Efficient MRF Deformation Model for Non-Rigid Image Matching ⋆

Alexander Shekhovtsov [a] Ivan Kovtun [b] Václav Hlaváč [a]

[a] *Czech Technical University in Prague, Center for Machine Perception, Czech Republic*

[b] *International Research and Training Center for Informational Technologies and Systems, Kiev, Ukraine*

Abstract

We propose a novel MRF-based model for deformable image matching (also known as registration). The deformation is described by a field of discrete variables, representing displacements of (blocks of) pixels. Discontinuities in the deformation are prohibited by imposing hard pairwise constraints in the model. Exact maximum a posteriori inference is intractable and we apply a linear programming relaxation technique.

We show that, when reformulated in the form of two coupled fields of x- and y- displacements, the problem leads to a simpler relaxation to which we apply the TRW-S (Sequential Tree-Reweighted Message passing) algorithm [Wainwright-03, Kolmogorov-05]. This enables image registration with large displacements at a single scale. We employ fast message updates for a special type of interaction as was proposed [Felzenszwalb and Huttenlocher-04] for the max-product Belief Propagation (BP) and introduce a few independent speedups. In contrast to BP, the TRW-S allows us to compute per-instance approximation ratios and thus to evaluate the quality of the optimization. The performance of our technique is demonstrated on both synthetic and real-world experiments.

Key words: Markov random fields, MRF, message passing, TRW-S, energy minimization, motion estimation, optical flow, image registration.

⋆ A preliminary version of this paper was presented at the International Conference on Computer Vision and Pattern Recognition, 2007 [1].

1 Introduction

Substantial advances in low-level vision problems like stereo, segmentation, image de-noising, etc., have been achieved by using random field models such as Markov Random Fields (MRFs). This is mainly due to new algorithms for inference in MRFs, in particular algorithms for finding maximum *a posteriori* (MAP) configurations. Szeliski *et al.* [2] gave an experimental comparison of recent optimization algorithms for popular vision problems. More background on inference and learning algorithms for MRFs can be found in, *e.g.*, [3].

In this work, we consider the problem of two-dimensional matching of non-rigid objects. We model the non-rigid deformation as a random field of discrete displacements with a prior model favoring continuous deformations. The optimal deformation is sought as a MAP configuration of the random field. Finding a globally optimal or even a reasonably good suboptimal solution of this problem is a difficult task. However a potential advantage is that harder matching problems can be solved thanks to the use of global optimization techniques. While this approach was investigated by several authors, we focus on the case where the displacement field vectors have a large admissible range and optimization is performed at a single scale. By considering an equivalent reformulation of the MRF model with separated variables and applying the LP-relaxation technique, we get a simpler relaxed problem. This allows us to achieve comparable speed to the coarse-to-fine methods while maintaining the desirable property that the method is a global convex relaxation (with a proviso that we solve the relaxation problem suboptimally using TRW-S for the sake of speed). We also discuss a better heuristic for obtaining a discrete solution from the relaxed problem for this specific application.

1.1 Related Work

Complex random field models for the optical flow problem were proposed as early as the simplest inference algorithms. Konrad and Dubois [4] and Heitz and Bouthemy [5] considered models of coupled vector-binary fields, where the vector field represents motion and the binary field represents motion discontinuities (*i.e.*, some lines along object boundaries). A prior model for discontinuities favors straight lines as object boundaries and prohibits their intersections, whereas a prior model for motion vector field imposes smoothness constraints in areas where discontinuities are not present. Zhang and Hanauer [6] considered an additional component: a binary field describes a segmentation identifying areas of uncovered background, for which no correct motion field can be computed. The optimization techniques used to derive the MAP solution (or an approximation to it) range from Iterated Conditional Modes [5] to simu-

lated annealing [4] and mean field approximation [6]. Even though computing power has increased significantly and new optimization methods have been developed, the difficult optimization problem arising in MAP estimation still represents a bottleneck.

Roy and Govindu [7] modeled orientation and magnitude fields of the flow as separate MRFs. They first solved for orientation field and then for the magnitude. Each of these problems is reduced to computation of max-flow under their model.

Boykov *et al.* [8] introduced α-β swap algorithm and considered the problem of 2D motion estimation as an example. Variables in their model take values from the product set of allowed displacements $\Delta X \times \Delta Y$. Optimization by α-β swap algorithm for this MRF becomes quickly intractable for larger displacements ΔX and ΔY.

Kumar *et al.* [9] computed piecewise rigid deformation of consecutive frames in a video sequence. Deformation was locally described by translation, rotation and scale. The MRF model with states being discretized deformations and pairwise potentials encouraging rigid motion of fragments was optimized using the sum-product Belief Propagation algorithm (BP). Due to the large state-space, a coarse-to-fine approach was applied to reduce memory and computational requirements.

Jiang *et al.* [10] used linear programming to solve a convexified image matching problem. Their formulation uses small number of basis matching pairs, which leads to a small scale linear program, which can be solved by a simplex method. They iteratively update the convex approximation around the current solution.

Glocker *et al.* [11,12] applied primal-dual approximation schema [13] based on the linear programming relaxation technique to the discrete search problem of finding a better configuration in the vicinity of the current one. This is efficiently a trust region optimization method where global optimization is applied in a local neighborhood, which has many advantages over the classical gradient-based methods. Larger deformations are handled via the coarse-to-fine scaling technique.

Felzenszwalb and Huttenlocher [14] applied max-product Belief Propagation to the optical flow problem. Their model had $\Delta X \times \Delta Y$ labels. The efficient updates and coarse-to-fine message propagation techniques allowed them to speed-up the computation significantly. We are using a similar algorithm (but with provable lower bounds and convergence properties). We apply some of their techniques and introduce additional speed-ups.

1.2 Outline

In Section 2, we introduce the matching problem, consider the standard random field model and the decomposed model [15] and show how the model can be made more robust by means of hard continuity constraints. Next we discuss the relaxation technique and the TRW-S algorithm and its efficient implementation in Section 3. We show how a message passing schedule can be modified to speed up information propagation in the model and how a suboptimal solution may be obtained by gradual greedy decisions on subsets of variables. In our experiments, presented in Section 5, we study the performance of the modified TRW-S algorithm and provide statistics of how well this suboptimal approach approximates the global optimization problem. Finally we demonstrate that qualitatively larger displacement ranges are feasible and that the solution provided is robust to noise and clutter. We conclude with possible extensions in Section 6.

2 Deformation Model

In this section, we review the simplest MRF model for the 2D deformation, considered in *e.g.* [8], and show how it can be alternatively formulated in terms of separated variables.

