Linear Intensity-based Image Registration by Markov Random Fields and Discrete Optimization

Darko Zikic[*,a], Ben Glocker[*,a], Oliver Kutter[a], Martin Groher[a], Nikos Komodakis[b], Ali Kamen[c], Nikos Paragios[d,e], Nassir Navab[a]

[a]*Computer Aided Medical Procedures (CAMP), Technische Universität München, Germany*
[b]*Computer Science Department, University of Crete, Greece*
[c]*Siemens Corporate Research (SCR), Princeton, NJ, USA*
[d]*Laboratoire MAS, Ecole Centrale Paris, Chatenay-Malabry, France*
[e]*Equipe GALEN, INRIA Saclay - Ile-de-France, Orsay, France*

Abstract

We propose a framework for intensity-based registration of images by linear transformations, based on a discrete Markov Random Field (MRF) formulation. Here, the challenge arises from the fact that optimizing the energy associated with this problem requires a high-order MRF model. Currently, methods for optimizing such high-order models are less general, easy to use, and efficient, than methods for the popular second-order models. Therefore, we propose an approximation to the original energy by an MRF with tractable second-order terms. The approximation at a certain point p in the parameter space is the normalized sum of evaluations of the original energy at projections of p to two-dimensional subspaces. We demonstrate the quality of the proposed approximation by computing the correlation with the original energy, and show that registration can be performed by discrete optimization of the approximated energy in an iteration loop. A search space refinement strategy is employed over iterations to achieve sub-pixel accuracy, while keeping the number of labels small for efficiency. The proposed framework can encode any similarity measure, is robust to the settings of the internal parameters, and allows an intuitive control of the parameter ranges. We demonstrate the applicability of the framework by intensity-based registration, and 2D-3D registration of medical images. The evaluation is performed by random studies and real registration tasks. The tests indicate increased robustness and precision compared to corresponding standard optimization of the original energy, and demonstrate robustness to noise. Finally, the proposed framework allows the transfer of advances in MRF optimization to linear registration problems.

Key words: Linear Image Registration, Intensity-based Registration, Markov Random Fields, Discrete Optimization

1. Introduction

In the last years, the solution of computer vision problems by Markov Random Fields (MRFs) [1, 2] and discrete optimization has become increasingly popular. Different image analysis problems have been solved by this approach, such as segmentation [3] or non-linear registration [4], just to name some. Recent advances [5, 6, 7, 8] in discrete optimization methods make this approach very attractive.

However, the MRF-based optimization has not yet been applied to the classical problem of intensity-based registration of images by linear transformations [9, 10, 11]. When analyzed closely, this problem turns out to be rather different in nature from many other problems which have so far been successfully addressed by MRF-based techniques. The differences are twofold.

For many MRF-based methods, the parameters can be assigned to certain points in the image domain at which the parameters are located, and the parameters have a limited local region of influence. In most cases, the parameters are directly identified with pixels and they only influence a direct neighborhood. That is, the parameters are *localized* in space and have

local effects. This is in strong contrast to the parameters of a linear transformation which act globally and cannot be assigned to a certain point in the image domain. Also, most MRF-based methods share a similar modeling approach. They employ a second-order MRF model which is composed of two terms: A first-order term which encodes a certain cost function (often referred to as *data term*), and a second-order term responsible for the regularization (*prior term*). Due to the mentioned fundamental difference in the nature of the parameters, the estimation of linear transformation parameters requires a different modeling, as will be discussed in Section 2.

The second difference originates from the fact that the solution for the linear registration problem cannot be determined by optimizing the single parameters independently of each other. In consequence, the original problem of estimating the linear transformation with n degrees of freedom (DOF) requires an nth-order MRF model. Compared to the second-order models, the current higher-order optimization algorithms are often complex to use or pose additional constraints on the class of energies they can be applied to. Also, the optimization of higher-order terms was until now mostly only demonstrated for regularization terms - a setting that does not translate to the linear registration problem. And finally, as of now, the efficiency of

---

*Corresponding authors and joint first authors.

```
Algorithm: Basic MRF-based Linear Registration

Input: images I,J
Output: transformation parameters p
0: set initial parameters p';
1: set initial discretization L;
2: set initialize parameter range refinement factor α;
3: until convergence do repeat
4:     Δp := optimize_E_approx(I,J,L,p');
5:     if (E(p'+Δp)<E(p')) //test for decrease of original energy
6:         p' := p' + Δp ;
7:     end
8:     L = refine_label_ranges(L,α);
9: end
```

Figure 1: Outline of the basic algorithm for linear intensity-based registration by optimization of a discrete MRF energy.

recent methods for second-order models cannot be matched by methods for higher-order models. This is a major point, since the single similarity evaluations for the linear registration problem are computationally demanding.

In order to be able to use the current efficient MRF optimization techniques for the intensity-based estimation of linear transformations, we present an approximation of the original energy by a second-order MRF model. This approximated MRF energy term is optimized by a suitable discrete optimization method inside an iteration loop. In every iteration, the range of the discrete search space for the single parameters is refined, eventually resulting in sub-pixel precision. In order to guarantee convergence, the descent of the original energy is ensured. The resulting algorithm is outlined in Figure 1.

The remainder of this paper is organized in the following way. In Section 2, the actual method is presented. After formulating the general problem of intensity-based estimation of linear transformations, we introduce the basic terms of discrete MRF models in Section 2.1, and argue that the estimation of the transformation parameters of the original energy requires a higher-order MRF model. Following this, in Section 2.2 we introduce the approximation to the original energy by a tractable second-order model. In order to complete the description of the actual resulting registration method, the Sections 2.3 - 2.6 provide details on parameterization, discretization of the search space, the employed discrete optimization method, and implementation. Section 2.7 summarizes and discusses the properties of the proposed method. In Section 3, we evaluate the quality of the proposed energy approximation, as well as the dependency of the method on the settings of the internal parameters. We demonstrate and evaluate an exemplary application to registration of 3D images in Section 4, and an application to 2D-3D registration of medical images in Section 5. Finally, Section 6 concludes the paper.

This paper extends and unifies our initial work from [12, 13]. We provide a revised presentation, including a new interpretation of the approximation, an evaluation of the approximation quality, as well as the influence of the internal parameters, extended experiments and a comparison to other optimization methods.

2. Intensity-based Estimation of Linear Transformations by Markov Random Fields

We consider the task to estimate a linear transformation T which aligns the source image I to the target image J, such that this alignment optimizes a suitable similarity measure ξ[1]. T_p denotes that the transformation T is parameterized by parameters p. Thus, the estimation can be written as

$$\widehat{p} = \arg \min_p \xi \left(I \circ T_p, J \right) \quad, \tag{1}$$

where $\widehat{p}$ are the optimal transformation parameters. Furthermore, we express the parameter p in terms of the initial estimate p' and an update Δp as $p = p' + \Delta p$, so that the optimal parameters are represented by the optimal update as $\widehat{p} = p' + \widehat{\Delta p}$.

For future reference we explicitly define the original energy E associated with the problem as

$$E(p) \equiv \xi \left(I \circ T_p, J \right) \quad. \tag{2}$$

2.1. Discrete Markov Random Fields

Let $\mathcal{V} = \{p_1, \ldots, p_n\}$ be a set of n parameter variables p_i, which are also represented by the parameter vector p. A *clique* is a subset of the parameters from $\mathcal{V}$, and is denoted either by explicitly stating the set, or by the usage of the multi-index $c \subseteq \{1, \ldots, n\}$, which contains the indices of the involved variables, such that $p_c \subseteq \mathcal{V}$. The set of all defined cliques is denoted as C. Then a Markov Random Field (MRF) can be defined as a hypergraph $\mathcal{G}$ in which the nodes correspond to the parameters, and the cliques constitute the hyperedges, that is $\mathcal{G} = (\mathcal{V}, C)$. In the case of discrete MRFs, each parameter variable p_i can take a discrete value (also referred to as *label*) from a discrete label set $\mathcal{L}_i$ with $\mathcal{L}_i = \{ x \mid x \in \mathbb{R} \}$. The number of labels per parameter is denoted by $|\mathcal{L}_i| = N$. In the context of MRFs, the term *labeling* describes a set of labels assigned to the parameters. A *potential* is a scalar-valued function ψ_c, which assigns a certain energy to a labeling of a clique p_c as $\psi_c : p_c \mapsto \psi_c(p_c) \in \mathbb{R}_0^+$. The general discrete MRF energy is a sum of the defined potentials

$$E_{\text{MRF}}(p) = \sum_{p_c \in C} \psi_c(p_c) \quad. \tag{3}$$

The *order* of the MRF model is the maximum size of the involved cliques.

For the following, the model containing exclusively all possible second-order terms will play an important role, thus we define it explicitly as

$$E_{\text{MRF-2}}(p) = \sum_{\{p_i, p_j\} \in C_2} \psi_{ij}(p_i, p_j) \quad. \tag{4}$$

Here, C_2 denotes the set of all possible second-order cliques, and correspondingly, cliques containing all possible k^{th}-order cliques will be denoted as C_k.

---

[1]We use the term similarity measure as a generic term for both similarity and difference measures.
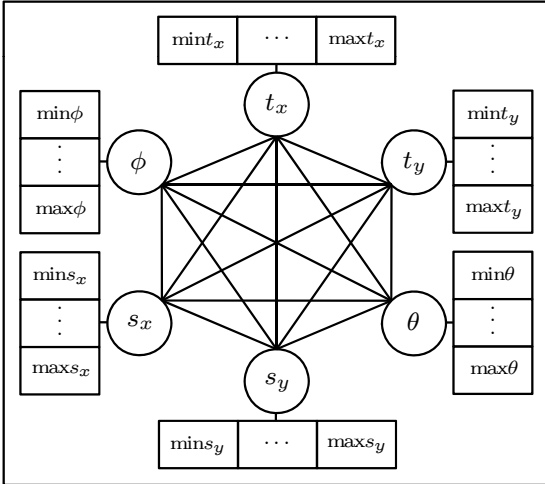
2

Figure 2: Topology of the proposed MRF model for an affine 2D transformation. The transformation parameters are represented by the nodes. The edges encode the image similarity costs associated with the variation of the labeling of the adjacent nodes.The MRF determines the parameters by optimizing the sum of the edge costs.

