

# Structured Decision Forests For Multi-modal Ultrasound Image Registration

Ozan Oktay<sup>1</sup>, Andreas Schuh<sup>1</sup>, Martin Rajchl<sup>1</sup>, Kevin Keraudren<sup>1</sup>, Alberto Gomez<sup>3</sup>, Mattias P. Heinrich<sup>2</sup>, Graeme Penney<sup>3</sup>, Daniel Rueckert<sup>1</sup>

<sup>1</sup> Biomedical Image Analysis Group, Imperial College London, UK  
o.oktay13@imperial.ac.uk

<sup>2</sup> Institute of Medical Informatics, University of Lübeck, Germany

<sup>3</sup> Imaging Sciences and Biomedical Engineering Division, King's College London, UK

**Abstract.** Interventional procedures in cardiovascular diseases often require ultrasound (US) image guidance. These US images must be combined with pre-operatively acquired tomographic images to provide a roadmap for the intervention. Spatial alignment of pre-operative images with intra-operative US images can provide valuable clinical information. Existing multi-modal US registration techniques often do not achieve reliable registration due to low US image quality. To address this problem, a novel medical image representation based on a trained decision forest named probabilistic edge map (PEM) is proposed in this paper. PEMs are generic and modality-independent. They generate similar anatomical representations from different imaging modalities and can thus guide a multi-modal image registration algorithm more robustly and accurately. The presented image registration framework is evaluated on a clinical dataset consisting of 10 pairs of 3D US-CT and 7 pairs of 3D US-MR cardiac images. The experiments show that a registration based on PEMs is able to estimate more reliable and accurate inter-modality correspondences compared to other state-of-the-art US registration methods.

## 1 Introduction

In cardiovascular minimally invasive procedures such as mitral valve repair and aortic valve implantation [12], pre-operative surgical plans and roadmaps may be complemented by intra-operative image guidance provided by ultrasound (US) and fluoroscopy. Computed tomography (CT) and magnetic resonance imaging (MRI) are widely used for planning of cardiovascular interventions since they can provide detailed images of the anatomy which are not usually visible in intra-operative cardiac US images. As the shape and pose of the heart changes over time during the intervention (e.g. due to respiration and patient motion), registration and fusion of pre- and intra-operative images can be a useful technology to improve the quality of image guidance during the intervention and subsequent clinical outcome. However, the difference in appearance of the heart in the different image modalities makes multimodal US image registration a challenging problem. In particular, the choice of an appropriate multimodal image similarity metric can be considered as the main challenge.

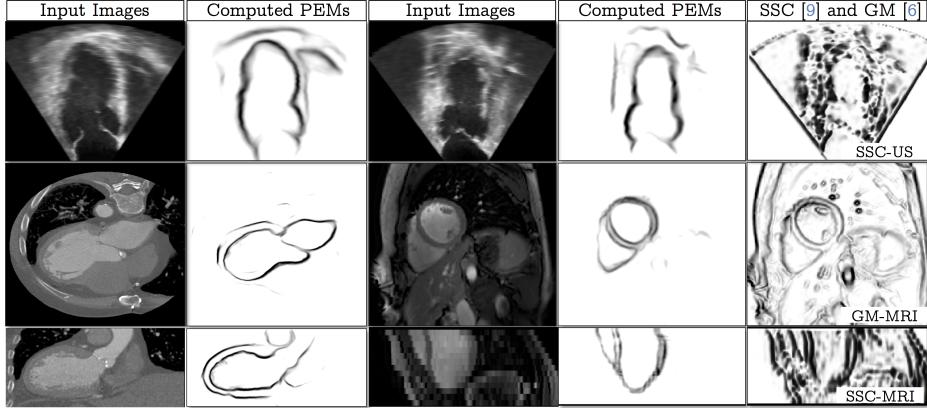


Fig. 1: Example of PEMs, obtained from cardiac images of different modalities, are compared to gradient magnitude (GM) and self-similarity descriptors (SSC). Since SSC is a vectorial descriptor, only a single component of SSC is visualized.