2.1 Maximum Likelihood Matching

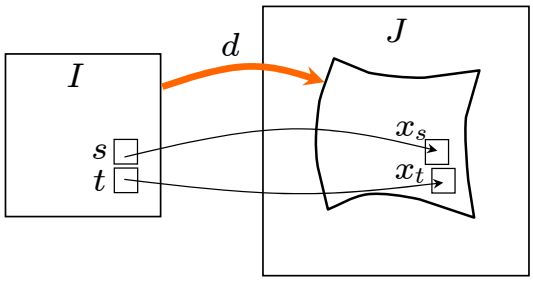

Fig. 1. A non-symmetric formulation of matching image I to image J. Two neighboring pixels are shown – the interaction potential on the pair of their displacements is set such that the cost of matching to spaced locations is penalized.

We formulate the image matching problem as a maximum *a posteriori* inference in a constructed probabilistic model (see *e.g.* [16]). Let I and J be two images to be registered. Let $I\colon T_1 \to \mathcal{C}$ and $J\colon T_2 \to \mathcal{C}$, where T_1 and T_2 are sets of locations and $\mathcal{C}$ is the set of signals. Let d be a mapping taking pixels from domain T_1 of I to domain T_2 of J. We assume that d is an injective

mapping, meaning that not all pixels from T_2 will have a preimage. This is illustrated in Fig. 1. This non-symmetric formulation is convenient since it allows image I to be matched to a subregion of J.

Treating images I, J and the mapping d as random variables, we model their statistical dependence as

$$p(I, d \mid J) = p(I \mid d, J)p(d). \tag{1}$$

Assuming that pixel signals in I are conditionally independent given d and J, we write

$$p(I \mid d, J) = \prod_{t \in T_1} p(I_t \mid d, J), \tag{2}$$

in addition, when pixel t is mapped to pixel $d(t)$, we consider that its signal I_t depends only on the corresponding signal $J_{d(t)}$ and does not depend on the rest of signals in J. This assumption is expressed probabilistically as

$$p(I_t \mid d, J) = p(I_t \mid J_{d(t)}). \tag{3}$$

The conditional distribution $p(c_1|c_2)$ will be assumed fixed and given by, *e.g.*, a Gaussian noise model. The prior distribution $p(d)$ expresses which deformations are more and which are less likely. Here, we model $p(d)$ as a pairwise MRF and consequently the conditional mode $p(I, d|J)$ is also a pairwise MRF.

2.2 Energy Minimization

Let us review basic definitions of MAP inference in MRF. More background on graphical models and MRFs can be found in, *e.g.*, [3,17,18]. Let $\mathcal{L} = \{1 \dots K\}$ be a set of labels. Let $G = (\mathcal{V}, \mathcal{E})$ be a graph, let $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ be antisymmetric and antireflexive, *i.e.*, $(s,t) \in \mathcal{E} \Rightarrow (t,s) \notin \mathcal{E}$. We will denote by st an ordered pair $(s,t) \in \mathcal{E}$. Let each graph node $s \in \mathcal{V}$ be assigned a label $x_s \in \mathcal{L}$ and let a *labeling* (or *configuration*) be denoted as $\mathrm{x} = \{x_s \mid s \in \mathcal{V}\}$. Let $\{\theta_s(i) \in \mathbb{R} \mid i \in \mathcal{L},\ s \in \mathcal{V}\}$ be *univariate* potentials (often referred to as data terms) and $\{\theta_{st}(i,j) \in \mathbb{R} \mid i,j \in \mathcal{L},\ st \in \mathcal{E}\}$ be *pairwise* potentials (or interaction terms). Let the *energy* of a configuration x be defined by:

$$E(\mathrm{x}|\theta) = \sum_{s \in \mathcal{V}} \theta_s(x_s) + \sum_{st \in \mathcal{E}} \theta_{st}(x_s, x_t)\,. \tag{4}$$

The probability distribution defined by $p(\mathrm{x}|\theta) \propto \exp(-E(\mathrm{x}|\theta))$ is Gibbs distribution which corresponds to a certain MRF. The problem of finding a maximum *a posteriori* configuration of this MRF corresponds to the *energy minimization*, $\min_{\mathrm{x}} E(\mathrm{x}|\theta)$.

2.3 Product Model

Let configuration x with components $x_s = (x_{s,1}, x_{s,2})$, $s \in \mathcal{V} = T_1$, be a 2D displacement field over T_1. Coordinates $x_{s,1}$ and $x_{s,2}$ denote first and second components of the displacement vector at the pixel s. Let both coordinates take values from $L = \{K_{\min}, \ldots, K_{\max}\}$, thus variables $x_s = (x_{s,1}, x_{s,2})$ take their values from the set $\mathcal{L} = L^2$. Deformation d maps locations s to locations $t = s + x_s$, where $s + x_s$ is a shorthand for a corresponding location displaced by x_s (see Fig. 1). Let univariate potentials be

$$\theta_s(x_s) = -\log p(I_s | J_{d(s)}). \tag{5}$$

Under the Gaussian noise model it is just the well-known sum of square differences (SSD) dissimilarity measure, $\rho(I_s - J_{s+x_s})^2$, where ρ is a distance metric in the color space $\mathcal{C}$.

The usual prior on d in MRF models is set as:

$$\theta_{st}(x_s, x_t) = \lambda \left\| x_s - x_t \right\|^2 = \lambda(x_{s,1} - x_{t,1})^2 + \lambda(x_{s,2} - x_{t,2})^2, \quad st \in \mathcal{E}, \tag{6}$$

where $\mathcal{E}$ is the set of all horizontally and vertically neighboring pairs of pixels. This term penalizes discontinuities in the deformation field x, so that close pixels are forced to match to close destinations.

2.4 Decomposed Model

We can exploit the property that continuity interaction (6) is separable (represented as a sum of x- and y- interactions) and reformulate the energy model (following [15]) as two interacting fields of scalar variables. Graph G in this model is constructed as follows (see Fig. 2): nodes form two layers $\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2$, with $\mathcal{V}_1 \sim \mathcal{V}_2 \sim T_1$ (here, $\sim$ denotes that sets $\mathcal{V}_1$, $\mathcal{V}_2$ and T_1 are isomorphic, i.e., copies of each other); edges $\mathcal{E} = \mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_{12}$, where $\mathcal{E}_1$ (resp. $\mathcal{E}_2$) is the set of vertically and horizontally neighboring pixel pairs in $\mathcal{V}_1$ (resp. $\mathcal{V}_2$), and $\mathcal{E}_{12} = \{(s_1, s_2) \mid s_1 \in \mathcal{V}_1, s_2 \in \mathcal{V}_2, s_1 \sim s_2\}$ is the set of *inter-layer* edges (here, $s_1 \sim s_2$ denotes that these elements correspond in the isomorphism $\sim$). Let $\mathrm{x} = \{x_{s_i} \mid s_i \in \mathcal{V}_i,\ i = 1, 2\}$. The data term in this model is encoded in the interaction pairs from $\mathcal{E}_{12}$, as

$$\theta_{st}(x_s, x_t) = -\log p(I_s \,|\, J_{s+(x_s,x_t)}), \quad s \in \mathcal{V}_1,\ t \in \mathcal{V}_2,\ st \in \mathcal{E}_{12}. \tag{7}$$

The continuity term is set identically for both layers as

$$\theta_{st}(x_s, x_t) = \lambda(x_s - x_t)^2, \quad st \in \mathcal{E}_i,\ i = 1, 2. \tag{8}$$

It can be seen that the resulting energy function $E(\mathrm{x}|\theta)$ is equivalent to that of the product model. In order to apply this alternative representation, it is necessary that the continuity terms are separable into x- and y- components. Let us also note that this decomposition can be applied also for 3D image registration, however in this case we would need to use ternary interactions to encode the local data similarities.

