# DETECTING MUTUALLY-SALIENT LANDMARK PAIRS WITH MRF REGULARIZATION

*Y. Ou*[a,d], *A. Besbes*[b,c], *M. Bilello*[d], *M. Mansour*[d], *C. Davatzikos*[a,d] *and N. Paragios*[b,c]

[a] Section of Biomedical Image Analysis (SBIA), [d] Dept. of Radiology, Univ. of Pennsylvania, USA
[b] Laboratoire MAS, Ecole Centrale Paris, Châtenay-Malabry, France
[c] Equipe GALEN, INRIA Saclay - Île-de-France, Orsay, France.

## ABSTRACT

In this paper, we present a framework for extracting mutually-salient landmark pairs for registration. Traditional methods detect landmarks one-by-one and separately in two images. Therefore, the detected landmarks might inherit low discriminability and are not necessarily good for matching. In contrast, our method detects landmarks pair-by-pair across images, and those pairs are required to be mutually-salient, i.e., uniquely corresponding to each other. The second merit of our framework is that, instead of finding individually optimal correspondence, which is a local approach and could cause self-intersection of the resultant deformation, our framework adopts a Markov-random-field (MRF)-based spatial arrangement to select the globally optimal landmark pairs. In this way, the geometric consistency of the correspondences is maintained and the resultant deformations are relatively smooth and topology-preserving. Promising experimental validation through a radiologist's evaluation of the established correspondences is presented.

***Index Terms—*** Landmark Matching, Image Registration, MRF, Mutual-Saliency

## 1. INTRODUCTION

Deformable image registration is one of the most challenging problems in medical imaging. The most natural classification of the prior work consists of iconic and geometric methods. Iconic (a.k.a. voxel-wise) methods define on the observation space a similarity criterion (like mutual information) and then seek the transformation that optimizes this criterion. These methods are efficient for intra-modal registration but fail miserably for modalities that are not related through the underlying statistical assumption. Geometric (a.k.a. landmark/feature-based) methods are based on the extraction of landmarks and the estimation of the corresponding transformation that creates a mapping between them.

In most landmark-based non-rigid registration methods, establishing landmark correspondences is a crucial component. Traditionally, landmark correspondences are established in two sequential and separate steps, namely, landmark detection (e.g., [1]) and landmark matching (e.g., [2]).

However, these two-step methods are usually limited in the following two respects.

First, those salient landmarks separately detected in two images might inherit low discriminability and therefore are not necessarily good for matching. As pointed out in [3], salient points in one image might not be present, or uniquely present in the other image. Therefore, instead of detecting landmarks one-by-one in two images separately, landmarks should be detected pair-by-pair, preferably as those pairs whose matching is unique across images.

Second, individually optimal correspondence for each landmark is not necessarily globally optimal for the deformation field. Most existing methods only consider landmark correspondences individually. As a result, the displacement vectors at two nearby landmarks might point to completely opposite directions, causing self-intersections in the resultant dense deformation field. To avoid this problem, an ideal method should have a systematic way to simultaneously consider the global smoothness when establishing landmark correspondences.

This paper presents a landmark correspondence establishment framework to cope with the aforementioned issues. To alleviate the first limitation, we simultaneously detect and match a number of mutually-salient candidate landmark pairs, i.e., pairs that are uniquely corresponding to each other across images. This is largely built upon the mutual-saliency measure [3], which quantifies the matching uniqueness of a pair of voxels. To alleviate the second limitation and to encourage global optimality of the correspondences, the correspondence problem is solved in a global manner through a Markov-random-field (MRF)-based formulation [4]. This formulation combines content similarities with geometric constraints (relative landmarks positions and displacement consistency). As a result, consistent landmark displacements are estimated, thereby avoiding potential self-intersections in the resultant deformation field. Results are evaluated by radiologists to demonstrate the necessities of the two components in our framework.

In the remainder of this paper, we present our framework in Section 2 and provide experimental validations in Section 3. The paper is discussed and concluded in Section 4.

## 2. METHODS

In this section, we first briefly describe the definition of mutual-saliency measure in section 2.1; interested readers are referred to [3] for more details. Based on that, we will elaborate the two components of our framework: the detection of mutually-salient landmark pairs in section 2.2, and the MRF formulation to find globally optimal correspondences in section 2.3.

