Learning Based Coarse-to-fine Image Registration

Jiayan Jiang
LONI, UCLA
jet.jiang@loni.ucla.edu

Songfeng Zheng
Dept. of Statistics, UCLA
sfzheng@stat.ucla.edu

Arthur W. Toga
LONI, UCLA
toga@loni.ucla.edu

Zhuowen Tu
LONI, UCLA
ztu@loni.ucla.edu

Abstract

This paper describes a coarse-to-fine learning based image registration algorithm which has particular advantages in dealing with multi-modality images. Many existing image registration algorithms [18] use a few designed terms or mutual information to measure the similarity between image pairs. Instead, we push the learning aspect by selecting and fusing a large number of features for measuring the similarity. Moreover, the similarity measure is carried in a coarse-to-fine strategy: global similarity measure is first performed to roughly locate the component, we then learn/compute similarity on the local image patches to capture the fine level information. When estimating the transformation parameters, we also engage a coarse-to-fine strategy. Off-the-shelf interest point detectors such as SIFT [12] have degraded results on medical images. We further push the learning idea to extract the main structures/landmarks. Our algorithm is illustrated on three applications: (1) registration of mouse brain images of different modalities, (2) registering human brain image of MRI T1 and T2 images, (3) faces of different expressions. We show greatly improved results over the existing algorithms based on either mutual information or geometric structures.

1. Introduction

Registration is an important step in medical image analysis [14]. The goal of image registration is to estimate a geometric transformation so that two images can be aligned properly. Registration algorithms often define a *similarity measure* with a distortion penalty, either local or global, which is then turned into an energy function. The algorithm then searches over the possible geometrical transformations to determine the best one which minimizes the total energy. Fig. (1) shows an image registration task for two modalities. A 2D slice of a mouse brain image is supposed to register to a carefully constructed template image where the genetic information is coded at specific locations.

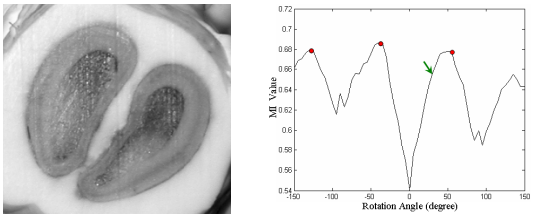
Information-theoretic similarity measures are widely used in the medical image registration community. These measures include mutual information (MI) [18, 25], the Kullback-Leibler distance [6, 10], Jésen-Renyi divergence [11], and cross entropy [30]. Pluim et al. [19] gave a survey of these measures with the corresponding algorithms.

(a) 2D brain image slice (b) template

Figure 1. Illustration of multi-modality image registration. (a) shows a 2D image slice of a mouse brain. (b) displays a specially constructed template to which the brain should register.

Despite the successes of the information-theory based measures, these approaches make independent identical distribution (i.i.d.) assumptions which may not be sufficient in many situations, especially for multi-modality images. For example, unsatisfactory results are obtained when we use mutual information based approaches to register a mouse brain to a template as illustrated in Fig. (2) (more examples in Fig. (6)). This is because the intensity patterns differ too much between corresponding points on the unregistered image and the template. More generally, information-theoretic measures tend to ignore the spatial structure information in the images, which are strong cues for registration. Some recent methods have attempted to take spatial structure into account [3, 17, 28], but they use a few carefully designed criterion, which may not be easy to generalize.

Instead, we push the learning aspect to learn the similarity functions by selecting and fusing a large number of features. The learning process is carried in a coarse-to-fine strategy. Intuitively, features computed from local image patches provide fine level correspondence but can only resolve the ambiguity to a certain degree. On the other hand, similarity measure based on global image scale can quickly align the structure to a rough position, but it has difficulty to compute accurate transformation. To reduce the complexity in both training and testing, we only learn fine level similar-



(a) (b)

Figure 2. (a) is a cropped image from Fig. (1.a). (b)) shows the mutual information between the two images (a) and Fig. (1.b), which is plotted as a function of the rotation angle. The correct rotation angle is $30°$ (green arrow), but the similarity measure has its maximum at $-50°$, with sub-maximal at $50°$ and $-120°$ (red points).

ity on the key structures and landmarks. However, off-the-shelf algorithms such as SIFT [12] and Canny detectors [5] have degraded results on medical images. We further push the learning idea to extract the main structures/landmarks by training the BEL detector [8] to capture the main image structures. In testing, coarse level similarity measure is applied to quickly locate the main image part in the source image, followed by fine level matching. Moreover, we adopt the coarse-to-fine strategy to compute the transformation also. The RANSAC algorithm [9] is used to estimate the affine transformation parameters first, followed by computing the detailed non-rigid transformation.

