HHS Public Access

Author manuscript

Med Image Anal. Author manuscript; available in PMC 2016 December 01.

Published in final edited form as:

Med Image Anal. 2015 December ; 26(1): 256–267. doi:10.1016/j.media.2015.10.001.

Building Dynamic Population Graph for Accurate Correspondence Detection

Shaoyi Du[1,2], **Yanrong Guo**[1], **Gerard Sanroma**[1], **Dong Ni**[3], **Guorong Wu**[1], and **Dinggang Shen**[1,4,*]

[1]Department of Radiology and Biomedical Research Imaging Center, University of North Carolina, Chapel Hill, NC 27599, USA

[2]Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, Shaanxi Province 710049, China

[3]The Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Department of Biomedical Engineering, Shenzhen University, Shenzhen 518060, China

[4]Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, Republic of Korea

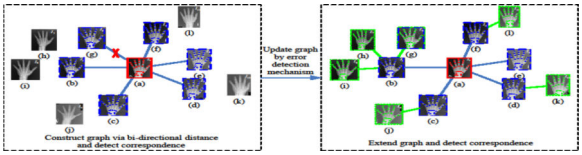
Abstract

In medical imaging studies, there is an increasing trend for discovering the intrinsic anatomical difference across individual subjects in a dataset, such as hand images for skeletal bone age estimation. Pair-wise matching is often used to detect correspondences between each individual subject and a pre-selected model image with manually-placed landmarks. However, the large anatomical variability across individual subjects can easily compromise such pair-wise matching step. In this paper, we present a new framework to simultaneously detect correspondences among a population of individual subjects, by propagating all manually-placed landmarks from a small set of model images through a dynamically constructed image graph. Specifically, we first establish graph links between models and individual subjects according to pair-wise shape similarity (called as *forward step*). Next, we detect correspondences for the individual subjects with direct links to any of model images, which is achieved by a new multi-model correspondence detection approach based on our recently-published sparse point matching method. To correct those inaccurate correspondences, we further apply an error detection mechanism to automatically detect wrong correspondences and then update the image graph accordingly (called as *backward step*). After that, all subject images with detected correspondences are included into the set of model images, and the above two steps of *graph expansion* and *error correction* are repeated until accurate correspondences for all subject images are established. Evaluations on real hand X-ray images demonstrate that our proposed method using a dynamic graph construction approach can

*Corresponding author. dgshen@med.unc.edu.

achieve much higher accuracy and robustness, when compared with the state-of-the-art pair-wise correspondence detection methods as well as a similar method but using static population graph.

Graphical abstract



Keywords

Correspondence detection; dynamic population graph; pair-wise matching; multi-models

1. Introduction

Correspondence detection (Shen et al., 1999) is a fundamental problem in medical image analysis with a wide range of applications, such as skeletal bone age estimation (BAE) (Martin-Fernandez et al., 2003; Thangam et al., 2012), medical image registration (Xue et al., 2006; Yang et al., 2008; Heimann and Meinzer, 2009; Tang et al., 2009; Wu et al., 2011), and organ detection and segmentation (Zhan et al., 2008; Shi et al., 2010; Zheng et al., 2010). For the case of computational anatomy, robust correspondence detection is a key step to allow the quantitative measurement of anatomical difference across individuals. However, it is still challenging for accurate correspondence detection, especially between the subjects with very large anatomical differences.

Given a set of landmarks in the model image (delineated either manually (Styner et al., 2003; Murphy et al., 2008; Castillo et al., 2009) or automatically (Criminisi et al., 2010; Ou et al., 2010; Paganelli et al., 2012)), the goal of correspondence detection is to determine the corresponding location of each landmark in the subject images. The existing pairwise correspondence detection methods can be roughly classified into two categories, according to the correspondence relationship and the expression of geometric transformation. The **first** category of methods determines point-to-point correspondences and geometric transformation simultaneously, such as iterative closest point algorithm (ICP) (Besl and McKay, 1992), thin plate spline–robust point match (TPS-RPM) (Chui and Rangarajan, 2003), etc. These methods mostly focus on point-to-point correspondences and model the whole transformation between the two point sets explicitly. The **second** category of methods takes into account pair-to-pair relationship. They mostly represent the geometric transformation implicitly, but focus on the correspondence searching, such as using graph-based methods. Graph-based methods have shown their potential in medical imaging applications by modeling *not only* the point-to-point correspondence (Zhang and Lu, 2004; Jiang et al., 2007) *but also* the pair-to-pair consistency (Leordeanu and Hebert, 2005; Zass and Shashua, 2008). Many scholars have developed various graph-based methods (Duchenne et al., 2011; Sanromà et al., 2012; Sanromà et al., 2012). Guo *et al*. (Guo et al., 2013) introduced a sparsity constraint into the conventional graph matching method for more accurate and robust correspondence detection.

Apart from these pairwise correspondence detection methods, some scholars also studied various groupwise models for correspondence detection. Donner *et al*. (Donner et al., 2009) proposed a MRF model for groupwise model learning. Adeshina and Cootes (Adeshina and Cootes, 2010a, b) and Zhang *et al*. (Zhang et al., 2012; Zhang and Cootes, 2012) proposed parts-based methods for initializing groupwise registration. Although these matching methods have been successfully used for correspondence detection, they still have limitations in detecting correspondences between images with large anatomical difference, such as in medical studies that often involve a large amount of images with variable anatomical shapes.

In the past, some methods have been proposed for addressing the above challenges. For example, in the case of image segmentation, Wolza *et al*. (Wolza et al., 2010) proposed the learning embeddings for atlas propagation (LEAP), which uses an intensity-based similarity to build a population graph for linking similar images in the neighboring nodes first. Then, they decomposed the task of segmenting all subjects into a series of easy segmentation tasks, i.e., 1) segmenting images with similar anatomical structures and 2) propagating segmentations forward to other nearby images, thus avoiding segmenting images with large structural discrepancies. However, the population graph built by LEAP does not prevent the propagation errors, since it is fixed throughout the whole segmentation procedure. Thus, this will lead to possible accumulation of segmentation errors due to the use of spurious graph links, since no mechanism is used to re-evaluate the population graph and then improve the segmentation results. More critically, the similarity measure used for building the population graph is simply computed with image intensity difference, which unfortunately has no direct relation with the shape of anatomical structure.

To address all the above issues, we propose using *dynamic population graph* for detecting correspondences among a set of individual images. Our method includes two steps, *forward step* and *backward step*, which are repeated until completion. In the ***forward step***, we investigate shape discrepancies between any pair of images, instead of simply using the traditional image intensity similarity. Then, we build a *population graph* with the *nodes* representing (model or subject) images and the *edges* linking pairs of *model-subject* images with small shape differences. Note that we here establish correspondences only for the subject images with direct links to any of the model images. For the case that one subject image is linked to multiple model images, we propose a *multiple model selection and fusion method* for correspondence detection. In the ***backward step***, we update the population graph according to the current correspondences just established, after removing correspondence detection errors in the *forward step* by using our proposed *error detection mechanism* which is built based on the shape similarity and the spatial relationship of all landmarks. As the result of this *backward step*, the inaccurate correspondence detection results are removed and also the set of model images is updated by including the subject images with their newly established correspondences. We repeat the above two steps until correspondences for all subject images are determined.

We have comprehensively evaluated our proposed correspondence detection method on the real hand X-ray images. Compared with the state-of-the-art methods, our correspondence

detection method (based on the *dynamic population graph*) achieves significant improvement, in terms of both accuracy and robustness.

The rest of this paper is organized as follows. In Section 2, we present our proposed method. In Section 3, we compare our method with the state-of-the-art methods. Finally, we draw the conclusion in Section 4.

2. Method

2.1. Overview of Our Proposed Method

Conventional methods often detect correspondences independently between each subject image and the pre-defined model image. Thus, their performances are limited when subject and model images have large structural discrepancies. To address this issue, we propose a novel correspondence detection method by dynamically updating the *population graph* during the iterative correspondence detection procedure. The framework of our proposed method is depicted in Fig. 1.

Specifically, each image is regarded as a node in the population graph. In the ***forward step***, we first build a population graph by linking between model images and their most similar subject images. Here, we use a novel shape similarity measurement, instead of simple image intensity similarity, as described in Section 2.2 for measuring the similarity of two images. Then, we determine correspondences for each subject image that is linked to any of model images by using the model image(s) linked to this subject image, as detailed in Section 2.3.1. For the case that a subject is linked to multiple model images, our proposed *multiple model selection and fusion method* based on the shape similarity are used, as described in Section 2.3.2. In the ***backward step***, we first detect wrong correspondences and then dynamically update the *population graph* by removing those graph edges linked to the subject images with wrong correspondences, as described in Section 2.4. In this way, only the subject images with reasonable correspondences will be included in the set of model images, thus avoiding the propagation of errors to connected subjects (under correspondence detection). By repeating the above *forward step* and *backward step*, correspondences on the remaining subject images can be detected.

Fig. 2 and Fig. 3 illustrate our approach in an intuitive way. Fig. 2 shows an initial population graph after linking the model image (a) with the most similar subject images (b)-(g), based on the shape similarity. Since these subject images are similar to the model image, we can determine their correspondences more accurately, compared to other subject images. By using the error detection mechanism, we detect the wrong correspondence results for subject image (g), and thus remove its respective edge and further update the population graph. In this way, we can prevent the propagation of correspondence detection errors to the rest of subject images. Next, we include all subject images (b)-(f) with reasonable correspondences into the set of model images. Fig. 3 shows a new population graph, which is built by linking the remaining subject images with the newly added models (i.e., subject images (b)-(f)). Then, we can perform correspondence detection for all the remaining subject images (g)-(l). It is worth noting that our described progressive correspondence

detection method can be straightforwardly extended to the case of multiple initial models, instead of only a single model as used for the example given in Fig. 2 and Fig. 3.

2.2. Population Graph Construction Based on the Shape Boundary Distance

Given a large number of subject images, it is inevitable to have some subject images with large anatomical dissimilarity *with respect to* the model images. To avoid the difficulty of detecting correspondences for those subject images in the initial stage, we propose to learn the distribution of all (model and subject) images in the manifold, so that we can identify, for each model image, its nearby subject images (with the most similar anatomical structures) in the manifold.

Specifically, we build a population graph, with each node representing an image and each edge linking two most similar images. For building a reasonable population graph, it is important to accurately measure image similarity. The conventional methods often use image intensity similarity to measure the similarity of two images (Jia et al., 2012). Recall that the goal in our application is to detect accurate landmark correspondences between the model image(s) and the subject images, where landmarks are often located on image boundaries. Hence, here we will use the shape boundary difference as a similarity measurement, instead of image intensity similarity.

Specifically, we compute the similarity between two images with the following 3 steps: (1) affinely register the two images by using Elastix (Klein et al., 2010; Shamonin et al., 2014), or Flirt in FSL toolbox (Smith et al., 2004; Woolrich et al., 2009; Jenkinson et al., 2012); (2) obtain the boundary point set in each image by applying the Canny edge detector; and (3) compute the bi-directional distance (Du et al., 2011) between the two point sets as explained below.

