



CENTER FOR  
MACHINE PERCEPTION



CZECH TECHNICAL  
UNIVERSITY

RESEARCH REPORT

ISSN 1213-2365

# Efficient MRF Deformation Model for Image Matching (Version 1.50)

Alexander Shekhovtsov<sup>1</sup>

Ivan Kovtun<sup>2</sup>

Václav Hlaváč<sup>1</sup>

shekhovt@cmp.felk.cvut.cz   kovtun@image.kiev.ua  
hlavac@cmp.felk.cvut.cz

<sup>1</sup>Center for Machine Perception  
Department of Cybernetics, Faculty of Electrical Engineering  
Czech Technical University, Prague

<sup>2</sup>International Research and Training Center  
for Informational Technologies and Systems, Kiev

CTU–CMP–2006–08

May 1, 2007

Available at  
<ftp://cmp.felk.cvut.cz/pub/cmp/articles/shekhovtsov/Shekhovtsov-TR-2006-08.pdf>

**Supervisor: Václav Hlaváč**

The authors were supported by the EC projects FP6-IST-004176 COSPAL, INTAS 04-77-7347 PRINCESS, MRTN-CT-2004-005439 VISIONTRAIN and by the Czech Ministry of Education under project 1M0567.

**Research Reports of CMP, Czech Technical University in Prague, No. 8, 2006**

Published by

Center for Machine Perception, Department of Cybernetics  
Faculty of Electrical Engineering, Czech Technical University  
Technická 2, 166 27 Prague 6, Czech Republic  
fax +420 2 2435 7385, phone +420 2 2435 7637, www: <http://cmp.felk.cvut.cz>

# Abstract

*We propose a novel MRF-based model for image matching. Given two images, the task is to estimate a mapping from one image to another, in order to maximize the matching quality. We consider mappings defined by discrete deformation field constrained to preserve 2-dimensional continuity. We approach the corresponding optimization problem by the TRW-S (sequential Tree-reweighted message passing) algorithm [11, 5]. Our model design allows for a considerably wider class of natural transformation and yields a compact representation of the optimization task. For this model TRW-S algorithm demonstrated nice practical performance in experiments. We also propose a concise derivation of the TRW-S algorithm as a sequential maximization of the lower bound on the energy function.*

**Keywords:** optical flow, registration, Energy minimization, MRF, message passing, early vision, graphical models.

## 1. Introduction

In many applications a flexible 2-dimensional deformation is highly desirable. In particular, to model flexible appearance, one incorporate such deformation as a component for joint segmentation-deformation [4] or joint segmentation-classification [13], where it serves to describe nonrigid motions (people, etc.), inclass variations (different cars), or both. The only natural constraint on the deformation field is that it has to be (piecewise) continuous. Applications also include computation of optical flow to estimate motion of a 3d scene, registration of images taken from different measuring devices, tracking in video sequences, etc. Great advance in early vision problems, like stereo, segmentation, image restoration and more, has been achieved due to models, based on the Markov Random Fields (MRFs), and development of global optimization algorithms. Recent comparison of optimization algorithms for popular vision problems can be found in [10]. More background on inference and learning algorithms for MRFs is given in [3].

Inspired by development of optimization algorithms for MRFs, we model the deformation as a discrete field and impose continuity constraints on it by means of interaction potentials. The corresponding problem of finding best deformation field is difficult and has not been addressed much (to our knowledge) in a global optimization framework. In [8] MRF model was applied to the optical flow problem. Flow orientation field and flow magnitude field are modeled as separate MRFs. It is first solved for orientation field and then for the magnitude. Each of these problems is reduced to computation of max-flow under their model. In [1] two-dimensional deformation field was modeled (as an example to  $\alpha$ - $\beta$  swap algorithm) as an MRF with unary variables taking values from the product set  $\Delta X \times \Delta Y$ . Optimization by  $\alpha$ - $\beta$  swap algorithm for this MRF quickly becomes intractable as  $\Delta X$  and  $\Delta Y$  grows, moreover, as a local search algorithm it could perform poorly without good initial guess. In [7] an algorithm is proposed 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. Experiments on real

images in [7] showed that it is hard to find a reasonably large persistent part of optimal solutions.

Our model is based on the model [7], we extend its flexibility by the use of blocks and special type of continuity constraints (Sect. 2). We apply recently developed [11, 5] TRW-S algorithm to our model. It is a variational optimization algorithm designed to solve certain LP-relaxation of the discrete energy minimization problem (this LP-relaxation was studied in e.g. original paper [9], review [12]). We rederive TRW-S in a concise way as a coordinate ascent algorithm for the dual to the LP relaxation problem (Sect. 3). The algorithm possess the same asymptotic behavior as algorithms in [6, 12], it may not solve the LP-relaxation problem, since its stationary points satisfy only a necessary optimality condition (studied in [9, 12, 5]). Yet it is very efficient in certain practical applications. We achieve significant improvement in the deformable image matching problem, as we show in our experiments (Sect. 4), demonstrating wide search range of pixel displacements ( $60 \times 60$  px range) and robustness to deformations and noise. We conclude with possible extensions in Sect. 5. Appendix A suggests a comparison of our decomposed relaxation to standard one.

## 2. Deformation model

In this section we consider previously used MRF models for the 2-dimensional deformation and introduce our *block model*. We discuss two important aspects of these models. First, their demands w.r.t TRW-S algorithm: number of variational variables (all components of all messages) and the complexity of passing of one message. Second, their modeling power to describe continuous deformations. We start with defining standard Energy minimization framework, following the notation of [11, 5]. For more background on graphical models and MRFs we refer to [3, 11].

### 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  $\mathbf{x} = \{x_s \mid s \in \mathcal{V}\}$ . 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  $\mathbf{x}$  be defined by:

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

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, corresponding to a certain Markov Random Field. The problem of finding maximum a posteriori (MAP) of this MRF corresponds to *energy minimization*:  $\min_{\mathbf{x}} E(\mathbf{x}|\theta)$ . Terms  $\theta_s(x_s)$ ,  $s \in \mathcal{V}$  can be viewed as vectors in  $\mathbb{R}^{\mathcal{L}}$  and terms  $\theta_{st}(x_s, x_t)$ ,  $st \in \mathcal{E}$  as vectors in  $\mathbb{R}^{\mathcal{L} \times \mathcal{L}}$ . The concatenated vector  $\theta \in \mathbb{R}^{(\mathcal{V} \times \mathcal{L}) \cup (\mathcal{E} \times \mathcal{L}^2)}$  is called a *parameter* vector.