2.1.1. Second- and Higher-order MRF Models

The arguably most common MRF model used in computer vision tasks is the second-order (pairwise) model containing at most cliques of size two (used e.g. for segmentation, image denoising, restoration, deformable registration). Often, second-order MRFs are used synonymously with MRF models, and the modeling consists of assigning the first-order potentials to the data term, and the second-order term to the regularization. Presumably, the major reason for this is that for the second-order MRFs, many efficient and general optimization methods have been proposed over the last years [5, 6, 7, 8].

Only recently, optimization methods for higher-order models have been proposed. Compared to the second-order approaches, these methods are still not as general and efficient, and not as easy to use. Also, these methods have in common that the higher-order models are applied for implementing regularization terms. While this is applicable to many computer vision problems, it is not obvious how this modeling and corresponding optimization can be transfered to the linear registration problem. Here, we give a brief overview of recent higher-order optimization methods and discuss their applicability to the problem at hand. Kohli *et al.* [14, 15] consider an efficient optimization for generalized higher-order Pott's model [16]. The Pott's model is an energy function applicable for regularization by enforcing label consistency in neighborhoods. However it is not applicable to modeling the linear registration problem with arbitrary similarity measures. The same issue regarding the regularization term holds also for the applications considered in [17, 18, 19]. Lan *et al.* [17] as well as Potetz and Lee [18] consider efficient versions of Belief Propagation (BP) algorithm for encoding label consistency in 2×2 neighborhoods for image denoising. Ishikawa [19] considers the same application with a different optimization strategy, by combining the reduction of higher-order terms with the fusion-move [8] and

quadratic pseudo-boolean optimization (QPBO) [20]. Rother *et al.* [21] address the optimization of sparse higher-order energies, using soft pattern-based representation of the energy functions, and by transforming the problem into an equivalent quadratic function minimization problem. The linear registration however - in contrast to many others in image analysis - is not a sparse problem. Komodakis and Paragios [22] employ a general framework for solving the higher-order model based on a master-slave decomposition - with many application examples for which the method can be applied. The proposed decomposition profits from a high degree of independence of the single slave problems. However, this independence is not given for the linear registration problem, and the efficient optimization of higher-order slave potentials is not straight-forward in the case of linear registration.

2.1.2. MRF Order for Linear Registration

The order of an MRF model expresses the degree of conditional dependence of parameters on each other [23]. Take for example a simple first-order model: in this case labeling one parameter with a certain value does not influence the choice for the labeling of the other parameters inside the MRF energy term. Thus, the labeling of the single parameters can be performed independently. Going one step further and looking at a second-order model, we see that the choice of the label for one parameter p_i will influence the choice for the labeling of all the parameters p_j, which form cliques together with p_i, that is $\{p_i, p_j\} \in C$. Finally, we observe that for an MRF model with n parameters, the most complex model is an nth-order model, in which all the nodes are directly dependent on all the other nodes. Due to the conditional dependence of the parameters for the linear registration problem in (2), it follows that in general the estimation of a linear transformation with n DOFs requires an nth-order clique model.

Since direct optimization of this high-order energy is difficult, in the next section we show how to circumvent this by introducing an approximation to the original energy E which involves only tractable second-order terms.

2.2. Energy Approximation for Estimation of Linear Transformations by Second-Order MRFs

The key idea of the proposed approximation is to use the second-order cliques to encode a cost of a simultaneous variation of two parameters, while the other parameters are fixed to their current values. This encodes the dependency of the similarity measure on the two respective parameters, while the dependency on all the other parameters is ignored. However, by simultaneously taking into account all possible combinations of parameter pairs, we construct the overall energy term such that the selection of one parameter value depends on all the others.

Formalized, this results in an MRF model represented by a fully-connected graph $\mathcal{G}^* = (\mathcal{V}, C_2)$, where the nodes $\mathcal{V}$ represent the transformation parameters, and the edges C_2 are the set of all possible parameter pairs (all possible second-order cliques). The potential $\widetilde{\psi}_{ij}$ determines the cost of a simultaneous variation of the transformation parameters p_i and p_j while

Type	2D			3D		
	Rig.	Sim.	Aff.	Rig.	Sim.	Aff.
#params (n)	3	5	6	6	9	12
#cliques ($|C_2|$)	3	10	15	15	36	66

Table 1: The number of cliques (or potentials/edges) in the approximated MRF model in Eq. (7), for rigid, similarity with anisotropic scaling, and affine transformations.

the other parameters are held fixed at their initial values (the so created parameter set is denoted as p^{ij}) and can be written as

$$\widetilde{\psi}_{ij}(p_i, p_j) = \xi\left(I \circ T_{p^{ij}}, J\right) \tag{5}$$

Here, the k-th entry of p^{ij} is explicitly defined as

$$p_k^{ij} = \begin{cases} p_i & , k = i \\ p_j & , k = j \\ p'_k & , k \neq i, j \end{cases}, \tag{6}$$

which simply means that the parameter vector p^{ij} takes the value of p_i as the i-th entry, and the value of p_j as the j-th entry, and has the corresponding values of the initial parameters p' at all other entries. Substituting (5) in (4), and normalizing by the number of cliques $|C_2|$, we define the approximative second-order MRF energy for estimation of linear transformations as

$$\widetilde{E}(p) = \frac{1}{|C_2|} \sum_{\{p_i, p_j\} \in C_2} \widetilde{\psi}_{ij}(p_i, p_j) \ . \tag{7}$$

In general, the number of cliques is $|C_k| = \binom{n}{k}$, where k is the order of the used terms so that for the second-order model we have $k = 2$. We give $|C_2|$ for relevant cases in Table 1.

In summary, the presented approximation allows us to model the conditional dependence of the parameters while efficient optimization methods for second-order models can be used. A simple example of the resulting MRF topology is illustrated in Figure 2. In Section 3.1, we evaluate the quality of the proposed approximation. By sampling of the original and the approximated energy, we demonstrate that they strongly correlate for reasonable initializations.

2.2.1. Interpretation of the Approximation

An interpretation of the approximative energy term (7) provides further insight into the proposed method. In Equation (7), the single potentials $\widetilde{\psi}_{ij}$ as defined in (5) are carving out two-dimensional subspaces in the original n-dimensional parameter space. These subspaces are parallel to the respective ij-planes of the parameter space and are all passing through the point p' representing the initial parameters. By construction, the subspaces are orthogonal to each other.

The approximated energy can now be compared to the original energy in the following way. The original energy $E(p)$ is simply the evaluation of the chosen similarity term at the point p. The approximated energy $\widetilde{E}(p)$ is computed by projecting the point p to the single subspaces and averaging the energy
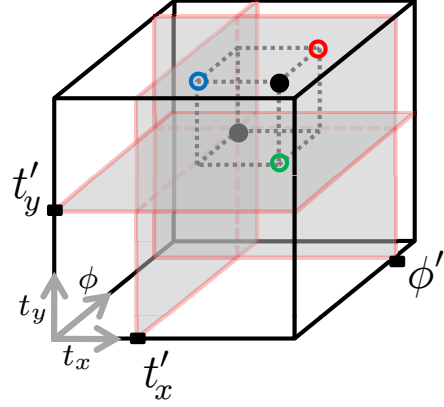

Figure 3: Visualization of the energy approximation by a second-order MRF model for rigid 2D registration. Here, t_x, t_y, ϕ denote the translation and rotation parameters, with initial values t'_x, t'_y, ϕ'. The evaluation of the original energy at the parameter point p (black) is approximated by the normalized sum of the energy evaluations at the projections of p to the 2D subspaces (red, green, blue). The subspaces are orthogonal and all pass through the initial point p' (gray).

evaluations at the projected points. If we denote the projections of the point p to the ij-subspace by $\Pi_{ij}(p)$, we can write the approximated energy from (7) also as

$$\widetilde{E}(p) = \frac{1}{|C_2|} \sum_{i,j} E(\Pi_{ij}(p)) \ . \tag{8}$$

This interpretation is illustrated for the simple case of a rigid 2D transformation in Figure 3.

It is important to notice that the quality of the approximation depends on the choice of the initialization p'. Obviously, if we evaluate the approximated energy at the initial point, that is $p = p'$, the approximated energy equals the original energy, since all the projections of p equal to the initial point p'. With increasing distance of p from p', the error made by the projection increases and deteriorates the approximation. This is demonstrated in Section 3.1 where the quality of the approximation is evaluated.

The stated interpretation can be used to reason about extending the proposed approximation, by utilizing cliques of higher orders. The approximation would then read

$$\widetilde{E}_k(p) = \frac{1}{|C_k|} \sum_{\{p_i, \dots, p_k\} \in C_k} \widetilde{\psi}_{i \dots k}(p_i, \dots, p_k) \tag{9}$$

$$= \frac{1}{|C_k|} \sum_{i, \dots, k} E(\Pi_{i \dots k}(p)) \ , \tag{10}$$

and the order k would equal the dimension of the subspaces. For example, using first-order terms would lead to projecting to lines parallel to the space axes and passing through the initial point, and third-order terms would imply projecting to three-dimensional volumes. This generalization might become interesting since using higher-order building blocks can be expected to increase the quality of the approximation. Also, efficient optimization for single third-order terms used as building blocks

N	3	5	7	9	11	13	15	17
FastPD	105	190	300	390	500	600	700	800
exhaust.	135	375	735	1215	1815	2535	3375	4335

Table 2: Average number of evaluations of similarity measure per iteration for FastPD, in comparison to exhaustive search. We give an example for the 6 DOF 3D rigid case with $k = 2$.

can be expected to arise earlier than for example for the 12$^\text{th}$-order model required for the full 3D affine transformation.