### 2.1. Definition of Mutual-Saliency Measure

Given two images $\mathcal{I}_1 : \Omega_1 \mapsto \mathcal{R}$ and $\mathcal{I}_2 : \Omega_2 \mapsto \mathcal{R}$ in the 3D image domains $\Omega_i (i = 1, 2) \subset \mathcal{R}^3$, a pair of voxels $\mathbf{u} \in \Omega_1$ and $\mathbf{v} \in \Omega_2$ is mutually-salient if they are similar to each other and meanwhile not similar to any other voxels in the neighborhood. In this case, as shown in Fig. 1, the similarity map between $\mathbf{u}$ and all voxels in the neighborhood of $\mathbf{v}$ should exhibit a delta-shaped function peaking at $\mathbf{v}$.
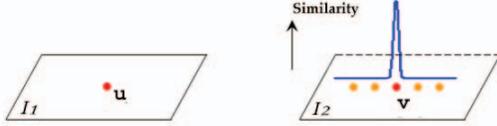


**Fig. 1**. The idea of mutual-saliency measure.

In practice, the mutual-saliency between $\mathbf{u}$ and $\mathbf{v}$, denoted as $ms(\mathbf{u}, \mathbf{v})$, is approximated via dividing the mean similarity in the central part of the neighborhood of $\mathbf{v}$ (denoted as $CN(\mathbf{v})$, by the mean similarity in the peripheral neighborhood of $\mathbf{v}$ (denoted as $PN(\mathbf{v})$).

$$ms(\mathbf{u}, \mathbf{v}) \overset{def}{=} \frac{\frac{1}{|CN(\mathbf{v})|} \sum_{\mathbf{w} \in CN(\mathbf{v})} [sim(\mathbf{u}, \mathbf{w})]}{\frac{1}{|PN(\mathbf{v})|} \sum_{\mathbf{w} \in PN(\mathbf{v})} [sim(\mathbf{u}, \mathbf{w})]} \quad (1)$$

Here, central neighborhood $CN(\cdot)$ and peripheral neighborhood $PN(\cdot)$ are defined as concentric rings around the voxel $\cdot$ and their radii are adaptive to the scales from which the multi-scale attributes are extracted (see [3] for details). Also in Eqn. 1, the similarity $sim(\mathbf{u}, \mathbf{w})$ is defined on the $d$-dimensional attribute vectors $A_i(\cdot)$ characterizing each voxel ($i = 1, 2$ for image $\mathcal{I}_i$), i.e.,

$$sim(\mathbf{u}, \mathbf{w}) \overset{def}{=} \frac{1}{1 + \frac{1}{d}\|A_1(\mathbf{u}) - A_2(\mathbf{w})\|^2}, \quad (2)$$

that is, smaller difference in their attribute vectors indicates higher similarity between the two voxels. Specifically, the attribute vector $A_i(\cdot)$ in our framework is constructed by incorporating the multi-scale and multi-orientation Gabor attributes. Here Gabor attributes are used to characterize each voxel because of their generally applicability in diverse images, their demonstrated successful applications in various tasks including image registration (e.g., [5, 3]), and their

multi-scale and multi-orientation nature, which is more likely to describe each voxel distinctively than the intensity attribute or other texture attributes [3, 1].

Fig. 2 shows examples of different voxel pairs having different mutual-saliency values. Here, similarity maps (c-e) are generated by calculating the attribute-based similarity between a specific voxel (denoted as red $+$, blue $\times$ and orange $\star$) in the subject image (a) and all voxels in the template image (b). It is observed that, the red point and its correspondence has the highest mutual-saliency (indicating most unique correspondence), followed by the blue point and its correspondence, and lastly the orange point and its correspondence.
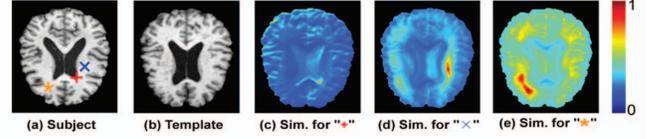


**Fig. 2**. Examples of different voxel pairs having different mutual-saliency values.

Note that, the mutual-saliency is defined on a pair of voxels, therefore it can be used to detect landmarks pair-by-pair, as those having high matching uniqueness across images. In this way, landmark detection and matching steps are naturally merged into a single step. When a landmark is detected, its correspondence is also automatically determined, which is guaranteed to be relatively unique.