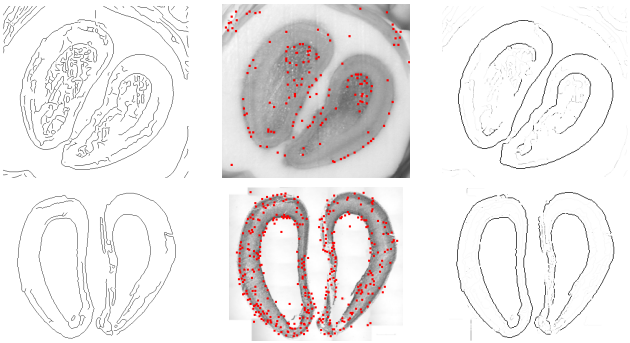
In medical image registration, the importance of applying different criterion on different situations has been noted [20, 17]. A learning based registration method [21] was proposed but they did not learn a similarity measure explicitly for multi-modality images, which is more challenging. Wu et al. [26] also used learning techniques for image registration. However, [26] used several manually carefully designed rules for measuring the correspondences, whereas ours is much more general and adaptive, and it fuses a large number of features (texture and shape) from a big candidate pool. Also, [26] is focused on images with the same modality, and the source and template have roughly the same size. Ours is capable of dealing with images of very different sizes and modalities. An algorithm for learning image similarity in computer vision [29] has some similarity to our approach, but their template and source images are similar and the details of the algorithm are different. Recently, several approaches have been proposed [4, 1] which are along the vein of learning metric for matching. However, Caetano et al. [4] carefully designed several functions for graph matching. Our fine level measure based on local patches is similar to that proposed in [1]. However, our approach differs to [1] in several aspects: (1) their focus was to learn a similarity function only, not for the overall task of image registration; (2) coarse-to-fine strategy was not used; (3) they did not have the structure learning part; (4) their method did not work on medical images of multi-modalities.

We apply our learning-based registration method to three applications. Firstly, we register a set of mouse brains to the corresponding template (atlas). This is a challenging case since the images are different. We show that our method outperforms alternatives which use similarity measures such as mutual information [25], landmarks by SIFT [15], and shape context [2]. We also register a set of T1 and T2 images. In this easier case, our methods performs as well as the state-of-the-art. In addition, we demonstrate our system for registering faces of different expressions.

2. The training process

In training, our approach consists of three main steps, namely, (1) learning salient structures, (2) learning the similarity measure at the coarse level, (3) learning the similarity measure at the fine level.

2.1. Salient Structure Extraction



(a) (b) (c)

Figure 3. (a) Canny edge maps of Fig. (2.a) and Fig. (1.b). (b) SIFT interest points. (c) Edges extracted by learned BEL detector.

Edge/landmark points contain structural information, which is important for image registration. The existing matching algorithms [1, 2, 15] mostly use off-the-shelf algorithms such as Canny edge [5] or SIFT interest point [12] detectors. Fig. (3.a) and Fig. (3.b) show the salient structures extracted by Canny and SIFT on the source image and the template respectively. As we can see, interest points extracted by SIFT are not consistent for the two images, which is a key for a landmark based registration algorithm to succeed. Also, Canny detector outputs too many non-informative edges which are distractions to the registration process. The reason is probably because existing edge/interest point detectors are based on some general assumptions about the image gradients and scale-invariance properties, which may not be adequate and adaptive to specific domains.

This motivates us to obtain an edge detector through learning, which enables us to capture the unique underlying statistics of images we are facing. We use a recently proposed edge detector, the Boosted Edge Learner (BEL) [8].

We assume that all the edge points will appear on the Canny edge map (with low threshold), and then train a boosting based classifier to classify the salient edge points and the rest of the edge points. The input for the training phase of BEL contains image pairs in which one is an intensity image and the other is a manual label map. BEL then automatically learns a set of features (Haars, gradients at different scales, etc.) to compute the probability of a pixel being on the edge or not. Please refer to [8] for details. The BEL detector performs significantly better and Fig. (3.c) shows the result by a trained BEL detector.