2.3. Parameterization

In this section we present the parameterization of linear transformations used in this work. Linear transformations in homogeneous coordinates can be written in the form

$$A = \begin{bmatrix} \widehat{A} \\ v^{\top} \quad 1 \end{bmatrix} \; , \qquad (11)$$

where $\widehat{A} \in \mathbb{R}^{d \times (d+1)}$ and $v \in \mathbb{R}^d$ for d-dimensional problems. Since our work is motivated by medical applications, in the following we do not consider projective transformations and we restrict ourselves to affine transformations [10] by assuming that $v = 0$. This results in 6 degrees of freedom in 2D and 12 DOF in 3D. We employ a parameterization in which the affine transformation is decomposed as

$$A = M_t \, R_\phi \, R_\theta^{-1} D_s R_\theta \; . \qquad (12)$$

Here, M_t represents a translation, R_ϕ a rotation, and $R_\theta^{-1} D_s R_\theta$ represents the shearing transformation. For the shearing, R_θ is a rotation and D_s is a diagonal matrix, representing anisotropic scaling. We parameterize the single matrices of Equation (12) by respective parameters, compare also [24]. The 3D rotation matrices are parameterized by Euler angles. The resulting parameter vectors for the 2D and 3D case are

$$p = \left[t_x, t_y, \phi, s_x, s_y, \theta \right] \qquad (13)$$

$$p = \left[t_x, t_y, t_z, \phi_x, \phi_y, \phi_z, s_x, s_y, s_z, \theta_x, \theta_y, \theta_z \right] \; . \qquad (14)$$

Please note that we always consider the parameter p in terms of an initial guess p' and an update Δp, such that $p = p' + \Delta p$.

With this representation, restricting the general affine model from (12) to a more constrained one is simply performed removing the corresponding nodes and the adjacent edges from the MRF model. Since our implementation performs these modifications automatically if the search range for one parameter is set to zero, there is no need for explicit implementation of the registration for the different transformation types.

2.4. Discretization of the Parameter Space

An important part of the proposed method is the discretization of the parameter search space, that is, the definition of the label space $\mathcal{L}$. On one hand, one would like to keep the number of labels small for efficient optimization of Equation (7). On the other hand, we want to avoid a too coarse sampling of
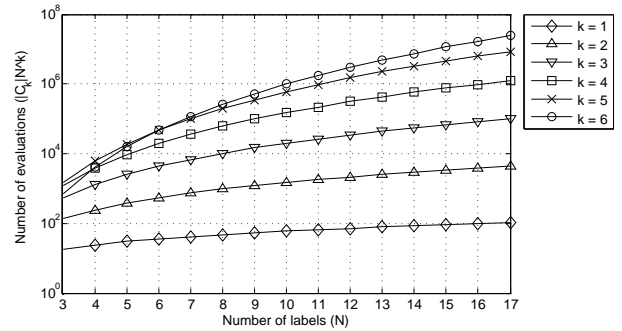

Figure 4: Number of evaluations of the similarity measure needed to compute the approximated energy at all points in the parameter space. We give an example for a transformation model with 6 parameters, depending on the order $k \in \{1, \ldots, 6\}$ of the potentials used for approximation, and with $N = \{3, \ldots, 17\}$ samples per dimension. The method proposed in this paper has $k = 2$, while the original energy is represented by $k = 6$. Please note that this is only a theoretical upper bound, and that the actual optimization methods require less computations.

the parameter search space since it might result in inaccurate registration results. To achieve a trade-off between these two objectives, we employ a successive strategy for the refinement of the search space. This results in refined parameter updates in every iteration.

We discretize $\mathcal{L}$ in the following manner. For each parameter p_i we define a search range which is discretized by uniformly sampling the range between preset values $\min p_i = p'_i + \min \Delta p_i$ and $\max p_i = p'_i + \max \Delta p_i$ about the initial value p'_i.

Since the algorithm is guaranteed to reduce the energy in every step (by explicitly assuring that the new solution decreases the original energy, cf. Figure 1), we can assume that the new estimate is closer to the locally optimal solution, and reduce the search ranges for the parameters. In iteration k, we rescale the range delimiters $\min \Delta p$ and $\max \Delta p$ by a parameter range refinement factor α^k ($\alpha < 1$) which defines the new possible values for the parameter updates. The iterative label space refinement allows us to keep the number of labels small and we can start with a large parameter range, while being able to achieve sub-pixel registration accuracy. In practice, the number of labels N ranges between 5 and 11 and the maximal possible values for the parameters are used to limit the parameter ranges (e.g. $\pm 180°$ for rotation), which allows for the fixed setting of the ranges. We evaluate the performance of the method for different choices of α as well as the number of labels N in Section 3.2.

An interesting observation is that the proposed approximation also strongly reduces the number of similarity evaluations which would be necessary to evaluate the energy at all points in the complete parameter space. The discretized n-dimensional parameter space with N sampling points per dimension contains altogether N^n points. Thus, the computation of the original energy at all points in the parameter space requires N^n similarity evaluations, since it is simply evaluated once at every point in the parameter space. On the other hand, the approximated energy at one point is computed as a sum of $|C_k|$ similarity evaluations at points in the two-dimensional subspaces,

5

while the subspaces require N^k evaluations each. This means that only $|C_k|N^k$ evaluations have to be performed (with $k = 2$ for our second-order terms).[2] From these evaluations, the approximated energy can be computed for every point by simply averaging the $|C_k|$ scalar values. The number of needed evaluations, depending on the order k of the potentials used as building blocks in the approximation, and the number of labels N, is illustrated in Figure 4. This implies that a brute force exhaustive search would be much more efficient for the approximated energy. For small label sets ($N \approx 5$) and low-order ($k = 2, 3$) this might even be a viable option. Please note that $|C_k|N^k$ is only a theoretical upper bound, and that the actual optimization methods (e.g. FastPD as used in our work) require less computations. For an example, please refer to Table 2. Advances in discrete optimization, leading to less evaluations, would have a direct impact on the proposed method.

2.5. Optimization

For the optimization of our MRF energy defined in Equation (7), we use the recently proposed FastPD algorithm[3] [7]. FastPD is able to handle non-submodular functions [25], such as general intensity-based similarity measures. Furthermore, FastPD performs well both in terms of computational efficiency and optimization quality [26]. Other optimization methods such as move algorithms based either on QPBO (Fusion Move [8]) or iterative graph-cuts (e.g Expansion Move, Swap Move [5]), or message-passing strategies [27, 28] (e.g. Belief Propagation, TRW-S [6]) might be considered as well. However, we should note that due to the non-submodularity of the pairwise energy terms, classical graph-cut based methods [5] have to be modified to be able to handle such energies (cf. [25]). On the other hand, message-passing algorithms can handle general energies, however with the cost of lower computational efficiency [29].

2.6. Implementation Details

The described registration method is implemented with a standard multi-resolution strategy realized by Gaussian image pyramids with a downsampling factor of two between levels. No downsampling is performed below 32 voxels per dimension. The method also supports random subsampling strategies, which are often used to improve the runtime, cf. e.g. [30] and references therein. In terms of performance, the evaluation of the given similarity measure presents the most computationally demanding part of the method. This is especially the case for large 3D data sets. To this end, we also implemented the evaluation of the 3D similarity measures on the GPU which results in a significant speedup of the method.

2.7. Method Discussion

Formulating the linear registration as an MRF problem shares the advantages of some standard optimization methods such as Nelder-Mead Simplex or Powell's method [31], that the similarity measure is easily interchangeable, since no derivatives of the measure are required. The integration of novel and more complex measures – where the derivatives might be complicated to compute – is simple to achieve in our framework. With freely available discrete optimization libraries, the implementation of the proposed approach is straight forward.

A possible alternative for decreasing the order of the MRF model is the usage of high-dimensional labels, where one label encodes the values for multiple parameters. This is done for example in [4] for encoding the local translations for deformable registration. For rigid registration for example, one could encode the translations in x/y/z direction as one parameter with a high-dimensional label consisting of the combinations of values for t_x, t_y, and t_z, and rotation as a second one for ϕ_x, ϕ_y, and ϕ_z. Such an MRF model would be of second-order and have only one clique, which would consist of the two parameters. The problem with this approach is that the size of the label spaces increases too much, such that no efficient solution is possible. For the above example, in which a high-dimensional label represents 3 ordinary labels, the size of the high-dimensional label space would be N^3. We performed several tests following this approach, however, the computation times were prohibitive.

A different possible future line of work would consist of testing different parameterizations. For example, the label values could be sampled from the ranges in a non-uniform manner, or different parameterizations of the rotation angles could be used.

In this work, we only used second-order potentials as building blocks for the energy approximation. One interesting option would be to also employ first-order terms, for example to encode the prior probability of the single parameters. Please note that this way, for the linear registration problem, the standard MRF philosophy in which the first-order terms encode the data term, and the second-order terms encode the prior knowledge is reversed. Also, a possible future modification as efficient, general, and easy to use optimization methods for third-order MRF models become available would be to build the proposed approximation on third-order instead of second-order potentials. This can be expected to improve the approximation quality. The generalization using k^{th}-order potentials for the approximation is discussed in Section 2.2.1.

