Efficient MRF Deformation Model for Non-Rigid Image Matching

Alexander Shekhovtsov¹ Ivan Kovtun² Václav Hlaváč¹

¹ Center for Machine Perception Czech Technical University, Prague ² International Research and Training Center for Informational Technologies and Systems, Kiev

Abstract

We propose a novel MRF-based model for deformable image matching. Given two images, the task is to estimate a mapping from one image to the other maximizing the quality of the match. We consider mappings defined by a discrete deformation field constrained to preserve 2D continuity. We pose the task as finding MAP configurations of a pairwise MRF. We propose a more compact MRF representation of the problem which leads to a weaker, though computationally more tractable, linear programming relaxation - the approximation technique we choose to apply. The number of dual LP variables grows linearly with the search window side, rather than quadratically as in previous approaches. To solve the relaxed problem (suboptimally), we apply TRW-S (Sequential Tree-Reweighted Message passing) algorithm [13, 5]. Using our representation and the chosen optimization scheme, we are able to match much wider deformations than was considered previously in global optimization framework. We further elaborate on continuity and data terms to achieve more appropriate description of smooth deformations. The performance of our technique is demonstrated on both synthetic and real-world experiments.

1. Introduction

Recently, substantial advances in low-level vision problems like stereo, segmentation, image de-noising, etc., have been achieved by using MRFs. This is mainly due to new algorithms for inference in MRF, in particular algorithms for finding maximum a posteriori (MAP) configurations. Szeliski *et al.* [12] give experimental comparison of recent optimization algorithms for popular vision problems. More background on inference and learning algorithms for MRFs is given *e.g.* in [3].

In this work we consider the problem of two-dimensional matching of non-rigid objects. Out of possible applications we would like to point some non-obvious ones, employing 2D deformation as a part of more complex generative models. For example, in learning object appearance jointly with motion [4, 8] or joint segmentation and classification [15] the deformation field is used to describe non-rigid motions (people in video), the inclass variations (different human faces, cars, etc.) or both. Classical applications include computation of the optical flow to estimate motion of a 3D scene, registration of images taken by different measuring devices, tracking in video sequences, etc.

1.1. Related work

Here, we survey existing works on modeling non-rigid deformation as a field of discrete displacements with continuity constraints imposed on pairs of neighboring displacements. The optimal deformation is sought as MAP configuration of the corresponding MRF. Finding a globally optimal or even a reasonably good suboptimal solution in this problem is a difficult task and such models of deformation have been addressed only rarely. The potential advantage however is that harder matching problems could be solved thanks to use of global optimization techniques.

Roy and Govindu [9] applied the MRF model to the optical flow problem. They modeled flow orientation field and flow magnitude field as separate MRFs. They first solve 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.* [1] 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 . Another disadvantage is that starting point is required and the solution is dependent on it — the algorithm finds local optimum w.r.t. *swap moves*, which might not be reasonable for the problem we consider.

Kumar *et al.* [8] computed piecewise rigid deformation of consecutive frames in a video sequence. Deformation was locally described by translation rotation and scale. An MRF model with states being discretized deformations and pairwise potentials encouraging rigid motion of fragments was optimizing using (sum-product) belief propagation algorithm. Due to large state-space, a coarse-to-fine approach was applied to reduce memory and computation requirements.

Winn and Jojic [15] employed 2D deformation as a part of their model of flexible appearance. They applied α -expansion algorithm [1] to find the best deformation (a truncation is needed to apply α -expansion to their non-*metric* [1] model).

Kovtun [7] proposed an algorithm for finding a guaranteed persistent part of optimal solutions. The approach was applied also to 2D deformation. The work suggests a more compact MRF model which we refer to as *decomposed* model and will discuss in more detail in the sequel. Kovtun's experiments on real images showed that it is hard to find a reasonably large persistent part of optimal solutions.