**Related work:** Current approaches for multi-modal US registration can be classified into two categories: (i) those using global and local information theoretic measures, and (ii) those reducing the multi-modal problem to a mono-modal problem. Early approaches in the first category suggested the use of normalized mutual information (NMI) [16] as a multimodal similarity metric. Recently, local similarity measures were proposed as an alternative to the global similarity measures. This reflects the fact that the US image characteristics show local variations due to wave reflections at tissue boundaries. Measures such as local NMI (LNMI) [10] and local correlation coefficient (LCC) [6] are designed to deal with this non-stationary behaviour. The second category of approaches converts the images from different modalities into a common feature space so that a mono-modal registration can be performed. In [11], fetal US and MR brain images are registered by generating pseudo-US images from MRI tissue segmentations. Other approaches, such as local phase [17], gradient orientations [4], and self-similarity descriptors (SSC) [9], generate similar structural representations from images to find correspondences between different imaging modalities.

**Contributions:** This paper proposes the use of probabilistic edge map (PEM) image representation for multi-modal US image registration. In contrast to local phase, SSC and gradient magnitude [6] (GM) representations, PEMs highlight only the image structures that are relevant for the image registration. This is because some of the anatomical structures visible in US images may not be visible in CT/MR images and vice versa, e.g. endocardial trabeculae, clutter, and shadowing artefacts. Therefore, extracting and registering only the left ventricle (LV) and atrial boundaries is a more robust approach than registering every voxel in the images. Moreover, PEMs generate a clear and accurate tissue boundary delineations, Fig. 1, in comparison to other image representations. This improvement produces better spatial alignment and increases the robust-

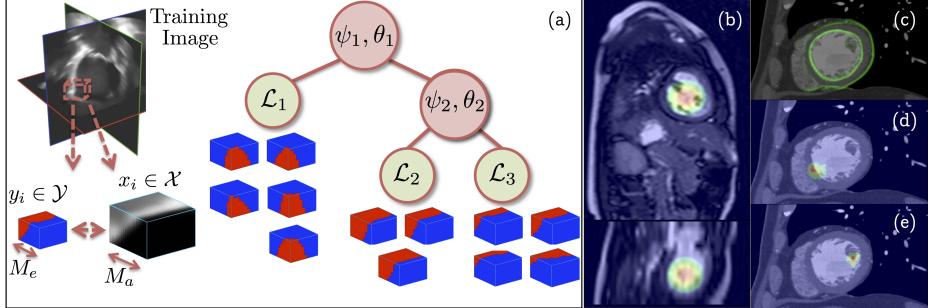


Fig. 2: Structured decision tree training procedure, label patches are clustered at each node split (a). Mid-ventricle (b), mid-septal (d) and mid-lateral (e) wall landmark localization by using PEMs (in green)(c) and regression nodes.

ness to noise which is important in US image registration. Additionally, PEMs are invariant under rigid transformations, unlike SSC descriptors. Thus they can cope with large rotational differences between images. Lastly, the proposed approach does not require tissue segmentations and can be generalized to other organs and modalities differing from the simulation based registration methods.

The proposed generic and modality independent representation is generated by a structured decision forest [5]. PEMs from different modalities are aligned with a block matching algorithm based on robust statistics followed by a deformable registration. The proposed registration framework is evaluated on US/CT and US/MR image pairs acquired from different pathological groups. The results show that PEM achieves lower registration errors compared to other image similarity metrics. This observation shows that PEM is a well-suited image representation for multimodal US image registration.

## 2 Methods

**Probabilistic Edge Map (PEM) Representation:** Cardiac US images mainly show the anatomy of the cardiac chambers and less detail of soft tissues. Therefore, further processing of these images, e.g. registration, requires a structural representation of the cardiac chamber boundaries such as the left ventricle (LV). In this work, a structural representation, PEM, is generated from a structured decision forest (SDF). Learning the image representation from training data increases the robustness in US images and allows the registration to focus on the organs of interest. As can be seen in Fig. 1, the PEMs outline only the LV and atrium while ignoring irrelevant anatomical boundaries. In addition the generated boundaries are cleaner and smoother.