3. Method Evaluation

In this section, we evaluate two important aspects of the proposed framework. In Section 3.1, we evaluate the quality of the proposed energy approximation from Equation (7). In Section 3.2, we test the influence of the internal parameters of the proposed method on the registration results.

All experiments are performed on a standard desktop system with an Intel® Core™ 2 Quad 2.83GHz, 3GB of RAM, and an NVIDIA® GeForce® GTX 285 GPU with 1GB RAM.

The following stopping criterion is used in all experiments, and for all levels of the image pyramid. The iterations are stopped when the relative decrease of the energy (compared to the energy at the initial state) falls below a threshold. We use a conservative threshold of 0.0001.

---

[2]Actually, slightly less evaluations than $|C_k|N^k$ are needed, since the single subspaces share points where they intersect.

[3]Available at: http://www.csd.uoc.gr/~komod/FastPD/.

Test Type	n	N	#eval=N^n	CC Test 1	CC Test 2
2D rigid	3	80	$\sim 0.5 \times 10^6$	0.9590	0.9687
2D similarity	5	14	$\sim 0.5 \times 10^6$	0.8925	0.9169
2D affine	6	9	$\sim 0.5 \times 10^6$	0.8595	0.8976
3D rigid	6	9	$\sim 0.5 \times 10^6$	0.8446	0.8510
3D similarity	9	4	$\sim 2.0 \times 10^6$	0.7864	0.8169
3D affine	12	4	$\sim 16.8 \times 10^6$	0.7961	0.8120

Table 3: Evaluation of the quality of the energy approximation $\widetilde{E}$. We compute the correlation coefficient (CC) between sampled versions of the original and the approximated energy. This is done for different transformations with n parameters, and with N sampling points per dimension, resulting in N^n energy evaluations per test. Transformation parameters increase from Test 1 to Test 2. We observe increased approximation quality for simpler models and "smaller" transformations.

3.1. Quality of Approximation

In this section we evaluate the quality of the energy approximation, that is how similar the approximated energy $\widetilde{E}$ from Eq. (7) is to the original energy E in Eq. (2). The idea is to compare the original and the approximated energy by sampling these functions and computing the correlation coefficient (CC) between them. This test assesses the error introduced by the approximation, which is independent of the particular method chosen for the optimization.

We perform this test for 2D and 3D images, and 3 types of transformations: rigid, similarity with anisotropic scaling, and general affine transformations.

In 3D, the tests are performed on an MR-T1 volume of the brain from The Retrospective Image Registration Evaluation Project (RIRE)[4] database [32]. The 2D tests are conducted on a slice from the above 3D volume. Since the random tests are mono-modal, we employ the sum of squared differences (SSD) as similarity measure. We sample the parameter space equidistantly about the initial parameters p' representing the identity transformation. Two different ranges are tested, **Test 1** representing the larger parameter ranges at the beginning of the registration process, while **Test 2** covers smaller parameter ranges and represents the setting towards the end of the registration. The ranges for the sampling of the transformation parameters for Test 1 are as follows: Translations result from ± 10mm, rotations lie between $\pm 45°$, scalings range between 1 ± 0.2, and the shearing angles θ are drawn from $\pm 20°$. For Test 2, the ranges are: ± 2mm for translations, $\pm 4°$ for rotations, 1 ± 0.04 for scaling, $\pm 4°$ for shearing angles. The number of sampling points per dimension N is set such that per test at least 500,000 energy evaluations are performed. In order to keep the computation time reasonable, the tests were performed on the third finest levels of the image pyramid, that is on $64 \times 64 \times 26$ images in 3D and 64×64 images in 2D. Altogether, approximately 21 million energy evaluations were performed. Details and results of the tests are summarized in Table 3.

The tests show that the proposed approximation correlates well for the tested settings. The average CC over all tests is

0.86, where CC=0 indicates that two signals are linearly unrelated and CC=1 indicate perfect linear correlation. As expected, the correlation is higher for simpler models, that is, increases from affine over similarity to rigid, and it is higher for the 2D than for 3D transformations. Also, the correlation is higher for the smaller parameter ranges.

These findings are supported by inspecting the energy logs of the original and approximated energy, acquired during the random registration experiments performed in Section 4.1. This results in sampling of the energies along the paths in the parameter space, which are being traced out by the single registrations. We observe higher correlation coefficient in the logs (CC > 0.98), compared to the coefficients from the 3D evaluation in Table 3 (CC < 0.85), caused by a large number of evaluations close to the optimum, which are performed towards the end of the registration. This shows that the proposed approximation improves during the registration process. A sample plot of the logs and the quantitative evaluation is given in Figure 7.

3.2. Dependency on Parameters

In this section we perform an evaluation on the sensitivity of the proposed method on the internal parameters: N, the number of labels for the discretization of the parameter space, and α the parameter range refinement factor.

We perform the evaluation by varying the setting of the parameters for 1000 random registration tests, and assessing the resulting performance. The random tests consist in registering a 2D MR-T1 source image to a randomly displaced target image. The target image is created by applying a random affine transformation to the source image. The random transformation is generated by uniformly drawing the transformation parameters from the following ranges: translations result from ± 10mm, rotations lie between $\pm 60°$, scalings range from 1 ± 0.2, and the shearing angles θ are drawn from $\pm 20°$. We test all possible combinations for N between 3 and 17, and α between 0.4 and 0.8, resulting in 40,000 registration runs. For faster computation, the registration is performed on a downsampled image.

The registration error is computed as the average distance (AD) of the corner points of a centered 200mm rectangle, which are warped once by the ground truth transformation and once by the estimated transformation. Please note that since the AD is measured at the corners, it will in general measure larger errors, than for the region of interest in the center of the volume.

Figure 5 summarizes the results of the experiment. Except for $N = 3$, we observe no large difference between the different parameter settings. For $N > 3$, the number of failed runs is similar very low (below 0.3%). For finer discretization of the search space (higher N) the precision is slightly improved, however, this comes at a higher computational cost. With respect to the resolution, the majority of results exhibit a sub-pixel precision. Based on the results of this evaluation, we set the values to $\alpha = 0.6$ and $N = 5$ keep them fixed throughout the registration experiments in Section 4.

On a side note, we also performed the same 1000 trials by registration with optimization of the original energy with the Nelder-Mead Simplex method [31] and otherwise same settings. This test resulted with a similar error distribution as the
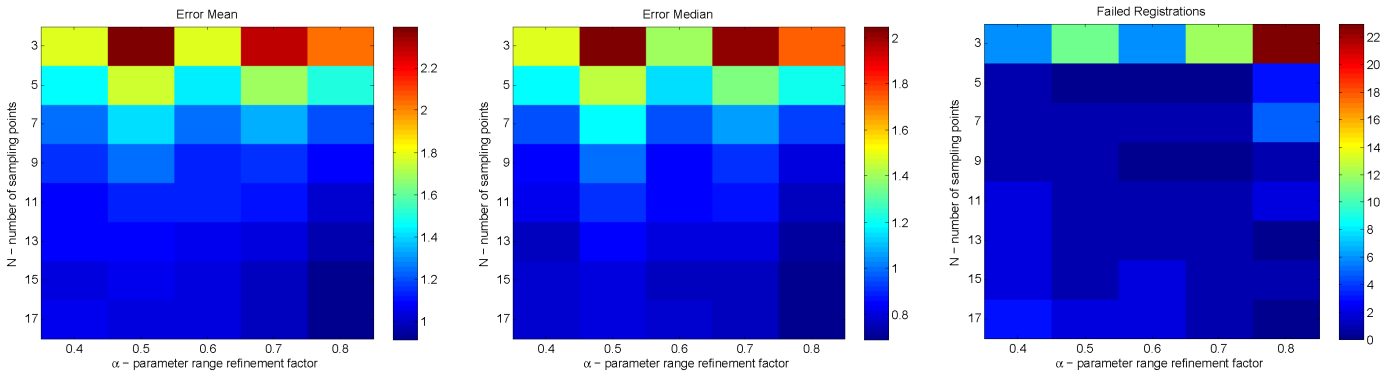
Figure 5: Evaluation of the performance of the proposed framework with respect to the internal parameters: the number of sampling points per dimension *N*, and the parameter range refinement factor α. For each different setting of *N* and α, 1000 random tests are performed. Except for *N* = 3, we observe no large difference between different parameter settings. The precision is slightly improved for finer discretization of the search space (higher *N*). This however comes at a higher computational cost.

proposed method, however it had 7 failed trials, which is inferior to any of the different settings of the MRF-based method for *N*>3.

4. Application: Linear Intensity-based Image Registration

In this section, we present results of the direct application of the proposed framework to the problem of linear intensity-based registration between 3D images. We test the proposed method by a random study on 3D brain data in Section 4.1, and by a series of 3D multi-modal registrations in Section 4.2. In this section we use 3D brain images (CT, MR-PD, MR-T1, MR-T2) provided by the RIRE data base. The CT image has a resolution of 512×512×29 and a physical voxel size of 0.65×0.65×4mm and the MR images have a resolution of 256×256×26 and a voxel size of about 1.25×1.25×4mm.

4.1. 3D Random Study

The 3D random study is performed by registering a source image to a target image generated by a random transformation. The source image is an MR-T1 volume. The target image is created by applying a random transformation to the source image, with transformation parameters uniformly drawn from the following ranges: Translation $t \in [-30, 30]$mm, rotation $\phi \in [-45, 45]°$, scaling $s \in [0.8, 1.2]$, and shearing angles $\theta \in [-20, 20]$mm.