1.2. Contribution and outline

Our model is based on the model [7], we extend its flexibility by using blocks and hard continuity constraints (Sect. 2). We apply TRW-S algorithm (recently developed [13, 5]) to our model. It is a variational optimization algorithm designed to solve the LP-relaxation of a discrete energy minimization problem. This LP-relaxation was studied in, e.g., original paper [10], review [14]). The TRW-S algorithm has the same asymptotic behavior as algorithms in [6, 14]. It may not solve the LP-relaxation problem, since its stationary points satisfy only a necessary optimality condition (studied in [10, 14, 5]). Yet the algorithm is very efficient in certain practical applications. We achieved significant improvement in the deformable image matching problem as we show in our experiments (Sect. 4), demonstrating that qualitatively larger displacement ranges are feasible and that the solution provided is robust to noise. We conclude with possible extensions in Sect. 5.

2. Deformation model

In this section we review previously used MRF models for the 2D deformation and introduce *block model*. We consider two important aspects of these models. The first aspect is demands of the models w.r.t. TRW-S algorithm: the number of variational variables (all components of all messages) and the complexity of elementary updates in one iteration (passing of one message). The second aspect is the modeling power to represent continuous deformations. We start with defining the standard energy minimization framework, following the notation of [13, 5]. We refer to [3, 13] for more background on graphical models and MRFs.

2.1. Energy minimization

Let $\mathcal{L} = \{1 \dots K\}$ be a set of labels. Let $G = (\mathcal{V}, \mathcal{E})$ be a graph with $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ antisymmetric and antireflexive, i.e. $(s,t) \in \mathcal{E} \Rightarrow (t,s) \notin \mathcal{E}$. In what follows 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 defined as $x = \{x_s \mid s \in \mathcal{V}\}$. Let $\{\theta_s(i) \in \mathbb{R} \mid i \in \mathcal{L}, s \in \mathcal{V}\}$ be *univariate* potentials and $\{\theta_{st}(i,j) \in \mathbb{R} \mid i, j \in \mathcal{L}, st \in \mathcal{E}\}$ be *pairwise* potentials.

Let *energy* of a configuration x be defined by:

$$E(\mathbf{x}|\theta) = \sum_{s \in \mathcal{V}} \theta_s(x_s) + \sum_{st \in \mathcal{E}} \theta_{st}(x_s, x_t) , \qquad (1)$$

where $\theta_s(\cdot)$ is also referred to as *data term* and $\theta_{st}(x_s, x_t)$ as pairwise *interaction* term. The probability distribution defined by $p(\mathbf{x}|\theta) \propto \exp(-E(\mathbf{x}|\theta))$ is a Gibbs distribution which corresponds to a certain Markov Random Field. The problem of finding a maximum a posteriori configuration of this MRF corresponds to the *energy minimization*, $\min_{\mathbf{x}} E(\mathbf{x}|\theta)$.

2.2. Product model

Let T^1 , T^2 be sets of pixels and I^1 : $T^1 \mapsto [0,1]^3$, I^2 : $T^2 \mapsto [0,1]^3$ be two images. We start with the simplest model for a 2D deformation which was considered in e.g. [1]. 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 x- and y- displacements of the pixel s. Let mapping D_x from image I^2 to image I^1 be defined by $(D_x I)_s = I_{s+x_s}$. Let both coordinates take values from $L = \{K_{\min}, \dots, K_{\max}\}$, thus variables $x_s = (x_s^1, x_s^2)$ take their values from the set $\mathcal{L} = L^2$. Let $\theta_s(x_s) = (I_s^1 - I_{s+x_s}^2)^2 / 2\sigma_{\rm I}^2$. This term corresponds to the statistical assumption of $p(({\rm D}_{\rm x}I^2)_s - I_s^1 \mid {\rm x}) \propto$ $\mathcal{N}^3(0,\sigma_{\mathrm{I}}^2), s \in \mathcal{V}$, which means that under fixed x deformed image $D_x I^2$ is a noisy observation of image I^1 assuming Gaussian noise in each color component with variance $\sigma_{\rm I}^2$. A usual setting for interaction potentials is: $\begin{array}{l} \theta_{st}(x_s,x_t) = \|x_s - x_t\|^2 / 2\sigma_{\mathrm{x}}^2 = (x_s^1 - x_t^1)^2 / 2\sigma_{\mathrm{x}}^2 + (x_s^2 - x_t^2)^2 / 2\sigma_{\mathrm{x}}^2, \, st \in \mathcal{E}, \, \text{where } \mathcal{E} \text{ is the set of all horizontally} \end{array}$ and vertically neighboring pairs of pixels. This term penalizes discontinuities in the deformation field x, so that close pixels are forced to go to close destinations (see Fig. 1).

