Nonrigid Image Registration Using Higher-Order MRF Model with Dense Local Descriptor

Dongjin Kwon[1], Kyong Joon Lee[1], Il Dong Yun[2], and Sang Uk Lee[1]
[1]School of EECS, ASRI, Seoul Nat'l Univ., Seoul, 151-742, Korea
[2]School of EIE, Hankuk Univ. of F. S., Yongin, 449-791, Korea
{djk, kjoon}@cvl.snu.ac.kr, yun@hufs.ac.kr, sanguk@ipl.snu.ac.kr

Abstract

In this paper, we propose a nonrigid image registration method using a Markov Random Field (MRF) energy model with higher-order smoothness priors and a dense local descriptor. The image registration is designed as finding an optimal labeling of the MRF energy model where the label corresponds to a discrete displacement vector. The proposed MRF energy model uses matching scores of dense local descriptors between images as a data cost. In this model, spatial relationships are constructed between nodes using higher-order smoothness priors. As the local descriptor is invariant to scale and rotation and also robust to changing appearances, this method can handle multi-modal images involving scale and rotation transformations. The higher-order smoothness priors can generate desired smoother displacement vector fields and do not suffer from fronto-parallel effects commonly occurred in first-order priors. The usage of higher-order priors in the energy model enables this method to produce more accurate registration results. In the experiments, we will show registration results using multi-modal Brain MRI Images and facial images with expression and light changes.

1. Introduction

Nonrigid image registration is the process of determining geometric transformation between two images which are not related by simple transformations such as rigid or affine. Over the last decade, many relevant works were proposed [19] including feature-based [13, 8] and image-based schemes [14, 3, 6]. The feature-based methods can easily cover large displacements, however they do not work well on the regions which have no features. On the other hand, the image-based methods use the entire area of images and work more accurately than feature-based methods. However, they become intractable to solve problems involving large displacements. In addition, it is not easy to define appropriate similarity measures for registering multi-modal images.

Recently, some approaches utilizing both feature- and image-based methods are introduced. Ou *et al*. [12] integrated a cost measure using Gabor filtering responses (attributes) for each pixels to the first-order MRF energy model. As full dimensional attributes have redundant information, they introduce a procedure for reducing the dimension of attributes by a learning-based method. Sotiras *et al*. [16] proposed a unified energy model of feature- and image-based schemes. They construct one graph for each scheme and connect these two graphs by appropriately designed edges. Liu *et al*. [9] used pixel-wise SIFT [10] descriptors for a data cost to the MRF energy model with first-order smoothness prior. They introduce a coarse-to-fine matching scheme to cover large displacements. This method is designed for robust matching across different appearances of scenes or objects.

Above mentioned methods model the deformation pattern as a discrete label set where labels correspond to displacements of control points (nodes) placed on the square mesh. The energy is constructed using the standard pairwise MRF model as follows

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

where $\mathcal{V}$ is the set of nodes, $\mathcal{E}$ is the set of edges incorporating neighborhood information of nodes, and x_s is the label of $s \in \mathcal{V}$. In this model, the data cost θ_s is computed using similarity measures between reference and input images and the smoothness cost θ_{st} is computed using displacement differences between s and t. For θ_{st}, the following truncated pairwise spatial prior is conventionally used

$$\theta_{st}(x_s, x_t) = \lambda_{st} \min\left(\|\mathbf{d}(x_s) - \mathbf{d}(x_t)\|_1, T_{st}\right) \qquad (2)$$

where λ_{st} is the regularization constant, $\mathbf{d}(x_s)$ represents the displacement vector corresponding to the label x_s, and T_{st} is a threshold for truncation. However, this pairwise potentials (2) penalize the global transformations of the mesh, such as rotation and scaling movements. Mesh nodes must

be remained at initial position or translated all together to get low energy.

Here, we propose a nonrigid image registration method integrating the strengths of the feature-based scheme into an MRF energy model with a higher-order smoothness prior. We use the densely sampled SIFT descriptor [10] for computing a data cost of the MRF energy model. Because the SIFT descriptor is invariant to scale and rotation as well as robust to appearance variation, the proposed method can handle largely deformed multi-modal images. For a smoothness cost, we use a mixed-order smoothness prior to take advantages of first- and second-order smoothness priors. This smoothness priors can generate desired smoother displacement vector fields and do not suffer from fronto-parallel effects commonly occurred in first-order priors. The usage of dense local descriptor and higher-order priors in the energy model enables the method to produce more accurate registration results.