## 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  $\mathbf{x}$  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  $D_{\mathbf{x}}$  from image  $I^2$  to image  $I^1$  be defined by  $(D_{\mathbf{x}}I)_s = I_{s+x_s}$ . 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  $\theta_s(x_s) = (I_s^1 - I_{s+x_s}^2)^2 / 2\sigma_I^2$ . This term corresponds to the statistical assumption of  $p((D_{\mathbf{x}}I)_s - I_s^1 | \mathbf{x}) \propto \mathcal{N}^3(0, \sigma_I^2), s \in \mathcal{V}$ , which means that the deformed image  $D_{\mathbf{x}}I^2$  is a noisy observation of image  $I^1$  under fixed  $\mathbf{x}$  assuming Gaussian noise in each color component with variance  $\sigma_I^2$ . The usual setting for the interaction potential is:  $\theta_{st}(x_s, x_t) = \|x_s - x_t\|^2 / 2\sigma_x^2 = (x_s^1 - x_t^1)^2 / 2\sigma_x^2 + (x_s^2 - x_t^2)^2 / 2\sigma_x^2, st \in \mathcal{E}$ , where  $\mathcal{E}$  is the set of all horizontally and vertically neighboring pairs of pixels. This term penalizes discontinuities in the deformation field  $\mathbf{x}$ , so that close pixels are forced to go to close destinations (see Fig. 1).

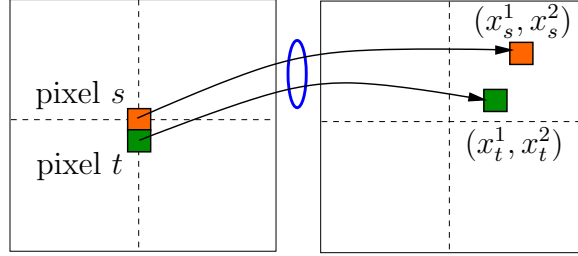


Figure 1. Deformation as pairwise MRF: data terms penalize color deviation of single pixels from hypothesized destination, pairwise terms penalize spatial deviations.

Minimizing energy, defined by  $\theta$ , is not known to be polynomially solvable problem for a general input (also, taking into account particular graph structure and particular interaction terms we do not have a proof that it is NP-hard). A message-passing algorithm for this problem (TRW-S, BP) would require keeping  $O(|\mathcal{V}||L|^2)$  variational variables – this is a number of variables in the dual of the LP-relaxation of the energy minimization. Also such algorithms would require  $O(|L|^4)$  operations for an elementary message update, which however may be reduced to  $O(|L|^2)$  if the special form of functions  $\theta_{st}$  is exploited, applying technique of [2].

## 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,  $\sim$  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 *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  $\theta_{st}(x_s, x_t) = (I_s^1 - I_{s+(x_s, x_t)}^2)^2 / 2\sigma_I^2$ ,  $s \in \mathcal{V}^1$ ,  $t \in \mathcal{V}^2$ ,  $st \in \mathcal{E}^{12}$ . 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  $E(x|\theta)$  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 however note, that the LP relaxation of the decomposed model is weaker then that one of the product model, see Appendix A.

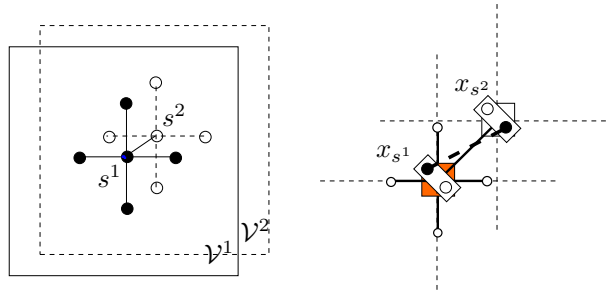


Figure 2. Decomposed model. Left: model consist of two layers  $\mathcal{V}^1$  and  $\mathcal{V}^2$ , the neighborhood structure is shown by nodes  $s^1$ ,  $s^2$  and interaction edges to their neighbors in  $\mathcal{E}$ . Right: inter-layer interaction is used to encode data term – 2D displacement 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).

## 2.4. Block model

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 also to affine transformations, unless it is a pure translation, in particular, scaling is poorly handled. It results in undesirable tradeoff between data term and continuity term: either the model will not follow the deformation observed by the data term, either it will allow discontinuities when the noise is present. We partially weaken the continuity term to allow for more flexibility and preserve overall continuity. We want the deformation field to be locally (in some small regions of pixels) affine. As we consider discrete models, in fact, we want the deformation field to be locally described by translations. We propose to group pixels in blocks and allow each block to have pixel wise displacements. We do not penalize relative displacements of  $\pm 1$  px in vertical and horizontal directions, and assign a large penalty to bigger displacements, see Fig. 3. Let the 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. We define the data term as

$$\theta_{st}(x_s, x_t) = \frac{1}{2\sigma_I^2} d(I_s^1, I_{s+(x_s, x_t)}^2), \quad s \in \mathcal{V}^1, t \in \mathcal{V}^2 : st \in \mathcal{E}^{12}, \quad (2)$$

where  $I_s^1$  is a fragment of image  $I^1$  on block  $s$ ;  $s + (x_s, x_t)$  is block  $s$  shifted by  $(x_s, x_t)$ , and  $d(\cdot, \cdot)$  is a dissimilarity measure (sum of squared differences across pixels of the

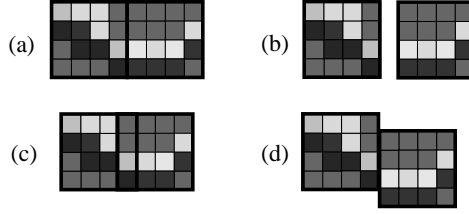


Figure 3. Block model: (a) two neighboring blocks of  $4 \times 4$  pixels; (b)-(d) examples of nonpenalized relative displacements. There is 9 nonpenalized relative displacements total.

corresponding image fragments or other). For all  $st \in \mathcal{E}^1 \cup \mathcal{E}^2$  we set the continuity term to

$$\theta_{st}(x_s, x_t) = \max(c_1(|x_s - x_t| - 1), c_2|x_s - x_t|), \quad (3)$$

where  $0 < c_2 \ll c_1$ , see Fig. 4. The subclass of transformations with low penalty (of factor of  $c_2$ ) naturally incorporate certain range of affine transformations (e.g. it include scale changes in the range 0.75–1.25, when blocks are  $4 \times 4$  px) and a certain degree of local flexibility, as could be seen from our experiments. Also, when the relative displacement of neighboring blocks equals  $-1$  in one or both of the coordinates (e.g. Fig. 3(d)), the blocks overlap and we count overlapping pixels twice. We choose to neglect this wrong counting, also treating it correctly would require using interactions of the 4th order.