Due to the mono-modality of the test, we use the SSD as the similarity measure. The similarity term for this test is performed on the GPU and without a sub-sampling strategy. A multi-resolution approach with 3 levels is used, and the registration is initialized by aligning the intensity masses of the two volumes. To speed up the computation and allow for a large number of tests, the registrations are performed only on the coarsest level of the image pyramids where the number of voxels in the first two dimensions is reduced by a factor of four while the number of slices is preserved. Since the tests are performed at lower pyramid levels at which noise is eliminated, we restrain from adding noise to the original images in

Test Type		Mean	Median	Std. Dev	Failed
Rigid	MRF	0.561	0.422	0.401	0%
	Simplex	0.526	0.432	0.403	0%
Similarity	MRF	0.935	0.657	0.805	0.2%
	Simplex	1.222	0.931	1.209	1.2%
Affine	MRF	4.277	3.876	2.279	11.2%
	Simplex	4.594	4.093	2.395	19.6%

Table 4: Summary of the results of the 3D random study from Section 4.1. The proposed MRF-based method is compared to Simplex optimization of original energy. Per transformation type 1000 trials were performed. Results with an AD error (given in mm) of more than 10mm are discarded as failed, and excluded from the computation of the statistics. (see also Figure 6). For all runs, the proposed method results in fewer failed runs and exhibits slightly better accuracy.

the first place. The registration error is computed as the average distance (AD), which in 3D is defined on the corner points of a centered 200mm cube. We perform the study for 3 types of transformations: rigid, rigid with anisotropic scaling, and affine. For each type we perform 1000 trials.

We compare the results of the proposed approach to the results obtained by using the Nelder-Mead Simplex optimization method [31] on the original energy function corresponding to the chosen similarity measure. To this end, the same general settings as for our approach are used. In all our experiments, the internal parameters of the Simplex method (step sizes for the transformation parameters) are set to the same values as the corresponding search space ranges of the MRF method. Convergence criteria are thresholds on the difference of function values at subsequent estimates (1e−20), and changes in parameters (1e−15), as well as the maximum number of iterations (500). We have tested these settings for different registration problems, with consistently good results. The results are summarized in Table 4 and the statistics are represented by box plots[5] in Figure 6. Please note that due to the chosen resolu-

---

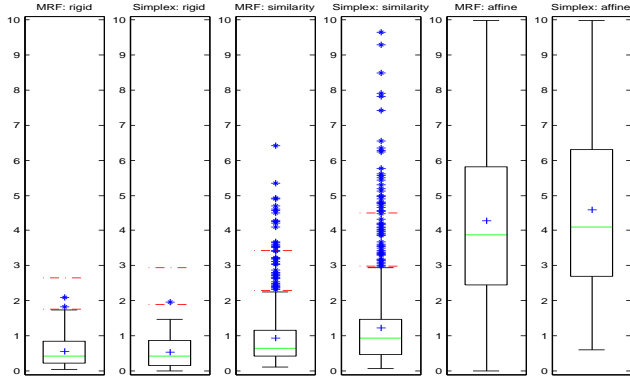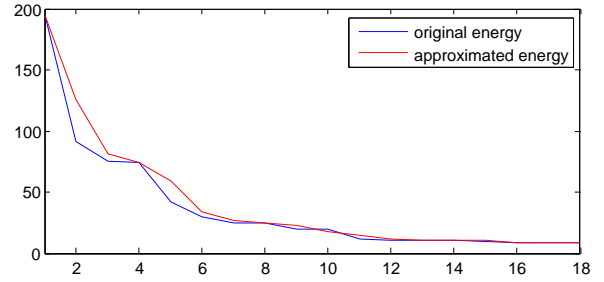[5]Box plot notation in the paper: The box denotes 50% of data between

Figure 6: Statistics of the AD error in mm for the 3D random study described in Section 4.1, represented by box plots (please see footnote 5 for description). Please refer to Table 4 for a brief description and the numerical representation of the results. With respect to resolution, all errors less than 4mm indicate sub-voxel accuracy.

tion, errors of less than 4mm are below the physical voxel size. In Figure 7, we give an exemplary energy log acquired over the iterations, in which we record the approximated and the original energy. Also, the correlation of the energy logs for all trials is analyzed.
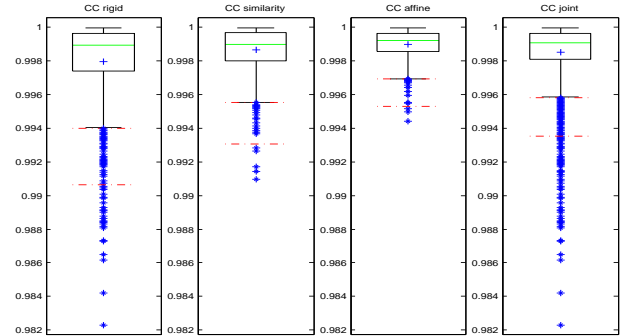
Two observations can be made from the results of the experiment. The first one is that the proposed MRF-based method seems to be more robust, such that more registrations are successful. We classify the registration as successful if the AD is below 10mm. With this threshold, all experiments for both, the Simplex and MRF-based optimization, for the rigid case are successful. However, for the case of similarity with anisotropic scaling, the Simplex method (1.2%) fails 6 times as often as the MRF-based method (0.2%). For the affine case, the Simplex method (19.6%) fails 1.75 times as often as MRF-based method (11.2%). Similar results are also obtained by using different thresholds. The second observation is that for the successful registrations, the MRF-based method yields higher precision. Generally, the MRF-based method exhibits lower mean and median errors, and the corresponding standard deviations are also lower. The only exception is the rigid test, in which the mean error for the Simplex method is lower. However, the more robust median error is lower for the MRF-based method also in this case.

With respect to the runtime, one must note that the MRF-based optimization of the approximated energy is more computationally intensive than the Simplex optimization of the original energy since it requires more similarity evaluations. Also, the difference is smaller for simpler models, since the number of evaluations needed for the evaluation of the approximated MRF energy is smaller. For the performed random tests, the average runtimes were 20.4 seconds for the rigid case, 45.6 sec-

(a) Energy logs over iterations.

(b) Distribution of the correlation coefficients.

Figure 7: Comparison between the original and approximated energy. **(a)** A typical energy log. **(b)** The CC statistics between the logs of all runs performed in Section 4.1 exhibit a high correlation (mean always > 0.98), and indicate that correlation (and thus the approximation quality) increases over iterations (cf. Section 3.1).

onds for similarity, and 85.6 seconds for affine transformations.

4.2. 3D Multi-Modal Rigid Registration

In this experiment we demonstrate the performance of the proposed approach for a real multi-modal rigid registration scenario, from the RIRE database. We perform the registration of CT images to MR-PD, MR-T1, and MR-T2 images. As the similarity measure, we use the entropy correlation coefficient (ECC) proposed in [33]. A multi-resolution approach with 4 pyramid levels is used. The transformations are initialized by aligning the intensity masses of the two volumes.

We perform two tests, the first one on data of one patient with available ground truth which can be used for training, and an extensive test on 10 patients, for which the evaluation is performed by the RIRE system. For the training data set, ground truth transformations between a CT volume and 6 MR images of one patient (original and rectified scans) are given. We compute the error by measuring the AD at the corner points of the CT volume, using the ground truth.

In order to assess the accuracy of the proposed method, we perform the same tests also by two other methods. The first one is the Simplex optimization on the original ECC energy term, as discussed in the previous sections. The second one is the module of rigid registration from the *Elastix* toolkit[6] [34]. With respect to Elastix parameters, we did our best to provide a setting

---

lower and upper quartiles Q1, and Q3, i.e. the medians of the upper and lower half of the sorted data; Median denoted by green line; Whiskers determined by extreme values within $Q1 - 1.5\,IRQ$ and $Q3 + 1.5\,IRQ$ (interquartile range $IRQ = Q3 - Q1$); The mean and outliers denoted by blue crosses; Red lines at $Q1 - 1.5\,IRQ$, $Q1 - 3\,IRQ$, $Q3 + 1.5\,IRQ$, $Q3 + 3\,IRQ$.

---

[6] Available at: http://elastix.isi.uu.nl/

Target	Initial	pre-al.	Simplex	Elastix	MRF
MR-PD	67.61	26.52	2.79	2.74	2.90
MR-PD rect.	49.64	22.69	4.10	2.85	3.43
MR-T1	46.18	22.49	3.41	3.02	1.82
MR-T1 rect.	65.00	15.63	2.32	1.93	2.66
MR-T2	56.13	22.00	4.08	2.52	3.60
MR-T2 rect.	62.94	22.87	4.43	4.08	4.28
mean	57.92	22.03	3.52	2.86	3.11
median	59.54	22.59	3.75	2.80	3.17

Table 5: Results of the 3D multi-modal CT-MR registrations on training data, described in Section 4.2. Given is the initial AD (in mm) before registration, after alignment of the intensity masses, and after registration using the Simplex, Elastix, and our MRF-based method.

as similar as possible to the other two methods, and followed the recommendations from the Elastix manual for other parameters. We use the same number of image pyramid levels, and employ the Normalized Mutual Information (NMI) [35] as the closest fit to the ECC. As the optimization method we selected the adaptive stochastic gradient descent [36], for which the internal parameters are estimated automatically. The maximum number of iterations was set to 250. We verified the validity of the settings on the training data. The results show a similar performance by the methods, with best average results obtained by Elastix, followed closely by the proposed methods, and then the Simplex based approach, please compare Table 5.