2. MRF Energy Model for Registration

For an input image, we construct a set $\mathcal{V}$ which consists of nodes placed on each pixel. Then we generate a factor graph [5] $\mathcal{G}_{\mathcal{F}} = (\mathcal{V}, \mathcal{F}_{\mathcal{P}}, \mathcal{F}_{\mathcal{H}})$ where $\mathcal{V}$, $\mathcal{F}_P$ and $\mathcal{F}_H$ are the set of nodes, factors for pairwise potentials (corresponding to $\mathcal{E}$ in (1)), factors for higher-order potentials, respectively. For each $s \in \mathcal{V}$, let x_s be a label taking values in some discrete set $\mathcal{L}$. A function $\mathbf{d} : \mathcal{L} \rightarrow \mathbb{R}^2$ is defined for mapping labels to 2-dimensional displacements where each label x_s corresponds to a displacement vector $\mathbf{d}(x_s) = \big(d_x(x_s), d_y(x_s)\big)$. In $\mathcal{G}_{\mathcal{F}}$, an unary potential $\theta_s(x_s)$ is defined for each node $s \in \mathcal{V}$, a pairwise potential $\theta_{st}(x_s, x_t)$ is defined for each factor $(s,t) \in \mathcal{F}_{\mathcal{P}}$ and a higher-order potential $\theta_{stu}(x_s, x_t, x_u)$ is defined for each factor $(s,t,u) \in \mathcal{F}_{\mathcal{H}}$.[1]

2.1. Higher-Order Smoothness Prior

In [2], deformation energy of the mesh is usually defined as a sum of squared second derivatives of its nodes. This deformation energy describes the natural representation of inherent deformedness of the mesh which depends only on the relative locations of mesh nodes. Following the analysis on [6], we add second-order smoothness priors into our energy model. To apply second-order smoothness priors, the ternary potential $\theta_{stu}(x_s, x_t, x_u)$ is constructed on every collinear three nodes s, t and u as follows

$$\theta_{stu}(x_s, x_t, x_u) = \lambda_{stu} \min\left(\|\mathbf{d}(x_s) - 2\mathbf{d}(x_t) + \mathbf{d}(x_u)\|_1, T_{stu}\right) . \quad (3)$$

In contrast to [6], we retain original first-order smoothness priors as following recent research [7] on higher-order smoothness priors: mixing different orders of smoothness priors produces better results. We can control regularization parameters λ_{st}, λ_{stu} relatively to emphasis the effects of each prior.

By applying higher-order smoothness priors, the final MRF energy model is described as follows:

$$E(x|\theta) = \sum\nolimits_{s\in\mathcal{V}} \theta_s(x_s) + \sum\nolimits_{(s,t)\in\mathcal{F}_{\mathcal{P}}} \theta_{st}(x_s, x_t) + \sum\nolimits_{(s,t,u)\in\mathcal{F}_{\mathcal{H}}} \theta_{stu}(x_s, x_t, x_u) \quad (4)$$

where $\mathcal{F}_{\mathcal{P}}$ and $\mathcal{F}_{\mathcal{H}}$ is defined on a set of 4-neighborhood nodes and a set of collinear three nodes, respectively.

2.2. Dense Local Descriptor

The unary term $\theta_s(x_s)$ of (4) which measures the cost when a node s has a label x_s is defined as

$$\theta_s(x_s) = f(s_x, s_y, s_x + d_x(x_s), s_y + d_y(x_s)) . \quad (5)$$

The cost is calculated by a dissimilarity measure f using two local image information centered on (s_x, s_y) in a source image and $(s_x + d_x(x_s), s_y + d_y(x_s))$ in a target image, respectively. For computing f, we apply L^1 distance measure on the SIFT descriptor [10] space. Among various local descriptors, we choose the SIFT descriptor, based on a recent study [9] which showed promising results. We extract SIFT descriptor on each pixel using following procedure. Firstly, we find the maximum gradient orientation using a gradient vector histogram constructed in the 16×16 window centered at each pixel location. Then we make gradient histogram having 4×4 grid rotated to the maximum of orientation, each bin has 4×4 window. The gradient orientation is quantized to 8 histogram bins and radially decreasing Gaussian weighting is applied to each corresponding bin [10]. Finally a 128-dimensional descriptor (grid size × histogram bin number) is constructed after L^2-norm normalization.