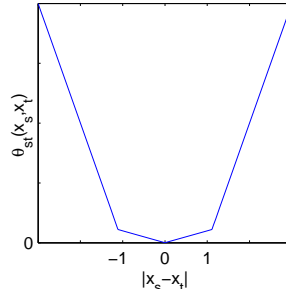


Figure 4. Continuity term of Eq. (3): shifts of  $\pm 1$ px are penalized by small regularization constant.

### 3. Optimization

In this section we give a compact review of the TRW-S algorithm developed in [11, 5]. We rederive it as (block-) coordinate ascent algorithm for maximization of the lower bound on the energy function, which is dual to the LP relaxation (originally, TRW algorithm was inferred [11] from necessary conditions of the maximum). This gives additional insight on the kind of suboptimality it possess w.r.t. the LP relaxation. We avoid using reparameterizations and max-marginal factorization in our inference and we use a simpler inequality for lower bounds, which is without coefficients of convex combination, as suggested for edge-based bound in [9, 12].

Let us consider graph  $G = (\mathcal{V}, \mathcal{E})$  and energy function of the from (1). Let  $\Omega_{G, \mathcal{L}} = \mathbb{R}^{(\mathcal{V} \times \mathcal{L}) \cup (\mathcal{E} \times \mathcal{L}^2)}$  be a state space of parameter vectors  $\theta$ , as defined in Sec. 2. Let  $E(x|\theta)$

be expressed as a scalar product in the space  $\Omega_{G,\mathcal{L}}$ :  $E(x|\theta) = \langle \theta, \mu(x) \rangle$ , where  $\phi : \mathcal{L}^{\mathcal{V}} \mapsto \Omega_{G,\mathcal{L}}$  is a mapping defined by  $[\mu(x)]_s(x) = \delta_{\{x_s=x\}}$ ,  $s \in \mathcal{V}$ ,  $x \in \mathcal{L}$ ;  $[\mu(x)]_{st}(x, x') = \delta_{\{x_s=x\}}\delta_{\{x_t=x'\}}$ ,  $st \in \mathcal{E}$ ,  $x, x' \in \mathcal{L}$ , where  $\delta_{\{A\}}$  is 1 if  $A$  is true and 0 otherwise.

**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  $\text{LB}(\boldsymbol{\theta}) = \sum_i \min_x \langle \theta^i, \mu(x) \rangle$  is a lower bound on the optimal energy:

$$\text{LB}(\boldsymbol{\theta}) \leq \min_x \langle \theta, \mu(x) \rangle. \quad (4)$$

*Proof.*  $\min_x \langle \theta, \mu(x) \rangle = \min_x \langle \sum_i \theta^i, \mu(x) \rangle = \min_x \sum_i \langle \theta^i, \mu(x) \rangle \geq \sum_i \min_x \langle \theta^i, \mu(x) \rangle$ . Alternatively could be seen from the Jensen's inequality as in [11] and linearity of  $\langle \cdot, \mu(x) \rangle$ .  $\square$

**Tree structured distributions.** The computation of  $\text{LB}(\boldsymbol{\theta})$  is tractable if all the problems  $\min_x \langle \theta^i, \mu(x) \rangle$  are tractable. For this purpose, each  $\theta^i$  can be chosen to define a tree-structured distribution [11]. Let  $\forall i \in I$ :  $T^i = (\mathcal{V}^i, \mathcal{E}^i)$  be a tree-structured subgraph of  $G$ . Let set  $\Omega_{T^i,\mathcal{L}}$  define the subspace of parameter vectors over the tree  $T^i$ , that is  $\Omega_{T^i,\mathcal{L}} = \{\tilde{\theta} \in \Omega_{G,\mathcal{L}} \mid \tilde{\theta}_{u;x_u} = 0, \forall u \in \mathcal{V} \setminus \mathcal{V}^i; \tilde{\theta}_{uv;x_{uv}} = 0, \forall uv \in \mathcal{E} \setminus \mathcal{E}^i\}$ . Let  $\theta^i \in \Omega_{T^i,\mathcal{L}}$ , the minimization problem  $\min_x \langle \theta^i, \mu(x) \rangle$  is tree structured and therefore is tractable.

Constraints  $\sum_i \theta^i = \theta$  require that trees  $\boldsymbol{T} = \{T^i \mid i \in I\}$  fully cover graph  $G$ :  $\bigcup_{i \in I} T^i = G$ . For each particular set of trees  $\boldsymbol{T}$  and choice of  $\boldsymbol{\theta}$  satisfying constraints, the lower bound  $\text{LB}(\boldsymbol{\theta})$  can be easily computed. Consider finding the tightest bound over possible choices of  $\boldsymbol{\theta}$ :

$$\begin{aligned} \text{LB} = \max_{\boldsymbol{\theta} = \{\theta^i \mid i \in I\}} \text{LB}(\boldsymbol{\theta}) & \quad \left( \leq \min_x \langle \theta, \mu(x) \rangle \right) \\ \text{s.t. } & \begin{cases} \theta^i \in \Omega_{T^i,\mathcal{L}}, i \in I; \\ \sum_i \theta^i = \theta. \end{cases} \end{aligned} \quad (5)$$

Function  $\text{LB}(\boldsymbol{\theta})$  is a concave function of  $\boldsymbol{\theta}$ , as it is a sum of minima of linear functions. Therefore maximization task (5) is a concave maximization. It was shown in [11] that dual problem to (5) constitutes an LP relaxation of the original energy minimization, independently of the trees  $\boldsymbol{T}$  chosen. Strong duality holds, therefore (5) does not depend on the choice of trees  $\boldsymbol{T}$  as well. This gives us a freedom to choose the trees in a convenient way.