For obtaining comparable runtimes, the similarity evaluation is performed on the CPU in this experiment. To decrease the running time, we perform a subsampling strategy for all methods, in which 10% of voxels are uniformly drawn and used for similarity computation. Our method performed on average 27 iterations per registration (ca. 7 per level), resulting in average runtimes of 190 seconds. Elastix on average took 200 seconds, and the Simplex method converged after 30 seconds. Please note that due to the differences in the implementation of the different methods, this gives only a rough estimate of the order of magnitude for the runtimes. Furthermore, since the focus of this work is not on the speed of registration, different tuning options were not used. For example, while we employ 10% of all voxels in the sub-sampling, it was shown in [37] that already 2000-3000 samples can yield accurate results, corresponding to ca. 0.15% of all voxels on the finest level.

For the actual multi-modal test, we used all data sets from the RIRE data base, for which the MR-PD, MR-T1, and MR-T2 data sets are available, resulting in 10 patients with 3 registrations per patient. All registrations were performed with the same parameters as for the training data set. The evaluation is performed by the RIRE system. The tests show a consistent performance of the proposed method. Table 6 summarizes the results and shows that the proposed method consistently slightly outperforms the other tested methods in terms of accuracy. Furthermore, we performed the registrations with the MRF-based method also with different settings of the internal parameters in order to support the findings from Section 3.2. For $N = 3$ and $\alpha = 0.6$ the error statistics (in mm) are: mean=1.845, me-

(a) PD Initial   (b) PD MRF Result

(c) T1 Initial   (d) T1 MRF Result

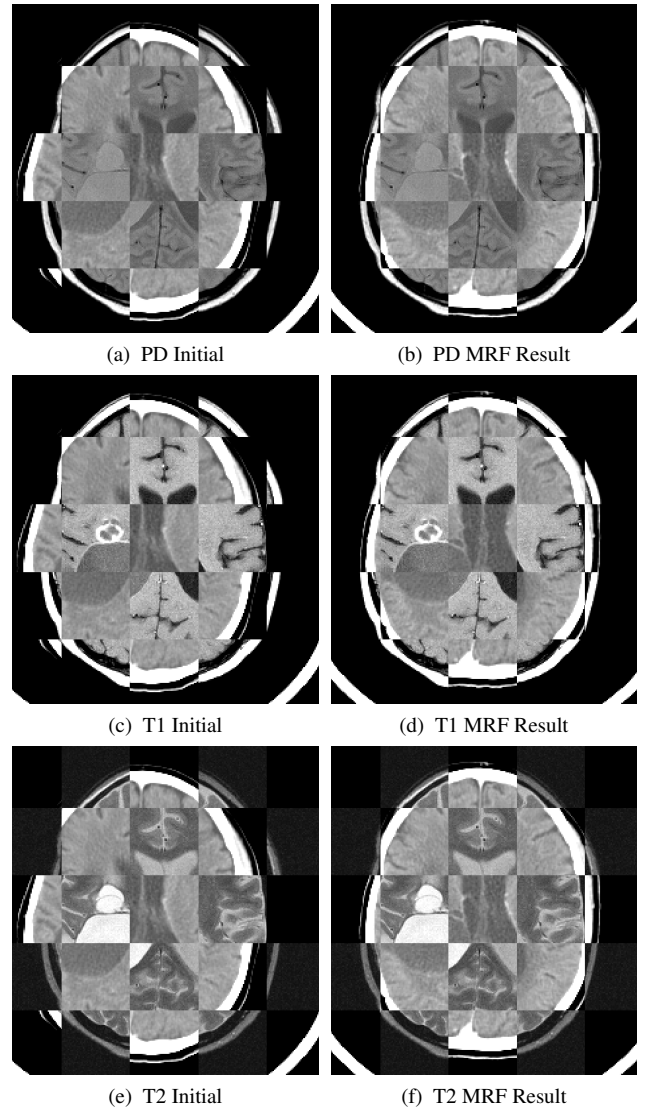(e) T2 Initial   (f) T2 MRF Result

Figure 8: Checkerboard visualization of the multi-modal CT-MR registration using MRFs on patient P04.

dian=1.708, max=4.409. On average, 22 iterations were performed per registration (ca. 5 per level), resulting in an average computation time of 86 seconds. For $N = 17$ with $\alpha = 0.8$, we have: mean=1.746, median=1.695, max=3.849. On average, 33 iterations were performed per registration (ca. 8 per level), resulting in an average computation time of 858 seconds.

In this scenario, the dependency on parameter settings is even less prominent than in Section 3.2 so that even the smallest possible label set (N=3) leads to accurate and efficient results.

5. Application: Intensity-based 2D-3D Registration

The second exemplary application is the 2D-3D registration of medical images, which shows the applicability of the proposed framework to variations of the standard registration problem. 2D-3D registration is a challenging task, since a transformation of 3D space is computed from 2D measurements, so

Target	Mean Error			Median Error			Maximal Error		
	Simplex	Elastix	MRF	Simplex	Elastix	MRF	Simplex	Elastix	MRF
PD - P01	1.831	2.200	1.870	1.743	2.042	1.781	2.387	2.807	2.501
PD - P02	0.972	1.237	1.001	0.952	1.226	0.996	1.460	1.826	1.479
PD - P03	16.438	16.784	16.407	16.569	16.550	16.472	17.427	18.690	17.495
PD - P04	3.359	4.749	3.226	3.260	4.663	3.133	4.052	6.310	3.884
PD - P05	1.706	1.493	1.746	1.724	1.434	1.733	2.047	1.783	2.029
PD - P06	1.534	1.769	1.587	1.506	1.589	1.525	2.819	3.079	2.902
PD - P07	1.880	2.026	1.891	1.924	2.036	1.938	2.047	2.880	2.026
PD - P08	3.069	3.300	3.063	2.959	3.255	2.962	3.711	4.287	3.716
PD - P09	2.591	2.400	2.574	2.398	2.297	2.385	3.668	3.915	3.634
PD - P10	2.180	1.999	2.190	2.262	2.061	2.263	2.416	2.362	2.416
PD overall	**2.067**	**2.226**	**2.078**	**2.005**	**2.018**	**1.986**	**4.052**	**6.310**	**3.884**
T1 - P01	1.218	1.520	1.215	1.313	1.487	1.090	1.680	2.444	1.688
T1 - P02	0.592	0.723	0.608	0.695	0.705	0.690	0.820	1.069	0.872
T1 - P03	1.306	1.205	1.423	1.424	1.286	1.484	1.880	1.937	2.006
T1 - P04	2.080	1.944	2.126	2.206	2.057	2.274	2.873	2.240	3.003
T1 - P05	0.933	0.917	0.956	0.929	0.961	0.953	1.064	1.215	1.084
T1 - P06	1.253	1.288	1.253	1.186	1.327	1.148	2.159	2.088	2.301
T1 - P07	0.908	1.012	0.907	0.874	1.058	0.863	1.054	1.655	1.136
T1 - P08	1.775	2.020	1.886	1.818	1.937	1.954	1.971	2.364	2.068
T1 - P09	1.569	1.737	1.521	1.405	1.553	1.465	2.495	3.172	2.589
T1 - P10	1.417	1.175	1.296	1.416	1.241	1.302	1.458	1.377	1.395
T1 overall	**1.275**	**1.334**	**1.286**	**1.259**	**1.230**	**1.154**	**2.873**	**3.172**	**3.003**
T2 - P01	2.909	2.099	2.247	2.826	1.957	2.268	4.271	2.707	2.835
T2 - P02	2.132	1.370	1.512	1.938	1.364	1.414	3.647	1.898	2.636
T2 - P03	1.224	2.018	1.160	1.236	1.938	1.113	1.536	3.060	1.518
T2 - P04	2.817	3.541	2.899	2.793	3.569	2.696	3.245	4.192	3.647
T2 - P05	1.614	1.828	1.652	1.743	1.879	1.799	2.376	2.678	2.299
T2 - P06	1.266	1.264	1.276	1.210	1.222	1.240	2.094	2.232	2.324
T2 - P07	1.952	1.986	1.842	1.945	1.875	1.840	2.168	2.750	1.877
T2 - P08	2.673	3.053	2.686	2.575	2.975	2.580	3.099	3.748	3.556
T2 - P09	2.218	2.220	1.899	2.104	2.072	1.735	3.331	3.598	2.737
T2 - P10	2.027	2.110	1.799	2.080	2.106	1.804	2.374	2.900	2.076
T2 overall	**2.053**	**2.085**	**1.856**	**1.979**	**1.950**	**1.853**	**4.271**	**4.192**	**3.647**
OVERALL	**1.789**	**1.870**	**1.729**	**1.743**	**1.815**	**1.739**	**4.271**	**6.310**	**3.884**

Table 6: Results of the 3D multi-modal CT-MR registrations on patient data, described in Section 4.2, with $\alpha = 0.6$ and $N = 5$. Given is the mean, median and maximal error (in mm) after registration using the Simplex, Elastix, and our MRF-based method. The results for PD - P03 are not taken into account, since all three methods failed in this case.

that for the 1-view case, the problem is ill conditioned in the projection direction. In this application we also demonstrate the robustness of the proposed method to high noise levels.

