Parallel Visual Motion Analysis
Using Multiscale Markov Random Fields

F. Heitz, P. Perez † and P. Bouthemy
IRISA/INRIA, †IRISA/CNRS
Campus Universitaire de Beaulieu, Rennes, France

Abstract

The use of Markov Random Field (MRF) models within the framework of global bayesian decision has recently brought new powerful solutions to visual motion analysis, [4, 8, 11]. The efficiency of MRF models for image sequence analysis has been proved on various classes of real-world sequences: outdoor and indoor scenes including several moving objects and camera motion. In this paper we extend this work by investigating new multiscale motion analysis algorithms based on MRF models. These algorithms are related to a new class of consistent multiscale MRF statistical models. The multiscale paradigm exhibits fast convergence properties towards quasi optimal estimates. Its performances are compared to standard relaxation in the case of optical flow measurement.

1 Introduction

Markov Random Fields (MRF) models have been successfully introduced in several important low-level issues of still image processing such as image restoration, [1, 6], edge detection, [5], image segmentation, [8], stereovision, computed tomography, or surface reconstruction. We have recently extended MRF models to the analysis of image sequences in motion detection, [11], visual motion estimation, [8] and motion-based segmentation, [4]. Markov Random Fields appear as an efficient and powerful framework for specifying non-linear spatial interactions between features of different nature (motion vectors and motion discontinuities for instance, [5]). When they are associated to the Maximum A Posteriori criterion, they lead to the minimization of a global energy function which may exhibit local minima, [6]. This minimization is generally performed using deterministic, [1] or stochastic, [6] relaxation algorithms.

Besides, it is known that "multigrid" methods can improve significantly the convergence rate of iterative relaxation schemes, [13]. As will be seen here, they may also be useful when the energy to be minimized presents many local minima. It has indeed been conjectured that multiscale analysis may, to a certain extent, smooth the energy landscape. Deterministic relaxation schemes can then be used at coarse scales to get a good initial guess, which may be refined over increasing resolution.

Multigrid relaxation algorithms have been considered early for visual motion analysis problems based on partial differential equations, [13] as well as on MRF, [10]. Many investigations have been pursued on classical multigrid methods for optical flow estimation, [3, 10]. Yet, in multigrid implementation of statistical models such as MRF, the key problem remains the derivation of the model parameters at different scales. When global mathematical consistency is not guaranteed, the parameters and the neighborhood structure associated to the model can only be adjusted over scale in an ad-hoc way. Gidas, [7] has recently described a consistent mathematical framework for multiscale Markov modeling, based on the renormalization group approach. Unfortunately, this standard technique in statistical physics does not yield tractable computational schemes (apart from simple cases such as the binary Ising model, [7].) Bouman et al., [2] have presented an attractive and tractable approach for multiresolution MRF-based estimation, with a general structure close to the model presented here, but the consistency problem is also addressed in an ad-hoc way.

The main contribution of this paper is the derivation of a multiscale algorithm suited to spatiotemporal MRF models which is both consistent and tractable. Models under concern are convenient for short and long range visual motion analysis. The algorithm (described in section 3) is related to a multiscale exploration of the set of solutions of the original optimization problem. This exploration is shown to be equivalent to a multigrid algorithm defined on a sequence of multiscale models. The approach extends a recent work of B. Levy, [12] on multiscale linear Gauss-Markov models to general non-linear Markov Random Field models. The multiscale method is compared to monoresolution schemes and classical multigrid algorithms in the case of optical flow computation (section 4). The multiscale algorithm exhibits fast convergence properties, comparable to multigrid determinis-
tic relaxation, and the final estimates are close to the near optimal solutions obtained by time consuming stochastic methods. Moreover, it is a good candidate for parallel implementation on regular arrays consisting of locally interconnected processors operating synchronously, [9].

2 Visual motion analysis and MRF models

The general mathematical framework under concern in this paper is based on global bayesian estimation along with MRF models. The goal of this theory is to extract a 2D field of discrete labels describing the information of interest in a given image processing task from a given set of observed variables. In the optical flow measurement issue, for instance, the observations correspond to the original image sequence, the labels to be estimated are displacement vectors, [8].