2.2. Learning the similarity measure

Figure 4. Training positive samples for similarity measure at the coarse level.

A key component in our algorithm is to obtain the similarity functions through learning, which implicitly combines both appearance and structural information. This is due to the availability of a large candidate feature pool we allow the learner to select from, such as different types of Haars [24], gradients, curvatures, all at different scales. To perform the registration task shown in Fig. (1), we learn similarity measure at both coarse and fine level. Intuitively, a coarse level similarity measure should be able to quickly locate the brain in Fig. (1.a) by capturing the similarity at a coarse scale. A similarity measure at the fine level then allow us get the fine level point correspondences.

2.2.1 coarse level

Fig. (4) shows some training positive samples at the coarse level. We use a cascade approach to train a sequence of boosting nodes, similar to [24]. The difference is that each positive/negative sample contains an image pair, (I_1, I_2). Let $y \in \{-1, +1\}$ be the label and $y = -1$ means that two images are not corresponding to each other and $y = +1$ says that the pair is a correct match. Each image is a 35×35 patch and we compute around $40,000$ features on both I_1 and I_2. The feature for the classifier is the difference between the two

$$F_i(I_1, I_2) = f_i(I_1) - f_i(I_2),$$

where f_i denotes the ith feature computed on I. We use the stump weak classifier in boosting

$$h(F_i(I_1, I_2), tr) = \begin{cases} +1 & \text{if } F_i(I_1, I_2) \geq tr \\ -1 & \text{otherwise} \end{cases} \quad (1)$$

The boosting node learns a discriminative function

$$q(+1|I_1, I_2) = \frac{\exp\{2\sum_t h_t(F(I_1, I_2))\}}{1 + \exp\{2\sum_t h_t(F(I_1, I_2))\}}.$$

The cascade can be formulated as a special case of probabilistic boosting-tree (PBT) [23], whose goal is to learn an overall discriminative model by

$$\begin{aligned} \tilde{p}(y|I_1, I_2) &= \sum_{l_1} \tilde{p}(y|l_1, I_1, I_2) q(l_1|I_1, I_2) \\ &= \sum_{l_1,..,l_n} \tilde{p}(y|l_n, ..., l_1), ..., q(l_2|l_1, I_1, I_2) q(l_1|I_1, I_2), \end{aligned}$$

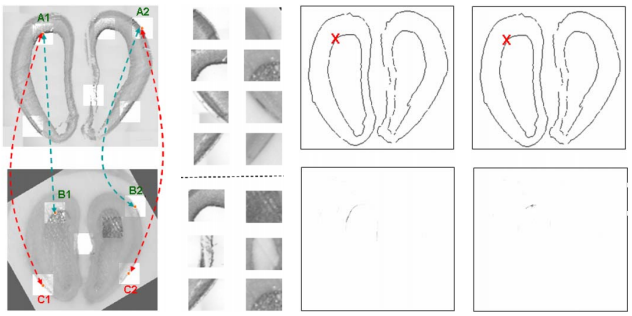
where $\tilde{p}(y|l_n, ..., l_1)$ is the empirical distribution at the leaf node and q outputs the discriminative probability for each boosting node in the cascade.

2.2.2 fine level

The coarse level measure can quickly locate the brain in the source image by searching over many locations, scales, and orientations, similar to the procedures in the object detection task [24]. This step takes the brain to roughly the same orientation and scale of the template. We then manually create correspondences on the landmarks sampled from the edges by BEL. Much like training the coarse level similarity measure, the corresponding pairs of landmarks in the source and template are labeled as $+1$ and any pair which is not matched to each other should be considered as negative -1.

The pair points that correspond will be the positive examples for the learning procedure, while the negative examples are the point pairs that do not correspond. For example, in Fig. (5.a), the point pair (A1, B1) is a positive example while (A1, C1) is a negative example.

For each point pair, we extract two 29×29 image patches centered at the landmarks. Examples are shown in Fig. (5.b). We apply a cascade of boosting algorithm to learn the classification model $p(y|I_1, I_2)$. Fig. (5.c,d) show examples of the similarity map. The learner automatically selects and combines different features extracted from the image patches. The features are drawn from a pool of $25,000$ features, similar to those described in the coarse learning part. The first five features selected are: Haar feature (width=16, height=16, top=7, left=7), Angle of derivative (scale=3), Angle of derivative (scale=1), Haar feature (width=4, height=16, top=7, left=13), Haar feature (width=4, height=16, top=7, left=18). We compare the mutual information (MI) similarity measure to our learned similarity measure for the image patches in Fig. (5.a). The similarity measures are listed in Table 1. Clearly, the MI measure has high values for both the correct and incorrect matches. But our learned similarity measure has high values only for the correct matches.