The task of 2D-3D registration is to recover a rigid 3D transformation T which relates the coordinate frame of the 3D volume I to the coordinate system of the 2D imaging devices, which generate the 2D projection images J_l. The transformation is estimated by optimizing the similarity measure ξ between the perspective projections of the transformed 3D image $P_l(I \circ T_p)$ and the m available 2D images J_l

$$E_{\text{2D3D}}(p) = \frac{1}{m} \sum_{l=1}^{m} \xi \left(P_l(I \circ T_p), J_l \right) , \quad (15)$$

We apply our MRF-framework to the 2D-3D problem, by using Equation (7) and approximate E_{2D3D} as

$$\widetilde{E}_{\text{2D3D}}(p) = \frac{1}{|C_2|} \sum_{\{p_i, p_j\} \in C_2} \widetilde{\psi}_{ij}(p_i, p_j) , \quad (16)$$

with

$$\widetilde{\psi}_{ij}(p_i, p_j) = \frac{1}{m} \sum_{l=1}^{m} \xi \left(P_l(I \circ T_{p^{ij}}), J_l \right) . \quad (17)$$

5.1. Evaluation

We assess the performance of the proposed approach by comparing it to the optimization of the original energy by the Simplex method. In order to perform the evaluation in a controlled setting, we conduct the 2-view tests on real 3D data, but with synthetically created projections, compare Section 5.1.2. This way, the choice of the similarity measure plays a smaller role and we can compare the performance of the optimization approaches. This also has the advantage that we can test on a large number of different views with known ground truth, rather than performing the tests on only a few real views by randomly disturbing the initialization. The robustness and applicability to real settings is demonstrated in experiments by adding noise to the projections. Furthermore, we test our algorithm in a more challenging real 1-view test scenario in Section 5.1.3, in which a real fluoroscopic image is used.

(a) 3DCBR   (b) DRR View 1   (c) DRR View 2

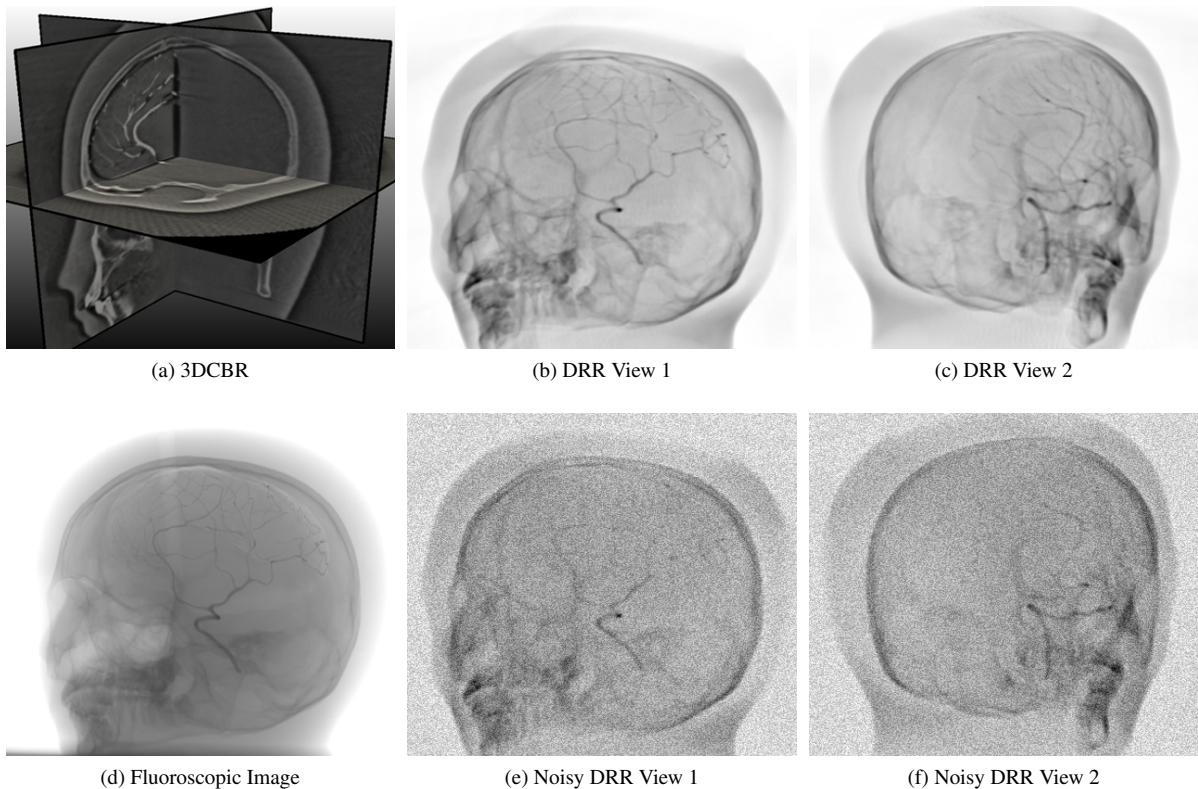(d) Fluoroscopic Image   (e) Noisy DRR View 1   (f) Noisy DRR View 2

Figure 9: Visualization of input data for the 2D-3D tests: **(a)** 3D cone beam reconstruction (3DCBR), which is used in all experiments; **(b,c)** Exemplary DRRs which are computed in the iterations of the algorithm; **(d)** Fluoroscopic image used as target in the 1-view test; **(e,f)** Exemplary DRRs with 20% uniform noise used as targets in the 2-view tests.

5.1.1. General Setting

As input data for the experiments we use a 3D cone beam reconstruction (3DCBR) of a phantom head with a cerebral vessel structure, computed from a single sweep of a monoplane stationary C-arm with flat-panel detector (Siemens Axiom Artis dTA). As 2D input we use a fluoroscopic image obtained by the same device, with ground truth transformation obtained by feature-based registration, and verified by careful inspection. An overview of the data is given in Figure 9.

The 2D projections required inside the algorithm and synthetically generated targets are computed as digitally reconstructed radiographs (DRRs) by GPU accelerated ray-casting. A conversion operator is used to remap the intensity values to X-ray energies.

All tests are performed with a set of random offset poses. The poses are generated by uniform random sampling of the parameters from certain ranges. For the 2-view test, we have $t \in [-30, 30]$mm, and $\phi \in [-80, 80]°$. For the 1-view test, we use $t \in [-15, 15]$mm, and $\phi \in [-40, 40]°$. We evaluate the results by the Target Registration Error (TRE). The TRE is computed as the mean of the distances between a set of points transformed by the ground truth transformation and the same point set transformed by the estimated transformation. The point set consists of points of a regular 10×10×10, 20cm large grid, centered at the phantom head.

As the similarity measure we use the local version of the Nor-malized Cross Correlation [38]. We also performed tests with the Gradient Difference measure [38], with very similar results.

The registrations are performed with a standard multi-resolution strategy using a Gaussian image pyramid. In all 2D-3D experiments, the initial search space ranges for the MRF-based method are ± 50mm for the translations and ± 90° for the rotations. We use $N = 7$ and $\alpha = 0.66$. For these settings, the average registration run takes about 4 minutes. Again, different parameter values were tested for the parameters, resulting in very similar registration results.

5.1.2. 2-View Test with DRRs as Target Image

We carry out the 2-view test by performing 400 runs. In each run, a new pair of orthogonal DRRs J_1, J_2 is generated by applying a random pose to the 3DCBR image. In order to test the robustness of the approach in a challenging scenario, uniform noise in the range of 20% of the intensities is added to all generated DRRs, please compare Figures 9e and 9f. The same test was also performed without noise, yielding comparable results. The tests show that both, the Simplex and the MRF approach perform well, with a small number of failed registrations only for very large initial misalignments, for details please compare Figure 10.

(a) TRE for 2-view test with noise
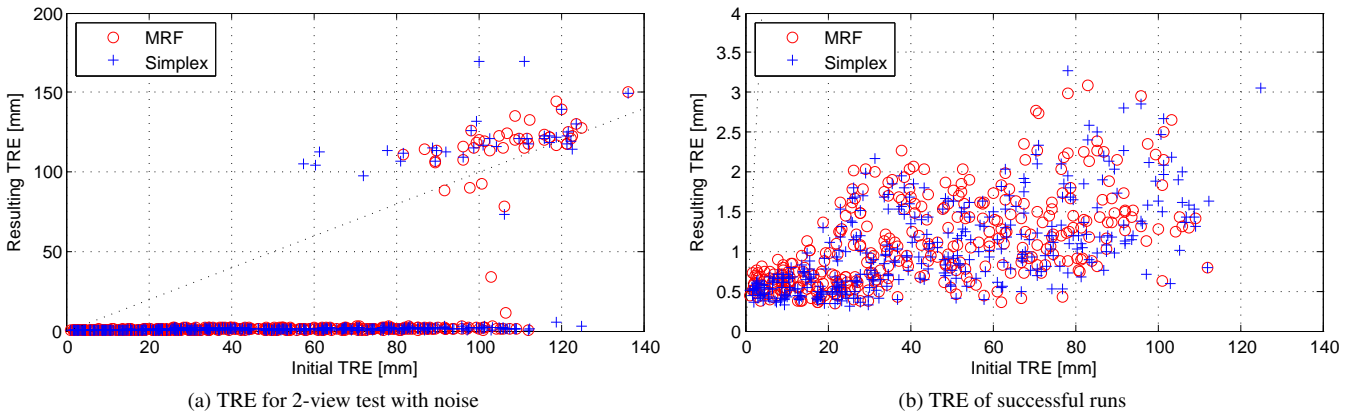

(b) TRE of successful runs

Figure 10: Results of 400 random 2-view test runs with 20% uniform noise. The x and y-coordinates of the graph points represent the TRE before and after registration. The diagonal is the line of no improvement. **(b)** shows a zoom of the area of successful test runs.