We let collection  $\boldsymbol{T} = \{(\mathcal{V}^i, \mathcal{E}^i) \mid i \in I\}$  to form an edge-disjoint cover of  $G$ , namely  $\mathcal{E}^i \cap \mathcal{E}^j = \emptyset$ ,  $i, j \in I$ ;  $\bigcup_{i \in I} T^i = G$ . In this case for each edge  $uv \in \mathcal{E}$  there is a single nonzero term in the LHS of the constraint  $\sum_i \theta_{uv}^i = \theta_{uv}$  – it is the term  $\theta_{uv}^i$  for which  $uv \in \mathcal{E}^i$ . Therefore  $\theta_{uv}^i = \theta_{uv} \delta_{\{uv \in \mathcal{E}^i\}}$  and we can rewrite constraints of (5) as follows:

$$\text{LB} = \max_{\boldsymbol{\theta}} \text{LB}(\boldsymbol{\theta}), \quad \text{s.t. } \begin{cases} \theta_s^i = 0, & s \in \mathcal{V} \setminus \mathcal{V}^i, i \in I \\ \sum_{i \in I_s} \theta_s^i = \theta_s, & s \in \mathcal{V} \\ \theta_{uv}^i = \theta_{uv} \delta_{\{uv \in \mathcal{E}^i\}}, & uv \in \mathcal{E} \end{cases}. \quad (6)$$

**Coordinate ascent.** Let  $I_s = \{i \mid s \in \mathcal{V}^i\}$  be the set of trees sharing the node  $s \in \mathcal{V}$ . Let us fix some  $s \in \mathcal{V}$  and consider maximization of  $\text{LB}(\boldsymbol{\theta})$  in (6) with respect to the

subset of variables  $\boldsymbol{\theta}_s = \{\theta_s^i \mid i \in I\}$ , while keeping the rest of variables, referred to as  $\boldsymbol{\theta}_{\mathcal{V} \setminus s}$ , fixed:

$$\text{LB}_s(\boldsymbol{\theta}_{\mathcal{V} \setminus s}) = \max_{\boldsymbol{\theta}_s} \sum_i \min_x \langle \theta_s^i, \phi(x) \rangle, \quad s.t. \begin{cases} \theta_s^i = 0, & i \in I \setminus I_s \\ \sum_{i \in I_s} \theta_s^i = \theta_s \end{cases}. \quad (7)$$

Iterative update of variables  $\boldsymbol{\theta}_s$  for each  $s \in \mathcal{V}$  yields a coordinate ascent algorithm, which builds a sequence of monotonically nondecreasing lower bounds. Generally, it does not converge to the optimal bound (5) (because function  $\text{LB}(\boldsymbol{\theta})$ , to which we apply sequential maximization is not smooth). We now derive the TRW-S algorithm as follows: we rearrange (7) as

$$\max_{\boldsymbol{\theta}_s} \sum_i \min_{x_s} \Phi_s^i(x_s), \quad s.t. \begin{cases} \theta_s^i = 0, & i \in I \setminus I_s \\ \sum_{i \in I_s} \theta_s^i = \theta_s \end{cases}, \quad (8)$$

where values  $\Phi_s^i(x_s) = \min_{x_{T \setminus s}} \langle \theta_s^i, \phi(x) \rangle = \theta_s^i(x_s) + \min_{x_{T \setminus s}} (\sum_{t \in \mathcal{V}^i \mid t \neq s} \theta_t^i(x_t) + \sum_{\tilde{st} \in \mathcal{E}^i} \theta_{\tilde{st}}^i(x_{\tilde{s}}, x_{\tilde{t}}))$  are the *min-marginals* [11] of the tree-structured energy  $\langle \theta^i, \phi(x) \rangle$ . Let us introduce also the *incomplete* min-marginals  $\Psi_s^i(x_s) = \Phi_s^i(x_s) - \theta_s^i(x_s)$ , which do not depend on  $\theta_s^i$ . Using incomplete min-marginals and incorporating constraints  $\theta_s^i = 0, i \in I \setminus I_s$ , we can express (8) as follows:

$$\max_{\{\theta_s^i \mid i \in I_s\}} \sum_{i \in I_s} \min_{x_s} (\theta_s^i + \Psi_s^i)(x_s) + \sum_{j \in I \setminus I_s} \Phi^j, \quad s.t. \sum_{i \in I_s} \theta_s^i = \theta_s, \quad (9)$$

where values  $\Phi^j = \min_{x_s} \Phi_s^j(x_s)$ ,  $j \in I \setminus I_s$  do not depend on the variables  $\{\theta_s^i \mid i \in I_s\}$ . To find a maximizer of (9) we apply inequality:

$$\sum_{i \in I_s} \min_{x_s} (\theta_s^i + \Psi_s^i)(x_s) \leq \min_{x_s} \sum_{i \in I_s} (\theta_s^i + \Psi_s^i)(x_s) = \min_{x_s} \left( \theta_s + \sum_{i \in I_s} \Psi_s^i \right)(x_s), \quad (10)$$

where we used  $\sum_{i \in I_s} \theta_s^i = \theta_s$ . The special case when (10) is satisfied with equality is when all min-marginals  $\Phi_s^i = (\theta_s^i + \Psi_s^i)$  are equal, which can be easily achieved by the following proposition:

**Min-marginal averaging.** To make all min-marginals  $\Phi_s^i$  equal for a fixed  $s$ , set

$$\theta_s^i = \frac{1}{|I_s|} \left( \theta_s + \sum_{j \in I_s} \Psi_s^j \right) - \Psi_s^i, \quad i \in I_s. \quad (11)$$

It is straightforward to verify that constraint  $\sum_{i \in I_s} \theta_s^i = \theta_s$  is satisfied and that resulting min-marginals  $\Phi_s^i = \theta_s^i + \Psi_s^i = \frac{1}{|I_s|} \left( \theta_s + \sum_{j \in I_s} \Psi_s^j \right)$  do not depend on  $i$  and therefore (10) holds with equality, which means that maximum in (9) is attained. Now we are ready to formulate the TRW-S algorithm.



*Algorithm 1: Simple TRW-S.*

1. Initialize  $\theta_s^i = \frac{1}{|I_s|}\theta_s, i \in I$ .
2. For each  $s \in \mathcal{V}$  sequentially perform:
  - (a) compute incomplete min-marginals  $\Psi_s^i, i \in I_s$  by dynamic programming on the tree  $T^i$ ;
  - (b) update  $\theta_s^i = \frac{1}{|I_s|} \left( \theta_s + \sum_{j \in I_s} \Psi_s^j \right) - \Psi_s^i, i \in I_s$ .
3. Compute the actual lower bound  $\text{LB}(\boldsymbol{\theta}^k) = \sum_i \min_x \langle \theta^i, \phi(x) \rangle$ .
4. Check stopping condition, based on convergence of  $\text{LB}(\boldsymbol{\theta}^k)$ . If not satisfied go to step 2.

The fixed point condition is characterized by all min-marginals  $\{\Phi_s^i \mid i \in I_s\}$  being equal. This allows one to choose a solution  $\mathbf{x}$  by local minimizers  $x_s \in \arg \min_{\mathcal{L}} \Phi_s^i(\cdot), i \in I_s$ . If all local minimizers are unique<sup>1</sup>, configuration  $\mathbf{x}$  is guaranteed to be optimal [11].