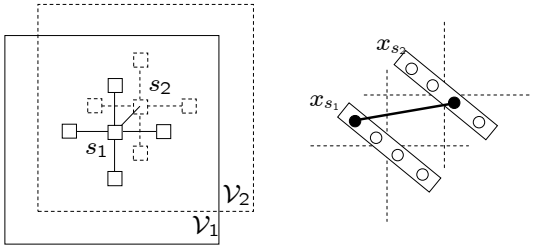

Fig. 2. Decomposed model. Left: the model consists of two layers $\mathcal{V}_1$ and $\mathcal{V}_2$, the neighborhood structure is shown by nodes s_1, s_2 and edges to their other neighbors in $\mathcal{E}$. Right: the inter-layer interaction is used to encode the data term. The 2D displacement of a pixel s is determined by a pair of labels x_{s_1}, x_{s_2} (filled circles) being the two spatial coordinates; the data fitness term for the displacement is encoded on the edge (x_{s_1}, x_{s_2}) (bold line).

2.5 Block Model with Hard Constraints

It is important in some applications to have hard constraints on the deformation that ensure its continuity no matter what input image data is used. In particular, it is important when clutter or unmodelled effects are present in the images. One of the advantages of the discrete formulation is that certain hard constraints of this type can be easily incorporated.

Another issue is that keeping a displacement variable for each separate pixel is usually too wasteful, while decreasing the image resolution is a loss of information and is undesirable.

We address this two issues by aggregating pixels into blocks and allowing each block to have displacements with a pixel precision. We do not penalize relative displacements of ± 1 pixels in vertical and horizontal directions, and completely forbid larger displacements. See Fig. 3 for examples of possible configurations if 4x4 blocks are considered. Let set T_1 be regularly subdivided into square blocks and let B be a set of these blocks. We assume the horizontal and vertical neighborhood of the blocks. We let $\mathcal{V}_1 \sim \mathcal{V}_2 \sim B$ and construct the graph G as in the decomposed model. The data term will just add up the

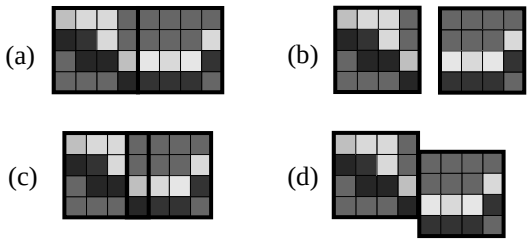

Fig. 3. Block model: (a) two neighboring blocks of 4×4 pixels; (b),(d) examples of nonpenalized relative displacements. There are nine nonpenalized relative displacements in total.

contribution from individual pixels,

$$\theta_{bb'}(x_b, x_{b'}) = \sum_{s \in b} \log p(I_s, J_{s+(x_b, x_{b'})}), \tag{9}$$

where $b \in \mathcal{V}_1, b' \in \mathcal{V}_2, bb' \in \mathcal{E}_{12}$. For all edges $bq \in \mathcal{E}^1 \cup \mathcal{E}^2$, we set the continuity term to

$$\theta_{bq}(x_b, x_q) = \theta^{\text{reg}}_{bq}(x_b, x_q) + \begin{cases} 0, & |x_b - x_q| \leq 1\,, \\ \infty, & |x_b - x_q| > 1, \end{cases} \tag{10}$$

where θ^{reg}_{bq} is a smoothness regularization determined by needs of a particular application. We currently use $\theta^{\text{reg}}_{bq} = 0$ in all our experiments, but as the remaining hard constraints basically restrict relative displacements to be within some intervals, an additional regularization might be needed to have better registration of textureless areas. The subclass of deformations x of finite energy naturally incorporates a certain range of affine transformations (*e.g.*, it includes scale changes in the range 0.75–1.25 when blocks are 4×4 pixels) and a certain degree of local flexibility (see Fig. 8 for an example of how rotation and scaling can be approximated by such block shifts). We can then choose the appropriate size of blocks, the appropriate allowed relative displacement in place of 1 in (10) and appropriate regularization θ^{reg} at our convenience.

3 Solving the Relaxed Energy Minimization

The problem of minimizing energy (4) is a hard one and we need to resort to approximate methods. We will follow a linear relaxation approach, in which energy minimization is formulated as an integer linear programming (ILP) and integrality constraints are relaxed. See [19,20,17] for details. To solve the relaxed linear program (suboptimally) we apply the TRW-S algorithm [17,18]. This algorithm operates on the dual LP problem, which can be written as an unconstrained maximization of a piecewise linear concave function. The TRW-S algorithm has the same asymptotic behavior as algorithms in [21,19]. It is a suboptimal method for the LP-relaxation problem, since its stationary points

satisfy only a necessary optimality condition (studied in [22,19,18]). Yet the algorithm is very efficient in certain practical applications.

3.1 Review of TRW-S algorithm

We review the algorithm [17,18] with certain simplifications. We assume each edge is covered by exactly one chain, so edge-wise updates are not needed and we omit the normalization step. We also assume for simplicity that weights of all trees in the collection are equal. These choices do not weaken the theoretical properties of the algorithm. The algorithm may be informally described as follows. It iteratively increases a lower bound on the energy function constructed as a combination of tractable subproblems on subtrees of the graph. When such subproblems share a vertex of the graph, there is a freedom to transfer univariate weights from one subproblem to another. By doing this, one can force the optimal configurations of the subproblems to agree about the assignment of label for the shared vertex. Loosely speaking, the more vertices are in the agreement across subproblems the tighter the bound is. And the bound is tight if and only if there exists a configuration which is optimal for all subproblems.