### 2.2. Detecting Mutually-Salient Landmark Pairs

Based on the aforementioned advantage of mutual-saliency measure, this sub-section describes extracting a number of mutually-salient landmark pairs. To encourage uniform registration accuracy in the image, it is desirable that the detected mutually-salient landmark pairs are scattered within the image space. Therefore, as sketched in Fig. 3, we regularly partition the subject image space $\Omega_1$ into $J$ regions, and in each region we select $K$ most mutually-salient landmark pairs across images. That is, from the $j^{th}$ ($j = 1, 2, \ldots, J$) region, we select $K$ pairs $(\mathbf{p}_j^k \in \Omega_1, \mathbf{q}_j^k \in \Omega_2)_{k=1}^{K}$, ranking by the similarity weighted by the mutual-saliency value, i.e., $[sim(\cdot, \cdot) \times ms(\cdot, \cdot)]$. Here $sim(\cdot, \cdot)$ encourages the detected pair to be similar, and $ms(\cdot, \cdot)$ reflects the uniqueness of their matching.
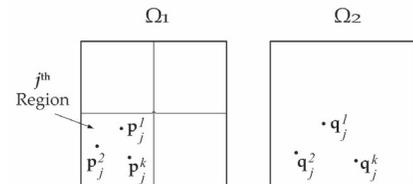


**Fig. 3**. Sketch of the detection of landmark pairs.

Note in Fig. 3 that, the template image space $\Omega_2$ is not partitioned, because at this stage, no transformation is conducted, therefore no corresponding regions can be assumed.

More importantly, from each region, we keep $K$ top-ranking pairs instead of simply choosing the highest ranking pair. This is because the highest ranking pair is only a locally optimal pair, not necessarily globally optimal for maintaining the smoothness of the resultant deformation field. To find the globally optimal pairs, a Markov-random-field (MRF)-based optimization is formulated in the next sub-section.

### 2.3. Finding Globally Optimal Landmark Pairs by MRF Formulation

Let us denote all the $J \times K$ landmark pairs detected in the previous step as a set $\mathcal{P} = \{(\mathbf{p}_j^k, \mathbf{q}_j^k) | j = 1, 2, \ldots, J; k = 1, 2, \ldots, K\}$. One should note that $K$ is not necessarily the same number for different regions; while in practice, we keep $K = 10$ top-ranking candidate pairs in each region.

Our goal in this section is to select one pair (out of $K$ pairs) from each region, i.e., $\mathcal{P}^{\star} = \{(\mathbf{p}_j^{l_j}, \mathbf{q}_j^{l_j}) | j = 1, 2, \ldots; l_j \in \{1, 2, \ldots, K\}\}$, such that they are globally optimal, in that they altogether maintain high mutual-saliency as well as the smoothness of the resultant deformation field.

We can formulate this task as a Markov-random-field (MRF)-based labeling problem. That is, to find a set of optimal labels $\mathbf{l}^{\star} = (l_1, \ldots, l_J)$, where $l_j \in \{1, 2, \ldots, K\}$ is the label (or index) of the globally optimal pair out of $K$ candidate pairs in the $j^{th}$ region. For that purpose, we construct a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where each node in $\mathcal{V}$ represents a region and each edge in $\mathcal{E}$ expresses the constraint of a local geometric compatibility between the landmark pairs.

In this formulation, our goal can be achieved by minimizing the following labeling energy on the graph,

$$\mathbf{l}^{\star} = \arg \min_{\mathbf{l}} E(l_1, \ldots, l_J), \qquad (3)$$

where the energy $E(l_1, \ldots, l_J)$ consists of two terms,

$$E(l_1, \ldots, l_J) = E_{data}(l_1, \ldots, l_J) + \alpha \, E_{reg}(l_1, \ldots, l_J). \quad (4)$$

Those two terms express the two criteria for selecting the globally optimal landmark pairs. The first term, $E_{data}$, encourages the selected pair in each region to exhibit high similarity weighted by high mutual-saliency,

$$E_{data}(\mathbf{l}) = \sum_{j=1}^{J} \exp \left( -\frac{ms(\mathbf{p}_j^{l_j}, \mathbf{q}_j^{l_j}) \cdot sim(\mathbf{p}_j^{l_j}, \mathbf{q}_j^{l_j})}{2\sigma^2} \right), \quad (5)$$

where $\sigma$ is a scaling factor, estimated as the standard deviation of the mutual saliency values of all the candidate pairs.