The stopping condition can be based on the convergence of  $\text{LB}(\boldsymbol{\theta})$ . The efficient implementation with chains [5] is as follows. Let  $T^i$  be a chain. Its min-marginals can be computed dynamically as  $\Phi_s^i = \overrightarrow{\Psi}_s^i + \theta_s^i + \overleftarrow{\Psi}_s^i$ , where  $\overrightarrow{\Psi}_s^i(x_s) = \min_{x_t} \{ \overrightarrow{\Psi}_t^i(x_t) + \theta_t^i(x_t) + \theta_{st}^i(x_s, x_t) \}$ , if  $t$  is the node previous to  $s$  in the chain  $T^i$  and  $\overrightarrow{\Psi}_s^i(x_s) = 0$ , if  $s$  is the first element of the chain. Analogously, values  $\overleftarrow{\Psi}_s^i = \min_{x_t} \{ \theta_{st}^i(x_s, x_t) + \theta_t^i(x_t) + \overleftarrow{\Psi}_t^i(x_t) \}$ , if  $t$  is next to  $s$  in the chain  $T^i$  and  $\overleftarrow{\Psi}_s^i = 0$  if  $s$  is the last in the chain. This computation can be efficiently combined with step 2(b) by substituting step 2(b) into the expression for  $\overrightarrow{\Psi}_s^i$ , which gives us the following update equation, known as *message passing* [5]:

$$\overrightarrow{\Psi}_s^i(x_s) = \min_{x_t} \left\{ \frac{1}{|I_t|} \left( \theta_t(x_t) + \sum_{j \in I_t} [\overrightarrow{\Psi}_t^j(x_t) + \overleftarrow{\Psi}_t^j(x_t)] \right) - \overleftarrow{\Psi}_t^i(x_t) + \theta_{st}(x_s, x_t) \right\}. \quad (12)$$

Computation order must guarantee that dependent values  $\overrightarrow{\Psi}_t^j$  are computed earlier. It is possible when the computation order matches orderings "previous to"  $(t, s)$  in the chain  $T^i$  for all  $i \in I$ . In this case<sup>2</sup> chains  $\mathbf{T}$  are said to be *monotonic* [5] w.r.t. ordering on  $\mathcal{V}$ .

**Convergence issues.** Let Algorithm 1 build a sequence  $\{\boldsymbol{\theta}^k\}_{k=0}^\infty$ . The algorithm performs sequential maximization of  $\text{LB}(\boldsymbol{\theta})$  w.r.t. variables  $\{\theta_s^i \mid i \in I_s\}$ , therefore sequence  $\{\text{LB}(\boldsymbol{\theta}^k)\}_k$  does not decrease. Since we have  $\text{LB}(\boldsymbol{\theta}^k) \leq \min_x \langle \boldsymbol{\theta}, \phi(x) \rangle \leq \infty$ , the sequence converges to some  $\text{LB}^*$ .

Fixed points of the algorithm possess the property:  $(\exists O_i \subset \arg \min_x \langle \theta^i, \mu(\mathbf{x}) \rangle, O_i \neq \emptyset, i \in I): \forall s \in \mathcal{V}: \forall i, j \in I_s: \{x_s \mid \mathbf{x} \in O^i\} = \{x_s \mid \mathbf{x} \in O^j\}$ , called WTA conditions in [5] or

<sup>1</sup>The initial problem is NP-complete in general, therefore we can not hope to have unique local minimizers and solve it to optimality.

<sup>2</sup>For the grid graph the natural choice is horizontal and vertical chains and row-major ordering on  $\mathcal{V}$ . Also it is satisfied for any graph and chains of length 1 w.r.t. any ordering, however the efficiency improvement is lost.

*local consistency* in [9, 12]. If this property does not hold for a point  $\theta^k$ , the lower bound  $\text{LB}(\cdot)$  is guaranteed to improve in finite number of iterations [5]. Continuity of the  $\text{LB}(\theta)$  and boundedness [5] of  $\theta^k$  implies that there exist a convergent subsequence  $\{\theta^{k_j}\}_j$ , with limit point satisfying WTA. Based on this condition local minimizers  $x_s$  can be chosen from  $\{\tilde{x}_s \mid \tilde{x} \in O^i\}$ , since these sets coincide for all  $i \in I_s$ .