Lower bounds. Let us consider a collection of parameter vectors $\boldsymbol{\theta} = \{\theta^i \in \Omega_{G,\mathcal{L}} \mid \ i \in I\}$, where I is a finite set. For each $\boldsymbol{\theta}$, such that $\sum_i \theta^i = \theta$, the value $LB(\boldsymbol{\theta}) = \sum_i \min_{\mathrm{x}} E(\mathrm{x}|\theta^i)$ is a lower bound on the optimal energy:

$$LB(\boldsymbol{\theta}) \leq \min_{\mathrm{x}} E(\mathrm{x}|\theta), \tag{11}$$

which is easily seen as the maximum of a sum is not greater than the sum of maxima or alternatively from Jensen's inequality (see [17]). Let $T^i = (\mathcal{V}^i, \mathcal{E}^i)$ be a tree subgraph of G, $i \in I$ and θ^i be a tree-structured distribution over the tree T^i:

$$\theta^i_s = 0 \ \forall s \in V \backslash \mathcal{V}^i, \quad \theta^i_{st} = 0 \ \forall st \in \mathcal{E} \backslash \mathcal{E}^i. \tag{12}$$

Let a collection of trees $\boldsymbol{T} = \{T^i \mid i \in I\}$ form an edge-disjoint cover of G. It follows from the constraint $\sum_i \theta^i = \theta$ that only univariate potentials θ^i_s are the free variables. Finding the tightest bound w.r.t. these free variables reads

$$\max_{\{\theta^i_s | i \in I, s \in \mathcal{V}\}} \sum_i \min_{\mathrm{x}} E(\mathrm{x} \mid \theta^i), \quad \text{s.t.} \sum_i \theta^i_s = \theta_s, \tag{13}$$

which is known to be a dual of the LP relaxation of energy minimization [17].

Algorithm. The TRW-S algorithm attempts to solve problem (13) by maximizing the objective sequentially w.r.t. $\{\theta^i_s \mid i \in I\}$ for a fixed $s \in \mathcal{V}$ (a block-coordinate ascent method). At each step, univariate potentials associated with one of the vertices are redistributed among subproblems in order

to make min-marginals of those subproblems equal, which corresponds to a maximum of the objective w.r.t. these potentials. It was shown [18] that in the case when all trees are selected to be simply chains, the computation of min-marginals can be efficiently combined with updating the univariate potentials (chains must be monotonic w.r.t. a total ordering on $\mathcal{V}$).

Let a total order on $\mathcal{V}$ be selected. It is assumed without loss of generality that the orientation of edges in the graph G matches the order: $st \in \mathcal{E} \Rightarrow s < t$. Let $\boldsymbol{T}$ be a collection of chains which are monotonic w.r.t. the ordering of $\mathcal{V}$: $st \in \mathcal{E}^i \Rightarrow s < t$, and each edge of G is covered exactly by one chain. Skipping some derivation, auxiliary variables $M^{\mathrm{fw}}_{st}(x_t)$ (forward messages) and $M^{\mathrm{bw}}_{st}(x_s)$ (backward messages), $st \in \mathcal{E}$, $x_s, x_t \in \mathcal{L}$ are introduced such that variables θ^i_s are represented as

$$\theta^i_s(x_s) = \frac{1}{n_s}\theta_s(x_s) + \sum_{t|ts\in\mathcal{E}^i} M^{\mathrm{fw}}_{ts}(x_s) + \sum_{t|st\in\mathcal{E}^i} M^{\mathrm{bw}}_{st}(x_s), \tag{14}$$

where n_s is number of chains sharing the node s. The block coordinate descent on $\{\theta^i_s \mid s \in \mathcal{V}, i \in I\}$ is then computed in terms of variables M as follows.

Algorithm 1: TRW-S on monotonic chains (algorithm [18, Fig.3]).

(1) Initialize $M^{\mathrm{fw}}_{st}(x_t) = 0,\ M^{\mathrm{bw}}_{st}(x_s) = 0 \quad st \in \mathcal{E},\ x_s, x_t \in \mathcal{L}$.
(2) Set LB $= 0$;
(3) For each $s \in \mathcal{V}$ selected in increasing order perform:
  (a) compute average min-marginals:
  $$\Phi_s(x_s) = \tfrac{1}{n_s}\Big(\theta_s(x_s) + \sum_{t|ts\in\mathcal{E}} M^{\mathrm{fw}}_{ts}(x_s) + \sum_{t|st\in\mathcal{E}} M^{\mathrm{bw}}_{st}(x_s)\Big);$$
  (b) for all t such that $st \in \mathcal{E}$ update:
  $$M^{\mathrm{fw}}_{st}(x_t) = \min_{x_s}\{\Phi_s(x_s) - M^{\mathrm{bw}}_{st}(x_s) + \theta_{st}(x_s, x_t)\}.$$
  (c) LB $=$ LB $+ n_{\mathrm{term}} \min_{x_s} \Phi_s(x_s)$;
(4) Reverse the ordering of $\mathcal{V}$, of all edges in $\mathcal{E}$, $\mathcal{E}^i$, swap M^{fw} with M^{bw} and go to step 2.

Here, n_{term} is the number of chains for which node s is the last node (*i.e.*, there is no such t that $st \in \mathcal{E}^i$). For such chains, the value $\min_{x_s} \Phi_s(x_s)$ correspond to the summand $\min_{\mathrm{x}} E(\mathrm{x}|\theta^i)$ in (13), as if it was computed after the sweep in step 3. During the sweep in step 3, only the values M^{fw} are updated, but after performing step 4, the next sweep will efficiently update values M^{bw} and so on.

Finding a Configuration. As an output of the TRW-S algorithm, we expect a locally consistent set of optimal configurations over the trees (WTA conditions) [18]. To assess convergence, we measure how much the vector of

messages changed during the last iteration. We choose the measure given by

$$\max_{st,x} |M_{st}^{\mathrm{fw(k)}}(x) - M_{st}^{\mathrm{fw(k-2)}}(x)| \frac{|\mathcal{E}|}{LB^{(k)}}, \tag{15}$$

where k is the iteration number. This expression is meaningful for non-negative lower bounds and is more or less independent of the graph size and of the dissimilarity measure scale. Though somewhat weaker convergence properties are proved for TRW-S [18] and we can not state (15) converges to zero, we perform iterations until (15) becomes less than a predefined $\varepsilon_0 > 0$ or a prespecified maximal number of iterations is exceeded.

There have been several heuristics proposed to fix a configuration based on the relaxed dual point. The simplest one is to select $x_s = \operatorname{argmin}_j \Phi_s(j)$, which would be optimal if the bound was tight and these local minimizers were unique. Generally, such a selection may violate our hard constraints, so we do not consider it. A more consistent selection [18] fixes variables in the order of $\mathcal{V}$ while propagating the updates to further variables. This can provide a globally optimal solution in the case when there are ties which can be resolved by a sequential selection. Generally, resolving the ties is a hard satisfiability problem.

We found the following greedy approach useful. Having reached ε_0 we do not make the decision about all variables x_s immediately, but only about a subset of them, and continue with updates by Alg. 1 to propagate constraints imposed by this partial fixation to the rest of the problem. We found it convenient to split the optimization problem into two independent subproblems by fixing an optimal configuration in the horizontal (vertical) chain in the middle of the grid graph. In the next iteration, it is divided into four independent subproblems and so on. The procedure is accomplished in $O(\log |\mathcal{V}|)$ splitting steps. Let us also note that we need to ensure that making greedy decisions on a subset of variables does not lead to violation of hard constraints. In our model, it is easily ensured by a single forward-backward pass of the message passing algorithm. After the first fixation is made, the value LB stops being a lower bound on the optimal energy. Instead, it becomes a lower bound on the conditional optimal energy under given fixation. When all variables are fixed to some values, the conditional LB will coincide with the energy of the fixed configuration.