In recent papers we have designed several new spatio-temporal MRF models for motion detection, [11], optical flow measurement, [8] and motion-based segmentation, [4]. These models describe the interactions between the spatio-temporal variations of the observed image sequence and the features of interest in a given problem: binary variables in the detection case, displacement vectors in the estimation case and region labels for the motion-based segmentation.

The solution of those "low-level" image processing issues is of course an important preliminary step in most applications involving a kinematical interpretation of the behavior of moving objects in a scene, [4].

Let \( O = \{0_s, s \in S\} \) designate an observation field defined on a rectangular lattice \( S \) (in image sequence analysis, observations are related to the spatiotemporal variations of the intensity function). Let \( E = \{E_s, s \in S\} \) denote the unobserved (hidden) motion label field, defined on the same lattice \( S \). Realizations of fields \( O \) (resp. \( E \)) will be denoted \( o = \{o_s, s \in S\} \) (resp. \( e = \{e_s, s \in S\} \)). Let \( \Lambda \) be the (discrete) state space of random variable \( e_s \) and \( \Omega \) the (finite) set of all possible label configurations \( e = \{e_s, s \in S\} \). If \( (E, O) \) is assumed to be a Markov random Field with respect to a chosen neighborhood system \( \mathcal{G} = \{G_s, s \in S\} \), it can be shown, [8] that the best estimate \( \hat{e} \) of \( e \) given \( o \) (according to the Maximum A Posteriori - MAP criterion) is obtained by minimizing a global energy function \( U(o, e) \):

\[
\hat{e} = \arg \min_{e \in \Omega} U(o, e) \quad (1)
\]

Thanks to the markovian property of the field, the energy decomposes as follows, [6]:

\[
U(o, e) = \sum_{e \in \mathcal{C}} V_c(o, e) \quad (2)
\]

\( \mathcal{C} \) denotes the set of cliques associated to neighborhood system \( \mathcal{G} \). Clique \( c \) are subsets of sites which are mutual neighbors. The potential function \( V_c \) is locally defined on clique \( c \) and expresses the local interactions between the different variables of the clique (see [6] for an extensive theory of MRF). The form of the potential functions \( V_c \) is of course problem dependent and defines the local and global properties of the model.

Minimizing the global energy function \( U \) is often an intricate problem: the number of possible label configurations is generally very large and, moreover, the global energy function \( U \) may contain local minima. Computationally demanding stochastic relaxation algorithms are therefore generally necessary to compute exact MAP solutions. Less cpu-time consuming deterministic relaxation algorithms such as ICM, [1], RCF (Highest Confidence First) can often be used instead, when a good initial guess is available (other deterministic approaches include Graduated Non Convexity and Mean field Annealing). Multigrid computational schemes have also proved very efficient for minimizing energy function \( U(e, o) \) in visual motion analysis applications, [10]. In section 3 we describe a new multiscale paradigm which has revealed particularly attractive as far as the minimization of \( U(e, o) \) is concerned.

A model for optical flow measurement

As an illustration, we consider a standard model for optical flow measurement which has been used in [10]. In this model the velocity vectors belong to a discrete state space \( \Lambda \). We choose a discrete formulation of the optical flow computation problem because it is known to lead to a complex energy landscape exhibiting numerous local minima. Hence this model appears as a difficult example as far as optimization is concerned and will be useful for comparison purpose between standard and multiscale models.

Let \( f_t(s) \) denotes the observed intensity function, where \( s = (x, y) \), \( s \in S \) designate the 2D spatial image coordinates and \( t \) the time axis. The velocity vector at point \( s \) is denoted \( \omega_s \), \( u_s = \frac{\partial f}{\partial t}(s) \), \( v_s = \frac{\partial f}{\partial x}(s) \) and \( \Omega = \{\omega_s, s \in S\} \). In the model considered here velocities are defined on the same grid \( S \) as the pixels and the velocities are discretized according to a discrete state space \( \Omega = (-u_{\max}, u_{\max}) \) with a step size of \( \delta \). The MRF model is associated to a 8-neighborhood and specified by the following energy function:

\[
U(f, \omega) = \sum_{s \in S} \{ f_t(s) - f_{s+t}(s + \omega_s, dt) \}^2 + \alpha^2 \sum_{(s,s) \text{ neighbors}} \|\omega_s - \omega_t\|^2 \quad (3)
\]

The first term in the energy (known as the "displaced frame difference") expresses the constant

\[
\frac{\partial f}{\partial t}(s) - f_{s+t}(s + \omega_s, dt)
\]
brightness assumption for a physical point over time. The second term balances the first one through the weighting parameter $\alpha$; it can be interpreted as a regularization term which favors smooth solutions.