## 4. Implementation and Experiments

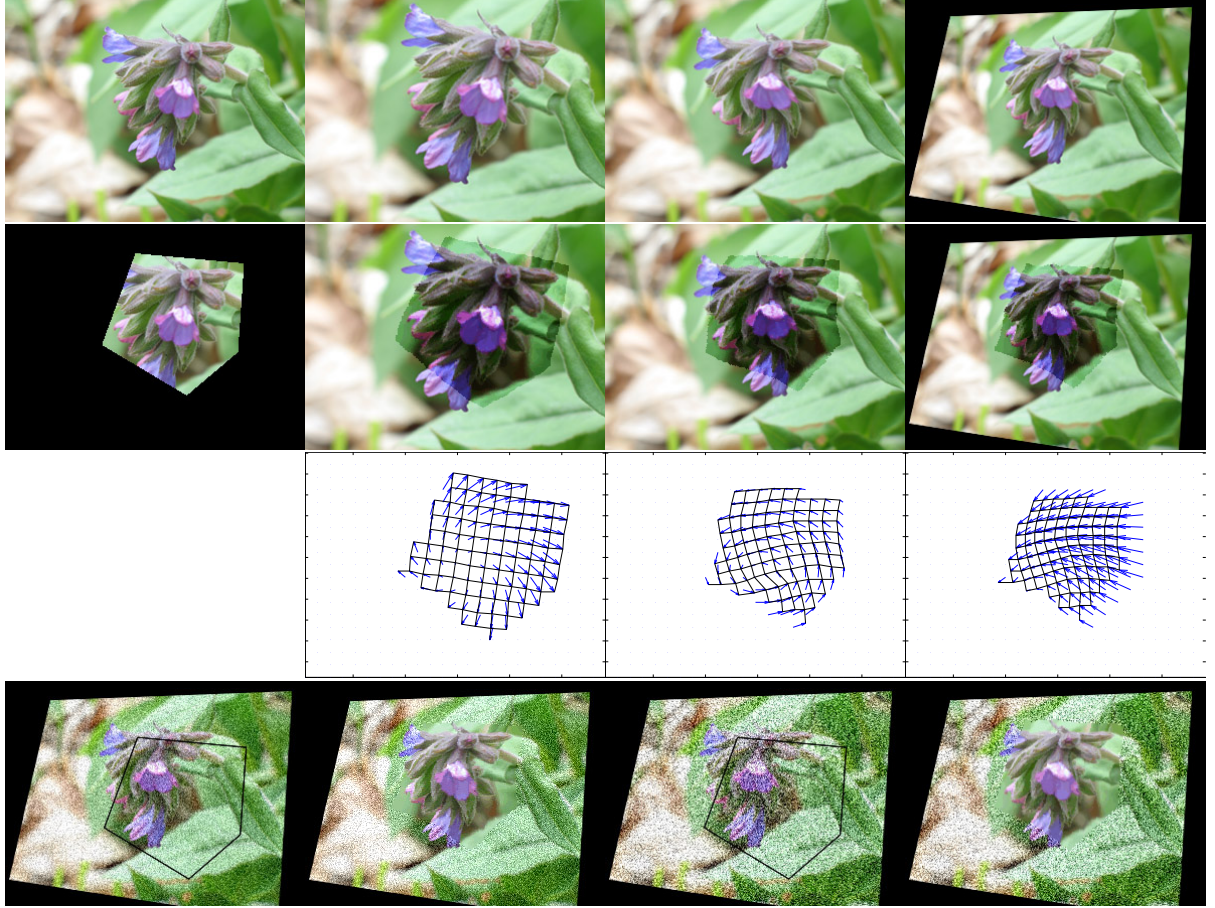


Figure 5. First row: initial picture, rotation+scaling, twirl, twirl + projective transform. Second row: searched fragment and its found transformations overlaid (with increased contrast) onto corresponding inputs. Third row: found deformation field (shown sparse) and a transformed regular grid. Fourth row: noised (0.1 gaussian) image of row 1d with overlaid contour of the searched fragment(see text), matched fragment, more noised (0.2 gaussian) image, matched fragment. See <http://cmp.felk.cvut.cz/~shekhovt/deform-match-mrf> for more examples and animated transformations.

**Implementation.** We have implemented TRW-S as suggested in [5] and described above. Our tree decomposition  $\mathbf{T}$  consists of all vertical and horizontal chains in layers  $\mathcal{V}^1$  and  $\mathcal{V}^2$  plus one-edge chains, corresponding 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  $O(|L|)$  method [2].

As inter-layer ( $\mathcal{E}^{12}$ ) updates are considerably slower, we modified the sequential schedule to perform more intra-layer iterations (this is compatible with constraints of TRW-S on the order of updates). For the data term we precompute block correlations. To speed up their computation we first cluster image color space and precompute the table of color comparisons. Also we stretched the quadratic form ( $F$ ) of the color comparison model along the  $(1, 1, 1)$  direction in the RGB space to improve robustness against brightness changes. We assess convergence based on WTA conditions, specifically we iterations when WTA conditions are satisfied with some precision  $\varepsilon$ , this is a good heuristic for our application.

**Constants.** We represent color space as  $[0, 1]^3$  and compute color correlation using stretched quadratic form  $F$  with eigenvalues  $(1, 1, 0.1)$ . Then we compute  $d(x, y)$  as  $\frac{1}{N} \sum_{i=1}^N F(x^i, y^i)$ , where  $x^i, y^i$  are colors of pixels in correspondence. We set  $\sigma_1 = 1$ ,  $c_1 = 10$ ,  $c_2 = 0.01 - c_1$  weights “hard” continuity terms,  $\sigma_1^2$  weights data term and  $c_2$  weights regularization term. When a pixel is mapped out of the field of view we assign it a penalty of 0.1. Experimental results were rather insensitive to these parameters, in fact any choice  $c_1 \gg 1/\sigma_1^2 \gg c_2$  is good.

**Experiments.** We propose synthetic experiments. To show the flexibility of the deformation field we applied several parametric transformations to an image (Fig.5) and searched for a matching of its subregion<sup>3</sup> to the deformed images. The chosen subregion does not align with visual edges of the image so the quality of the matching could be seen when overlayed. All images are  $300 \times 225$  px, we used  $L = \{-30 \dots 30\}$  for  $x$  and  $y$  allowed displacements, blocks were  $4 \times 4$ . Computation took  $\sim 40$  sec./ image, from which 11.2 sec were spent on precomputing block correlations. We keep small translation component, since the optimization is invariant to it, as far as the search range is wide enough. We illustrate this on noisy images, for which finer components of the matching are less accurate. We also demonstrate how our method works on real images Fig.6-8.

## 5. Conclusion

We 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. We apply the TRW-S algorithm to maximize the the dual of the linearly relaxed energy minimization problem. We gave an additional insight on the TRW-S algorithm by interpreting it as a coordinate ascent to maximize the dual and proposed its short derivation. As we inherit the structure of the model from [7], our variational optimization problem possess the 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 deformation fields compared to considered previously in the global optimization framework. Our experiments demonstrates high visual accuracy within wide class of continuous deformations and robustness to a high degree of noise.

As in stereo, we could possibly use truncated continuity term. However, as we are tied to blocks, the discontinuity boundary would be poorly modeled in this case. Another option is to combine our model with segmentation, as it is done in [13]. We actually could incorporate pixel-wise segmentation, while keeping the deformation model operating with

---

<sup>3</sup>We need the subregion to assume that all pixels from  $T^1$  have their correspondences in  $T^2$



Figure 6. Morphing face expressions using found deformation: interpolated intermediate frames from top-left image to bottom-right image. For the intermediate frames the deformation and colors of pixels in correspondence are linearly interpolated.



Figure 7. Morphing a rumpled newspaper: interpolated intermediate frames from top-left image to bottom-right image. For the intermediate frames the deformation and colors of pixels in correspondence are linearly interpolated.

blocks. To cope with larger global translations (or scale, rotation etc.) one may consider adding a common, roughly discretized, transformation variable. This however leads to multiple ambiguous solutions in our model so the optimization by TRW-S becomes badly conditioned. Another related issue is the choice of a solution based on min-marginals, if it is done by local minimizers, it might happen that output solution will violate hard





Figure 8. Deformation for face expression: interpolated intermediate frames following the found deformation from top-left image to bottom-right image.

continuity constraints. 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.