4 Implementation

Relaxation of the Product Model vs. Decomposed Model. We now turn to a discussion why we prefer to solve the linear relaxation of the decomposed model. Consider the following table:

	Product Model	Decomposed Model
Number of discrete variables	$\lvert\mathcal{V}\rvert$	$2\lvert\mathcal{V}\rvert$
Number of relaxed dual variables	$\lvert\mathcal{V}\rvert\lvert L\rvert^2$	$2\lvert\mathcal{V}\rvert\lvert L\rvert$
Complexity of passing of one message	$\lvert L\rvert^4$	$\lvert L\rvert^2$

At a closer look, however, it appears that messages could be passed faster for certain special types of pairwise interaction functions. In particular, to compute the update in step 3(b) for the product model in the case of separable in x- and y- components quadratic functions $\theta_{st}(x_s, x_t)$, the distance transform can be applied (see *e.g.* [23]) and complexity is then reduces to $O(|\mathcal{L}|) = O(|L|^2)$. However for the decomposed model there are two types of interactions: intra-layer and inter-layer. The complexity of passing a message is respectively $O(|L|^2)$ and $O(|L|)$. So the real difference in speed of updates for the two models is by a constant factor.

We consider computing the relaxation of the product model to be too memory-demanding and therefore we prefer the decomposed model. Let us note, however, that LP relaxation of the decomposed model is weaker than that of the product model (this is because interactions in the product model couple $(x_{s,1}, x_{s,2})$ and $(x_{t,1}, x_{t,2})$, which introduce in the relaxation efficiently *fourtuple* constraints if viewed in the decomposed variables, which tighten the relaxation, see details in [24]). Nevertheless relaxation of the decomposed model performs quite well in our experiments.

On the Computation Order. For the graph in the decomposed model, we suggest selecting set $\boldsymbol{T}$ to be composed of all vertical and all horizontal chains in both layers, plus all inter-layer chains of length 1 (see Fig. 4). It can be shown that one can perform several scans of message-passing updates (step 3(a,b)) in one layer without destroying the correctness of the algorithms and only then switch to the other layer. This achieves faster spatial information propagation by more frequent cheap updates inside the layers. The speedup effect of this schedule is demonstrated in experiments.

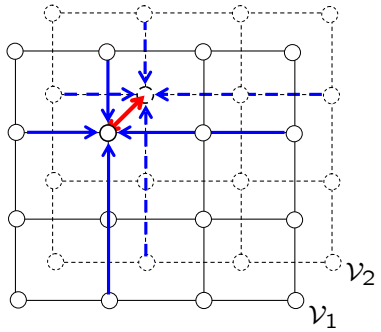

Fig. 4. Schedule of message-passing for the layered grid graph: messages inside each of the layers (blue) are updated more often then between layers (red).

Details. We have implemented Alg. 1 with the schedule of updates set to perform message updates inside the layers 5 times more often than between layers. For the data term, we precomputed block dissimilarities (this is not necessary and the technique can be implemented purely in $O(|\mathcal{V}||L|)$ memory complexity). As a color dissimilarity measure, we used the sum of squared differences and the sum of absolute differences for experiments with a large amount of clutter (Fig 5, 11). Block size is a free parameter chosen according to the desired flexibility of deformation, ε_0 is a "precision" parameter. Our C++ implementation with a MATLAB interface is available at http://cmp.felk.cvut.cz/~shekhovt/deform-match-mrf.

5 Experiments

The performance of our method builds on two factors: how well the energy model suits the posed task, and how well the energy is optimized. To assess the latter we have collected statistics on a large database of natural images. These statistics determine how close we come to solving the optimization problem in practical applications.

To see how well our discrete model for image registration performs, we evaluated it quantitavely on a set of synthetically generated instances and qualitatively on several illustrative synthetic and real-world instances.

Optimization evaluation. We sampled random deformations and noise applied to an image and computed different performance statistics. The test image shown in Fig. 5(a) is of size 160×140 and with color space [0 1]. First, it was deformed by a smooth deformation created using uniformly randomly displaced by up to 10px control points of a regular grid and random global translation in the range $[-5\ 5]$px. Both input images were then perturbed with noise. We modeled Gaussian $\mathcal{N}(0, 0.1^2)$ noise (exemplified in Fig. 5(b)) and a clutter-type noise $\mathcal{B}(\sigma^2)$ (exemplified in Fig. 5(c)), which generates 100 random color spots at random positions. The spots are created by superimposing random colors with a Gaussian transparency mask. The size of the blobs is thus controlled by the variance σ of the Gaussian mask. The algorithm was run with 4×4 blocks, SAD dissimilarity measure and $\varepsilon_0 = 0.005$.

At each run, we computed the best TRW-S lower bound LB guaranteed not to be greater than the quality of an optimal configuration Q^* [17]. Let the quality of the found solution be Q. We calculated the per-instance approximation ratio as

$$1 + \alpha = Q/\mathrm{LB} \geq Q/\mathrm{Q}^* \geq 1, \tag{16}$$

where we refer to α as the approximation error. Let us note that the optimization is invariant to adding a constant to the energy function but not the

approximation ratio (16). It is specific to our chosen energy function, which, however, does not include a constant term and is purely the SAD data measure in this case. Boykov *et al.* [8] give a theoretic approximation guarantee for α-expansion algorithm w.r.t. metric energies (which our energy is, if viewed in the equivalent product model formulation). Unfortunately it does not apply since our energy contain ∞ edges. The goal is to verify how well the weaker relaxation of the decomposed model performs and if the proposed labeling fixation heuristic is indeed important. Results are shown in Fig. 5.

While we have not compared LP relaxation of the decomposed model with that of the product model, the approximation errors are acceptable, and our results clearly show that the gradual fixation of labels is a successful and practical strategy.