For the two affinely-aligned images I_i and I_j($i \neq j$), we apply the Canny edge detector to extract their own boundary point sets, as denoted by $S_i=\{s_{i,k}|k = 1,\ldots,L_i\}$ and $S_j = \{S_{j,l}|l = 1, \ldots,L_j\}$ for images I_i and I_j, respectively. Here, L_i and L_j are the numbers of boundary points in S_i and S_j, respectively. The bi-directional distance between images I_i and I_j is then defined as:

$$d(I_i, I_j)=\frac{1}{L_i+L_j}\sum_{k=1}^{L_i} d(s_{i,k}, S_j)+\frac{1}{L_i+L_j}\sum_{l=1}^{L_j} d(S_i, s_{j,l}) \quad (1)$$

where the first and second terms represent the point set distances from S_i to S_j and S_j to S_i respectively. In Eq. (1), $d(s_{i,k},S_j)$ denotes the shortest Euclidian distance between the point $s_{i,k}$ and the point set S_j, which is defined as $d(s_{i,k}, S_j)=\min_{c(k)\in\{1,\ldots,L_j\}} \|s_{i,k} - S_{j,c(k)}\|_2^2$. The distance $d(S_i,s_{j,l})$ can be defined similarly. Therefore, the computational complexity between every two images is $O(L_iL_j)$.

Fig. 4 shows an example demonstrating the superiority of using the shape boundary difference over the simple image intensity difference. As we can see from this figure, the accurate correspondences between model image (a) and subject image (b) can be established, while it fails to establish correspondences for the subject image (c) by the model

image (a). But if using the simple image intensity difference (even after histogram matching and normalization), subject image (c) is more similar to the model image (a) than the subject image (b). Obviously, this result does not reflect the underlying anatomical similarity, as indicated by the correspondence detection results.

As a result, we compute the bi-directional shape boundary distances between all current model images and subject images to estimate their respective similarities/differences. In this way, we can build a population graph by linking all pairs of *subject-model* images if their respective distances are below a certain threshold r_0. Specifically, after computing the distance $d(I_i, I_j)$ between every pair of images I_i and I_j, a similarity matrix can be obtained. Then, based on this similarity matrix, we can compute the minimum value among the maximum values along columns of this similarity matrix, i.e., $r_m = \min_j \max_i d(I_i, I_J)$, which is the minimum threshold to connect all other images for a certain image. Next, we use this value as a reference, and further define a searching range around r_m to finally determine our threshold r_0, according to the experimental results on the training samples. Fig. 2 and Fig. 3 show such two graphs built in two iterations of our approach.

2.3. Correspondence Detection Based on Sparse Point Matching

2.3.1. Sparse Point Matching—In the scenario of graph matching (Cour et al., 2006), multiple correspondences are allowed and each potential correspondence is considered as the graph node. Then, graph edge is used to measure the pointwise and pairwise agreement between possible correspondences, which includes the similarity of the geometric relationships and the local descriptors. After encoding the graph into an affinity matrix, the task of correspondence detection becomes the optimization problem of finding the set of one-to-one correspondences that produce the maximal pairwise agreement in the affinity matrix. The graph matching methods make full use of different image information including point-to-point and pair-to-pair correspondences, which have been intensively studied.

As mentioned, we will propagate landmarks from model images to subject images through the determination of correspondences between any pair of model image and subject image that are linked together in the *population graph* (as detailed above in Section 2.2). Note that, in the model image, we already have a set of manually-placed landmarks, e.g., 30 landmarks as shown in Fig. 5. For the subject image, we need to first determine the candidates of feature points, which will be used for locating landmarks for the subject image (since all landmarks in the model images are manually placed on the joints (edge) of hand bones). To do this, **1)** we will first detect edge points in the subject image by Canny edge detector. **2)** To reduce the unnecessarily large number of feature points generated by Canny edge detector, we keep only the edge points that are close to model landmarks (after affine alignment of subject image with the model image) and also have distinctive local appearances. With this operation, we can reduce from > 3000 feature points generated by Canny edge detector to $\sim$ 700 feature points, as demonstrated in Fig. 5(c).

Finally, we determine landmarks for the subject image by the correspondence detection method, as described below. Fig. 5 shows an example of detecting correspondences on a hand X-ray image.

We use the sparse point matching method in (Guo et al., 2013) for detecting correspondences between the model image and the subject image. Assume that we have N_L landmarks (i.e., equal to 30 in Fig. 5) on the model image, which can be represented by $M=\{m_i|i= 1,\ldots,N_L\}$. Similarly, assume that we have N_F feature points on the subject image, denoted by $F = \{f_i|j = 1,\ldots,N_F\}$. We represent correspondences between these two sets of landmarks/points by the *assignment matrix* $\mathbf{X}=[x_{i,j}]_{N_L\times N_F}$, where $x_{i,j} \in \{0,1\}$ indicates whether the *i*-th model landmark corresponds to the *j*-th subject feature point or not. To obtain one-to-one correspondences, we impose the constraint that each landmark m_i must be matched to one feature point in *F*, i.e., $\sum_{j=1}^{N_F} x_{i,j}=1$. By reshaping the assignment matrix **X** into a N_LN_F dimensional *assignment vector* $\vec{x}$, we can reformulate the one-to-one constraint in an affine way as follows: $\|\mathbf{C}\vec{x} - 1^T\|_2^2 \to 0$, where **C** is a matrix used to ensure the sum of each row in **X** close to 1 (Maciel and Costeira, 2003). To solve the assignment vector $\vec{x}$, we construct an $N_LN_F \times N_LN_F$ affinity matrix **M** for describing the compatibility between all pairs of correspondences between the model image and the subject image, whose computational complexity is $O((N_LN_F)^2)$. We further take into account the coherence in distances, angles, and local appearance, to compute such compatibilities. Finally, the sparse point matching method formulates correspondence detection as the following quadratic optimization problem:

$$\max_{\vec{x}} \vec{x}^T\mathbf{M}\vec{x} - \gamma\|\vec{x}\|_1 - \lambda\|\mathbf{C}\vec{x} - 1^T\|_2^2 \; s.t. \vec{x} \in [0,1]^{N_LN_F} \quad (2)$$

where parameters γ and λ control sparsity and one-to-one matching constraint, respectively. Note that the sparsity constraint helps suppress ambiguous correspondences in the assignment vector $\vec{x}$. Moreover, the maximization of the objective function (2) is a quadratic programming problem, whose computational complexity is $O((N_LN_F)^{3.2})$(Du et al., 2010).

2.3.2. Multi-model Correspondence Detection—In the above, we have provided a method for detecting correspondences for the subject image, if it is connected with one model image. In many situations, for a subject image, it can be linked with multiple model images. In such case, we detect correspondence for this subject image with each model image separately, and then apply our multi-model selection method proposed below to fuse the most confident correspondence candidates for achieving the final accurate result.

In our multi-model selection method, we assume that the overall distribution of the detected correspondences in the subject image should be similar to the global shape of landmarks in each model image. With this assumption, we can remove incorrect landmarks placed in the subject image, as detailed next. **First**, we use an affine transform to map all landmarks of each model image onto the subject image space, to best match with the detected correspondences in the subject image. **Then**, we keep only those detected correspondences as candidates for fusion, if their distances to the mapped landmarks of the model image are small.

More formally, assume that we have *K* landmark sets from *K* model images, $M_i = \{m_{i,n}|n = 1,\ldots,N_L\}$, $(i = 1,\ldots,K)$, each with the same N_L landmarks. The detected landmarks for the

subject image by each of K model images can be denoted by $D_i = \{d_{i,n} | n = 1, \ldots, N_L\}$, ($i = 1, \ldots, K$). As mentioned in the last paragraph, we select the candidate landmarks that are only close to the affine aligned counterparts of the model image. Here, the affine transformation to best match the detected correspondences D_i in the subject image and the mapped landmarks M_i in the model image is estimated as follows:

$$\min_{\mathbf{A}_i, t_i} \sum_{n=1}^{N_L} \|\mathbf{A}_i m_{i,n} + t_i - d_{i,n}\|_2^2 \; s.t. \det(\mathbf{A}_i) \neq 0 \quad (3)$$

where $\mathbf{A}_i$ is the 2×2 matrix (including rotation, scaling and shearing), and t_i is the translation vector. The objective function in Eq. (3) can be easily solved according to (Du et al., 2008), as given below:

$$\begin{cases} t_i = \frac{1}{N_L} \sum_{n=1}^{N_L} (d_{i,n} - \mathbf{A}_i m_{i,n}) \\ \mathbf{A}_i = (\sum_{n=1}^{N_L} u_{i,n} v_{i,n}^T)(\sum_{n=1}^{N_L} u_{i,n} v_{i,n}^T)^{-1} \end{cases} \quad (4)$$

where $u_{i,n} = m_{i,n} - \frac{1}{N_L} \sum_{n=1}^{N_L} m_{i,n}$ and $v_{i,n} = d_{i,n} - \frac{1}{N_L} \sum_{n=1}^{N_L} d_{i,n}$.

Using Eq. (4), we can obtain the affine transformation between subject image and each model image, denoted as $\{\mathbf{A}_i, t_i\}$. Then, we can compute the n-th transformed landmark of the i-th model as $\mathbf{A}_i m_{i,n} + t_i$, and further compute its distance to the corresponding detected landmark of subject image as $\| \mathbf{A}i m_{i,n} + t_i - d_{i,n} \|_2$. Since the shapes of both the model image and the subject image are often similar, the distance $\| \mathbf{A}_i m_{i,n} + t_i - d_{i,n} \|_2$ should be small as well. Thus, we keep this detected landmark as a candidate landmark for the subject image, only if its computed distance is less than a certain threshold r_1, which depends on both image resolution and the distribution of landmarks. Specifically, the threshold r_1 is used to search the candidate landmarks according to the global shape similarity, which is often set to a relatively large value since the global affine transformations usually does not accurately register the point sets. The final threshold r_1 can be determined based on the statistics estimated from images with similar hand shapes.

Afterwards, we can combine all detected landmark candidates from all model images that are linked to the subject image under study. Specifically, for each landmark, we can use a mean-shift method to estimate the actual location from all landmark candidates. In particular, we can first compute the mean of all landmark candidates and then remove the one with the largest distance to the calculated mean location. Next, we repeat the same procedure until only two corresponding locations remain. Finally, the estimated landmark location for the subject image can be computed as the mean of the two remaining landmark candidates.

Note that our above multi-model correspondence detection method can help remove *gross errors*, not the small errors in the local parts, because of using global affine transformation. However, these remaining small errors will be corrected by our *error detection mechanism* as proposed in the next section (Section 2.4).

2.4. Graph Update by Error Detection Mechanism

As described above, we can detect correspondences for each subject image if it is connected to one or more model image(s) in the previously built population graph. Then, the subject images with reasonably detected correspondences can be included into the set of model images, and further used to expand the previous population graph in the next iteration, by computing the links from the newly added model images to the remaining subject images. By iteratively repeating this procedure, the correspondences for all subject images can be eventually detected. Since the population graph connects the model images only to the subject images with similar global shapes, we have high confidence on the detected correspondences for the linked subject images in the population graph. However, since our above-proposed approach ignores local similarity between model image and subject image, some local errors may still exist, as shown in Fig. 6(b) where the model image (a) and subject image (b) are globally similar but have some discrepancies in the local regions, i.e., in thumb and forefinger. Such errors could be easily propagated through the population graph, if not corrected immediately. Accordingly, in the following, we propose an error detection mechanism for updating the population graph, thus avoiding the propagation of even those small errors.