Classical implementations of a multiresolution algorithm based on this model have been reported in [10]. We now present our own approach to the problem.

### 3 Multiscale MRF models

Our approach is based on a "multiscale exploration" of the space of solutions of the optimization problem (2). We show that this exploration is equivalent to a multigrid algorithm defined on a set of multiscale MRF models. This multigrid algorithm has several attractive properties.

- It allows to propagate local MRF interactions more efficiently between distant points, yielding faster convergence;
- The energy functions at different scales do not present the same "landscape". As a matter of fact, it has been conjectured, [10] that in low resolution images the energy landscape is "smoother" and may contain fewer local minima than the energy associated to the original full resolution image. Whereas the deterministic relaxation algorithm gets stuck in the first local minimum near the starting configuration, estimates closer to the global optimum may be expected from a multigrid method starting at a low resolution level where less local minima exist.

In the following we will focus on a particular class of MRF models with the following energy function:

$$U(e, o) = U_1(e, o) + U_2(e)$$

with:

- $U_1(e, o) \triangleq \sum_{s \in S} V_1(e_s, o_s)$;
- $U_2(e) \triangleq \sum_{c \in C} V_2(e)$

where a 8-neighborhood is adopted for $G$. The cliques $c \in C$ associated to this neighborhood structure contain at most four sites. The energy function presented in Section 2 obviously belongs to this class. Moreover, this class comprises most usual energy functions used in still and dynamic image analysis problems, [6, 8] (the extension to larger neighborhoods is straightforward, although not only a notational matter).

The multiscale MRF model can be described as follows. First recall that the MRF model assigns an energy to all possible configurations of a field $e = \{e_s, s \in S\}, e \in \Omega$. Let us assume that the size of the full resolution lattice is $2^m \times 2^m$. To generate a "coarse grid" MRF model, let us divide the initial grid $S$ into cells (blocks) of $4 \times 2 \times 2$ pixels. Among all configurations in $\Omega$ we consider a subset $\Omega_i$ where $e$ is constant over each of the above defined 2x2 cells (Fig. 1).

These configurations will describe the MRF at scale 1, at which the size of a "constant value cell" is $2^1 \times 2^1$.
This partition yields the following decomposition for energy function $U_2(e)$:

$$U_2(e) = \sum_{B_k^* \in B^*} V_k^*(e_k^*)$$

$$+ \sum_{\{B_{k^1}, B_{k^2}\}_k} V_{k^1, k^2}^*(e_{k^1}, e_{k^2})$$

$$+ \sum_{\{B_{k^1}, B_{k^2}, B_{k^3}\}_k} V_{k^1, k^2, k^3}^*(e_{k^1}, e_{k^2}, e_{k^3})$$

$$+ \sum_{\{B_{k^1}, B_{k^2}, B_{k^3}, B_{k^4}\}_k} V_{k^1, k^2, k^3, k^4}^*(e_{k^1}, e_{k^2}, e_{k^3}, e_{k^4})$$

$$\text{with: } V_k^*(e_k^*) \triangleq \sum_{e \in e_k^*} V_k(e)$$

$$V_{k^1, k^2}^*(e_{k^1}, e_{k^2}) \triangleq \sum_{e \in e_{k^1, k^2}} V_k(e)$$

$$V_{k^1, k^2, k^3}^*(e_{k^1}, e_{k^2}, e_{k^3}) \triangleq \sum_{e \in e_{k^1, k^2, k^3}} V_k(e)$$

$$V_{k^1, k^2, k^3, k^4}^*(e_{k^1}, e_{k^2}, e_{k^3}, e_{k^4}) \triangleq \sum_{e \in e_{k^1, k^2, k^3, k^4}} V_k(e)$$

Equivalent coarse resolution models

$$\Phi^i : \Omega \rightarrow \Omega$$

$$\Phi^i (e^i)$$

$$\Phi^{i+1} (e^{i+1})$$

$$e^0$$

$$e^i$$

$$e^{i+1}$$

Figure 3: Isomorphism $\Phi^i$ between a configuration in $\Omega_i$ on which the coarse model is defined and a configuration in $\Omega_{i+1}$