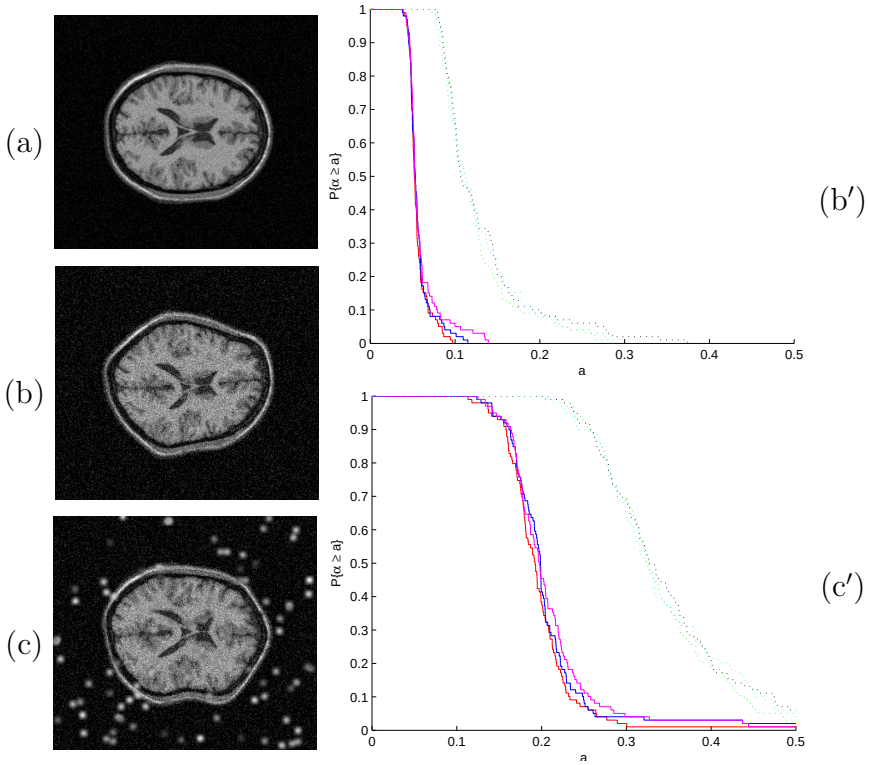

Fig. 5. Approximation error statistics of the optimization scheme on synthetic deformations. (a) Original image. (b) Randomly deformed and noised with Gaussian $\mathcal{N}(0, 0.1^2)$. (c) Deformed noised with $\mathcal{B}(2^2)$ and $\mathcal{N}(0, 0.1^2)$. Image in (a) is also noised though the examples are not shown. (b′) Estimated probability that the approximation error α will exceed value a for inputs exemplified in (b). (c′) Similar for inputs exemplified in (c). Solid plots correspond to our proposed fixation strategy (different number of iterations), dashed plots correspond to the single-pass strategy [18]. Limits on number of iterations were 200, 500 and 1000, the corresponding plots are almost indistinguishable, which indicate that when large enough this number is not important.

We compare against the iterative multi-scale discrete search based method of Glocker *et al.* [11,12] and elastic method based on free-form deformations using B-splines and local gradient descent optimization implemented by the Image Registration Toolkit software (ITK) [25,26]. We run all the three methods with SSD data measure. The regularization parameters for Glocker *et al.* and ITK (only λ_1, the smoothness parameter) were set to achieve the smallest maximal error on independent training samples for each noise level separately. Our current model uses only the hard regularization, so no learning is applied. Results of the comparison are proposed in Table 1. Only the errors over the region of interest (the object) were taken into account. It is seen that on

$\mathcal{N}(0, 0.02^2)$

	Our	Glocker *et al.*	ITK (0.000)
AE mean	0.22	0.30	**0.19**
AE median	0.16	0.22	**0.07**
AE std	2.05	**1.78**	2.70
AE max	**0.23**	0.30	0.35
MOD mean	0.65	0.89	**0.53**
MOD median	0.54	0.73	**0.24**
MOD max	3.43	**3.42**	4.79
MOD std	**0.47**	0.63	0.75

$\mathcal{N}(0, 0.05^2)$

	Our	Glocker *et al.*	ITK (0.000)
AE mean	0.27	0.31	**0.25**
AE median	0.20	0.22	**0.14**
AE std	2.09	**1.84**	2.98
AE max	**0.25**	0.30	0.36
MOD mean	0.79	0.91	**0.71**
MOD median	0.68	0.75	**0.47**
MOD max	3.65	**3.54**	5.23
MOD std	**0.53**	0.65	0.76

$\mathcal{B}(1)$ and $\mathcal{N}(0, 0.1^2)$

	Our	Glocker *et al.*	ITK (0.004)
AE mean	0.51	**0.33**	0.62
AE median	0.37	**0.24**	0.45
AE std	3.17	**2.10**	3.64
AE max	0.49	**0.32**	0.58
MOD mean	1.46	**0.94**	1.77
MOD median	1.23	**0.78**	1.51
MOD max	5.81	**3.69**	6.24
MOD std	1.00	**0.66**	1.18

$\mathcal{B}(2^2)$ and $\mathcal{N}(0, 0.1^2)$

	Our	Glocker *et al.*	ITK (0.008)
AE mean	0.48	**0.33**	0.77
AE median	0.34	**0.24**	0.61
AE std	3.07	**2.07**	3.59
AE max	0.46	**0.33**	0.66
MOD mean	1.38	**0.96**	2.27
MOD median	1.14	**0.78**	2.09
MOD max	5.73	**3.81**	6.45
MOD std	0.99	**0.69**	1.32

Table 1
Statistics of Angular Error (AE, degrees) and Magnitude of Difference (MOD, pixels) errors for different amount of noise. For ITK the value in brackets is the best found value for regularization parameter λ_1. The measures reported are averaged over 100 test samples for each noise level.

this test data the method of Glocker *et al.* with properly tuned regularization parameter outperforms other methods. The best value for the regularization parameter appeared to be stable over the different noise levels. A lower mean error for this method can be achieved but at the cost of significantly higher max error. When the amplitude of deformations is smaller, the performance of ITK is close to that of Glocker *et al.* with both methods outperforming our's. We believe our optimization scheme may be used in conjunction with a spline model and proper regularization to achieve a better performance.

Convergence Speed. The dependence of the performance on the value of ε_0, the threshold at which a greedy fixation is done, is illustrated in Fig. 6 (a). To determine if it is indeed beneficial to perform updates inside layers more often, we studied how fast the lower bound was increased with a different ratio: Fig. 6 (b). Note that plots are time-based rather than iteration-based. The test problem instance for this experiment is shown in Fig. 10. The image size is 500×332, blocks are 4×4, the search window is 80×80.

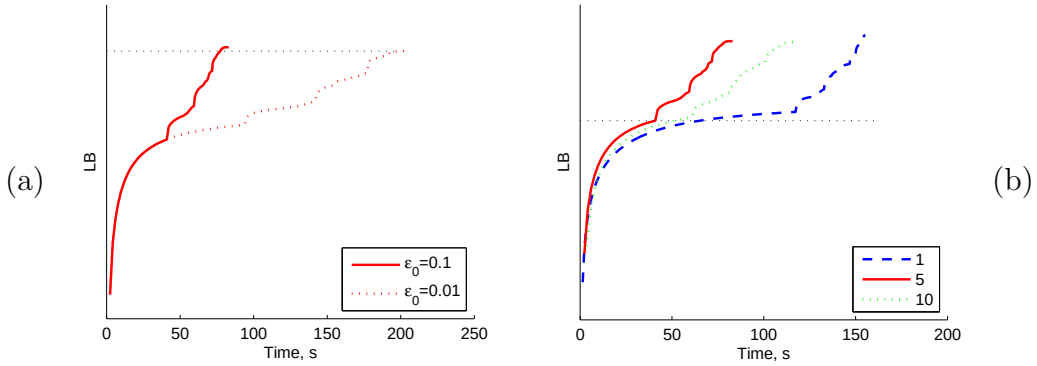

Fig. 6. (a) Effect of threshold ε_0: later greedy fixation normally leads to convergence at a lower energy. (b) Effect of modified message-passing schedule: progress in the lower bound for different ratios of within-layer to inter-layer updates. Jumps in the LB plots indicate places where a greedy fixation of a part of variables was performed.