3. Optimization Strategy

Optimizing the proposed energy model (4) directly is time-consuming when a large search space $\mathcal{L}$ should be applied to cover large displacements. To reduce computational burden, we apply a decomposed scheme with a coarse-to-fine scheme.

3.1. Decomposed Scheme

Based on the scheme introduced in [15], we generate a new graph $\mathcal{G}^D$ by making two layers of nodes $\mathcal{V}^x$ and $\mathcal{V}^y$ corresponding to x and y displacements from the original nodes $\mathcal{V}$. This is only applicable when potentials for spatial priors are summation of a potential for each dimensional displacement. Using the upper bound approximation, we decouple smoothness priors (2) and (3) as follows

$$\theta_{st}(x_s, x_t) \leq \theta^x_{st}(x^x_s, x^x_t) + \theta^y_{st}(x^y_s, x^y_t) , \quad (6)$$

$$\theta_{stu}(x_s, x_t, x_u) \leq \theta^x_{stu}(x^x_s, x^x_t, x^x_u) + \theta^y_{stu}(x^y_s, x^y_t, x^y_u) \quad (7)$$

[1] When a factor a connects nodes s and t and a factor b connects nodes s, t and u, we use $(s,t) \in \mathcal{F}_{\mathcal{P}}$ or $(s,t,u) \in \mathcal{F}_{\mathcal{H}}$ to represent $a \in \mathcal{F}_{\mathcal{P}}$ or $b \in \mathcal{F}_{\mathcal{H}}$ when we do not need to use factor representations explicitly.

where

$$\theta^i_{st}(x^i_s, x^i_t) = \lambda_{st} \min(|d_i(x_s) - d_i(x_t)|, T_{st}) ,$$

$$\theta^i_{stu}(x^i_s, x^i_t, x^i_u) = \lambda_{stu} \min(|d_i(x_s) - 2d_i(x_t) + d_i(x_u)|, T_{stu})$$

for all $i \in x, y$. In the graph $\mathcal{G}^D$, factor sets include $\mathcal{F}^x_{\mathcal{P}}$, $\mathcal{F}^x_{\mathcal{H}}$ and $\mathcal{F}^y_{\mathcal{P}}$, $\mathcal{F}^y_{\mathcal{H}}$ for intra-layer interaction potentials and $\mathcal{F}^{xy}_{\mathcal{P}}$ for inter-layer interaction potentials defined as

$$\theta^{xy}_{st}(x^x_s, x^y_t) = f(s_x, t_y, s_x + d_x(x_s), t_y + d_y(x_t)) . \quad (8)$$

Then, the MRF energy on $\mathcal{G}^D$ is defined as follows

$$E(x|\theta) = \sum_{(s,t)\in\mathcal{F}^D_{\mathcal{P}}} \theta_{st}(x_s, x_t) + \sum_{(s,t,u)\in\mathcal{F}^D_{\mathcal{H}}} \theta_{stu}(x_s, x_t, x_u) \quad (9)$$

where $\mathcal{F}^D_{\mathcal{P}} = \mathcal{F}^x_{\mathcal{P}} \cup \mathcal{F}^y_{\mathcal{P}} \cup \mathcal{F}^{xy}_{\mathcal{P}}$ and $\mathcal{F}^D_{\mathcal{H}} = \mathcal{F}^x_{\mathcal{H}} \cup \mathcal{F}^y_{\mathcal{H}}$.

3.2. Coarse-to-Fine Scheme

To cover large displacements efficiently, we compute multi-level displacement in a coarse-to-fine manner [14]. For a pyramidal dense descriptor representation, we smooth and downsample SIFT descriptors of finer level to generate SIFT descriptors of the coarser level instead of extracting SIFT descriptor on the image of lower resolution [9]. The displacements are computed on each level with propagated displacement offset which is scaled from the coarser levels. To achieve a sub-pixel accuracy, we perform the sub-pixel refinement on the finest level after final level computation.

3.3. Discrete Optimization Method

To optimize the proposed energy model (4) or (9), TRW message passing [18, 4] is applied. The TRW message passing provides the lower bound guaranteed not to decrease. A recent comparative study shows the TRW gives the state-of-the-art performances among the various discrete optimization methods [17]. As the TRW theory is built on the pairwise MRF, we need to convert factor graphs representations to pairwise interactions to use TRW. The detailed conversion procedure is referred to [6]. As the proposed energy model uses higher-order smoothness priors, converting to the hierarchical gradient node graph [7] is possible.

