Learning Distance Function for Regression-Based 4D Pulmonary Trunk Model Reconstruction Estimated from Sparse MRI Data

Dime Vitanovski[a,c], Alexey Tsymbal[c], Razvan Ionasec[a], Bogdan Georgescu[a], Shaohua K. Zhou[a], Joachim Hornegger[b] and Dorin Comaniciu[a]

[a] Medical Informatics, Siemens Corporate Research, Princeton, NJ, USA
[b] Pattern Recognition Lab, Friedrich-Alexander-University, Erlangen, Germany
[c] Biomedical Informatics, Siemens Corporate Technology, Erlangen, Germany

ABSTRACT

Congenital heart defect (CHD) is the most common birth defect and a frequent cause of death for children. Tetralogy of Fallot (ToF) is the most often occurring CHD which affects in particular the pulmonary valve and trunk. Emerging interventional methods enable percutaneous pulmonary valve implantation, which constitute an alternative to open heart surgery. While minimal invasive methods become common practice, imaging and non-invasive assessment tools become crucial components in the clinical setting. Cardiac computed tomography (CT) and cardiac magnetic resonance imaging (cMRI) are techniques with complementary properties and ability to acquire multiple non-invasive and accurate scans required for advance evaluation and therapy planning. In contrary to CT which covers the full 4D information over the cardiac cycle, cMRI often acquires partial information, for example only one 3D scan of the whole heart in the end-diastolic phase and two 2D planes (long and short axes) over the whole cardiac cycle. The data acquired in this way is called sparse cMRI. In this paper, we propose a regression-based approach for the reconstruction of the full 4D pulmonary trunk model from sparse MRI. The reconstruction approach is based on learning a distance function between the sparse MRI which needs to be completed and the 4D CT data with the full information used as the training set. The distance is based on the intrinsic Random Forest similarity which is learnt for the corresponding regression problem of predicting coordinates of unseen mesh points. Extensive experiments performed on 80 cardiac CT and MR sequences demonstrated the average speed of 10 seconds and accuracy of 0.1053mm mean absolute error for the proposed approach. Using the case retrieval workflow and local nearest neighbour regression with the learnt distance function appears to be competitive with respect to "black box" regression with immediate prediction of coordinates, while providing transparency to the predictions made.

Keywords: sparse MRI reconstruction, pulmonary trunk, discriminative distance function, random forest, intrinsic random forest similarity, regression, case retrieval

1. DESCRIPTION OF PURPOSE

Congenital heart defects (CHD) are abnormalities in the heart's structure which often affect the right side of the heart and are the major cause of pulmonary trunk dysfunction which encompasses pulmonary vale stenosis, artesia and regurgitation. As CHD are common diseases for newborns, multiple interventions are required as the patient grows. Hence, interventional approach to the management of such complex conditions is becoming increasingly recognized as the preferred treatment option.[1] Percutaneous pulmonary valve intervention (PPVI) alternatives to classic cardiac surgery with important benefits: less invasive, reduced risks associated with cardiopulmonary bypass, bleeding, infections and reduced expenses for postoperative intensive care.[2]

Cardiac Computer Tomography (CT) or cardiac Magnetic Resonance Imaging (cMRI) are the modalities of choice for PPVI intervention planning.[3] In contrary to CT which covers the full 4D information over the cardiac cycle, the so called sparse 4D cMRI, encompasses only one 3D end-diastolic (ED) image of the whole heart and two orthogonal cine projections over the cardiac cycle (2D+t), short axis (SA) and long axis (LA). LA passes

---

through the main pulmonary artery and the descending aorta, while SA is aligned with the pulmonary valve, perpendicular to the LA (Fig. 1).