Visual evaluation of the results. We applied several parametric transformations (Fig. 7) to a natural image to show the flexibility of the deformation field. We searched for a matching subregion to the deformed images (Figs. 7, 8). We cropped the subregion to ensure that all pixels from T^1 are matched; however, the algorithm also works well without this cropping. The chosen subregion does not align with visual edges of the image so the quality of the matching could be seen visually when superimposed. All images are 300×225 pixels, we used $L = \{-30 \dots 30\}$ for x and y allowed displacements, blocks were 4× 4. Computation took ∼40 sec per image, from which 11.2 sec were spent on precomputing block correlations (PIII, 1.5GHz). We keep a small translation component since the optimization is invariant to translations provided the search range is wide enough. The robustness to noise is tested with synthetic Gaussian noise (Fig. 9). Finally, we illustrate how our method works on real images Fig.10-11.

6 Conclusions

We have designed a novel model for image matching based on the MRF optimization framework. Our constraints allow flexible local deformations and impose a hard penalty on discontinuities. By varying the size of blocks, we can

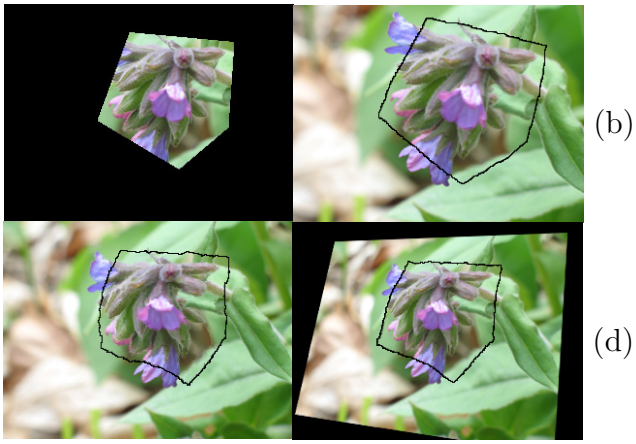

Fig. 7. Searched fragment (image I) and its found deformations superimposed onto target input images (J).

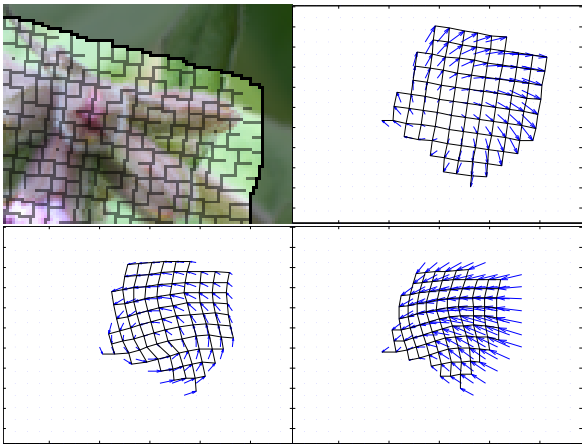

Fig. 8. (a) Close-up of rotation-scaling deformation shown without fills between blocks. (b-d) Found deformation field (shown sparse) and a transformed regular grid w.r.t. Fig. 7.

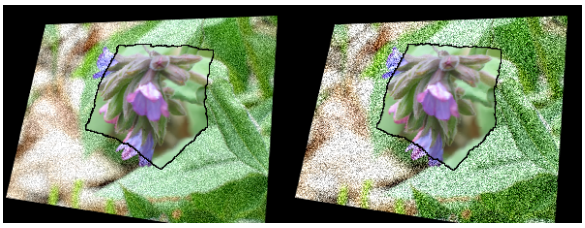

Fig. 9. (a) Test image perturbed with noise $\mathcal{N}(0, 0.1^2)$ and the matched fragment superimposed. (b) The same with noise $\mathcal{N}(0, 0.2^2)$.

control the class of allowed deformations and avoid overfitting. We have applied the TRW-S algorithm to maximize the dual of the linearly relaxed energy minimization problem. As we have used the structure of the model from [15], our relaxed optimization problem has a number of variables which scales linearly with the search window. This can be contrasted with the commonly

Fig. 10. Morphing face expressions using found deformation: interpolated intermediate frames from top-left image to bottom-right image. The deformation and colors of pixels in correspondence are linearly interpolated for the intermediate frames.

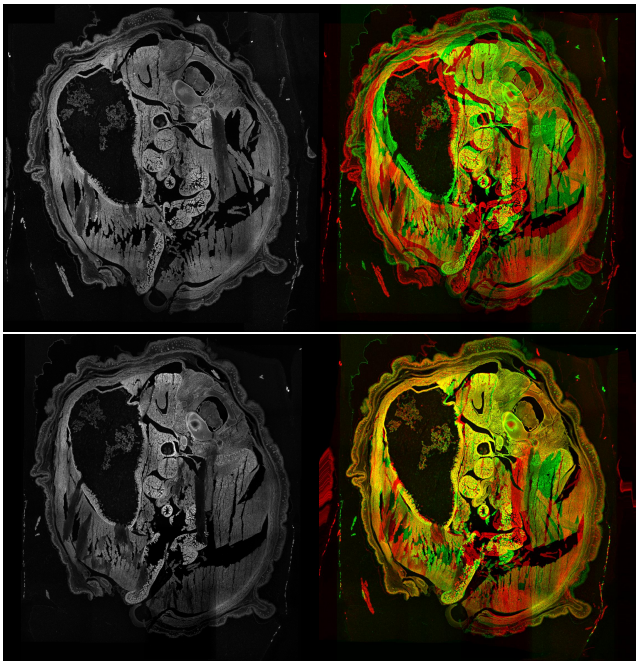

Fig. 11. Registration of microscopy images. Left: input images, courtesy of the Institute of Physiology, Academy of Sciences of the Czech Republic (images contain sections through a middle part of a 21-day-old embryo of a Norway rat; the sections were created by mosaicing 25 partial fields of view captured by a confocal laser scanning microscope Leica SP2 AOBS). Right: first image (red) overimposed onto second (green) before (upper) and after (lower) registration.

used model where the number of variables scales quadratically. This allows us to deal with much wider local displacements of the deformation fields than considered previously feasible within the global optimization framework. Our experiments demonstrate high visual accuracy within a wide class of continuous deformations and robustness to a high degree of noise.