PEMs are modality independent; as such the same representation can be generated from different modalities. More importantly, a single training configuration can be applied to train SDFs for all modalities, including the same image feature types and patch sizes. While SDFs are very similar to decision forests,

they possess several unique properties. In particular, in SDFs the output space  $\mathcal{Y}$  is defined by structured segmentation labels  $\mathbf{y}_i \in \mathcal{Y}$  rather than a single label  $y_i \in \{0, 1\}$ . SDF decision trees can be trained as long as the high-dimensional structural labels can be clustered into two or more subgroups at each tree node split  $\psi_j$  as illustrated in Fig. 2. This is achieved by mapping each patch label to an intermediate space ( $\Theta : \mathcal{Y} \rightarrow \mathcal{Z}$ ) where label clusters can be generated based on their pairwise Euclidean distances in  $\mathcal{Z}$  (cf. [5]). Once the training is completed, each leaf node  $\mathcal{L}_j$  stores a segmentation patch label  $\mathbf{y}_i \in \mathbb{Z}^{(M_e)^3}$  of size  $(M_e)^3$ . Additionally, the corresponding edge map  $\mathbf{y}'_i \in \{0, 1\}^{(M_e)^3}$  is stored as well. This allows predictions to be combined simply by averaging during inference.

Similar to decision forests, the input space  $\mathcal{X}$  for the SDFs is characterized by the high dimensional appearance features ( $\mathbf{x}_i \in \mathcal{X}$ ) extracted from image patches of fixed size  $(M_a)^3$ . These features are computed in a multi-scale fashion and correspond to intensity values, gradient magnitudes, six HoG-like channels, and local phase features. The weak learner parameters ( $\theta_j$ ), e.g. stump threshold value and selected feature channel id, are optimized to maximize Shannon entropy  $\mathcal{H}_1$  based information gain  $\mathcal{I}(S_j) = \mathcal{H}_1(S_j) - \sum_{k \in \{L, R\}} \frac{|S_j^k|}{|S_j|} \mathcal{H}_1(S_j^k)$  at each node split. Here  $S_j = \{\mathcal{T}_i = (\mathbf{x}_i, \mathbf{y}_i)\}$  is a set of patch labels and features reaching node  $j$ . During the PEM generation (testing time), each input voxel accumulates  $N_t \times (M_e)^3$  overlapping binary edge votes  $\mathbf{y}'_i$  from  $N_t$  number of trained decision trees. These predictions are later averaged to yield a probabilistic edge response. Spatial aggregation of a large number of edge votes results in a smoother and accurate delineation of the structures of interest.

**Structured Regression Forest Based Initialization:** The presented SDF can be modified to allow additionally for simultaneous voting for predefined landmark locations such as the apex, mid-lateral and mid-septal walls, as shown in Fig. 2. Each landmark point is detected independently in both target and source images. This can be used to obtain an initial rigid alignment as a starting point for the registration. Similar to Hough forests [7], classification nodes  $\psi_j$  are combined with regression nodes  $\Lambda_j$  in the tree structure. In this way, in addition to the patch labels  $\mathbf{y}_i$ , each leaf node contains an average location offset  $\mathbf{d}_i^n$  and confidence weight  $\sigma_i^n$  for each landmark  $n \in \{1, \dots, N\}$  (cf. [3]). Inclusion of these nodes does not introduce any significant computational cost at testing time but adds additional information about the location of the landmarks.

The set of offsets  $\mathcal{D}_i = (\mathbf{d}_i^1, \dots, \mathbf{d}_i^N)$  in the training samples  $\mathcal{T}_i = (\mathbf{x}_i, \mathbf{y}_i, \mathcal{D}_i)$ , is defined only for the voxels close to the boundaries of the ventricle. At each regression node  $\Lambda_j$ , the offset distributions are modelled with a multivariate Gaussian. The differential entropy of these distributions [3] is used as the uncertainty measure  $\mathcal{H}_2$  in the training of  $\Lambda_j$ . In training, classification and regression nodes are selected randomly [7]. During testing, landmark location votes are weighted by the confidence weights and edge probabilities of the voxels.