Specifically, we divide each (model or subject) image into several different local regions and then check their similarities in the model and subject images, respectively. In particular, the hand image is divided here into six local regions, consisting of five fingers and palm. Then, in each local region, we can use the landmarks of each model image to examine whether the detected corresponding landmarks in the subject image are correct or not. As there often exist different numbers of landmarks in each local region, we need to use two different methods for evaluating the respective detection results, as detailed below. Note that, for correcting the local correspondences in the subject image, we often need to compute affine transformation from at least 3 pairs of correspondences established between model and subject images, and then we can use the computed affine transformation to possibly verify other detected correspondences in the local region of subject image. By repeating this step, we can possibly verify all correspondences in the local region; of course, with more pairs of correspondences for computing affine transformation, the computed transformation can be more robust and accurate and then the verification results can be better. Since there are different numbers of landmarks in each local region, we adopt two different strategies for evaluating the detection results:

- In the **first** case, when the number of pairs of correspondences is more than five, such as for the palm region, we can use Eq. (5) to compute the affine transformation between palms of the model and subject images, and then transform all model landmarks onto the subject image for landmark correction. Specifically, we consider a detected landmark as correct if its distance to the corresponding transformed landmark of the model image is less than a certain threshold r_2. Otherwise, we consider it as wrong detection result.
- In the **second** case, if the number of pairs of correspondence is less than or equal to five, such as for each of five finger regions, the computed affine transformation could be not good enough to validate the correspondence detection results. Also,

for this case, the incorrectly detected landmarks are often clustered together. To identify this situation, we will compute the distance between any two detected landmarks in each finger region. If we find the distance is below a certain threshold r_3, we will consider these two detected landmarks are clustered together and thus should regard them as false detections.

Note that, for both above cases, their respective thresholds r_2 and r_3 could be determined according the image resolution and also the landmark distribution of each region. Specifically, the transformation of the palm region is always smaller than that of the finger region, so r_2 is always smaller than r_3. To determine these two parameters, we perform the experiments based on the training samples by varying these parameters within a suitable range and finally select their optimal values for the testing samples. Moreover, if any correspondence/landmark is determined to be wrong, this subject image will be removed from the population graph, and its correspondence detection will be postponed to the next iterations, after it is included again into the future population graph.

Fig. 7 gives an example demonstrating the idea proposed above, with two iterations. In the first iteration, as shown in the left panel of Fig. 7, landmarks of two subject images (b) and (g) (which are the same subjects as in Figs. 2 and 3) are detected by applying the sparse point matching method and using the model image (a). Note that the detected landmarks are shown in blue points, while model landmarks are shown in red points. Then, the above-proposed error detection mechanism is applied to verify the landmark detection results for the two subject images (b) and (g), and find out that the detection results for subject image (g) are incorrect, while the detection results for subject image (b) are correct. In this way, only the subject image (b) will be included into the set of model images, but the subject image (g) will be removed from the population graph by deleting its link with the model image (a). In the second iteration, as shown in the right panel of Fig. 7, the image (g) is connected to the subject image (b) (now used as a new model image) during the updating of population graph. Then, the landmarks for the subject image (g) can be correctly detected by the newly added model image (b).

2.5. Summary of Our Proposed Method

Our method can be summarized as follows. **(i)** Given the set of model images and the set of subject images, we establish a population graph by linking the model images with the subject images according to their shape similarities, as described in Section 2.2. **(ii)** In the *forward step*, we detect correspondences for the subject images included in the current population graph via our sparse point matching method, as described in Section 2.3. **(iii)** In the *backward step*, correspondence detection results are evaluated by our error detection mechanism, and the subject images with false detection results are removed from the current population graph to prevent their error propagation to other subject images, as described in Section 2.4. Note that only the subject images with reasonable correspondences will be included into the set of model images. The above 3 steps are repeated until all subject images are detected with landmarks, or no further subjects could be connected to the model images. The whole method is also summarized in Algorithm 1.

Algorithm 1

Correspondence detection based on dynamic population graph construction

Input: A set of model images, $\mathbb{M}$ with manually-placed landmarks; and also a set of subject images, $\mathbb{S}$ for correspondence/landmark detection.

Repeat

Forward Step

for each subject image $I_j \in \mathbb{S}$

1. Compute distance $d(I_i,I_j)$ between this subject image I_j and each model image $I_i \in \mathbb{M}$ and then connect this subject image I_j to the model image I_i, if $d(I_i, I_j) < r_0$. In this way, we can build a population graph.
2. Use the correspondence detection method in Section 2.3 to detect correspondences/landmarks for the subject image I_j, if it is connected to the model image(s) in the above population graph.

end for

Backward Step

Evaluate the correspondence detections for each subject image connected to the model image(s) in the population graph, by using the error detection mechanism in Section 2.4. All subject images with reasonable detected correspondences will be included into the set of model images, $\mathbb{M}$ and then deleted from the set of subject images, $\mathbb{S}$

until $|\mathbb{S}| = 0$, or no any subjects could be connected to the model images.

3. Experimental Results

In this section, we evaluate the performance of our correspondence detection method on 50 hand X-ray images (Cao et al., 2003). The image size varies across the image set with a width ranging from 1100 to 1800 pixels and a height ranging from 1800 to 2400 pixels, and the resolution of the image is 0.1 mm per pixel. We compare our method with both the spectral matching with affine constraint (SMAC) (Cour et al., 2006) and the sparse point matching (SPM) (Guo et al., 2013). Moreover, we compare our method with the idea of learning embedding for atlas propagation (LEAP) (Wolza et al., 2010), to demonstrate the advantage of dynamically constructing the population graph (in our method) than simply using the fixed population graph in the whole procedure. In the experiments, we set the maximum number of connections to the model images to be 5 , which is 10% of the total number of images. Then, we set the distance thresholds as follows: r_0=2.5 mm, r_1=12 mm, r_2=8 mm, r_3=10 mm. Also, we found that, when varying r_0, r_1, r_2 and r_3 by 0.1, 1.5, 1 and 1, respectively, the obtained results are similar, indicating the robustness of our method.

3.1. Evaluation with Single Model

In this section, we evaluate the performances of all comparison methods (SMAC, SPM, and our method) in detecting landmarks by using *a single* (randomly selected) image as the model image. Both SMAC and SPM directly detect correspondences between the model image and the remaining 49 subject images, whereas our method detects correspondences for the subject images through a dynamic population graph. Table 1 shows the mean and standard deviation of landmark detection errors between the ground truth and the estimated correspondences by SMAC, SPM and our method, respectively. To avoid any bias introduced by the selection of the model image, we repeat our experiment for 5 times by randomly selecting a model image at each time to detect landmarks for the remaining subject images.

As we can see, our method achieves much higher accuracy than other two compared methods. Moreover, our method obtains a smaller standard deviation, indicating more stability of our method. To further analyze the performances, we compare the maximum error among all the detected landmarks on each subject image by each method. Thus, we can obtain a maximum landmark error for all subjects, and then report in Table 2 the average of maximum landmark errors of all subject images for each method.

As shown in Table 2, our method achieves considerably lower maximum landmark errors than both SMAC and SPM methods, which indicates that our method can achieve considerably more accurate results. Fig. 8 shows a typical correspondence detection result. As we can see from this figure, the detection accuracies of both SMAC and SPM are limited, due to large shape discrepancies between the model image and the subject image. However, our method overcomes this issue, because of using the dynamically-constructed population graph, as well as a progressive correspondence detection strategy. In Fig. 8, our method also shows the correspondence detection results for an *intermediate subject image* (d), which is much closer to the model image (a) (than the subject image (e)) and thus has its correspondence determined first. Then, the detection results of this intermediate subject image are used to detect correspondences for the subject image (d), thus obtaining much better results compared to those by SMAC and SPM in (b) and (c), especially for the highlighted regions.

3.2. Evaluation with Multiple Models

In this section, we further demonstrate the performance of our correspondence detection method by using multiple model images, instead of only one model image. For each selected number of model images, we similarly repeat the experiment 5 times by randomly selecting model images from the dataset. For fair comparison with SPM, we use their multi-model based approach as well (Guo et al., 2013), i.e., first computing correspondences between each subject and all model images and then fusing all results by using a mean-shift technique. Table 3 shows the mean and standard deviation of the landmark detection errors, as well as both maximum error and mean maximum error, for SPM and our method, respectively. Note that, since SMAC has no respective strategy for handling multiple model images, we did not include it for comparison here. Results in Table 3 show that our method achieves *not only* more accurate *but also* more stable correspondence detection results, compared to SPM. Specifically, our method achieves lower maximum errors by the use of the dynamic population graph, which demonstrates that the proposed approach can prevent the incorrect results effectively. One typical example is also shown in Fig. 9, demonstrating more accuracy and robustness of our method than SPM.

To give more visual insight on the performance of these comparison methods, Fig. 10 shows the average results of the above 5 random testing experiments, with respect to the use of different number of model images. Our approach achieves much better performance than SPM.

3.3. Evaluation on Each Anatomical Part

Here, we give more detailed performance comparison for the palm and fingers, separately. First, we randomly select 5 model images, and then use them to detect correspondences for the 45 remaining subject images. There are a total of 19 landmarks on the fingers, and 11 landmarks on the palm. To avoid any bias induced by random selection, we again repeat the experiment 5 times. We calculate the mean error of all subjects for all 5 experiments. Fig. 11 shows the average of all mean errors.

As we can see, our method obtains much better results at the fingers. This is because fingers undergo much larger variability, compared to the palms. In such case, our built population graph can effectively reduce the error, as reflected by the bar plot in Fig. 11. On the other hand, the palm gets worse results than fingers. The main reason is that landmarks on the palm are too close to each other and their appearances are also more similar. Therefore, it is difficult to identify them. In Fig. 12, we report the average error for each landmark across the whole dataset.

3.4. Evaluation of Dynamic vs. Static Population Graph

Here, we will demonstrate the advantage of using the dynamic population graph, compared to the use of the static graph in LEAP (Wolza et al., 2010). Although LEAP used an intensity-based similarity measurement for construction of the population graph, for fair comparison, we combine the idea of LEAP with our proposed landmark-based similarity measurement to construct a static population graph and then use SPM for correspondence detection. Here, we denote this method as the LEAP-like method.

In the experiment, we again select 5 model images randomly, and then repeat the experiment for 5 times. As these two methods work in a very similar way, only a few results from LEAP-like are worse, so the mean and median errors of our method are just a little better than LEAP-like method. However, due to the use of *backward step*, our method is superior at preventing errors, as reflected by the lower standard deviations. In addition, we also compute the maximum landmark error for each subject image, and further compare 95th percentile value of the maximum landmark errors for all methods (Zhang and Cootes, 2012). Table 4 shows the results by LEAP-like method and our method in each of 5 experiments. We can see that the use of the dynamic graph construction produces much smaller standard deviation and considerably lower maximum errors. Specifically, both the maximum landmark errors and the 95th percentile values of the maximum landmark errors by our method are much smaller than those by the LEAP-like method. This is mainly because the *backward step* in our method helps effectively prevent error propagation. On the other hand, both our method and LEAP-like method obtain better results than SPM, indicating the benefit of establishing the population graph for propagating correspondences across subject images.