(a) (b) (c) (d)

Figure 5. (a) Positive and negative examples: (A1,B1) and (A2,B2) are positive examples while (A1,C1) and (A2,C2) are negative examples; (b) The first four rows show the positive image patches from (a) (template on the left, image on the right), and the last three rows show the negative image patches from (a); (c) above: a point sampled from the BEL edge map of template image; below: In the BEL edge map of unregistered image, the corresponding point with probabilities after the first layer of the cascade. (d) Same convention as (c), after full cascade.

Image Patch Pair	(A1, B1)	(A2, B2)	(A1, C1)	(A2, C2)
MI value	2.7	2.9	2.5	2.8
Learned similarity	0.89	0.81	0.02	0.01

Table 1. The MI value between the image patches in Fig. 5(a), and the similarity (as probability) measure computed after the first layer of PBT.

It is important to understand that the similarity measure between I_1 and I_2 will vary if we rotate the unregistered image. However, this condition is not hard to meet since we can search for the best matched sub-image in the source image using the measure learned from the coarse level.

2.3. Registration

After the similarity measure is learned, we sample 500 points from the edge map of the template image, denoted as $\{(u_i, v_i) : i = 1, ..., 500\}$. Given an unregistered image, we search for the sub-image which best matches with the template. Searching is performed on all the locations, different scales, and rotations of the source image. Once the best match is found, we then rotate and re-scale it to roughly the same size as the template. We then run the learned BEL detector to obtain the edge map points $\{(x_j, y_j)\}$. Computing the similarity measure learned at the fine scale for each pair of image patches based on the landmarks $I(u_i, v_i)$ and the edge points $I(x_j, y_j)$ gives

$$p(+1|I(u_i, v_i), I(x_j, y_j)).$$

If we set up a threshold, we can obtain a list of matched points to each landmark on the template. The fifth column of Fig. (7) shows those matched points.

Using learned similarity measure on the patch pairs, we obtain a list of matched ones, some of which however should not correspond to each other. This is because locally they do look similar to the positive training pairs. Therefore, we carry out the process of estimating the transformation parameters also in a coarse-to-fine manner. We adopt the RANSAC algorithm [9] to find the affine transformation parameters for point correspondences. The algorithm randomly samples the points to estimate the parameters by voting, and thus is able to remove the outliers. Since the RANSAC algorithm has been widely used in the area of structure from motion and matching, we refer to [9] for details of the implementation. Fig. (7.f) shows refined matched points by the RANSAC algorithm, which are much cleaner than those shown in Fig. (7.e). Based on the refined matched pairs, we estimate the transformation parameters using a least-square fitting.

Finally, the overall transformation is a combination of global affine with local non-rigid thin plate splines transformation, as used in [2].

$$u(x, y) = a_1 + a_x x + a_y y + \sum_{i=1}^{n} w_i^u U(||(x_i, y_i) - (x, y)||),$$

$$v(x, y) = b_1 + b_x x + b_y y + \sum_{i=1}^{n} w_i^v U(||(x_i, y_i) - (x, y)||),$$

$$\sum_{i=1}^{n} w_i = 0 \ \ and \ \ \sum_{i=1}^{n} w_i x_i = \sum_{i=1}^{n} w_i y_i.$$

where $U(r) = r^2 \log r^2$ is the kernel function, and $(a_1, b_1, a_x, b_x, a_y, b_y)$ are the parameters for the affine transformation.

2.4. Outline of the Algorithm

The outline of the algorithm is given below:

Training: Given a set of unregistered images and template images: (1) Train a coarse level similarity measure based on roughly aligned (cropped, rotated, and re-scaled) image with the template as shown in Fig. (4). (2) Train a BEL edge detector on the roughly aligned brain image to detect salient edges. (3) Learn similarity measure on the manually annotated image patch pairs, as shown in Fig. (5).