**Global Alignment of PEMs:** Similar to [11], a block matching approach is used to establish spatial correspondences between the PEMs (Fig. 3), where the normalized cross correlation (NCC) is used as a measure of similarity. As

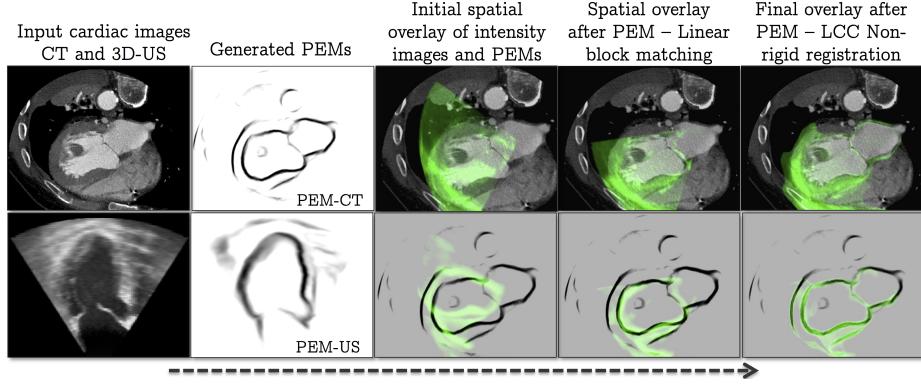


Fig. 3: A block diagram of the proposed multi-modal image registration method.

suggested in [13], the set of displacement vectors computed between the corresponding PEM blocks are regularized with least-trimmed squared regression before estimating the global rigid transformation  $\mathbf{R}$ . In this way, the influence of image blocks with no correspondences is removed, reducing the potential registration error introduced by shadowing or limited field-of-view in US images.

**Non-rigid Alignment of PEMs:** To correct for the residual misalignment, due to cardiac and respiratory motion, between target ( $P^T$ ) and source ( $P^S$ ) PEMs, B-spline FFD [14] based non-rigid alignment follows the global rigid registration. The total energy function, minimized with conjugate-gradient descent optimization, is defined as:  $E(\mathbf{T}) = -LCC(P^T, P^S \circ \mathbf{T} \circ \mathbf{R}) + \lambda BE(\mathbf{T})$ , where  $\lambda$  is the trade-off between the local correlation coefficient [2] (LCC) similarity metric and bending energy regularization ( $BE$ ). As the SDF classifier makes use of intensity values, local intensity variations in US images influence the edge probabilities in PEMs. For this reason, a local similarity measure is more suitable for PEMs than a global measure such as sum of squared differences. For similar reasons, LCC was used in [6] to align US-MR gradient magnitude images.

### 3 Experiments and Results

The proposed PEM registration framework is evaluated on 3D US-CT (10 pairs) and 3D US-MR (7 pairs) cardiac images. The CT images were acquired pre-operatively from adults after a contrast injection. The corresponding trans-esophageal US images were acquired during the cardiac procedure. In the second dataset, lower quality trans-thoracic US acquisitions are registered with multi-slice, cine-MR images (slice thickness 8 mm). Five of these pairs were acquired from children diagnosed with hypoplastic left heart syndrome and the rest from healthy adults. As a preprocessing step, all the images are denoised using non-local means [1] and resampled to isotropic voxel size of 0.80 mm per dimension.

In both experiments, PEM-LCC registration performance is compared with LNMI [10] and SSC [9] similarity based registration methods. Implementations of

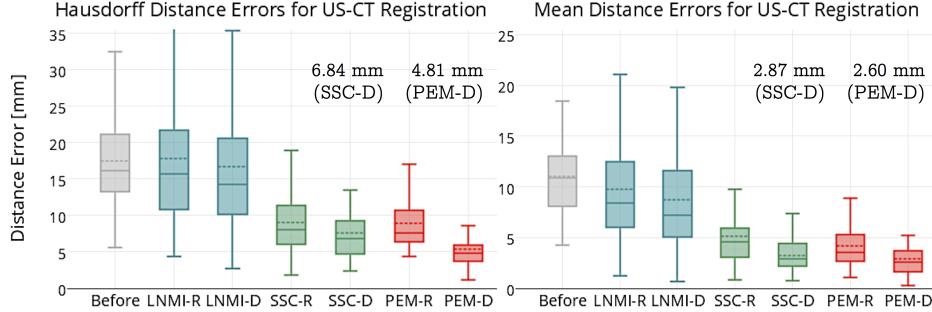


Fig. 4: US/CT registration errors after rigid (*R*) and deformable (*D*) alignment. The mean and median values are shown in dashed and solid lines respectively.

these methods were obtained from their corresponding authors. Particularly, in the SSC method, linear registration of the descriptors is performed by using least-trimmed squares of point-to-point correspondences for improved robustness. In the local alignment of SSC, a discrete optimization scheme (deeds [8]) is used to optimize SSC energy function. In LNMI, histograms with 64 bins are built locally and the registration is optimized using stochastic gradient descent. The number of bins is optimized for best performance.