4. Conclusion

In this paper, we have proposed a dynamic correspondence detection method for the subject images with large anatomical differences. Specifically, to improve correspondence detection

results, we fully use the distribution information of all images to dynamically construct a population graph, and then progressively determine correspondences (for selected subjects at each time) in an iterative manner. Specifically, our method alternates two steps, *forward step* and *backward step*. In the *forward step*, we construct a population graph by linking only the subject images with similar shapes to the current model images. Here, we further integrate a sparse point matching into a multiple model detection approach for detecting correspondences of each subject image that is connected to multiple model images. In the *backward step*, we propose an error detection mechanism for identifying subject images with wrongly detected correspondences and further remove them from the population graph for avoiding error propagation. Then, we include those subject images with reasonable correspondences into the new set of model images, and repeat the above two steps until the correspondences for all subject images are detected. We have demonstrated higher accuracy of our method than the state-of-the-art methods on correspondence detection of 50 hand X-ray images.

The main contribution of this paper is: **1)** The bi-directional distance for boundary points is introduced to measure the shape similarity between every two images, instead of directly using the simple image intensity similarity. The proposed similarity measure used for building the graph can better capture the manifold structure of the data. **2)** A graph-based approach is used for correspondence detection, where the graph is built to connect similar images. By using the proposed way of simultaneous graph construction and correspondence detection, we can decompose the initial large deformation between subject images into several small ones, thus progressively achieving more accurate correspondence detection results. **3)** A novel fusion method is also proposed to combine multiple correspondence detection results for a subject image if it is linked to multiple model images. **4)** Finally, an error detection mechanism is further proposed in our method to limit the propagation of potential detection errors.

In the future, we will integrate our proposed correspondence detection method into image registration framework, and then validate its performance in various medical imaging applications.

Acknowledgements

This work was supported in part by the National Institutes of Health (NIH) under Grant Nos. EB006733, EB008374, EB009634, MH100217, AG041721, and AG042599, the National Natural Science Foundation of China under Grant Nos. 61573274 and 61473190, and the 973 Program of China under Grant No. 2015CB351703.

References

Adeshina SA, Cootes TF. Constructing part-based models for groupwise registration. IEEE International Symposium on Biomedical Imaging: From Nano to Macro. IEEE. 2010a:1073–1076.

Adeshina SA, Cootes TF. Evaluation of Performance of Part-based Models for Groupwise Registration. MIUA. 2010b:221–251.

Besl PJ, McKay HD. A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1992; 14(2):239–256.

Cao F, Huang HK, Pietka E, Gilsanz V, Dey PS, Gertych A, Pospiech-Kurkowska S. Image Database for Digital Hand Atlas. Medical Imaging 2003: PACS and Integrated Medical Information Systems: Design and Evaluation. 2003:461–470.

Castillo R, Castillo E, Guerra R, Johnson VE, McPhail T, Garg AK, Guerrero T. A Framework for Evaluation of Deformable Image Registration Spatial Accuracy Using Large Landmark Point Sets. Physics in Medicine and Biology. 2009; 54(7):1849–1870. [PubMed: 19265208]

Chui H, Rangarajan A. A new point matching algorithm for non-rigid registration. Computer Vision and Image Understanding. 2003; 89(2):114–141.

Cour T, Srinivasan P, Shi J. Balanced Graph Matching. Advances in Neural Information Processing Systems (NIPS). 2006:313–320.

Criminisi A, Shotton J, Robertson D, Konukoglu E. Regression Forests for Efficient Anatomy Detection and Localization in CT Studies. MICCAI 2010 Workshop MCV. 2010:106–117.

Donner R, Wildenauer H, Bischof H, Langs G. Weakly supervised group-wise model learning based on discrete optimization. Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer. 2009:860–868.

Du S, Zheng N, Meng G, Yuan Z. Affine Registration of Point Sets Using ICP and ICA. IEEE Signal Processing Letters. 2008; 15:689–692.

Du S, Zheng N, Ying S, Liu J. Affine Iterative Closest Point Algorithm for Point Set Registration. Pattern Recognition Letters. 2010; 31(9):791–799.

Du S, Zhu J, Zheng N, Zhao J, Li C. Isotropic Scaling Iterative Closest Point Algorithm for Partial Registration. Electronics Letters. 2011; 47(14):799–800.

Duchenne O, Bach F, Kweon I, Ponce J. A tensor-based algorithm for high-order graph matching. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2011; 33(12):2383–2395. [PubMed: 21646677]

Guo Y, Wu G, Jiang J, Shen D. Robust Anatomical Correspondence Detection by Hierarchical Sparse Graph Matching. IEEE Transaction on Medical Imaging. 2013; 32(2):268–277.

Heimann T, Meinzer H. Statistical shape models for 3D medical image segmentation: A review. Medical Image Analysis. 2009; 4(13):543–563. [PubMed: 19525140]

Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. NeuroImage. 2012; 62:782–790. [PubMed: 21979382]

Jia H, Yap P, Shen D. Iterative Multi-Atlas-Based Multi-Image Segmentation with Tree-Based Registration. NeuroImage. 2012; 59(1):422–430. [PubMed: 21807102]

Jiang H, Drew MS, Li Z. Matching by linear programming and successive convexification. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2007; 29(6):959–975. [PubMed: 17431296]

Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. Elastix: a toolbox for intensity based medical image registration. IEEE Transactions on Medical Imaging. 2010; 29(1):196–205. [PubMed: 19923044]

Leordeanu, M.; Hebert, M. A Spectral Technique for Correspondence Problems Using Pairwise Constraints; Tenth IEEE International Conference on Computer Vision (ICCV); 2005. p. 1482-1489.

Maciel J, Costeira JP. A Global Solution to Sparse Correspondence Problems. IEEE Trans. Pattern Analysis and Machine Intelligence. 2003; 25(2):187–199.

Martin-Fernandez M, Martin-Fernandez M, Alberola-Lopez C. Automatic bone age assessment: A registration approach. Medical Imaging 2003: Image Processing. International Society for Optics and Photonics. 2003:1765–1776.

Murphy K, van Ginneken B, Pluim JPW, Klein S, Staring M. Semi-automatic Reference Standard Construction for Quantitative Evaluation of Lung CT Registration. Medical Image Computing and Computer Assisted Intervention (MICCAI). 2008:1006–1013.

Ou Y, Besbes A, Bilello M, Mansour M, Davatzikos C, Paragios N. Detecting Mutually-salient Landmark Pairs with MRF Regularization. 2010 IEEE International Symposium on Biomedical Imaging (ISBI). 2010:400–403.

Paganelli, C.; Peroni, M.; Pennati, F.; Baroni, G.; Summers, P.; Bellomi, M.; Riboldi, M. Scale Invariant Feature Transform as Feature Tracking Method in 4D Imaging: A Feasibility Study; 34th Annual International Conference of the IEEE EMBS; 2012. p. 6543-6546.

Sanromà G, Alquézar R, Serratosa F. A new graph matching method for point-set correspondence using the EM algorithm and Softassign. Computer vision and image understanding. 2012; 116(2): 292–304.

Sanromà G, Alquézar R, Serratosa F, Herrera B. Smooth point-set registration using neighboring constraints. Pattern Recognition Letters. 2012; 33(15):2029–2037.

Shamonin DP, Bron EE, Lelieveldt BPF, Smits M, Klein S, Staring M. Fast Parallel Image Registration on CPU and GPU for Diagnostic Classification of Alzheimer's Disease. Frontiers in Neuroinformatics. 2014; 7(50):1–15.

Shen D, Wong W, Ip HH. Affine-invariant image retrieval by correspondence matching of shapes. Image and Vision Computing. 1999; 17(7):489–499.

Shi F, Yap P, Fan Y, Gilmore JH, Lin W, Shen D. Construction of multi-region-multi-reference atlases for neonatal brain MRI segmentation. Neuroimage. 2010; 51(2):684–693. [PubMed: 20171290]

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage. 2004; 23(S1):208–219.

Styner MA, Rajamani KT, Nolte L, Zsemlye G, Szekely G, Taylor CJ, Davies RH. Evaluation of 3D Correspondence Methods for Model Building. Information Processing in Medical Imaging (IPMI). 2003:63–75.

Tang S, Fan Y, Wu G, Kim M, Shen D. RABBIT: rapid alignment of brains by building intermediate templates. NeuroImage. 2009; 47(4):1277–1287. [PubMed: 19285145]

Thangam P, Mahendiran T, Thanushkodi K. Skeletal Bone Age Assessment - Research Directions. Journal of Engineering Science and Technology Review. 2012; 5(1):90–96.

Wolza R, Aljabar P, Hajnal JV, Hammers A, Rueckert D. LEAP: Learning Embeddings for Atlas Propagation. NeuroImage. 2010:1316–1325. [PubMed: 19815080]

Woolrich MW, Jbabdi S, Patenaude B, Chappell M, Makni S, Behrens T, Beckmann C, Jenkinson M, Smith SM. Bayesian analysis of neuroimaging data in FSL. NeuroImage. 2009; 45:S173–S186. [PubMed: 19059349]

Wu G, Jia H, Wang Q, Shen D. SharpMean: groupwise registration guided by sharp mean image and tree-based registration. NeuroImage. 2011; 56(4):1968–1981. [PubMed: 21440646]

Xue Z, Shen D, Davatzikos C. Statistical representation of high-dimensional deformation fields with application to statistically constrained 3D warping. Medical Image Analysis. 2006; 10(5):740–751. [PubMed: 16887376]

Yang J, Shen D, Davatzikos C, Verma R. Diffusion tensor image registration using tensor geometry and orientation features, Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008. Springer. 2008:905–913.

Zass, R.; Shashua, A. Probabilistic graph and hypergraph matching. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2008. 2008. p. 1-8.

Zhan Y, Zhou XS, Peng Z, Krishnan A. Active Scheduling of Organ Detection and Segmentation in Whole-Body Medical Images. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI). 2008:313–321.

Zhang D, Lu G. Review of shape representation and description techniques. Pattern Recognition. 2004; 37(1):1–19.

Zhang P, Cootes TF. Automatic construction of parts+ geometry models for initializing groupwise registration. IEEE Transactions on Medical Imaging. 2012; 31(2):341–358. [PubMed: 21947520]

Zhang P, Yap P, Shen D, Cootes TF. Initialising Groupwise Non-rigid Registration Using Multiple Parts+ Geometry Models. Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer. 2012:156–163.

Zheng Y, John M, Liao R, Boese J, Kirschstein U, Georgescu B, Zhou SK, Kempfert J, Walther T, Brockmann G, Hide DC. Automatic Aorta Segmentation and Valve Landmark Detection in C-Arm CT: Application to Aortic Valve Implantation. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI). 2010:476–483.

Highlights

- A dynamic graph based on multi-models is built for accurate correspondence detection across subject images with large anatomical differences
- Bi-directional distance with boundary points is presented to measure the image similarity.
- Error detection mechanism is proposed for updating graph to avoid error propagation.
- Our method achieves much higher accuracy on correspondence detection.