The main advantage of discriminative distance learning is the opportunity to combine the performance of strong learners with the transparency of nearest neighbor (NN) classification and regression, which may be desirable in particular in decision support.[4] A notable example of discriminative distance learning that can be applied both in classification and regression is the intrinsic random forest (RF) dissimilarity.[5] The main objectives of our study in this paper are to analyse the benefits of RF-based non-metric distance learning in comparison to using the orthodox Euclidean distance for NN regression, and also to empirically compare the performance of RF-based NN regression with the plain RF regression, in the context of solving a regression problem between the CT and the sparse cMRI data for full 4D model reconstruction from sparse cMRI.
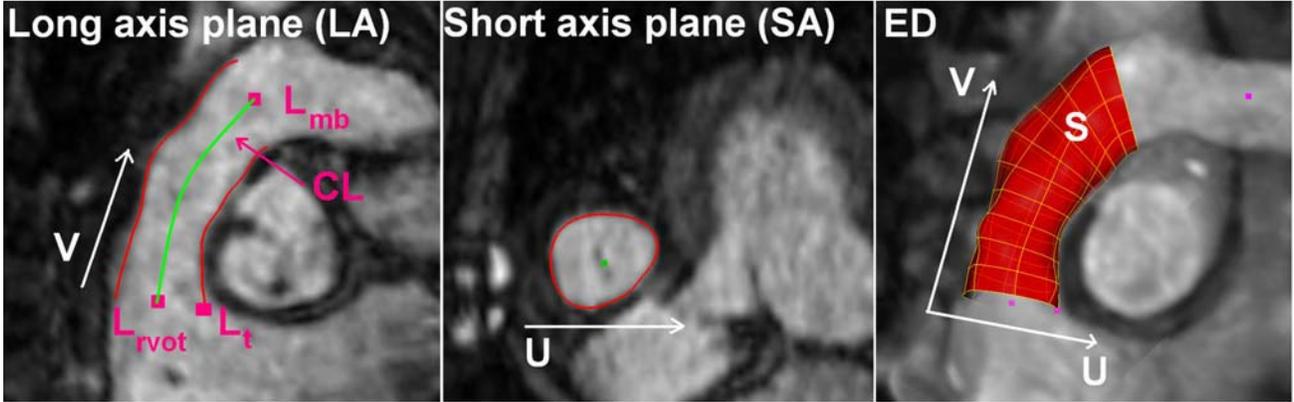
Figure 1. 3D MRI scan of the whole heart in the ED phase (**a**). 2D long axis (LA) plane (**b**) and short axis (SA) plane (**c**) of the pulmonary artery over the cardiac cycle

2. METHODS

The sparse 4D cMRI acquisition protocol provides dynamic information only in the two 2D orthogonal projections, as opposed to 4D cardiac CT, which provides full dynamics over the cardiac cycle. The reconstruction task is defined as a regression problem between the full dynamic model of the pulmonary trunk extracted from 4D CT data and the sparse one extracted from the sparse cMRI data:

$$Y(Model_{full4D}) = \hat{\Re}(X(Model_{sparse4D})) + \epsilon \tag{1}$$

We learned a non-metric distance function (Sec. 2.2) to solve the regression problem and to infer the incomplete temporal information characteristic to the sparse cMRI acquisition.

2.1 Physiological Patient-Specific Pulmonary Trunk Modelling

In this section we represent our anatomy driven model of the pulmonary trunk and the RVOT capable to express a large spectrum of morphological and dynamical variations. Similar to Vitanovski at al.,[6] the anatomical complexity is reduced by employing a coarse to fine parameterization which includes: anatomical landmarks, pulmonary artery center line and the full surface model of the pulmonary trunk. A set of clinical relevant anatomical landmarks (Trigone (L_t), RVOT (L_{rvot}) and Main-Bifucation (L_{mb}) form the first coarse to fine layer (Fig. 1(a)). Constrained by the landmarks and parameterized by 12 points, $CL = CL_0 \dots CL_{11}$, the centre line CL passes through the center of the pulmonary trunk. The considered physiological compliant model S, as illustrated in Fig. 1(c), is defined by a structured grid, spanned along two anatomical directions, $u - circumferential$ and $v - longitudinal$.

