing an edge site into the motion field must also be considered. In effect, we favour the situation when motion boundaries exist where intensity edges were identified. Here a spatial edge map obtained by some edge detecting operator has been used as a constraint on the moving discontinuities. This map also serves as an initial estimate for the moving edge detector module.

3.3 Error Criterion

In order to make use of the mean field theory, the error criterion is defined here to be the Minimum Mean Squared Error (MMSE). It is well known from the literature that in this case, the Bayesian estimator's rule is equivalent to determine the conditional mean value, or in our case the mean field of the *a posteriori* function. This is where the mean field approximation comes into view.

4 MEAN FIELD APPROXIMATION

Given the MRF model defined in (1), its mean field is defined as follows:

$$\bar{\mathbf{u}}_{ij} = Z^{-1} \sum_{\mathbf{u}} \mathbf{u}_{ij} e^{-\beta U(\mathbf{u})} \tag{5}$$

It is clear that in order to compute the mean value of the field at a given site, all possible configurations must be taken into consideration. The mean field approximation, however, states that we can reasonably determine this value by assuming that the influence of the other sites can be fixed, that is, approximated by their mean values when evaluating the above expression [7]. In physical terms, this assumption is justified when a field is in the equilibrium state, where fluctuations are supposed to cancel each other, and this makes the influence of the different sites be globally well approximated by their means.

It can be pointed out that following this method, the computation at a given site reduces to:

$$\bar{\mathbf{u}}_{ij} \approx {Z_{ij}^{mf}}^{-1} \sum_{\mathbf{u}_{ij}} \mathbf{u}_{ij} e^{-\beta U_{ij}^{mf}(\mathbf{u}_{ij})} \tag{6}$$

where U_{ij}^{mf} is the *local* mean field energy function that depends only on $\mathbf{u}_{ij}$ and the mean values of its neighbor sites. It turns out from (6) that in order to compute the mean value at a site, only as many configurations are to be considered as the number of possible values the given site can take, and only the mean values of the neighbor sites are required. This fact substantially reduces the computational load represented by the multivariable optimization problem, and a closer look at the concept reveals that the computation process can be easily implemented in parallel. The whole process by its nature is deterministic and usually few iterations are needed to obtain the final estimate.

5 EXPERIMENTAL RESULTS

We use the Gaussian pyramid as our primary data structure in software implementation. This chosen approach is based on the following considerations: ability to deal with wide range of motion velocities, better convergence characteristics, and robustness to noise (e.g. derivative noise). We follow the coarse-to-fine propagation strategy. It saves the computational cost, and prevents the estimator from getting confused with details on

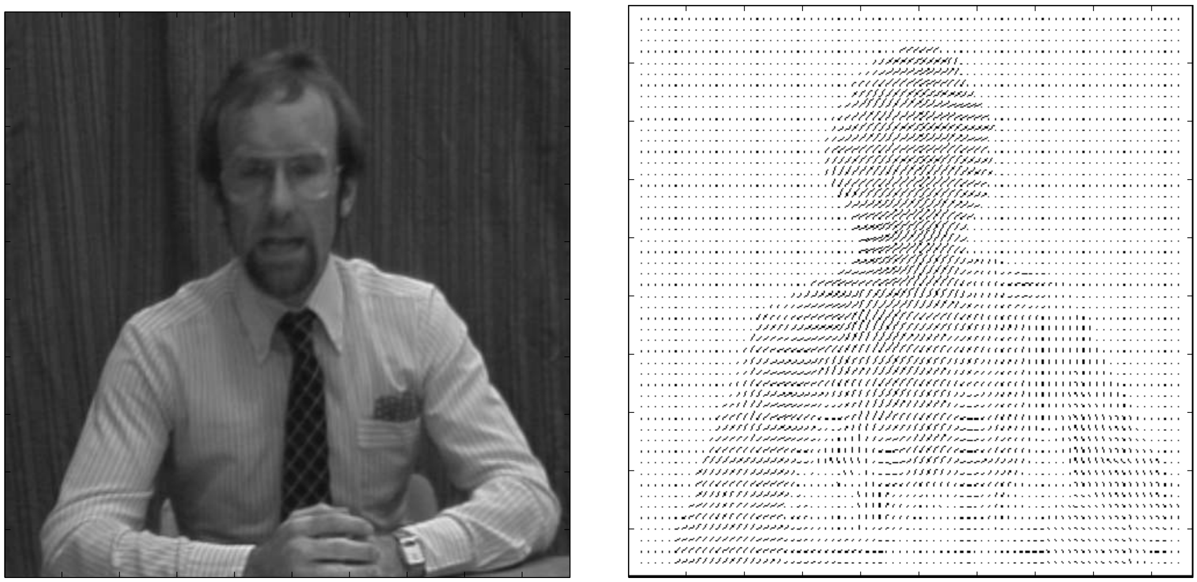

Figure 1: Simulation result with "Trevor White" test sequence: Only one frame and the corresponding displacement field are shown. Only every 4th motion vector is displayed.

finer levels (e.g. texture details can easily falsify the derivative information). The test images were taken from the standard "Trevor White" sequence. One typical frame and the corresponding computed motion field are shown here.

6 CONCLUSION

A new deterministic optimization technique, the mean field approximation, has been investigated and applied to the multimodal motion estimation problem. Experimental results have shown superior performance in terms of both convergence speed and quality of the recovered motion field. The computational cost can be further reduced significantly by using the multiresolution approach and parallel computation in implementation. Our future work involves the optimization of different regularization parameters, and the implementational issues.

7 REFERENCES

[1] P. Bouthemy, "A Maximum Likelihood Framework for Determining Moving Edges," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 499-511, May 1989.

[2] S. Geman, D. Geman, "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721-741, Nov. 1984.

[3] F. Heitz, P. Bouthemy, "Multimodal Estimation of Discontinuous Optical Flow Using Markov Random Fields," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 1217-1232, Dec. 1993.

[4] B.K.P. Horn, B.G. Schunck, "Determining Optical Flow," *Artificial Intelligence*, vol. 17, pp. 185-203, 1981.

[5] J. Konrad, E. Dubois, "Estimation of Image Motion Fields: Bayesian Formulation and Stochastic Solution," in *Proc. ICASSP'88*, pp. 1072-1075, 1988.

[6] G. Winkler, *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, Springer–Verlag, 1995.

[7] J. Zhang, G.G. Hanauer, "The Application of Mean Field Theory to Image Motion Estimation," *IEEE Trans. on Image Processing*, vol. 4, pp. 19-33, Jan. 1995.