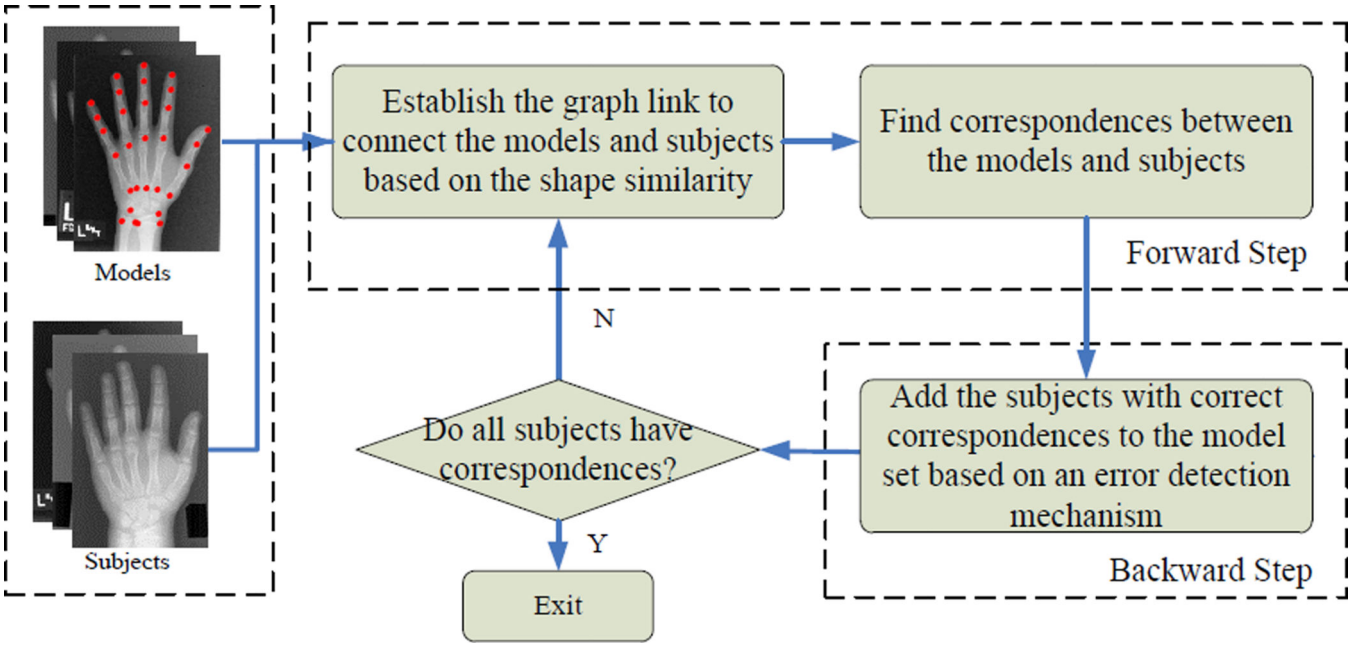


Fig. 1.
Correspondence detection based on a dynamic population graph.

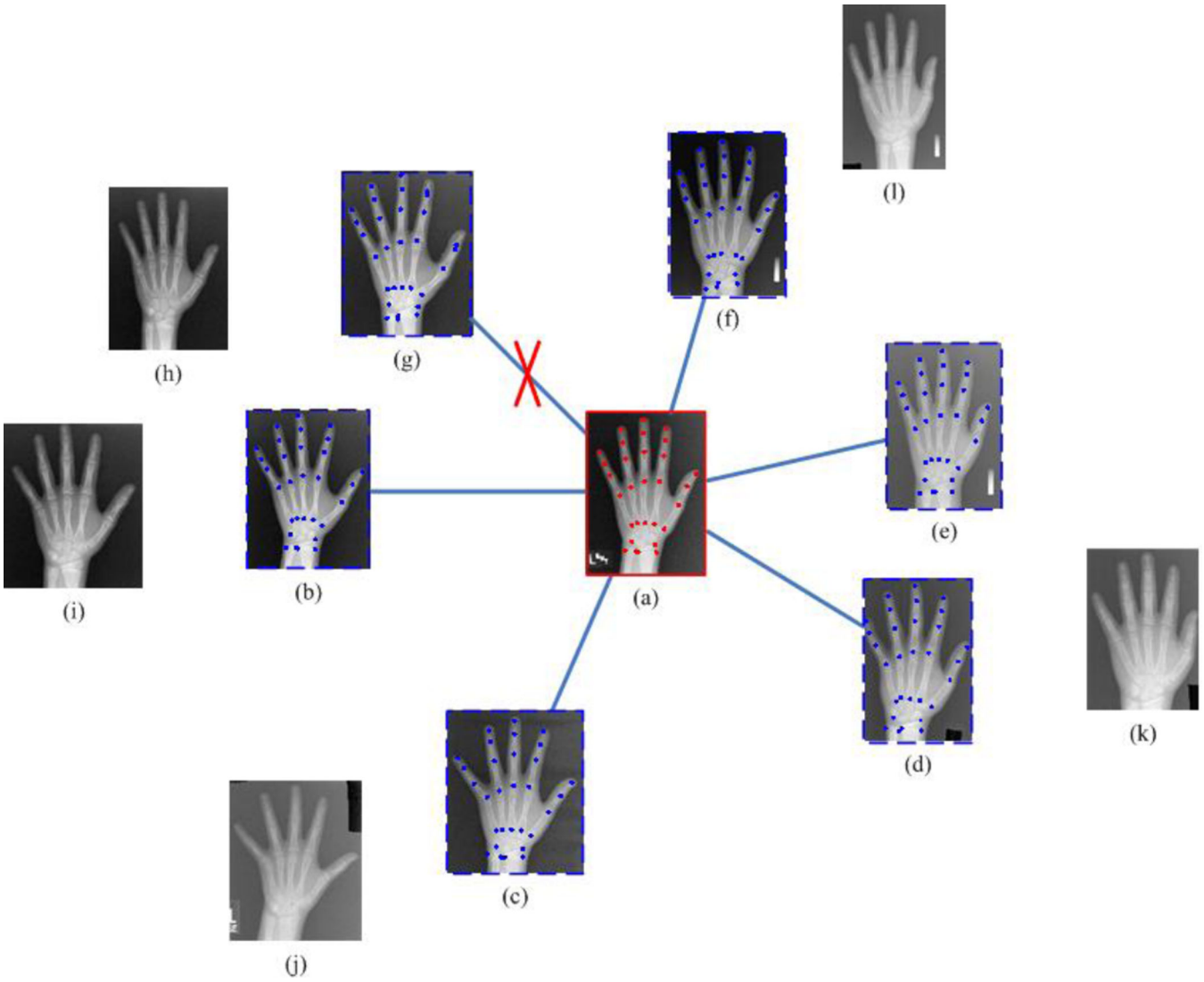


Fig. 2.

Correspondence detection results after completing the 1st iteration in our approach. (a) Model image with manual landmarks (red points). (b)-(f) Subject images with direct links to the model image, each of them having reasonable correspondences (thus these subject images will be eventually included into the set of model images) (blue points). (g) Subject image with wrong correspondences (thus it will not be included into the set of model images). (h)-(l) Subject images not linked to the model image in the 1st iteration.

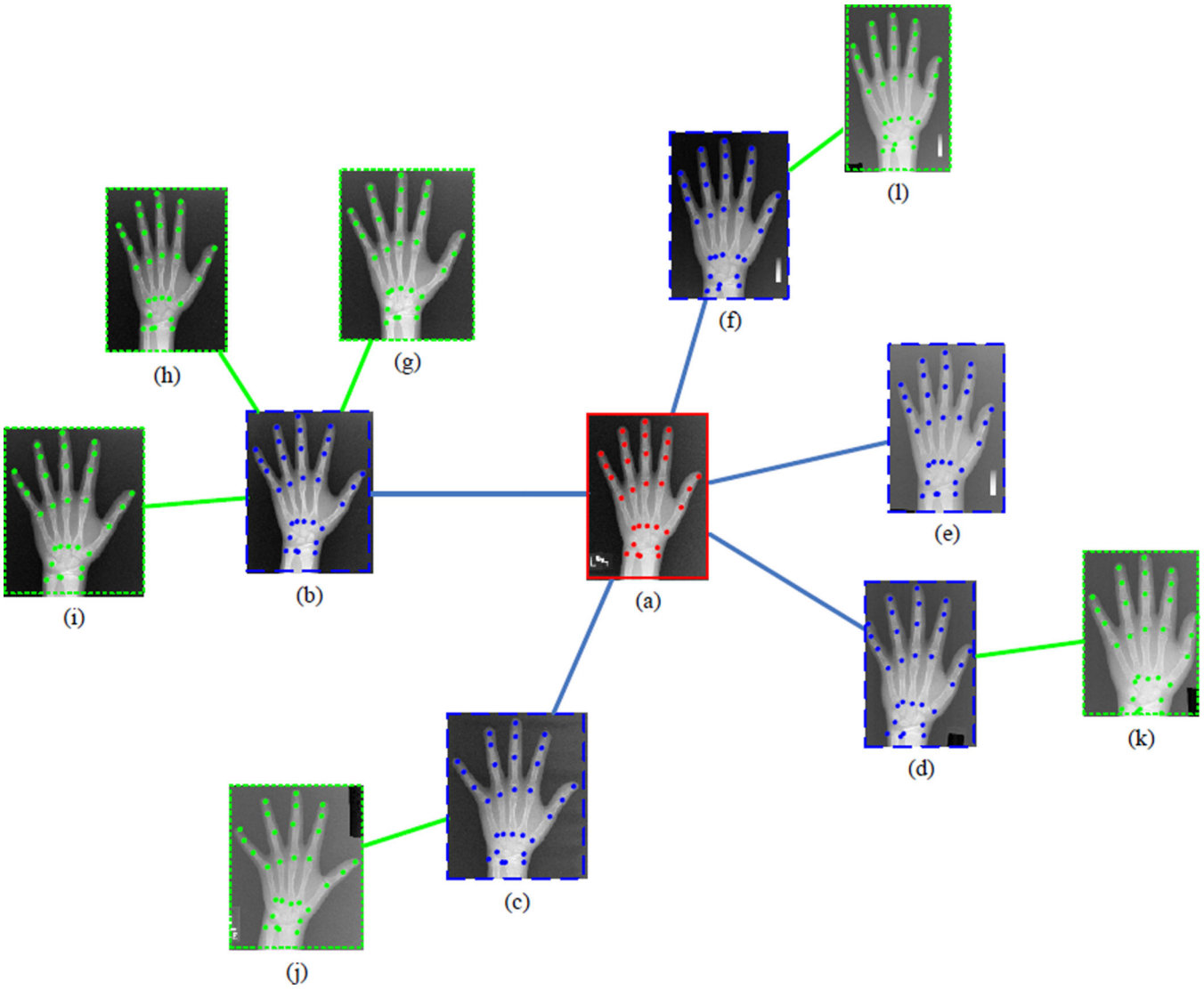


Fig. 3.

Correspondence detection results after completing the 2nd iteration in our approach. (a) Model image with manual landmarks (red points). (b)-(f) Additional model images (blue points), which are actually the subject images included into the set of model images in the previous iteration. (g)-(l) Newly annotated subject images (green points) using the (additional) model images (b)-(f).