Registration: Given an source image and a template image to which the source image should register: (1) Search for the sub-image which best matches the template using the similarity learned at the coarse level. (2) Perform a rough alignment by cropping, rotating, and re-scaling the sub-image to the same orientation and size of the template. (3) Use learned BEL edge detector to extract salient edges. (4) Search for the matched points on the extracted edges to the landmarks on the template. (5) Use the RANSAC algorithm to remove the outliers. (6) Use least-square fitting to estimate the affine transformation parameters. (7) Use thin plate splines transformation to obtain the final detailed registration.

3. Experiments

We apply the proposed algorithm on three applications: (1) mouse brain image registration of different modalities, (2) human brain image registration of T1 and T2, and (3) face expression matching.

3.1. Mouse brain registration

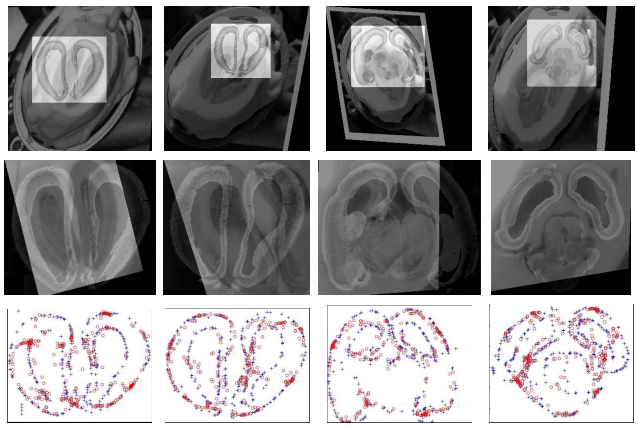


Figure 6. Results by mutual information based registration algorithm [25] and shape context [2] on the extracted salient structures. The first row shows the result if we register the entire source image to the template using [25]. The second displays transformed images if we use the detected brain images. The third row shows the registered result by shape context on the extracted landmarks of template (blue dots) and of source image (red dots).

We first show the results on a problem of mouse brain registration. In this experiment, we have 25 unregistered/template image pairs of mouse brains. We used 12 pairs for training and the remaining 13 for testing. In this dataset, the unregistered image and template image have different modalities and the template image is noisy with missing parts. Each image represents a slice in a sequence of maps. There are multiple templates and a template can match to several images. Having a single template is a special case and it is actually simpler than this situation. It takes about half day to manually annotate the correspondences and the training stage requires about 5 hours on a PC with 2.4 GHz CPU and 1.0GB memory. Standard code optimization can reduce the training time significantly.

Fig. 7 shows some of our results. Visually, the coarsely registered results are already quite close to the template images. To see the details clearly, we overlapped the registered image and the template image as shown in Fig. (7.c). Fig. (7.d) is a detailed registered version of the brain in Fig. (7.c). Fig. (7.e) shows the matched pairs using learned fine level similarity measure and Fig. (7.f) are the matches after the RANSAC algorithm.

To compare our algorithm to some state-of-the-art registration algorithms, we use the mutual information based algorithm [25] in the widely used ITK package. We first perform the registration on the original image and the first row of Fig. (6) shows the overlayed registered image to the template, which has a very big offset to the right solution. To further compare the learned fine level similarity with the mutual information measure, we use the same program, but on the roughly aligned image by the first step of our algorithm. The second row of Fig. (6) displays the overlayed images, which are still not quite satisfactory. Since the salient structure extraction is also an important step in our algorithm, we applied the shape context algorithm [2] which uses the shape information solely on the extracted structures. The third row of Fig. (6) shows the matched points. The results are not as good as those given by the learning based method, in that our learned similarity measure exploits a much richer feature pool than the mere structure information employed by shape context. We also tested the SIFT detector based registration algorithm [15] which completely miss the right orientation of the template due to the reasons mentioned before: SIFT detector can not consistently detect the salient interest points for the medical images we are dealing with. Table (2) shows the error measure in terms of the average distance from the transformed salient points to the correct landmarks. We compare the mean errors by MI based algorithm on the roughly aligned, shape context on the extracted structures, and our overall proposed method. The errors for MI based algorithm on the original image and SIFT based registration algorithm [15] are much bigger and thus, are not included in the comparison.

MI based (on roughly aligned)	shape context	proposed algorithm
43.6	13.6	7.3

Table 2. Average distance of the registered image to the template.