Minimizing energy defined by θ for a general input is not known to be polynomially solvable. A message-passing algorithm for this problem (TRW-S, belief propagation) would require keeping $O(|\mathcal{V}||L|^2)$ variational variables this is the number of variables in the dual of the LPrelaxation of the energy minimization. Generally, such algorithms would require $O(|L|^4)$ operations for an elementary message update. In our case, it may be reduced to $O(|L|^2)$ if the special form of functions θ_{st} is exploited (applying distance transform, see *e.g.* [2]). Let us also note that if θ_{st} was chosen to be $||x_s - x_t||$ the energy would satisfy metric properties and the α -expansion algorithm could be applied.



Figure 1. Deformation as pairwise MRF: data terms penalize color deviation of singe pixels from hypothesized destination, pairwise terms penalize spatial deviations.

2.3. Decomposed model

The model, proposed in [7], represents x and y displacements by two interacting fields which we refer to as layers. Graph G in this model is constructed as follows (see Fig. 2): nodes $\mathcal{V} = \mathcal{V}^1 \cup \mathcal{V}^2$, with $\mathcal{V}^1 \sim \mathcal{V}^2 \sim T^1$ (here, ~ 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 *intra-layer* edges (here, $s^1 \sim s^2$ denote that these elements correspond in the isomorphism \sim). Let $\mathbf{x} = \{x_{s^i} \mid s^i \in \mathcal{V}^i, i = 1, 2\}$. The data term in this model is encoded in the interaction pairs form \mathcal{E}^{12} , as $\theta_{st}(x_s, x_t) = (I_s^1 - I_{s+(x_s, x_t)}^2)^2 / 2\sigma_{\mathbf{I}}^2$, $s \in \mathcal{V}^1$, $t \in \mathcal{V}^2$, $st \in \mathcal{E}^{12}$. The continuity term is set identically for both layers: $\theta_{st}(x_s, x_t) = (x_s - x_t)^2 / 2\sigma_x^2$, $st \in \mathcal{E}^i$, i = 1, 2. It can be seen that the resulting energy function $E(\mathbf{x}|\theta)$ is equivalent to that of the product model.

For this model we need to keep only $O(|\mathcal{V}||L|)$ variational variables, with complexity of elementary update of $O(|L|^2)$ for pairs $st \in \mathcal{E}^{12}$ and O(|L|) for pairs $st \in \mathcal{E}^1 \cup \mathcal{E}^2$ (the latter is due to special form of interactions, $(x_s - x_t)^2$, [2]). Let us note that LP relaxation of the decomposed model is weaker than that of the product model (see details in [11]), however this did not seem critical for our instances.

2.4. Block model, crisp discontinuity term

A drawback of the two above models is that the continuity term $||x_s - x_t||^2 / 2\sigma_x^2$ assigns a nonzero penalty to affine transformations except pure translations, in particular, scaling is very poorly handled. The solution therefore is biased towards pure translations which is not only undesirable but leads to a sensitive tradeoff between the data term and the continuity term: either the model will not follow the deformation observed by the data term or it will allow discontinuities in the case noise is present.