The second term, $E_{reg}$ is a regularization term for the goodness of resultant deformation field. The regularization imposes constraints on both the spatial positions and the displacement directions on the selected landmark pairs. Specifically, spatial position constraint encourages those selected pairs to scatter in the image space other than being close to each other. Displacement direction constraint encourages displacement vectors on adjacent pairs to be consistently oriented, therefore avoiding self-intersection of the resultant deformation field. Those two constraints are similar to the distortion terms in [6] and [7]. Mathematically, those two constraints are expressed in a unified regularization term,

$$E_{reg}(\mathbf{l}) = \sum_{(m,n) \in \mathcal{E}} \left\| (\mathbf{p}_m^{l_m} - \mathbf{p}_n^{l_n}) - (\mathbf{q}_m^{l_m} - \mathbf{q}_n^{l_n}) \right\|. \quad (6)$$

One should note that such a framework is not invariant to scale. However, since we have assumed that a rigid registration step was performed prior to the deformable fusion this is not a main concern. We refer to [4] for a variant of this concept that is also scale-invariant.

To implement the data and the regularization criteria, the total energy in Eqn. (3) is minimized using the sequential tree-reweighted message passing (TRW-S) algorithm [8] [1].

### 3. EXPERIMENTAL RESULTS

Our framework is applied to finding mutually-salient landmark pairs in brain and cardiac images. Results in the following sub-sections aim to demonstrate the necessities of the two components in our framework.

Results for Detecting Mutually-Salient Landmark Pairs. Fig. 4 shows a landmark pair detected across subjects based on the mutual-saliency measure. Had we used image intensity only to characterize each voxel and used existing methods (like edge/corner detectors) to extract landmarks separately in two images, this pair of voxels are less likely to be detected in the first place, simply because they are not edges or corners or surface boundaries. The detection of this pair in our framework is largely due to 1) the use of Gabor attributes, which characterize each voxel relatively distinctively identifiable, as shown in similarity maps (c,d); and due to 2) the mutual-saliency measure, which effectively quantifies the matching uniqueness, as shown in similarity maps (e,f). Overall, this example shows the advantage of detecting landmarks in pairs other than one-by-one and separately from the two images.

Results for Finding Globally Optimal Landmark Pairs. To demonstrate the necessity of MRF-based regularization, landmark pair detections without and with the MRF regularization are compared. They are respectively denoted as Method 1 (M1) and Method 2 (M2) in the following comparisons.

The first comparison is in terms of the smoothness of the resultant dense deformation field (based on thin-plate-spline interpolation). The avoidance of self-intersection of the deformation field in Fig. 5(b) demonstrates the need for a global MRF-based optimization as we described in Section 2.3.

---

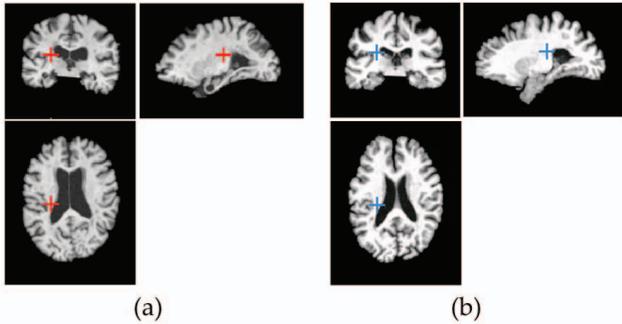[1]The code is available on the author's webpage.

Fig. 4. An example landmark pair (denoted by red and blue crosses) detected based on mutual-saliency measure. (a) Subject image $\mathcal{I}_1$; (b) Template image $\mathcal{I}_2$. Similarity maps are generated (c) between the red cross point and all voxels in $\Omega_1$; (d) between the blue cross point and all voxels in $\Omega_2$; (e) between the red point and all voxels in $\Omega_2$; and (f) between the blue cross point and all voxels in $\Omega_1$.
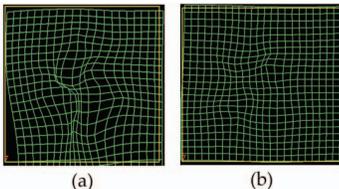


Figure 5. Dense deformation fields generated by (a) M1 – no MRF regularization and (b) M2 – with MRF regularization.