3.2. T1-T2 Brain Image Registration

We also tested our algorithm on a set of T1-T2 image pairs. We used 3 image pairs as training data and 2 as testing data for illustration. Fig. (8) shows the results and our method performs equally well as [25] and [15].

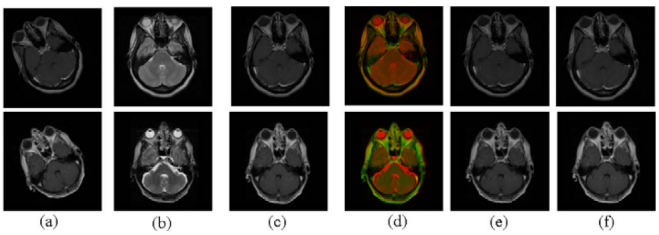


Figure 8. Results on the T1-T2 brain image pairs. (a) shows the original images. (b) are the template images. (c) are the registered image by our algorithm and (d) show the overlay. (e) and (f) are the results by [25] and [15] respectively.

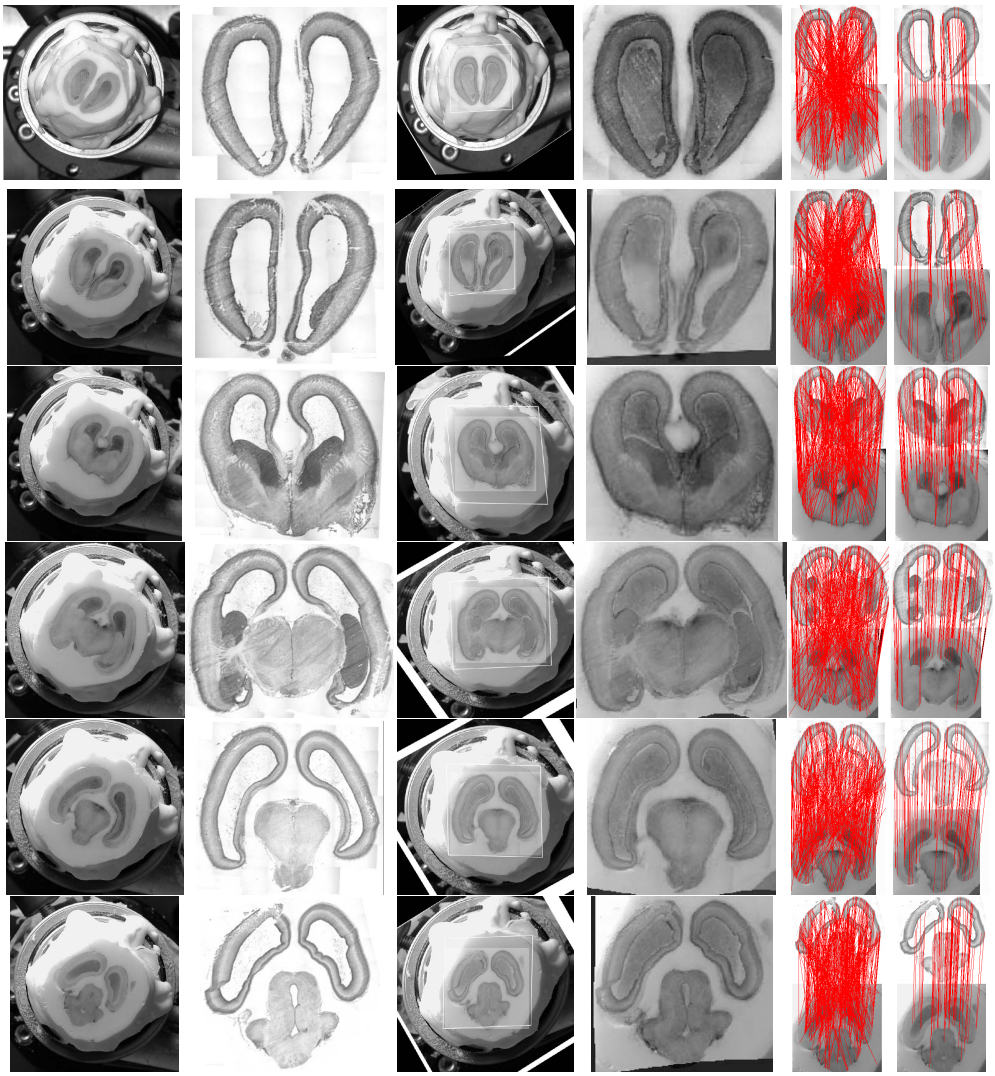


Figure 7. Registration results on the 2D mouse brain images. (a) shows the original image. (b) displays the template image. (c) gives the coarsely registered image with the template (affine model). (d) is a detailed registered version (thin plate splines model) of (c). (e) shows the matched pairs based on the learned similarity measure. (f) illustrates the corresponding patches after RANSAC (affine model).