Considering equations (5) and (7), it turns out that the energy of a configuration $e \in \Omega_i$ can be decomposed as a sum of potential functions, which are defined on mutually neighboring cells $B_k^*$ and only depend on the values $e_k^*$ in those cells and on the whole observation field $\omega$. Thanks to that decomposition, the energy of a configuration $e \in \Omega_i$ can be computed very efficiently using an equivalent coarse model corresponding to a label field whose resolution has been reduced by a factor given by the cell size $2^i \times 2^i$ (Fig. 3). The resulting coarse grid, denoted $S^i$ is isomorphic to $B^i$. The coarse label field $E' = \{E_s' \in S^i\}$ is associated to a reduced configuration space $\Gamma_i = \Lambda^{|S^i|}$ isomorphic to $\Omega_i$, with $\Gamma_0 = \Omega_0 = \Omega$. Let $\Phi^i$ denote the bijection from $\Gamma_i$ in $\Omega_i$:

$$\Phi^i : \Gamma_i \rightarrow \Omega_i$$

$$\Phi^i (e^i)$$

defined as an interpolation from the coarse grid $S^i$ to the fine grid $S^0 = S$ using simple repetition of the label estimates on the fine grid $S$ (Fig. 3). The sequence of models on grid $S^i$, $i = n, \ldots, 0$ defines a set of consistent multiscale Markov Random Field models, whose energy functions $U^i(e^i, \omega)$ are given by:

$$U^i(e^i, \omega) = U_1(\Phi^i(e^i), \omega) + U_2(\Phi^i(e^i))$$

$$U^i(e^i, \omega) = U_1(\Phi^i(e^i), \omega) + U_2(\Phi^i(e^i))$$

(10)

Let us note that the coarse models have the same neighborhood structure as the original fine grid Markov Random Field (equ(7)) and the potential functions at coarse resolution are derived easily from the characteristics of the original model, by equations (6) and (8). This approach to multiscale MRF can be interpreted as a constrained exploration of subsets of the original (generally very large) configuration space $\Omega$. In these subsets $\Omega_i$ configurations are constrained to remain constant in cells of size $2^{i-1} \times 2^{i-1}$

A multigrid relaxation algorithm

To take benefit from the sequence of multiscale models previously defined, instead of handling directly the original optimization problem:

$$\hat{e} = \arg\min_{e \in \Omega} U(e, \omega)$$

(11)

we consider the following sequence of problems:

$$\hat{e}^i = \arg\min_{e' \in \Gamma_i} U(e^i, \omega), \quad i = n, \ldots, 0$$

(12)

These are equivalent to a minimization on the coarse MRF models, i.e.:

$$\hat{e}^i = \arg\min_{e' \in \Gamma_i} U^i(e^i, \omega), \quad i = n, \ldots, 0$$

(13)

(with $e' = \Phi^i(e')$). This multigrid estimation strategy is a usual coarse to fine one: starting from a coarse scale $n$, the optimization problem (equ. (11)) is first solved in subset $\Omega_n$ (equ. (12)) by solving the equivalent problem (13). An estimate of $e^i$ is obtained by a deterministic relaxation algorithm known as ICM, [1]. At level $i$ if $e^i$ designates the estimate of $e^i$ (obtained after convergence of the deterministic relaxation at that level), the algorithm at resolution level $i-1$ is initialized by $[\Phi^{i-1}]^\top \circ \Phi^i(e^i)$, which is just an interpolation of $e^i$ on the finer grid $S^{i-1}$

This approach to MRF-based multiresolution image analysis has several advantages.
it is globally consistent and can easily be interpreted as constrained optimization,
- as already said, it involves only a label pyramid but does not require a pyramid of observations to be constructed: the observations at full resolution are used,
- it exhibits fast convergence properties applied to our non-linear optimization problems,
- as will be seen in the following section, it yields better results (convergence to lower energy values) than classical multigrid methods.

4 Application to motion measurement

The multiscale paradigm has been implemented for different MRF models. We present here the performances for the optical flow measurement model (equ. 3) (results on motion detection, [11] may be found in an extended version of this paper, [9]).