## Appendix A: Weaker Relaxation

Consider graph  $G = (\mathcal{V}, \mathcal{E})$  and energy function  $E(\mathbf{x}|\theta)$  as defined in Sec. 2. The *LP-relaxation* of energy minimization is the following lower bound:

$$\min_{\mathbf{x} \in \mathcal{L}^{\mathcal{V}}} E(\mathbf{x}|\theta) = \min_{\mathbf{x} \in \mathcal{L}^{\mathcal{V}}} \langle \theta, \mu(\mathbf{x}) \rangle = \min_{\mu \in \mathcal{M}_{G,\mathcal{L}}} \langle \theta, \mu \rangle \geq \min_{\mu \in \Lambda_{G,\mathcal{L}}} \langle \theta, \mu \rangle, \quad (13)$$

where  $\mathcal{M}_{G,\mathcal{L}} = \text{conv}\{\mu(\mathbf{x}) \mid \mathbf{x} \in \mathcal{L}^{\mathcal{V}}\}$  is a *marginal* polytope and  $\Lambda_{G,\mathcal{L}} \supseteq \mathcal{M}_{G,\mathcal{L}}$  is a *local* polytope defined as a set of locally compatible unary and pairwise distributions:

$$\Lambda_{G,\mathcal{L}} = \left\{ \mu \in \Omega_{G,\mathcal{L}} \left| \begin{array}{ll} \mu_{st}(x_s, x_t) \geq 0, & st \in \mathcal{E}, x_{st} \in \mathcal{L}^2 \\ \sum_{x_t \in \mathcal{L}} \mu_{st}(x_s, x_t) = \mu_s(x_s) & st \in \mathcal{E}, x_s \in \mathcal{L} \\ \sum_{x_s \in \mathcal{L}} \mu_{st}(x_s, x_t) = \mu_t(x_t) & st \in \mathcal{E}, x_t \in \mathcal{L} \\ \sum_{x_s \in \mathcal{L}} \mu_s(x_s) = 1 & s \in \mathcal{V}, x_s \in \mathcal{L} \end{array} \right. \right\} \quad (14)$$

We show below that whereas LHS in (13) is preserved when the problem is reformulated via decomposed model, the RHS is not, so the LP-relaxation of the decomposed model is weaker than that of product model.

### Energy equivalence

Let  $\mathcal{L} = L \times L$ . Let  $x_s = (x_s^1, x_s^2) \in L \times L$ ,  $s \in \mathcal{V}$ ,  $st \in \mathcal{E}$ . Let  $\theta_{st}(x_s, x_t) = \theta_{st}^1(x_s^1, x_t^1) + \theta_{st}^2(x_s^2, x_t^2)$ .

Recall the construction of decomposed model (Sec. 2). Let us denote the graph of decomposed model as  $\tilde{G} = (\tilde{\mathcal{V}}, \tilde{\mathcal{E}})$ , with set of vertices  $\tilde{\mathcal{V}} = \tilde{\mathcal{V}}^1 \cup \tilde{\mathcal{V}}^2$ , where  $\tilde{\mathcal{V}}^1 \sim \tilde{\mathcal{V}}^2 \sim \mathcal{V}$  and set of edges  $\tilde{\mathcal{E}} = \tilde{\mathcal{E}}^1 \cup \tilde{\mathcal{E}}^2 \cup \tilde{\mathcal{E}}^{12}$ , where  $\tilde{\mathcal{E}}^1 \sim \tilde{\mathcal{E}}^2 \sim \mathcal{E}$ ,  $\tilde{\mathcal{E}}^{12} = \{(s^1, s^2) \mid s^1 \in \tilde{\mathcal{V}}^1, s^2 \in \tilde{\mathcal{V}}^2 : s^1 \sim s^2\}$ . Let  $\tilde{x} = \{\tilde{x}_s \mid s \in \tilde{\mathcal{V}}\} \in L^{\tilde{\mathcal{V}}}$  be a configuration in the decomposed model. Clearly  $\{(\tilde{x}_{s^1}, \tilde{x}_{s^2}) \mid s \in \mathcal{V}\} \in \mathcal{L}^{\mathcal{V}}$  is then a configuration in the product model, we will say that  $x \equiv \tilde{x}$  when  $(x_s^1, x_s^2) = (\tilde{x}_{s^1}, \tilde{x}_{s^2})$  for all  $s^1 \sim s^2 \sim s \in \mathcal{V}$ .

The parameter vector  $\theta$  of the decomposed model is defined as follows: for all  $i, j \in L$ :

$$\begin{aligned}\tilde{\theta}_s(i) &= 0, & s \in \tilde{\mathcal{V}} \\ \tilde{\theta}_{s^1 s^2}(i, j) &= \theta_s((i, j)), & s^1 \sim s^2 \sim s \in \mathcal{V} \\ \tilde{\theta}_{s^1 t^1}(i, j) &= \theta_{st}^1(i, j), & s^1 \sim s, t^1 \sim t, st \in \mathcal{E} \\ \tilde{\theta}_{s^2 t^2}(i, j) &= \theta_{st}^2(i, j), & s^2 \sim s, t^2 \sim t, st \in \mathcal{E}\end{aligned}\tag{15}$$

Energy function in the decomposed model  $\tilde{E}(\tilde{x}|\tilde{\theta}) = \langle \tilde{\theta}, \tilde{\mu}(\tilde{x}) \rangle$  is easily seen to be equal to  $E(x|\theta)$  for all  $x \equiv \tilde{x}$ . However LP-relaxations of these energies do not necessarily coincide, since they depend on the underlying graph. Indeed, we have:

$$\min_{x \in \mathcal{L}^{\mathcal{V}}} E(x|\theta) \geq \min_{\mu \in \Lambda_{G, \mathcal{L}}} \langle \theta, \mu \rangle \quad \text{and} \quad \min_{\tilde{x} \in L^{\tilde{\mathcal{V}}}} \tilde{E}(\tilde{x}|\tilde{\theta}) \geq \min_{\tilde{\mu} \in \Lambda_{\tilde{G}, L}} \langle \tilde{\theta}, \tilde{\mu} \rangle\tag{16}$$

### Weaker relaxation

We will show that in general the LP-relaxation of  $\tilde{E}$  is weaker:

$$\min_{\mu \in \Lambda_{G, \mathcal{L}}} \langle \theta, \mu \rangle \geq \min_{\tilde{\mu} \in \Lambda_{\tilde{G}, L}} \langle \tilde{\theta}, \tilde{\mu} \rangle\tag{17}$$

*Proof.* We prove “ $\geq$ ” by constructing a mappings:  $\Lambda_{G, \mathcal{L}} \rightarrow \Lambda_{\tilde{G}, L}$  preserving the equality  $\langle \theta, \mu \rangle = \langle \tilde{\theta}, \tilde{\mu} \rangle$ . Then we provide a counterexample showing that in general equality is not attained.