It is worth to mention that our framework is not constrained by 2D and it automatically fuses shape and texture features (either 2D or 3D). We plan to include 3D registration in the future to further illustrate the advantage of this algorithm.

3.3. Face registration

We also applied our algorithm on registering two faces of the same person but with different expressions. A subset of FERET database [16], which includes 25 individuals, was used to demonstrate our learning based similarity measure. We used 15 image pairs for training and the remaining 10 for testing. Since frontal faces can be reliably detected in most cases [24], we can use a face detection algorithm directly. Also, we use the edges extracted by Canny directly in this case. Therefore, the major task in this application is to learn the similarity measure between the local image patches. In the stage of registration, we run Canny edge detector on two frontal faces of the same person. Again, 500 points are sampled from the edge map of the template, which are used to search for their corresponding points on the source image. We note that our similarity measure (after thresholding) is sufficient to give accurate correspondence results in this case. Then a thin plate splines transformation is obtained from the correspondences and the source image is finally morphed to the template image. Fig. (9) shows some results. There is a rich body of work in the face expression analysis domain [22, 27, 7] but our main purpose of this experiment is to show the effectiveness of the proposed learning based method in term of automatic registration.

(a) source (b) template (c) registered (d)

Figure 9. Registering faces of different expressions. (a) shows the source image and (b) is the template image. (c) displays the registered image from the source to the template. (d) shows the matched points from our learning based similarity measure (after thresholding).

4. Conclusions and Future Work

In this paper, we have presented a new image registration method through coarse-to-fine learning. Unlike the standard information-theoretic based methods, which predefine a similarity measure, our similarity measure is learned from a set of manually annotated images. This provides more flexibility in registering images in different modalities, which is hard for existing approaches.

The method was applied to register a set of mouse brains and corresponding templates. This is a challenging registration problem because the two types of images are from different modalities and have different intensity properties. We showed that our method gives good registration results while the other methods performed poorly. On the easier task of registering T1-T2 image pairs, our method performs as well as the alternatives. We also show its application on registering faces with different expressions.

The disadvantages of our algorithm are that there are a couple of training processes involved and it requires training for each different types of data. In the future, we will use more training images and test our algorithm on a larger testing dataset. Also, a more thorough comparison with the existing registration algorithms will be conducted. In this paper, we only illustrated the system on 2D images. Our learning-based approach is general, and we are planning to apply it to registering 3D images as well.

Acknowledgments This work is funded by NIH Grant U54 RR021813 entitled Center for Computational Biology. We thank Alan Yuille for valuable discussions.

References

[1] B. Babenko, P. Dollar, and S. Belongie, "Task Specific Local Region Matching", *Proc. of ICCV*, Brazil, 2007. 2

[2] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching and Object Recognition Using Shape Contexts", *IEEE Trans. on PAMI*, vol. 24, no. 24, 2002. 2, 4, 5

[3] T. Butz, and J. P. Thiran, "Affine Registration with Feature Space Mutual Information", *Proc. of MICCAI* 2001, pp. 549-556. 1

[4] T. Caetano, L. Cheng, Q. Le, and A. Smola, "Learning Graph Matching", *Proc. of ICCV*, Brazil, 2007. 2

[5] J. F. Canny, "A Computational Approach to Edge Detection," *IEEE Trans. PAMI*, vol. 8, no. 6, pp. 679-698, Nov. 1986. 2

[6] A. C. S. Chung, W. M. Wells III, A. Norbash, and W. E. L. Grimson, "Multi-modal image registration by minimising Kullback-Leibler distance", In *MICCAI* 2002, pp. 525C532. 1

[7] T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active Shap Models-Their Training and Application", *CVIU*, vol. 61, no. 1, 1995. 6

[8] P. Dollár, Z. Tu, and S. Belongie, "Supervised learning of edges and object boundaries", In *Proc. of CVPR*, 2005. 2, 3