Four algorithms have been tested and compared:
1) a standard monoresolution deterministic relaxation algorithm (DR) known as ICM, [1];
2) a monoresolution stochastic relaxation (SR) algorithm based on the Gibbs sampler, [6];
3) a classical coarse to fine multigrid relaxation (MGR) algorithm in which the same model is considered at each resolution (same parameters, same neighborhood system and same potential functions) and a pyramid of observations is constructed. Similar implementations of this third version may be found in [10];
4) the proposed multiscale relaxation (MSR), in which the energy function at each resolution level are computed in a consistent way as explained previously.

The temperature schedule used in the stochastic relaxation algorithm was: \( T(j) = T_0 \cdot A^j \), \( A = 0.97 \) and \( T_0 = 300 \), where \( j \) designates the number of sweeps on the image. The same parameters were chosen in every case for the finest resolution model. Four resolution levels are considered in the multigrid and multiscale methods. For these schemes the same deterministic algorithm (ICM, [1]) was run at the different resolutions. The relaxation was performed until convergence was reached on the label fields (at each resolution). As a matter of fact, the process was stopped as soon as the number of sites changing their state between two complete image sweeps went below some specified threshold \( \lambda \).

We do not reproduce here the equations giving the multiscale energies for motion measurement: they are obtained from equations (5), (6), (7), (8) and (10).

Our experiments clearly show that the discrete formulation of motion measurement (equ. 3) yields an energy function which contains many local minima. The results reported here have been obtained with random initialization for the label field. We present results on a real world sequence (the highway sequence, Fig. 4) and on a benchmark composed of 41 sequences of two 64x64 frames obtained by applying the same synthetic motion to different real images. The synthetic motion included translation, rotation and dilation on the same image. In Fig. 5 are drawn the velocity vectors estimated by each of the four algorithms for the highway sequence. The multiscale approach supplies fields close in quality to the one obtained by stochastic relaxation. Table 1 presents the number of iterations and the final energy value reached by each algorithm. On this sequence, the multiscale approach performed even better than the time consuming stochastic relaxation. The standard methods (DR,MGR) remain stuck in rather high energy local minima. Those results are confirmed on the 41 sequences-benchmark. Fig. 6 presents the ratio of the final energy value reached by stochastic relaxation (SR) to the one obtained by the different other methods. The corresponding number of iterations required in each case is plotted in Fig. 7. In the average, the

![Figure 4: The "highway sequence"](image)

![Figure 5: Optical flow fields ("highway sequence")](image)

<table>
<thead>
<tr>
<th>algorithm</th>
<th>SR</th>
<th>DR</th>
<th>MGR</th>
<th>MSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>nb-iter</td>
<td>259</td>
<td>32</td>
<td>16.88</td>
<td>7.46</td>
</tr>
<tr>
<td>final energy</td>
<td>760427</td>
<td>1117945</td>
<td>811468</td>
<td>749471</td>
</tr>
</tbody>
</table>

Table 1: Motion measurement: number of iterations required to reach convergence (nb-iter: computational equivalent to the number of complete sweep through the image at the finest resolution)
MSR method finds configurations close to the best estimates obtained by SR, with an improvement of up to two orders of magnitude on the number of iterations.

Those results demonstrate the efficiency and the robustness of the proposed multiscale paradigm when an energy function, presenting a great number of local minima, is to be minimized. We are currently implementing the method on a coupled MRF model, with a complex energy landscape, enabling both continuous motion measurement ([8]) and motion-based segmentation ([4]). Besides, we also study general parallelization approaches associated to the multiscale algorithm. This algorithm is indeed a good candidate for parallel implementation on regular arrays consisting of locally interconnected processors; [9].

Acknowledgments

This work has been partly supported by MRT and CNRS in the context of the PRC Program "Man Machine Interface" under contract PMFE 88F 548 and by Brittany County Council under contribution to student grant. We thank B. Levy for fruitful discussions.

References


5 Conclusion

In this paper we have presented a new multiscale approach for visual motion analysis based on Markov Random Field modeling. It presents several attractive features when compared with standard multigrid methods: mathematical consistency, fast convergence toward robust and quasi-optimal estimates. The performances of the multiscale method have been demonstrated in optical flow estimation and may be applied to other visual motion analysis problems.

Figure 6: Ratio of the final energy value reached with SR to the one obtained with the other methods

Figure 7: Number of iterations at convergence