It is reasonable that e.g. small affine transformations and scaling receive the same penalty as pure translations. One might think that this could be ensured by replacing the soft interaction term with a crisp one, which would assign zero



Figure 2. Decomposed model. Left: model consist of two layers \mathcal{V}^1 and \mathcal{V}^2 , the neighborhood structure is shown by nodes s^1 , s^2 and interaction edges to their neighbors in \mathcal{E} . Right: inter-layer interaction is used to encode data term – 2D displacement of a pixel s is determined by pair of labels x_{s^1}, x_{s^2} (shown black) being the two coordinates; data fitness term for the displacement is encoded on edge (x_{s^1}, x_{s^2}) (shown bold dashed).

penalty to small relative displacement of ± 1 pixel and infinite penalty to larger displacements. However, it turned out in experiments that this model is too flexible and causes overfitting, which results in many irregularly spaced gaps in the displacement field.

To avoid overfitting we require the deformation field to be locally affine (in some small regions of pixels). As we consider discrete models, we want the deformation field to be described locally by translations. We propose to aggregate pixels into blocks and allow 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. Let set T^1 be regularly subdivided into square



Figure 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.

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. We define the data term as

$$\theta_{st}(x_s, x_t) = \frac{1}{2\sigma_1^2} d(I_s^1, I_{s+(x_s, x_t)}^2),$$
(2)

where $s \in \mathcal{V}^1, t \in \mathcal{V}^2, st \in \mathcal{E}^{12}$; I_s^1 is a fragment of image I^1 over block s; $s + (x_s, x_t)$ is the block s shifted by (x_s, x_t) , and $d(\cdot, \cdot)$ is a dissimilarity measure between the

two image fragments (sum of squared differences or other). For all edges $st \in \mathcal{E}^1 \cup \mathcal{E}^2$ we set the continuity term to

$$\theta_{st}(x_s, x_t) = \begin{cases} 0, & x_s = x_t, \\ c_r, & |x_s - x_t| \le 1, \\ \infty, & |x_s - x_t| > 1, \end{cases}$$
(3)

where $0 < c_r \ll 1/\sigma_1^2$ is a regularization term that enforces uniqueness of solution. The subclass of transformations with low penalty (less than c_r for each continuity edge) 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, as can be seen from our experiments.

3. Implementation

We have implemented TRW-S as proposed in [5], i.e. using monotonic chains and the efficient update scheme (see also short derivation of TRW-S in [11]). Our tree decomposition consists of all vertical and horizontal chains in layers \mathcal{V}^1 and \mathcal{V}^2 plus one-edge chains, correspond-ing to inter-layer edges \mathcal{E}^{12} . For intra-layer (edges from $\mathcal{E}^1, \mathcal{E}^2$) updates corresponding to continuity term (3), we used method [2] giving complexity O(|L|). Because interlayer (\mathcal{E}^{12}) updates are considerably slower, we modified the sequential schedule to perform more intra-layer iterations (this is compatible with TRW-S's constraints on the order of updates). For the data term, we precomputed block similarities (also this is not necessary and the technique can be implemented purely in $O(|\mathcal{V}||L|)$ memory complexity). To speed up this computation, we first cluster image color space and precompute the table of color comparisons. For the color comparison we use $F_{\lambda}(c_1, c_2) =$ $\lambda^2 \frac{\langle c_1 - c_2, c_2 \rangle}{||c_2||^2} + (c_1 - \langle c_1, c_2 \rangle \frac{c_2}{||c_2||^2})^2$, which scales down the projection of $c_1 - c_2$ onto c_2 by λ and provides robustness against brightness changes: when $\lambda = 1$, F_{λ} becomes squared Euclidean distance $(c_1 - c_2)^2$, when $\lambda < 1$, then component of $(c_1 - c_2)$ along the direction defined by c_2 is scaled down by λ .

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) [5]. To assess convergence, we measure how well WTA conditions are satisfied: we estimate the smallest ε such that ε -optimal configurations of the trees satisfy WTA, where ε -optimal configurations are those having quality within ε from the optimum. It follows from convergence properties of TRW-S [5] that for a given ε_0 there exists an iteration number such that $\varepsilon < \varepsilon_0$ is satisfied. Having reached predefined ε_0 , we need to pick a solution based on the set of configurations satisfying WTA, which is hard (NP-complete) constraint satisfaction problem and its solution need not even exist. Instead we resort to the following greedy heuristic. Hav-

ing reached ε_0 we split the optimization problem into two independent ones by fixing an optimal configuration at the horizontal (vertical) chain in the middle of our grid graph. This fixation can increase ε for each of the two subproblems so we continue with TRW-S updates and the whole procedure for the subproblems until we fix solution at all nodes. The procedure is accomplished in $O(\log |\mathcal{V}|)$ splitting steps and it is guaranteed that no ∞ edge will be encountered at the output solution.