4. Experiments

In this section, we show registration results using multi-modal Brain MRI Images and facial images with expression and light changes. We compare the results with a feature-based method [8] using sparse SIFT features and SIFT flow [9] using dense SIFT descriptors with a first-order MRF energy model. In experiments, $\lambda_{st} = 2$, $T_{st} = 20$, $\lambda_{stu} = 0.4$ and $T_{stu} = 20$ are used. For label widths, we use $d_i \in \{-10, \ldots, 10\}$ for the coarsest level, $d_i \in \{-(n+1), \ldots, (n+1)\}$ for n^{th} level ($n = 1$ for the finest level) and $d_i \in \{-2.0, -1.8, \ldots, 1.8, 2.0\}$ for the sub-pixel refinement where $i \in \{x, y\}$. All parameters are empirically chosen.

Table 1. Registration Errors for Brain MRI Images (RMSE)

Method	T1 - T1		T1 - T2	
	σ = 6	σ = 9	σ = 6	σ = 9
Feature-Based [8]	3.45	6.90	38.67	58.42
SIFT Flow [9]	1.56	3.69	3.40	5.06
Proposed	1.36	3.44	3.23	4.89

4.1. Brain MRI Images

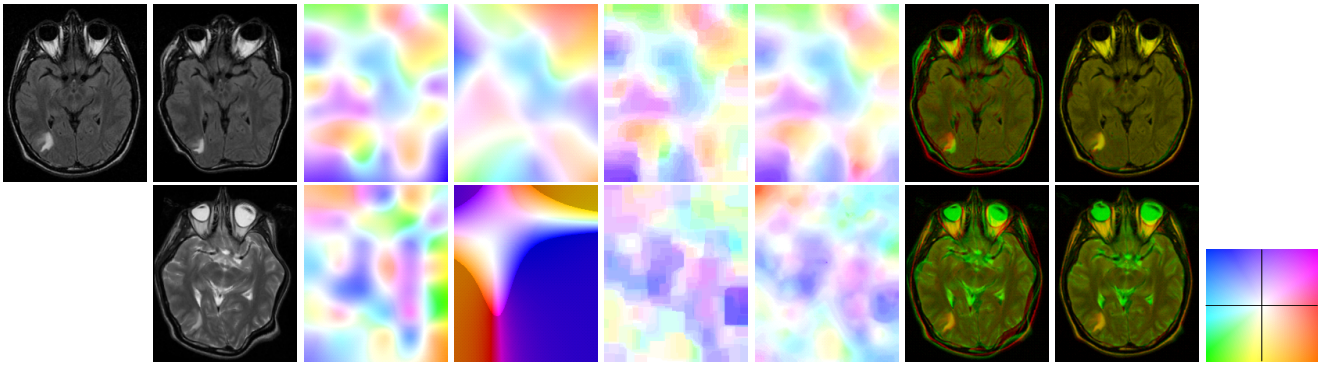
We generate synthetically deformed data sets (10 images for each σ) given T1 and T2 weighted brain MRI images (Fig.1(a) and (b)), respectively. The deformed image is generated by TPS [2] warping with control points perturbed with random variation $[-\sigma, \sigma]$. More detailed procedure for generating synthetic images are described in [6]. The registration is performed between T1 and warped T1 images (uni-modal) and T1 and warped T2 images (multi-modal), respectively. As intensity characteristics between T1 and T2 images are not consistent, registrations between T1 and T2 are more challenging problem. In Table 1, it is shown root mean square error (RMSE) between the ground truth and the registration result. One can see the performances of the proposed method are better than those of other methods. The feature-based method for registering T1 and T2 images do not perform well as they are not converged to solutions due to small numbers of right SIFT matchings. In Fig. 1, we show some registration results when $\sigma = 9$. In the figure, the proposed method produces smoother displacements with higher accuracy than other methods.

4.2. Facial Images

We test registration methods on some facial images from the AR face database [11]. In Fig. 2, we show registration results between neutral faces with bright light (a) and a smiling face with normal light (b). This is a challenging example as there are mixed variations of facial expression and appearance. While the proposed method (e) produces smoother displacements than that of SIFT flow [9] (d), the feature-based method [8] (c) is failed to generate detailed displacements. The image differences for before and after registration using the proposed method showed in (f) and (g), respectively. One can see smiling faces are overlapped well with neutral faces after registrations.