**Table 1**. Radiologist's evaluation on results generated by two methods (M1 and M2 for without and with MRF optimization). $\gg$ (or $\ll$) means "obviously better (or worse)", $>$ (or $<$) means "slightly better (or worse)", and $\approx$ means "almost equivalent".

|           | M1 $\gg$ M2 | M1 $>$ M2 | M1 $\approx$ M2 | M1 $<$ M2 | M1 $\ll$ M2 |
|-----------|-------------|-----------|-----------------|-----------|-------------|
| Dataset1  |             |           | 68%             | 12%       | 20%         |
| Dataset2  |             | 4%        | 80%             | 16%       |             |
| Dataset3  | 4%          | 4%        | 68%             | 20%       | 4%          |
| Dataset4  |             | 4%        | 60%             | 24%       | 12%         |

The second comparison involves the radiologist's evaluation on the landmark correspondences finally determined without and with MRF regularization (M1 v.s. M2). Four datasets are used: *Dataset1* (size $256 \times 256 \times 171$) and *Dataset2* (size $192 \times 236 \times 171$) for two different sets of intra-modality brain MR images across-subjects; *Dataset3* (size $192 \times 236 \times 171$) for a pair of multi-modality brain MR images across-subjects; and *Dataset4* (size $150 \times 150 \times 49$) for a pair of intra-modality cardiac MR images across-subjects. For each dataset, 25 random pairs out of all the finally selected pairs (typically hundreds) are evaluated in terms of uniqueness and accuracy of correspondences. Evaluation results in Table 1 demonstrate the advantage of incorporating MRF regularization (M2) in all datasets.

## 4. CONCLUSION

This paper presents a framework for detecting mutually-salient landmark pairs. The work is built upon the mutual-saliency measure that quantifies the matching uniqueness between a pair of voxels. Our framework conducts landmark detection and matching simultaneously. In this way, landmarks are detected pair-by-pair, and when they are detected, their correspondences are naturally determined with matching uniqueness. A second contribution is the incorporation of the MRF-based regularization. The MRF formulation systematically finds landmark correspondences that altogether maintain mutual saliency and the smoothness of the resultant dense deformation field. Experimental results on brain and cardiac images have demonstrated the necessity of both components in our framework.

To further demonstrate the advantage of detecting landmarks pair-by-pair than one-by-one and separately from two images, more comparisons with classic landmark detection methods (such as SIFT [9]) are needed.

Future work also includes the handling of missing correspondences. It is possible that even if individual landmark content can be very similar, the correspondence problem cannot be solved for them within the global framework. This limitation can be addressed by introducing to the MRF the notion of missing correspondences. Furthermore, the combination of geometric and iconic registration methods through a graphical model is a very promising direction, where both problems can mutually benefit from their corresponding solutions. The ability to simultaneously and consistently solve both the landmark correspondences and the dense registration problems in a unified framework can be a major breakthrough in the field of deformable image fusion.

## 5. REFERENCES

[1] T. Kadir and M. Brady, "Saliency, scale and image description," *Intl J. Comp. Vis.*, vol. 45, no. 2, pp. 83–105, 2001.

[2] S. Joshi and M. Miller, "Landmark matching via large deformation diffeomorphisms," *IEEE TIP*, vol. 9, no. 8, pp. 1357–1370, 2000.

[3] Y. Ou and C. Davatzikos, "DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting," in *IPMI*, 2009, pp. 50–62.

[4] A. Besbes, N. Komodakis, G. Langs, and N. Paragios, "Shape priors and discrete MRFs for knowledge-based segmentation," *CVPR*, pp. 1295–1302, 2009.

[5] J. Liu, B. C. Vemuri, and J. L. Marroquin, "Local frequency representations for robust multimodal image registration.," *IEEE TMI*, vol. 21, no. 5, pp. 462–469, 2002.

[6] M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints," in *ICCV*, 2005, pp. 1482–1489.

[7] L. Torresani, V. Kolmogorov, and C. Rother, "Feature correspondence via graph matching: Models and global optimization," in *ECCV*, 2008, pp. 596–609.

[8] V. Kolmogorov, "Convergent tree-reweighted message passing for energy minimization," *IEEE T-PAMI*, vol. 28, no. 10, pp. 1568–1583, 2006.

[9] David Lowe, "Object recognition from local scale-invariant features," in *ICCV*, 1999.