In all three modalities, PEMs are generated and evaluated with the same training, testing and registration parameters. A PEM-SDF classifier was trained for each modality with set of cardiac images (50-80 images/modality) that is disjoint from the test set (17 pairs). It is important to highlight that the images from different modalities are not required to be spatially registered or come from the same subjects because the classifiers are trained separately for each modality. This significantly increases the availability of training data. The training data contains labels for myocardium, LV endocardium and atrium. The registration results are evaluated based on mean and Hausdorff distances of seven landmark points in the US images to closest points on the manually annotated LV endocardial surface in CT and MR images. The landmark points correspond to: apex, apical (2), basal (2), and mid-ventricle (2). For each method, the distances are computed after rigid and deformable alignment, and they are reported together with their values before the registration. For each image pair, 10 different registrations were performed with random global transformations for initialization.

**US/CT Evaluation Results:** The registration errors, provided in Fig. 4, show that local intensity statistics based similarity measures do not perform consistently and in some cases the registration fails to converge to the correct solution. This suggests that structural representation methods, such as PEMs and SSC, are more reliable measures in US/CT image registration. On the other hand, after local alignment, PEMs achieve lower registration errors compared to SSC, which can be linked to the accurate and smooth boundary representation provided by PEMs. As shown in Fig. 1, PEMs are less sensitive to noise and follow the anatomical boundaries more accurately.

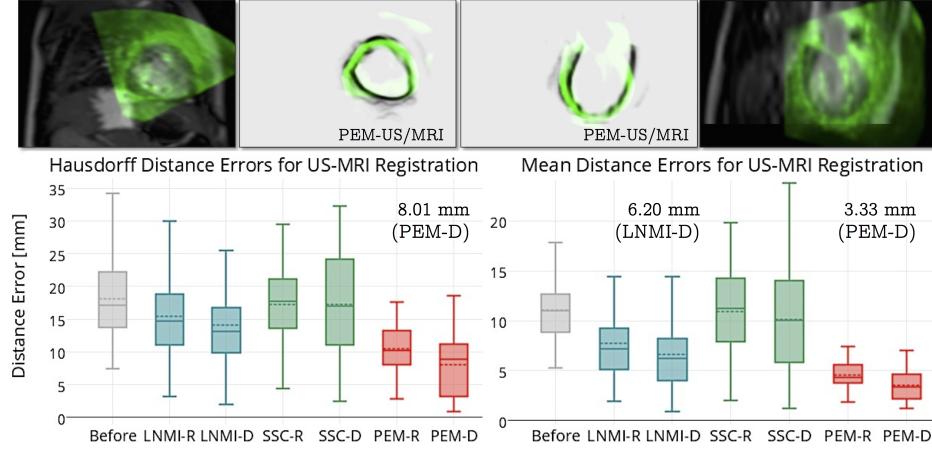


Fig. 5: Global alignment of US/MR images (top). The boxplot shows the US/MR landmark distance errors after rigid (*R*) and deformable (*D*) alignment.

**US/MRI Evaluation Results:** It is observed that the SSC descriptors do not provide enough guidance in alignment of MR images acquired from children. As shown in Fig. 5, the images contain more texture due to smaller size of the heart and trabeculae in transthoracic acquisitions. In contrast to this, PEM relies on annotations of the boundaries used in the training. It is able to generate a more accurate structural representation that is better matching with the myocardial boundary in MR images. The performance difference, shown in Fig. 5, is explained with this observation.

**Experimental Details:** All experiments were carried out on a 3.00 GHz quad-core CPU. The average computation time per registration was 73s for non-rigid registration, 21s for rigid alignment and 20s to compute each PEM. The training of the SDFs (70 min/tree) was performed offline prior to the registrations.