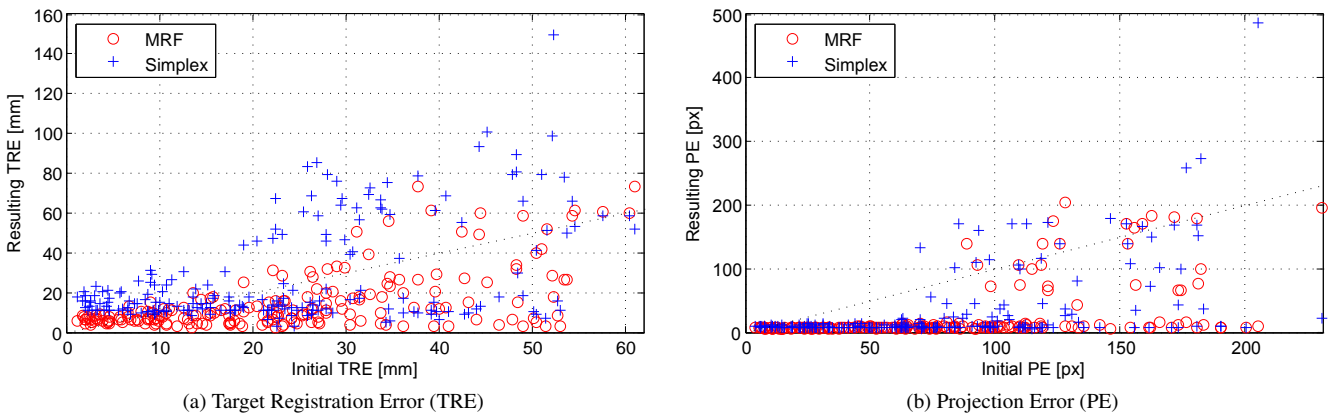(a) Target Registration Error (TRE)

(b) Projection Error (PE)

Figure 11: Results of the 200 random 1-view test runs with fluoroscopic image. In **(a)**, the TRE is shown, while **(b)** illustrates the projection error.

5.1.3. 1-View Test with Fluoroscopic Image as Target

The second series of experiments is performed in a 1-view scenario by registering the 3DCBR volume to a real fluoroscopic image. Here, we use 200 random poses as initial offsets to the 3D transformation. Similar to Section 4.1, we employ only the two coarsest levels of a five-level image pyramid to speed up the computation (resulting in projection images of size 128×128). Additionally to the TRE, we also evaluate the Projection Error (PE), which is computed by projecting the transformed 3D samples used for TRE computation to the image and computing their mean distance. The PE measures the pixel error visible in the image, such that it is not very sensitive to misalignments in the viewing direction. By comparing the TRE and PE measures in Figure 11, we observe that while the PE is approximately the same for both approaches, the TRE (in mm) is slightly better for the MRF-based approach (Simplex: mean=27.48, median=15.43, std.=25.09; MRF: mean=14.7, median=8.86, std.=14.8). With respect to the used resolution, the minimal errors are within sub-voxel accuracy. This suggests that the MRF optimization is able to better recover the translation along the viewing direction.

6. Overall Conclusion

In this paper we present an MRF-based framework for linear intensity-based registration of medical images. The key idea is to approximate the original energy corresponding to the chosen similarity measure by tractable second-order terms. To our best knowledge, this is the first time that an approach for linear intensity-based registration based on MRFs and discrete optimization is presented. The proposed model deviates from the common MRF problem modeling since it involves only pairwise terms, which do not encode a regularization term but are used as building blocks to represent the cost function. While the number of parameters is lower than for classical MRF problems, the parameters are not localized in space, and they influence the problem globally and not only in their direct neighborhood. The experimental evaluation confirms the quality of the proposed approximation and shows that the method is not sensitive to the setting of the internal parameters.

Furthermore, we demonstrate the applicability of the method for standard mono- and multi-modal image registration, as well as 2D-3D image registration. The registration experiments indicate increased robustness and precision of the proposed method

in comparison to standard optimization of the original energy function. The 2D-3D experiments show the applicability of the framework also to less well-conditioned problems, such as 1-view 2D-3D registration. These tests also demonstrate the robustness of the method to high noise levels.

Regarding the computation time, we find that our method is less efficient than the Simplex method due to the higher number of similarity evaluations. However, our running times are similar to the stochastic gradient descent when using the same number of sub-samples. Further tuning options, such as reducing the number of sub-samples were shown to decrease the computation time of the stochastic gradient descent with similar accuracy. This might also be an option for our method.

Due to the active development in discrete optimization, the proposed method has a strong further potential as it allows to transfer the advances in MRF optimization to linear registration problems, for example the integration of faster methods, or upcoming methods for third-order models.

References

[1] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.

[2] Stan Z. Li, *Markov random field modeling in image analysis*, Springer-Verlag New York, Inc., 2001.

[3] Y.Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images," in *IEEE International Conference on Computer Vision*, 2001.

[4] B. Glocker, N. Komodakis, G. Tziritas, N. Navab, and N. Paragios, "Dense image registration through mrfs and efficient linear programming," *Medical Image Analysis*, vol. 12, pp. 731–741, 2008.

[5] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, 2001.

[6] Vladimir Kolmogorov, "Convergent tree-reweighted message passing for energy minimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1568–1583, 2006.

[7] Nikos Komodakis, Georgios Tziritas, and Nikos Paragios, "Fast, approximately optimal solutions for single and dynamic mrfs," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[8] V. Lempitsky, C. Rother, and A. Blake, "Logcut - efficient graph cut optimization for markov random fields," in *IEEE International Conference on Computer Vision*, 2007.

[9] J. Maintz and M. Viergever, "A survey of medical image registration," *Medical Image Analysis*, vol. 2, no. 1, pp. 1–36, 1998.

[10] J.V. Hajnal, D.J. Hawkes, and D.L.G. Hill, *Medical Image Registration*, CRC Press, 2001.

[11] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[12] Ben Glocker, Darko Zikic, Nikos Komodakis, Nikos Paragios, and Nassir Navab, "Linear image registration through mrf optimization," in *IEEE International Symposium on Biomedical Imaging*, 2009.

[13] Darko Zikic, Ben Glocker, Oliver Kutter, Martin Groher, Nikos Komodakis, Ali Khamene, Nikos Paragios, and Nassir Navab, "Markov random field optimization for intensity-based 2d-3d registration," *SPIE Medical Imaging*, 2009.

[14] Pushmeet Kohli, Pawan Mudigonda, and Philip Torr, "P and beyond: Solving energies with higher order cliques," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[15] Pushmeet Kohli, Lubor Ladicky, and Philip H.S. Torr, "Robust higher order potentials for enforcing label consistency," *International Journal of Computer Vision*, vol. 82, no. 3, pp. 302–324, 2009.

[16] R. Potts, "Some generalized order-disorder transformation," in *Proceedings of the Cambridge Philosophical Society*, 1952, vol. 48.

[17] Xiangyang Lan, Stefan Roth, Daniel Huttenlocher, and Michael J. Black, "Efficient belief propagation with learned higher-order markov random fields," in *European Conference on Computer Vision*, 2006.

[18] Brian Potetz and Tai Sing Lee, "Efficient belief propagation for higher-order cliques using linear constraint nodes," *Computer Vision and Image Understanding*, vol. 112, pp. 39–54, 2008.

[19] Hiroshi Ishikawa, "Higher-order clique reduction in binary graph cut," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[20] E. Boros and P.L. Hammer, "Pseudo-boolean optimization," *Discrete Applied Mathematics*, vol. 123, pp. 155–225, 2002.

[21] Carsten Rother, Pushmeet Kohli, Wei Feng, and Jiaya Jia, "Minimizing sparse higher order energy functions of discrete variables," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[22] Nikos Komodakis and Nikos Paragios, "Beyond pairwise energies: Efficient optimization for higher-order mrfs," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[23] C.M. Bishop, *Pattern recognition and machine learning*, Springer, 2006.

[24] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2003.

[25] Vladimir Kolmogorov and Carsten Rother, "Minimizing nonsubmodular functions with graph cuts-a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 7, pp. 1274–1279, 2007.

[26] Nikos Komodakis, Georgios Tziritas, and Nikos Paragios, "Performance vs computational efficiency for optimizing single and dynamic mrfs: Setting the state of the art with primal-dual strategies," *Computer Vision and Image Understanding*, vol. 112, no. 1, pp. 14–29, 2008.

[27] J. Pearl, *Probabilistic Reasoning*, San Francisco, CA: Morgan Kaufmann, 1988.

[28] Y. Weiss and W.T. Freeman, "On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs," *IEEE Transactions On Information Theory*, vol. 47, no. 2, pp. 736–744, 2001.

[29] 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 with smoothness-based priors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 1068–1080, 2008.

[30] JPW Pluim, JBA Maintz, and MA Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 986–1004, 2003.

[31] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C*, Cambridge University Press, 1993.

[32] J. West, J.M. Fitzpatrick, M.Y. Wang, B.M. Dawant, C.R. Maurer Jr, R.M. Kessler, R.J. Maciunas, C. Barillot, D. Lemoine, A. Collignon, et al., "Comparison and evaluation of retrospective intermodality brain image registration techniques," *Journal of Computer Assisted Tomography*, vol. 21, no. 4, pp. 554, 1997.

[33] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 187–198, 1997.

[34] S. Klein, M. Staring, K. Murphy, M.A. Viergever, and J.P.W. Pluim, "elastix: a toolbox for intensity-based medical image registration," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 196–205, 2010.

[35] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3d medical image alignment," *Pattern Recognition*, vol. 32, no. 1, pp. 71–86, 1999.

[36] S. Klein, J. P. W. Pluim, M. Staring, and M.A. Viergever, "Adaptive stochastic gradient descent optimisation for image registration," *International Journal of Computer Vision*, vol. 81, no. 3, pp. 227–239, 2009.

[37] S. Klein, M. Staring, and J. P. W. Pluim, "Evaluation of optimisation methods for nonrigid medical image registration using mutual information and b-splines," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2879–2890, 2007.

[38] Graeme P. Penney, Jürgen Weese, John A. Little, Paul Desmedt, Derek L. G. Hill, and David J. Hawkes, "A comparison of similarity measures for use in 2D-3D medical image registration," *IEEE Transactions on Medical Imaging*, vol. 17, no. 4, pp. 586–595, 1998.