5. Conclusion

In this paper, we proposed a nonrigid registration method using the MRF model which consists of a higher-order smoothness prior and a dense local descriptor based data cost. Using the SIFT descriptor [10] for a data cost, we can register largely deformed images with appearance variation more effectively. Moreover, the higher-order prior enables us to generate smoother displacement fields with better accuracy. In experiments, we showed the proposed method



(a) Source (b) Target (c) Ground Truth (d) Feature-Based [8] (e) SIFT Flow [9] (f) Proposed (g) Initial (h) Final (i) Color Codes

Figure 1. Some results for brain MRI images. In (a), original T1 weighted scan image is shown, and in (b), warped T1 (upper) and T2 (lower) weighted scan images are shown. In the upper row, we show registration results between T1 and warped T1 images, and in the lower row, results between T1 and warped T2 are shown. In (i), color coded displacements [1] which encode their direction and magnitude as a color. In (c)-(f), displacement vectors are represented using color codes in (i). The final difference (h) is a combined image of (a) (red) and warped (b) using the proposed method (green).

(a) Source (b) Target (c) Feature-Based [8] (d) SIFT Flow [9] (e) Proposed (f) Initial (g) Final

Figure 2. Results for facial images. In (c)-(e), displacement vectors are represented using color codes in Fig.1(h). The final difference (g) is a combined image of (a) (red channel) and warped (b) (green channel) using the proposed method.

outperformed a feature-based method [8] and SIFT flow [9] using brain MRI images and facial images.

References

[1] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski. A Database and Evaluation Methodology for Optical Flow. *Int. J. Comput. Vision*, 92(1):1–31, 2011.

[2] F. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(6):567–585, 1989.

[3] B. Glocker, N. Komodakis, N. Paragios, G. Tziritas, and N. Navab. Inter and Intra-modal Deformable Registration: Continuous Deformations Meet Efficient Optimal Linear Programming. In *IPMI*, 2007.

[4] V. Kolmogorov. Convergent Tree-Reweighted Message Passing for Energy Minimization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(10):1568–1583, 2006.

[5] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger. Factor Graphs and the Sum-Product Algorithm. *IEEE Trans. Inf. Theory*, 47(2):498–519, 2001.

[6] D. Kwon, K. J. Lee, I. D. Yun, and S. U. Lee. Nonrigid Image Registration Using *Dynamic* Higher-Order MRF Model. In *ECCV*, 2008.

[7] D. Kwon, K. J. Lee, I. D. Yun, and S. U. Lee. Solving MRFs with Higher-Order Smoothness Priors Using Hierarchical Gradient Nodes. In *ACCV*, 2010.

[8] D. Kwon, I. D. Yun, K. H. Lee, and S. U. Lee. Efficient Feature-Based Nonrigid Registration of Multiphase Liver CT Volumes. In *BMVC*, 2008.

[9] C. Liu, J. Yuen, and A. Torralba. SIFT Flow: Dense Correspondence across Scenes and Its Applications. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(5):978–994, 2011.

[10] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.

[11] A. Martinez and R. Benavente. The AR Face Database. *CVC Technical Report #24*, 1998.

[12] Y. Ou, A. Sotiras, N. Paragios, and C. Davatzikos. DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting. *Medical Image Analysis*, in press.

[13] K. Rohr. Image Registration Based on Thin-Plate Splines and Local Estimates of Anisotropic Landmark Localization Uncertainties. In *MICCAI*, 1998.

[14] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid Registration Using Free-Form Deformations: Application to Breast MR Images. *IEEE Trans. Medical Imaging*, 18(8):712–721, 1999.

[15] A. Shekhovtsov, I. Kovtun, and V. Hlavác. Efficient MRF Deformation Model for Non-Rigid Image Matching. In *CVPR*, 2007.

[16] A. Sotiras, Y. Ou, B. Glocker, C. Davazikos, and N. Paragios. Simultaneous Geometric - Iconic Registration. In *MICCAI*, 2010.

[17] 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*, 2006.

[18] M. J. Wainwright, T. Jaakkola, and A. S. Willsky. MAP Estimation Via Agreement on Trees: Message-Passing and Linear Programming. *IEEE Trans. Inf. Theory*, 51(11):3697–3717, 2005.

[19] B. Zitova and J. Flusser. Image registration methods: a survey. *Image and Vision Comput.*, 21(11):977–1000, 2003.