Parameters. We represent color space as $[0, 1]^3$ and compute color similarity using F_{λ} with $\lambda = 0.1$. For d(x, y) we use $\frac{1}{N} \sum_{i=1}^{N} F_{\lambda}(x^i, y^i)$, where x^i, y^i are colors of pixels in correspondence. We set $\sigma_{\rm I} = 1$ and $c_r = 0.001$. When a pixel is mapped out of the field of view, we assign it a penalty of 0.01. Experimental results were not very sensitive to these parameters. Block size is a free parameter chosen accordingly to the desired flexibility of deformation, ε_0 is a "precision" parameter.

4. Experiments

We ran the optimization scheme on many real images and gathered statistics to evaluate its performance. We also optimized on synthetically deformed images to learn how well the deformation field is recovered and how robust it is to noise. Results on large real-world images and the animated deformations are available at http://cmp.felk.cvut.cz/ ~shekhovt/deform-match-mrf.



Figure 4. First line: example of input-output for the *matching* set: input image pair (template and target images) and the deformed template overimposed on target image. The deformation is found by the algorithm. Second line: the same for the *non-matching* set (notice: traffic signs are different). Images are 100×100 , search regions are 10×10 and blocks are 4×4 . Black background of the template images is excluded from the data term.

Optimization evaluation. We evaluated the optimization on two sets of real world images: one contains truly matching image pairs and the second contains pairs of nonmatching images, Fig. 4 gives representative examples. In the set of non-matching pairs many instances contain really different objects, which could not be matched by a smooth deformation, however it does not prevent us from running the algorithm on them. At each run, we compute the best TRW-S lower bound, LB, guaranteed not to be greater than the quality of an optimal configuration Q^* [13]. Let the quality of found solution be Q. We calculate the approximation ration as $\alpha = Q/\text{LB} \ge Q/Q^*$. The statistics over 100 test pairs from both sets is given in Table 1. Let us note that the optimization is invariant to adding a constant to the energy function, however, the approximation ration is not. Therefore it is specific to our particularly chosen energy function.

	ε_0	10^{-2}	10^{-3}	10^{-4}
(a)	mean $\alpha - 1$	0.72%	0.43%	0.28%
	$\max \alpha - 1$	2.14%	1.64%	1.5%
	std of α	0.43%	0.33%	0.3%
	avg time	0.92s	1.4s	4.2s
(b)	mean $\alpha - 1$	4.04%	3.5%	3.34%
	$\max \alpha - 1$	19.5%	18.7%	18.7%
	std of α	4.5%	4.2%	4.3%
	avg time	0.97s	2.1s	7.5s

Table 1. Statistics of approximation ratio over 100 test instances of : (a) matching traffic signs; (b) non-matching traffic signs.

Our observations are as follows: 1) Waiting longer for better TRW-S convergence gives only a slight improvement at the cost of significantly more computations. 2) More difficult matching tasks (ambiguous, clutter, etc.) lead to harder optimization problems – the approximation ratio is worse and the required time to achieve the same convergence precision is higher.

Visual evaluation of the results. We applied several parametric transformations to an image (Fig. 5) to show the flexibility of the deformation field, and searched for a matching of its subregion to the deformed images (Figs. 6, 7). We cropped the subregion to ensure that all pixels from T^1 are matched; however, the algorithm works well also 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 $L = \{-30...30\}$ for x and y allowed displacements, blocks were 4×4 . Computation took \sim 40 sec. per image, from which 11.2 sec were spent on precomputing block correlations. We keep a small translation component since the optimization is invariant to it provided the search range is wide enough. The robustness to noise is tested with synthetic gaussian noise (Fig. 8).