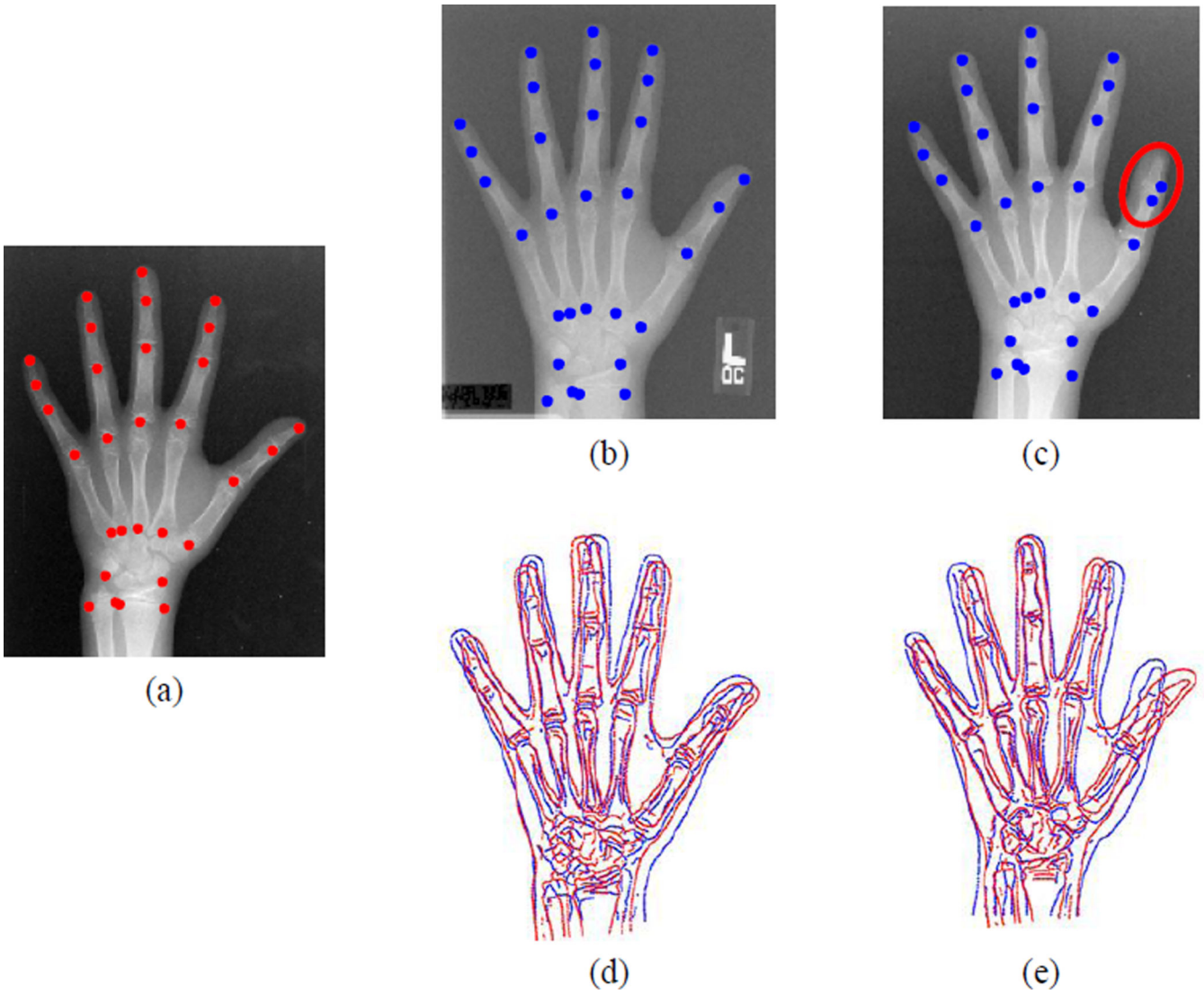


Fig. 4.

Advantage of using the shape boundary difference compared to the simple image intensity difference. (a) Model image with manual landmarks (red points). (b)-(c) Subject images with correspondence detection results (blue points). Averaged intensity difference is 0.0251 between (a) and (b) and 0.0139 between (a) and (c), after histogram matching and further normalizing the intensity to the range [0, 1]. (d)-(e) Shape boundaries between model image (red) and subject image (blue). The root mean square error for the shape boundary difference is 2.00 mm between (a) and (b) and 3.14 mm between (a) and (c). This example shows the importance of using shape boundary difference for comparing two images.

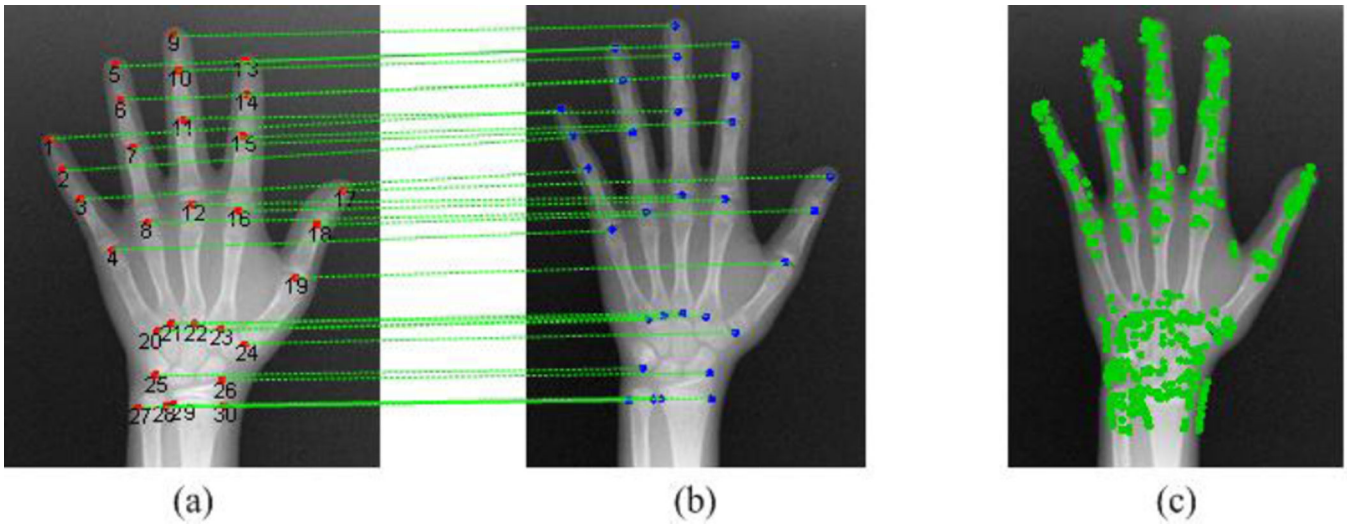


Fig. 5.

An example of the correspondence on X-ray hand image. (a) Model image with 30 landmarks. (b) Subject image with detected correspondences. (c) Subject image with feature points.

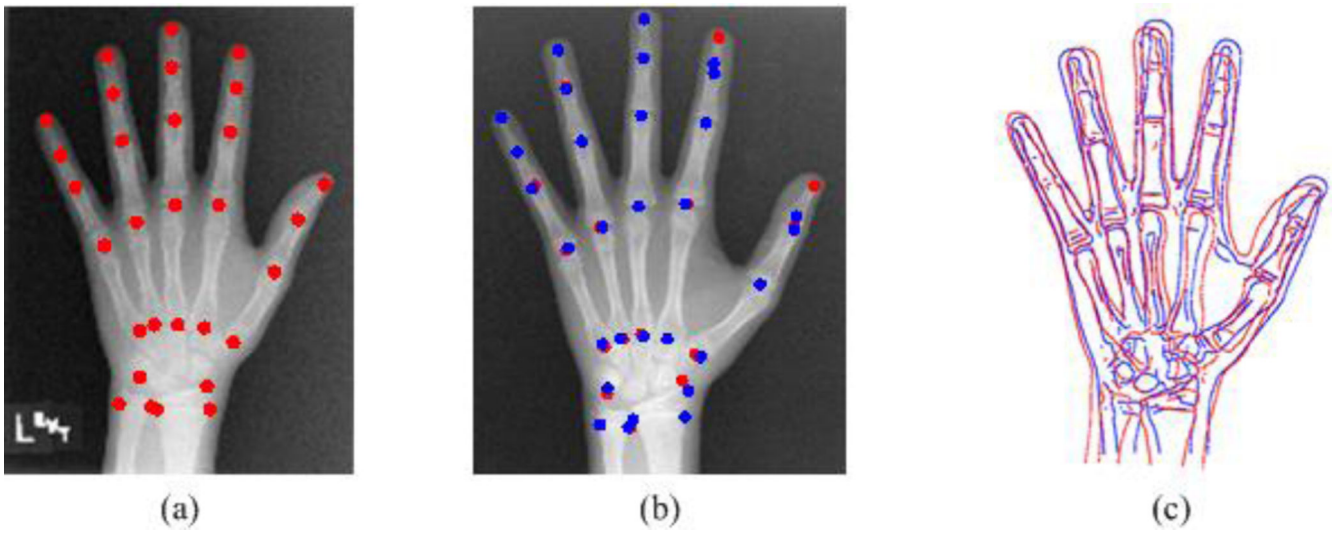


Fig. 6.

Correspondence detection results by sparse point matching. (a) Model image with manual landmarks (red points). (b) Detected corresponding landmarks (blue points) in the subject image, along with the ground-truth landmarks (red points). (c) Shape discrepancies between model (red) and subject (blue) images.

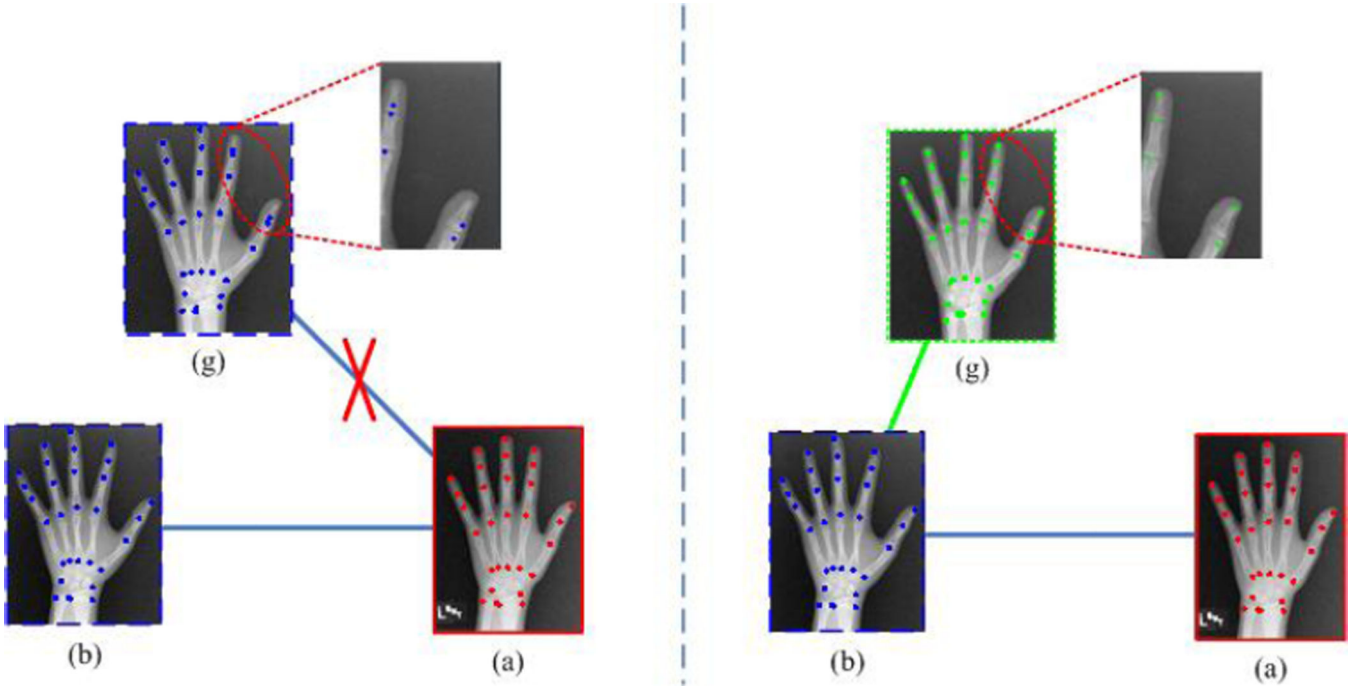


Fig. 7.