Extensions. As in stereo, we could possibly use a truncated continuity term and gradient cues to model piecewise continuous deformations, and a truncated data term to model outliers (clutter, occlusions, etc.). Another option is to couple our model with other MRF models, *e.g.*, a segmentation model, as in [27]. A sparse set of corresponding pixels can be incorporated easily in the data term as a prior. This is likely to stabilize the result significantly. One could imagine a user-guided tool for image morphing: first, images are matched fully automatically, then the user constraints correspondence by a mouse in one or several places and re-runs. To cope with larger global translations (or scale, rotation *etc.*), one may consider adding a common, roughly discretized, transformation variable. This, however, leads to many ambiguous solutions so the optimization by TRW-S becomes badly conditioned. Another related issue is the choice of a solution based on min-marginals. We consider pairing this algorithm with a local search algorithm, *e.g.*, graph cuts, so the latter would have a good initial point to start.

Acknowledgment

The authors thank the European Commission grant 215078 (DIPLECS) and the Czech government grant MSM6840770038 for support. We thank Tomas Werner for discussions, Ben Glocker for assistance with their software and Hannah Dee for careful proofreading. We also thank reviewers for their useful comments.

References

[1] A. Shekhovtsov, I. Kovtun, V. Hlavac, Efficient MRF deformation model for non-rigid image matching, in: International Conference on Pattern Recognition, IEEE Computer Society, 2007, pp. 1–6.

[2] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, C. Rother, A comparative study of energy minimization methods for Markov random fields, in: European Conference on Computer Vision, Vol. 2, 2006, pp. II: 16–29. 2

[3] N. Jojic, A comparison of algorithms for inference and learning in probabilistic graphical models, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (9) (2005) 1392–1416. 2, 5

[4] J. Konrad, E. Dubois, Bayesian estimation of motion vector fields, IEEE Transactions on Pattern Analysis and Machine Intelligence 14 (9) (1992) 215–228. 2, 3

[5] F. Heitz, P. Bouthemy, Multimodal estimation of discontinuous optical flow using Markov random fields, IEEE Transactions on Pattern Analysis and Machine Intelligence 15 (12) (1993) 1217–1232. 2

[6] J. Zhang, J. Hanauer, The mean field theory for image motion estimation, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 5, 1993, pp. 197–200. 2, 3

[7] S. Roy, V. Govindu, MRF solutions for probabilistic optical flow formulations, in: International Conference on Pattern Recognition, IEEE Computer Society, 2000, p. 7053. 3

[8] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (11) (2001) 1222–1239. 3, 4, 14

[9] M. P. Kumar, P. H. S. Torr, A. Zisserman, Learning layered motion segmentation of video, in: International Conference on Computer Vision, IEEE Computer Society, 2005, pp. 33–40. 3

[10] H. Jiang, M. Drew, Z. Li, Matching by linear programming and successive convexification, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (6) (2007) 959–975. 3

[11] B. Glocker, N. Komodakis, N. Paragios, G. Tziritas, N. Navab, Inter and intra-modal deformable registration: Continuous deformations meet efficient optimal linear programming, in: Information Processing in Medical Imaging, 2007, pp. 408–420. 3, 15

[12] B. Glocker, N. Komodakis, G. Tziritas, N. Navab, N. Paragios, Dense image registration through MRFs and efficient linear programming, Medical Image Analysis(in press), 2008. 3, 15

[13] N. Komodakis, G. Tziritas, N. Paragios, Fast, approximately optimal solutions for single and dynamic MRFs, in: IEEE Transactions on International Conference on Pattern Recognition, 2007, pp. 1–8. 3

[14] P. F. Felzenszwalb, D. P. Huttenlocher, Efficient belief propagation for early vision, in: Proc. 2004 IEEE Computer Soc. Conf. Computer Vision and Pattern Recognition, Vol. 1, IEEE Computer Society Press, 2004, pp. 261–268. 3

[15] I. Kovtun, Image segmentation based on sufficient conditions of optimality in NP-complete classes of structural labelling problem, Ph.D. thesis, IRTC ITS National Academy of Sciences, Ukraine, in Ukrainian (2004). 4, 6, 17

[16] A. Roche, G. Malandain, N. Ayache, Unifying maximum likelihood approaches in medical image registration, International Journal of Imaging Systems and Technology 11 (1) (2000) 71–80. 4

[17] M. Wainwright, T. Jaakkola, A. Willsky, MAP estimation via agreement on (hyper)trees: Message-passing and linear-programming approaches, IEEE Transactions on Information Theory 51 (11) (2005) 3697–3717. 5, 8, 9, 13

[18] V. Kolmogorov, Convergent tree-reweighted message passing for energy minimization, IEEE Transactions on Pattern Analysis and Machine Intelligence 28(10) (2006) 1568–1583. 5, 8, 9, 10, 11, 14

[19] T. Werner, A linear programming approach to max-sum problem: A review, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (7) (2007) 1165–1179. 8, 9

[20] C. Chekuri, S. Khanna, J. Naor, L. Zosin, A linear programming formulation and approximation algorithms for the metric labeling problem, SIAM Journal on Discrete Mathematics 18 (3) (2005) 608–625. 8

[21] V. Koval, M. Schlesinger, Two-dimensional programming in image analysis problems, Automatics and Telemechanics 2 (1976) 149–168, in Russian. 8

[22] M. Schlesinger, Syntactic analysis of two-dimensional visual signals in noisy conditions, Kibernetika 4 (1976) 113–130, in Russian. 9

[23] P. F. Felzenszwalb, D. P. Huttenlocher, J. M. Kleinberg, Fast algorithms for large-state-space HMMs with applications to web usage analysis, in: S. Thrun, L. Saul, B. Schölkopf (Eds.), Advances in Neural Information Processing Systems, MIT Press, Cambridge, MA, 2003. 12

[24] A. Shekhovtsov, I. Kovtun, V. Hlaváč, Efficient MRF deformation model for image matching, Research Report CTU–CMP–2006–08, Center for Machine Perception, Czech Technical University (October 2006). 12

[25] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, D. Hawkes, Non-rigid registration using free-form deformations: Application to breast MR images, IEEE Transactions on Medical Imaging 18 (8) (1999) 712–721. 15

[26] J. A. Schnabel, D. Rueckert, M. Quist, J. M. Blackall, A. D. Castellano-Smith, T. Hartkens, G. P. Penney, W. A. Hall, H. Liu, C. L. Truwit, F. A. Gerritsen, D. L. G. Hill, D. J. Hawkes, A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2001, pp. 573–581. 15

[27] J. Winn, N. Jojic, LOCUS: Learning object classes with unsupervised segmentation, in: International Conference on Computer Vision, Vol. 1, 2005, pp. 756–763. 19

[28] A. Kannan, N. Jojic, B. Frey, Generative model for layers of appearance and deformation, in: AI and Statistics, Society for Artificial Intelligence and Statistics, 2005, pp. 166–173.