**Discussion:** In the experiments, after spatially aligning US/CT and US/MR images, we observed that the endocardial annotations in US images underestimate the LV volume compared to the annotated CT and MR images. In a clinical study [15], US volume measurement errors and underestimation were linked to the visible endocardial trabeculae. Therefore, to obtain matching PEM correspondences between the modalities, the US annotations are morphologically dilated before the training of the SDF. In contrast, the other structural representation techniques are expected to fail to achieve optimal local alignment accuracy as these tissues are not visible in CT and MR images (Fig. 5).

## 4 Conclusion

In this paper, a novel PEM image representation technique and its application on US multi-modality image registration has been presented. The experimental

results show that PEM provides a more accurate and consistent registration performance between different modalities compared to the state-of-the-art methods. Therefore, we can conclude that PEM is more suitable for image-guided cardiac interventions where alignment of pre-operative images to inter-operative images is required. Moreover, PEM is a generic and modular image representation method; it can be applied to any other US, CT and MR image analysis problem.

## References

1. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 60–65 (2005)
2. Cachier, P., Pennec, X.: 3D non-rigid registration by gradient descent on a Gaussian windowed similarity measure using convolutions. In: IEEE Workshop on Mathematical Methods in Biomedical Image Analysis. pp. 182–89 (2000)
3. Criminisi, A., Shotton, J., Robertson, D., Konukoglu, E.: Regression forests for efficient anatomy detection and localization in CT studies. In: Medical Computer Vision. Recognition Techniques and Applications in Med Imag, pp. 106–17 (2011)
4. De Nigris, D., Collins, D.L., Arbel, T.: Multi-modal image registration based on gradient orientations of minimal uncertainty. *IEEE TMI* 31(12), 2343–54 (2012)
5. Dollár, P., Zitnick, C.L.: Structured forests for fast edge detection. In: IEEE International Conference on Computer Vision. pp. 1841–48 (2013)
6. Fuerst, B., Wein, W., Müller, M., Navab, N.: Automatic ultrasound - MRI registration for neurosurgery using the 2D and 3D LC<sup>2</sup> metric. *Media* (2014)
7. Gall, J., Lempitsky, V.: Class-specific hough forests for object detection. In: IEEE Conference on Computer Vision and Pattern Recognition (2009)
8. Heinrich, H., Jenkinson, M., Brady, M., Schnabel, J.A.: MRF-based deformable registration and ventilation estimation of lung CT. *IEEE TMI* pp. 1239–48 (2013)
9. Heinrich, M.P., Jenkinson, M., Papiež, B.W., Brady, M., Schnabel, J.A.: Towards realtime multimodal fusion for image-guided interventions using self-similarities. In: MICCAI, pp. 187–94 (2013)
10. Klein, S., van der Heide, U.A., Lips, I.M., van Vulpen, M., Staring, M., Pluim, J.P.: Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. *Medical Physics* 35(4), 1407–17 (2008)
11. Kuklisova-Murgasova, M., Cifor, A., Napolitano, R., Papageorghiou, A., Quaghebeur, G., Rutherford, M.A., Hajnal, J.V., Noble, J.A., Schnabel, J.A.: Registration of 3D fetal neurosonography and MRI. *Media* 17(8), 1137–50 (2013)
12. Nguyen, C.T., Lee, E., Luo, H., Siegel, R.J.: Echocardiographic guidance for diagnostic and therapeutic percutaneous procedures. *Cardiovasc Diagn Ther* (2011)
13. Ourselin, S., Roche, A., Pennec, X., Ayache, N.: Reconstructing a 3D structure from serial histological sections. *Image and Vision Computing* 19(1), 25–31 (2001)
14. Rueckert, D., Sonoda, L., Hayes, C., Hill, D.L., Leach, M., Hawkes, D.J.: Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE T Med Imaging* 18(8), 712–21 (1999)
15. Shernand, S.: Evaluation of the left ventricular size and function. In: Comprehensive Atlas of 3D Echocardiography. Lippincott Williams and Wilkins (2012)
16. Studholme, C., Hill, D.L., Hawkes, D.J.: An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition* 32(1), 71–86 (1999)
17. Zhang, W., Noble, J.A., Brady, J.M.: Spatio-temporal registration of real time 3D ultrasound to cardiovascular MR sequences. In: MICCAI, pp. 343–50 (2007)