[9] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", *Comm. of the ACM*, 24: 381–395, June 1981. 2, 4

[10] C. Guetter, C. Xu, F. Sauer, and J. Hornegger, "Learning Based Non-rigid Multi-modal Image Registration Using Kullback-Leibler Divergence", Proc. of MICCAI, 2005. 1

[11] Y. He, A. B. Hamza, and H. Krim, "A generalized divergence measure for robust image registration", *IEEE Trans. on Signal Processing*, vol. 51, no. 5, May 2003. 1

[12] D. G. Lowe, "Distinctive Image Features From Scale-invariant Keypoints", *Int'l Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004. 1, 2

[13] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality Image Registration by Maximization of Mutual Information", *IEEE Trans. on Medical Imaging*, vol. 16, no. 2, pp. 187-198, Feb. 1997.

[14] J. B. Maintz and M. A. Viergever, "A Survey of Medical Image Registration", *Medical Image Analysis*, vol. 2, no. 1, pp. 1-36, 1998. 1

[15] M. Moradi, P. Abolmaesoumi and P. Mousavi, "Deformable Registration Using Scale Space Keypoints",*Medical Imaging 2006: Image Processing. Proceedings of the SPIE*, vol. 6144, pp. 791-798, 2006. 2, 5

[16] P. Phillips, H. Moon, S. Rizvi, and P. Rauss, "The FERET Evaluation Methodology for Face-Recognition Algorithms", *IEEE Trans. on PAMI*, vol. 22, no. 10, pp. 1090-1104, 2000. 6

[17] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Image registration by maximization of combined mutual information and gradient information", *IEEE Trans. on Medical Imaging*, vol. 19, no. 8, pp. 809-814, Aug. 2000. 1, 2

[18] J. P. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based Registration of Medical Images: A Survey", *IEEE Trans. on Medical Imaging*, vol.22, no.8, pp. 986-1004, Aug. 2003. 1

[19] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "f-Information measures in medical image registration", *IEEE Trans. on Medical Imaging*, vol. 23, no. 12, pp. 1506-1518, Dec. 2004. 1

[20] A. Roche, G. Malandain, and N. Ayache, "Unifying Maximum Likelihood Approaches in Medical Image Registration", *Int'l J. of Ima. Sys. and Tech.*, 11(1):71-80, 2000. 2

[21] L. Teverovskiy, and Y. Liu, "Learning-based Neuroimage Registration", tech. report CMU-RI-TR-04-59, Robotics Institute, Carnegie Mellon University, October, 2004 2

[22] Y.L. Tian, T. Kanade, T., J.F. Cohn, "Recognizing Action Units for Facial Expression Analysis", *IEEE PAMI*, Vol. 23, No. 2, pp.97-116, 2001. 6

[23] Z. Tu, "Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering", *Proc. of ICCV*, 2005. 3

[24] P. Viola and M. Jones, "Robust Real-Time Face Detection", *Int'l J. of Com. Vis.*, vol. 57, no. 2, 2004. 3, 6

[25] W. M. Wells III, P. Viola, H. Atsumi, S. Nakajima, R. Kikinis. "Multi-modal Volume Registration by Maximization of Mutual Information", *Medical Image Analysis*, vol. 1, no. 1, pp. 35-52, 1996. 1, 2, 5

[26] G. Wu, F. Qi, and D. Shen, "Learning-Based Deformable Registration of MR Brain Images", *IEEE Trans. Medical Imaging*, 25(9):1145-1157, Sept. 2006. 2

[27] J. Xiao, S. Baker, I. Matthews, and T. Kanade, "Real-Time Combined 2D+3D Active Appearance Models", *Proc. of CVPR*, pp. 535-542., June, 2004. 6

[28] J. Zhang, and A. Rangarajan, "A unified feature-based registration method for multimodality images", In *Proc. of ISBI*, 2004, pp. 724-727. 1

[29] S. K. Zhou, J. Shao, B. Georgescu, and D. Comaniciu, "BoostMotion: Boosting a Discriminative Similarity Function for Motion Estimation", *Proc. of CVPR*, New York, NY, 2006 2

[30] Y. Zhu, "Volume Image Registration by Cross-Entropy Optimization", *IEEE Trans. Med. Imaging*, vol. 21, no. 2, pp. 174-180, 2002. 1