Two iterations of graph updating with the use of error detection. (a) Model image with manual landmarks (red). (b) Subject image with detected corresponding landmarks in the 1st iteration and kept in the 2nd iteration (blue). (g) Subject image with detected corresponding landmarks in the 1st iteration (blue) and the 2nd iteration (green), respectively.

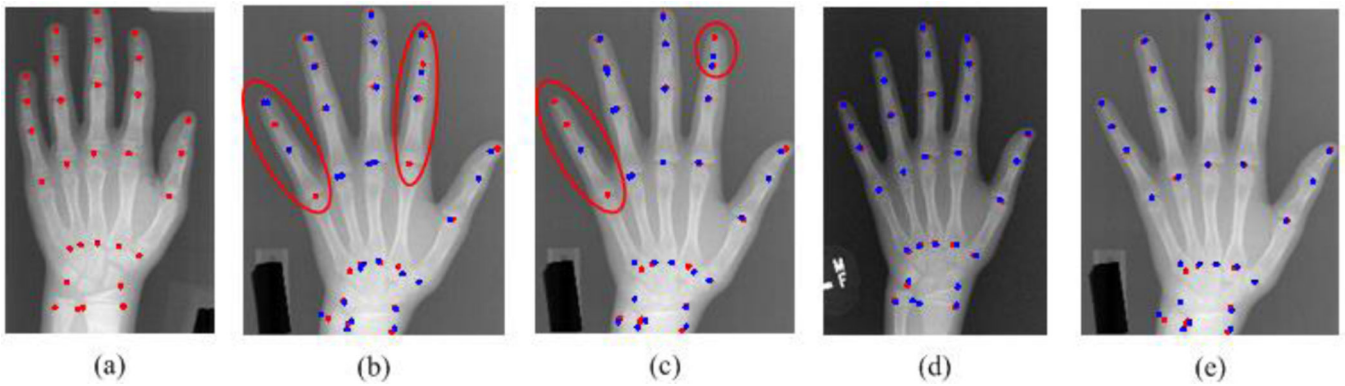


Fig. 8.

Comparison between ground-truth correspondences (red points) and the automatic correspondences (blue points) estimated by SMAC, SPM, and our methods, respectively. (a) Model image. (b) Results by SMAC. (c) Results by SPM. (d) Results by our method obtained first for an *intermediate subject image*. (e) Results by our method obtained finally for the subject image. (Note that the same subject image is shown in (b), (c), and (d), with correspondence detection results by three different methods.)

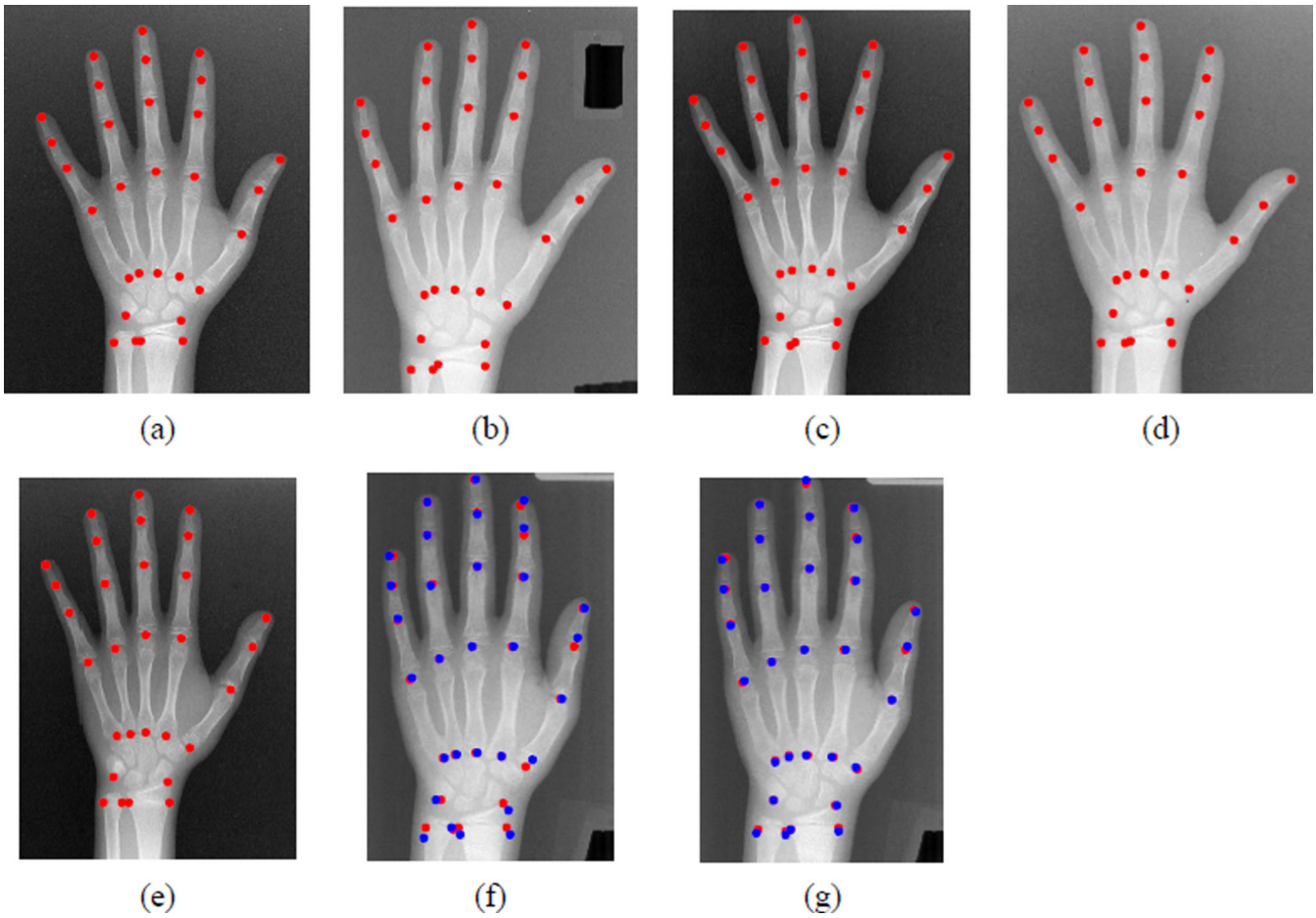


Fig. 9.

Comparison between the ground truth (red points) and the automatic landmarks (blue points) detected by SPM and our method. (a)-(e) Model images. (f) Results by SPM. (g) Results by our method.

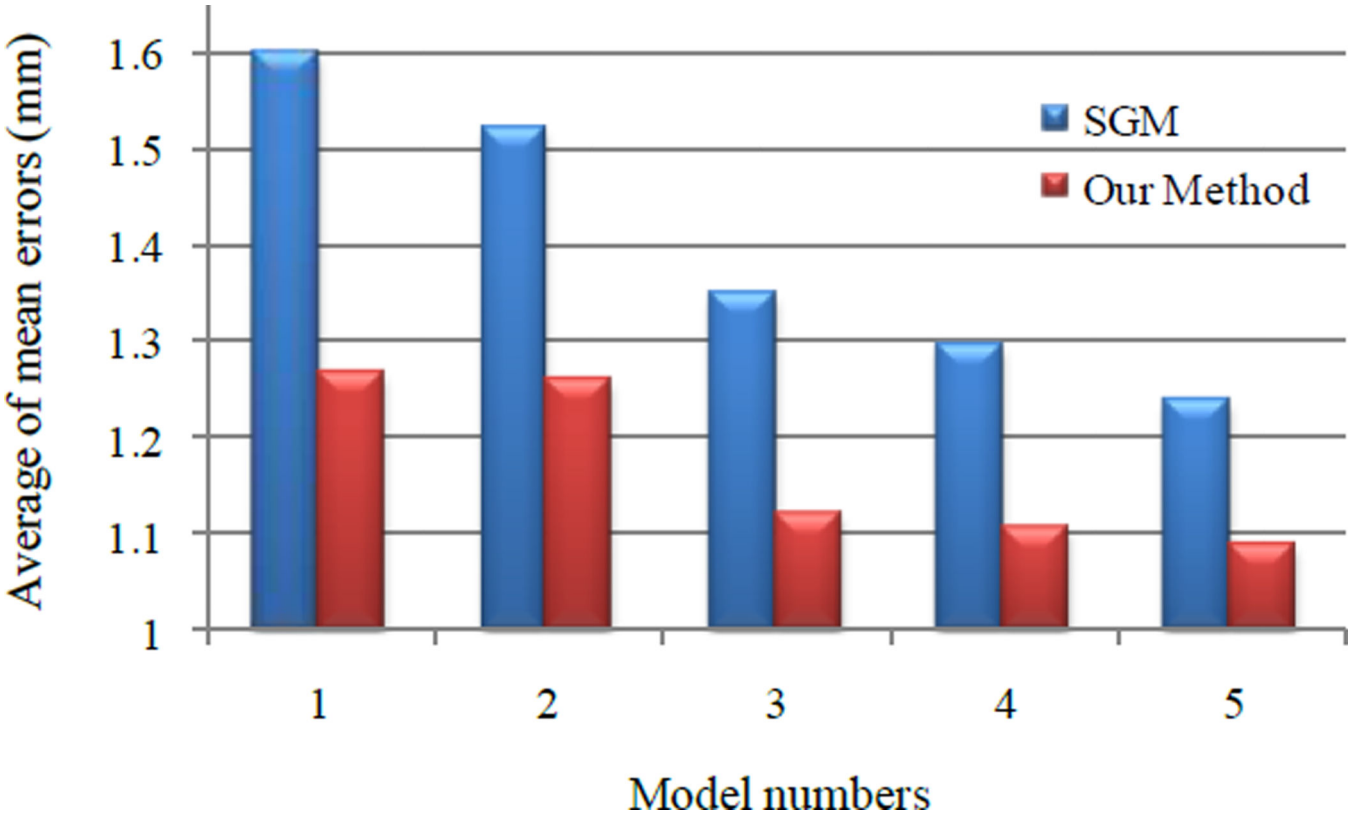


Fig. 10.

Average of mean errors of SPM and our method, with respect to the use of different number of model images.