- Mapping  $\Lambda_{G, \mathcal{L}} \rightarrow \Lambda_{\tilde{G}, L}$ . Let  $\mu \in \Lambda_{G, \mathcal{L}}$  is fixed. We chose  $\tilde{\mu} \in \Lambda_{\tilde{G}, L}$  as follows:

$$\begin{aligned}\tilde{\mu}_{s^1 s^2}(i, j) &= \mu_s((i, j)), & s^1 \sim s^2 \sim s \in \mathcal{V} \\ \tilde{\mu}_{s^1 t^1}(i, i') &= \sum_{j, j' \in L} \mu_{st}((i, j), (i', j')), & s^1 \sim s, t^1 \sim t, st \in \mathcal{E} \\ \tilde{\mu}_{s^2 t^2}(j, j') &= \sum_{i, i' \in L} \mu_{st}((i, j), (i', j')), & s^2 \sim s, t^2 \sim t, st \in \mathcal{E} \\ \tilde{\mu}_{s^1}(i) &= \sum_{j \in L} \mu_s((i, j)), & s^1 \in \tilde{\mathcal{V}}^1 \\ \tilde{\mu}_{s^2}(j) &= \sum_{i \in L} \mu_s((i, j)), & s^2 \in \tilde{\mathcal{V}}^2\end{aligned}\tag{18}$$

Now we verify that  $\tilde{\mu} \in \Lambda_{\tilde{G}, L}$ . In particular all the following marginalization constraints must hold:

$$\begin{aligned}\tilde{\mu}_{s^1}(i) &= \sum_{j \in L} \tilde{\mu}_{s^1 s^2}(i, j), & s^1 \in \mathcal{V}^1, s^1 s^2 \in \tilde{\mathcal{E}}^{12} & \text{(a)} \\ \tilde{\mu}_{s^1}(i) &= \sum_{i' \in L} \tilde{\mu}_{s^1 t^1}(i, i'), & s^1 t^1 \in \tilde{\mathcal{E}}^1 & \text{(b)} \\ \tilde{\mu}_{s^2}(j) &= \sum_{i \in L} \tilde{\mu}_{s^1 s^2}(i, j), & s^2 \in \mathcal{V}^2, s^1 s^2 \in \tilde{\mathcal{E}}^{12} & \text{(c)} \\ \tilde{\mu}_{s^2}(j) &= \sum_{j' \in L} \tilde{\mu}_{s^2 t^2}(j, j'), & s^2 t^2 \in \tilde{\mathcal{E}}^2 & \text{(d)}\end{aligned}\tag{19}$$

Case (a) follows from the definition of  $\tilde{\mu}_{s^1}(i)$ . Case (b) is verified as follows:  
 $\sum_{i' \in L} \tilde{\mu}_{s^1 t^1}(i, i') = \sum_{i' \in L} \sum_{j, j' \in L} \mu_{st}((i, j), (i', j')) = \sum_{j \in L} \sum_{i', j' \in L} \mu_{st}((i, j), (i', j')) = \sum_{j \in L} \mu_s((i, j)) = \tilde{\mu}_{s^1}(i)$ . Cases (c) and (d) are analogous to (a) and (b). Other constraints of  $\Lambda_{\tilde{G}, L}$  are straightforward.

The equality  $\langle \theta, \mu \rangle = \langle \tilde{\theta}, \tilde{\mu} \rangle$  is shown as follows:

$$\begin{aligned}
\langle \theta, \mu \rangle &= \sum_{s \in \mathcal{V}} \sum_{(i, j) \in \mathcal{L}} \theta_s((i, j)) \mu_s((i, j)) + \sum_{st \in \mathcal{E}} \sum_{\substack{(i, j) \in \mathcal{L} \\ (i', j') \in \mathcal{L}}} \theta_{st}((i, j), (i', j')) \mu_{st}((i, j), (i', j')) \\
&= \sum_{s^1 s^2 \in \tilde{\mathcal{E}}^{12}} \sum_{i \in L} \tilde{\theta}_{s^1 s^2}(i, j) \tilde{\mu}_{s^1 s^2}(i, j) + \sum_{st \in \tilde{\mathcal{E}}^1} \sum_{i, i' \in L} \theta_{st}^1(i, i') \tilde{\mu}_{st}(i, i') + \sum_{st \in \tilde{\mathcal{E}}^2} \sum_{j, j' \in L} \theta_{st}^2(j, j') \tilde{\mu}_{st}(j, j') \\
&= \sum_{st \in \tilde{\mathcal{E}}} \sum_{i, j \in L} \theta_{st}(i, j) \tilde{\mu}_{st}(i, j) = \langle \tilde{\theta}, \tilde{\mu} \rangle
\end{aligned} \tag{20}$$

- Example of a gap. Because we represent each interacting pair  $(x_s, x_t)$ ,  $st \in \mathcal{E}$  as a four tuple  $(\tilde{x}_{s^1}, \tilde{x}_{s^2}, \tilde{x}_{t^1}, \tilde{x}_{t^2})$  in the decomposed model, it is easy to see that there exists an example for which LP-relaxation on this four tuple is not tight, see Fig. 9.

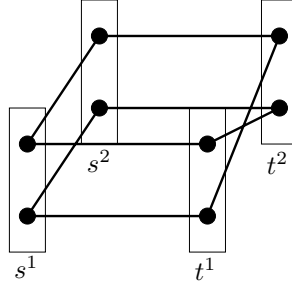


Figure 9. Part of decomposed graph for interacting pair  $st$ : a four tuple  $\tilde{\mu}_{s^1 s^2}, \tilde{\mu}_{s^1 t^1}, \tilde{\mu}_{s^2 t^2}, \tilde{\mu}_{t^1 t^2}$  is shown (having value  $\frac{1}{2}$  on bold edges and nodes), which can not be represented as marginals of a valid distribution  $\mu_{st}((x_{s^1}, x_{s^2}), (x_{t^1}, x_{t^2}))$  over the four tuple.

□

## References

- [1] 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 (AISTATS’05)*, pages 166–173, 2005.
- [5] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. In *AI and Statistics (AISTATS’05)*, pages 182–189, 2005.
- [6] V. Kovla and M. Schlesinger. Two-dimensional programming in image analysis problems. *Automatics and Telemekhanics*, 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] S. Roy and V. Govindu. MRF solutions for probabilistic optical flow formulations. In *ICPR’00*, pages Vol III: 1041–1047, 2000.
- [9] M. Schlesinger. Syntactic analysis of two-dimensional visual signals in noisy conditions. *Kibernetika, Kiev*, 4:113–130, 1976. In Russian.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] J. Winn and N. Jojic. Locus: Learning object classes with unsupervised segmentation. In *ICCV*, volume 1, pages 756–763, 2005.