Figure 5. Synthetic deformations: initial picture, rotation+scaling, twirl, twirl + projective transform.



Figure 6. Searched fragment and its found deformations superimposed (inside the outline) onto corresponding inputs.



Figure 7. (a) Close-in at rotation-scaling deformation shown without fills between blocks. (b-d) Found deformation field (shown sparse) and a transformed regular grid w.r.t. Fig. 6.



Figure 8. (a) Image from Fig. 5(d) perturbed with noise $\mathcal{N}(0, 0.1)$ with the matched fragment superimposed. (b) The same with noise $\mathcal{N}(0, 0.2)$.

5. Conclusion

We have designed a novel model for image matching based on the MRF optimization framework. Our constraints allow flexible local deformations and impose hard penalty on discontinuities. By varying size of blocks we can 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 used the structure of the model from [7], our relaxed optimization problem needs only a number of variables, which is linear in the sizes of the search window (not quadratic as it would be with the commonly used model). This allows us to deal with much wider local displacements of the deformation fields compared to ones considered previously in 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 [15]. A sparse set of corresponding pixels can be easily incorporated in the data term as a prior. This is likely to greatly stabilize the result. One could imagine a user-guided tool for image morphing: first, images are matched fully automatically, then the user constraints correspondence by 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. We also noted that many states of variables x_s could be discarded as non-optimal if the computation of lower bound is repeated for each of them and compared to the quality of the best solution found, however the challenge is to organize this computation efficiently.

Acknowledgment

This work has been supported by the European Commission project FP6-IST-004176 (COSPAL). We thank Tomas Werner for discussions and valuable advices; the reviewers for their comments.

References

- Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222–1239, Nov. 2001.
- [2] P. F. Felzenszwalb, D. P. Huttenlocher, and J. M. Kleinberg. Fast algorithms for large-state-space HMMs with applications to web usage analysis. In *NIPS*, 2003.
- [3] N. Jojic. A comparison of algorithms for inference and learning in probabilistic graphical models. *PAMI*, 27(9):1392– 1416, 2005. Senior Member-Brendan J. Frey.
- [4] A. Kannan, N. Jojic, and B. Frey. Generative model for layers of appearance and deformation. In AI and Statistics (AIS-TATS'05), pages 166–173, 2005.
- [5] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. In AI and Statistics (AIS-TATS'05), pages 182–189, 2005.
- [6] V. Kovla and M. Schlesinger. Two-dimensional programming in image analysis problems. *Automatics and Telemechanics*, 2:149–168, 1976. In Russian.
- [7] I. Kovtun. Image segmentation based on sufficient conditions of optimality in NP-complete classes of structural labelling problem. PhD thesis, IRTC ITS National Academy of Science Ukraine, 2004. In Ukrainian.
- [8] M. P. Kumar, P. H. S. Torr, and A. Zisserman. Learning layered motion segmentation of video. In *ICCV* '05, pages 33–40, 2005.
- [9] S. Roy and V. Govindu. MRF solutions for probabilistic optical flow formulations. In *ICPR'00*, pages Vol III: 1041– 1047, 2000.
- [10] M. Schlesinger. Syntactic analysis of two-dimensional visual signals in noisy conditions. *Kibernetika, Kiev*, 4:113–130, 1976. In Russian.
- [11] A. Shekhovtsov, I. Kovtun, and 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] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. A comparative study of energy minimization methods for markov random fields. In *ECCV*, volume 1, 2006.
- [13] M. Wainwright, T. Jaakkola, and A. Willsky. Exact MAP estimates by (hyper)tree agreement. In Advances in Neural Information Processing Systems 15, pages 809–816. 2003.
- [14] T. Werner. A linear programming approach to max-sum problem: A review. Research Report CTU–CMP–2005–25, Center for Machine Perception, Czech Technical University, Dec. 2005.
- [15] J. Winn and N. Jojic. Locus: Learning object classes with unsupervised segmentation. In *ICCV*, volume 1, pages 756– 763, 2005.