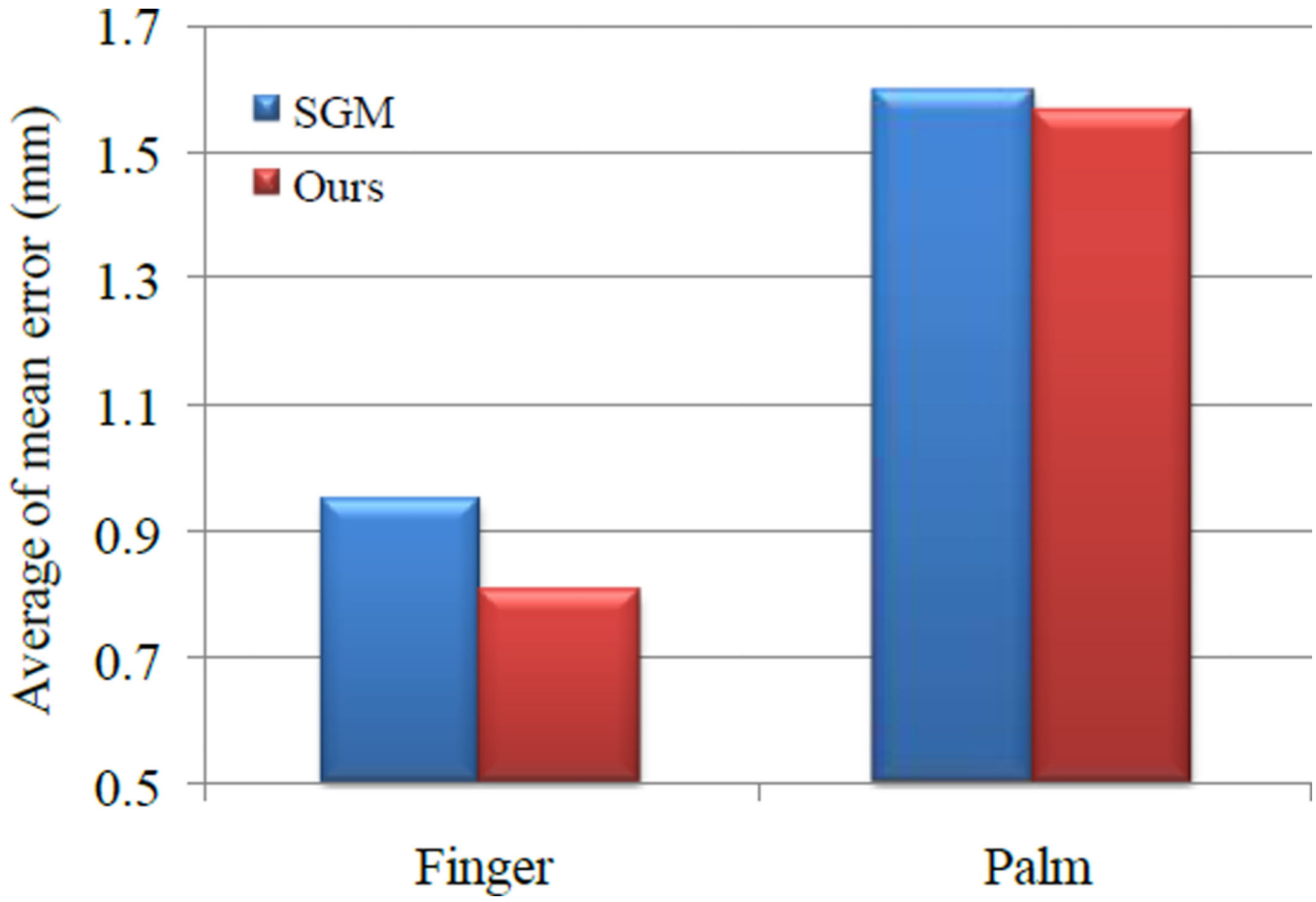


Fig. 11.
Average of mean errors of SGM and our method separately for the fingers and the palm.

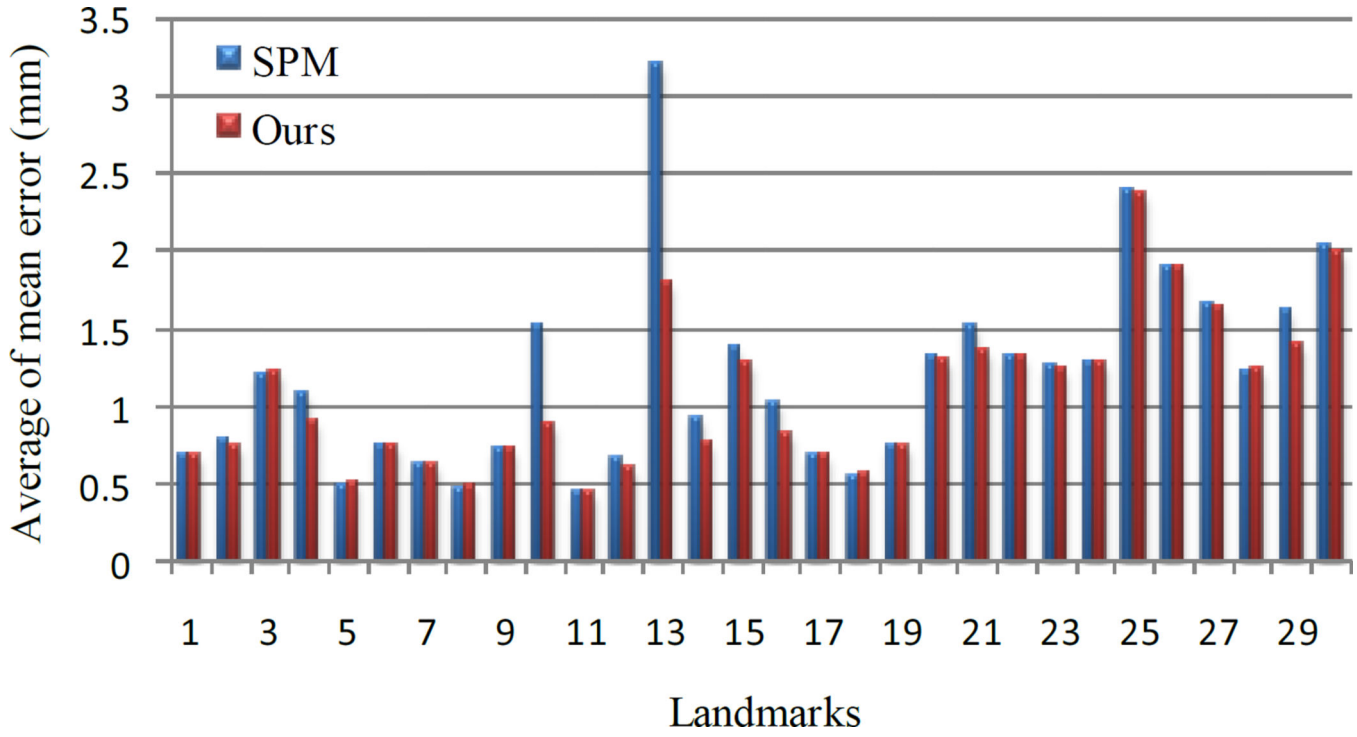


Fig. 12.
Average error for each landmark obtained by SPM and our method.

Table 1

Mean and standard deviation of landmark detection errors between the ground truth and the estimated correspondences by SMAC, SPM, and our method in each of the 5 experiments. (Unit: mm)

Test No.	SMAC	SPM	Our Method
1	2.16±3.81	1.69±2.77	1.26±1.32
2	1.55±2.23	1.49±2.43	1.17±1.27
3	1.86±3.76	1.61±2.72	1.26±1.28
4	1.71±3.00	1.66±2.86	1.26±1.29
5	1.99±4.27	1.56±2.25	1.39±1.26
Avg.	**1.85±3.41**	**1.60±2.61**	**1.27±1.28**

Table 2

Maximum error and mean of maximum errors for all subject images by SMAC, SPM, and our method, respectively. (Unit: mm)

	SMAC		SPM		Our Method	
Test No.	**Maximum error**	**Mean of maximum errors**	**Maximum error**	**Mean of maximum errors**	**Maximum Error**	**Mean of maximum errors**
1	46.11	12.72	45.22	11.16	12.61	4.78
2	22.16	7.32	21.26	7.57	11.39	4.29
3	49.96	13.67	23.16	7.80	6.00	3.93
4	52.59	10.69	52.59	9.94	8.08	4.43
5	66.69	11.43	27.03	7.09	9.40	4.42
Avg	**47.50**	**11.17**	**33.85**	**8.71**	**9.50**	**4.37**

Table 3

Mean and standard deviation of landmarks detection errors between the ground truth and the estimated correspondences by SPM and our method. Also, both the maximum error and the mean of maximum errors for all subject images are provided. (Unit: mm)

Number of model images	Test No.	SPM			Our Method		
		Mean and standard deviation	Maximum error	Mean of maximum errors	Mean and standard deviation	Maximum error	Mean of maximum errors
2	1	1.39±1.86	23.16	6.88	1.20±1.20	9.83	4.40
	2	1.69±3.62	51.60	9.93	1.28±1.26	6.94	4.10
	3	1.43±2.74	47.30	8.99	1.14±1.18	9.88	4.13
	4	1.64±2.15	23.70	7.21	1.39±1.28	9.35	4.59
	5	1.47±2.32	51.60	8.41	1.30±1.26	8.12	4.65
	Avg.	**1.52±2.54**	**39.47**	**8.28**	**1.26±1.24**	**8.82**	**4.37**
3	1	1.46±2.20	22.11	6.88	1.18±1.17	7.43	4.19
	2	1.25±1.83	18.61	5.21	1.13±1.15	7.03	4.04
	3	1.33±2.99	60.47	8.94	1.07±1.16	6.33	3.80
	4	1.57±2.69	29.20	6.88	1.13±1.16	7.60	3.99
	5	1.37±1.73	18.82	5.90	1.20±1.22	6.18	3.96
	Avg.	**1.40±2.29**	**29.84**	**6.76**	**1.14±1.17**	**6.91**	**4.00**
4	1	1.52±2.42	22.11	7.59	1.19±1.19	7.43	3.94
	2	1.21±1.85	19.69	5.78	1.05±1.09	6.49	3.80
	3	1.24±1.95	24.49	6.65	1.05±1.08	7.29	3.88
	4	1.27±1.74	17.85	5.81	1.11±1.13	6.38	4.11
	5	1.22±1.72	21.26	6.32	1.09±1.12	6.98	3.88
	Avg.	**1.29±1.94**	**21.08**	**6.43**	**1.10±1.12**	**6.91**	**3.92**
5	1	1.38±2.21	22.11	6.68	1.15±1.16	6.33	3.95
	2	1.30±2.10	22.16	5.95	1.05±1.16	6.49	3.78
	3	1.14±1.57	16.61	4.84	1.05±1.09	6.33	3.73
	4	1.19±1.89	39.14	6.66	1.06±1.15	6.91	3.69
	5	1.11±1.41	18.63	4.92	1.05±1.11	6.18	3.74
	Avg.	**1.22±1.84**	**23.73**	**5.81**	**1.07±1.15**	**6.45**	**3.78**

Table 4

Mean and standard deviation of landmarks detection errors between the ground truth and the estimated correspondences by LEAP-like method and our method, respectively. Also, both the maximum error and the mean of maximum errors for all subject images are provided. (Unit: mm)

Test No.	LEAP-like Method			Our Method		
	Standard deviation	Maximum error	95%	Standard deviation	Maximum error	95%
1	1.26	16.53	10.88	1.16	6.33	6.18
2	1.31	15.21	12.00	1.16	6.49	6.42
3	1.15	12.00	10.87	1.09	6.33	5.96
4	1.16	8.63	7.89	1.15	6.91	6.33
5	1.21	15.01	11.99	1.11	6.18	5.88
Avg.	**1.22**	**13.48**	**10.73**	**1.13**	**6.45**	**6.15**