Given the different modalities supported and the characteristic imaging protocols, we differentiate among two dynamic extensions of the proposed physiological model. The definition of a full 4D model, which can be directly estimated from 4D cardiac CT data, is rather straightforward and realized by concatenating a time variable t:

$$Model_{full4D} = \{L_t, L_{rvot}, L_{mb}, CL, S\}_t \qquad (2)$$

However, given the sparse 4D acquisition, common to MRI exams, the extension to a temporal model includes two additional representations: LA (Fig. 1(a)) and SA (Fig. 1(b)). Hence, the sparse dynamic model is parameterized as follows:

$$Model_{sparse4D} = \{L_t, L_{rvot}, L_{mb}, CL, S\}_{ED} + \{LA, SA\}_t \qquad (3)$$

2.2 Learning non-metric distance for regression problems

During the last three decades, the importance of the distance function in machine learning has been gradually acknowledged.[7] However, for many years only canonical distance functions or hand-crafted distance functions were used. Even today in many areas including computer vision the Euclidean distance is one of the most widely used, though it is well known that its use is justified only when the feature data distribution is Gaussian.

There are several reasons that motivate the studies in the area of learning distance functions and their use in practise.[7] First, it is easy to show that choosing an optimal distance function makes classifier learning redundant. Next, learning distance functions breaks the learning process into two sequential stages (distance function learning followed by clustering or classification or regression), fostering the creation of more modular and thus more flexible systems, supporting component reuse. For example, the same distance functions learning module can be used with various classification or regression techniques and vice versa. Another important benefit is the opportunity for inductive transfer between similar tasks. While knowledge sharing at the level of creating a classifier is often non-trivial, one solution is to learn the same or similar distance function for each task. This approach is often used in computer vision applications; see e.g.[8] Learning a distance function helps to combine the power of strong learners with the transparency of NN classification or regression, a feature often desirable in decision support. Moreover, learning a proper distance function was shown to be especially helpful for high-dimensional data with correlated, weakly relevant and irrelevant features, where most traditional techniques would fail.

While historically the research on distance function learning has started from supervised learning of distance functions for k-nearest neighbour classification in the original "feature vector-object label" representation,[9] today by far the most commonly used representation, especially in computer vision, is the one based on so called equivalence constraints.[10] Interestingly, the distance learnt from labels is almost always metric, while learning from weak representations such as equivalence constraints usually provides more flexibility for learning arbitrary functions.[11] Usually, equivalence constraints are represented using triplets (x_1, x_2, y), where x_1, x_2 are data points in the original space and $y \in (+1, -1)$ is a label indicating whether the two points are similar (from the same class or cluster) or dissimilar.[7] Learning from these triples is also often called learning in the product space (i.e. with pairs of points as input); see[12],[13],[7] for examples. While learning in the product space is perhaps a more popular form of learning from equivalence constraints, yet another common alternative is to learn in the difference space, the space of vector differences; see[14],[15] for examples. The difference space is normally used with homogeneous high-dimensional data, such as pixel intensities or their PCA coefficients in imaging. While both representations demonstrate promising empirical results in different contexts, there is no understanding which representation is better and when.

Thus, learning from equivalence constraints usually results in a binary distance function, predicting only whether the two objects are similar or dissimilar. Though often, in order to avoid this limitation, the signed margin of margin-based classifiers such as SVM and boosting is used as the required distance function There are two essential reasons that motivate the use of equivalence constraints in learning distance functions; their availability in some learning contexts and the fact that they are a natural input for optimal distance function learning.[7] It can be shown that the optimal distance function for classification is of the form $p(y_i \neq y_j | x_i, x_j)$. Under the i.i.d. assumption the optimal distance measure can be expressed in terms of generative models $p(x|y)$ for each class as follows:[8]

$$p(y_i \neq y_j | x_i, x_j = \sum_y p(y|x_i)(1 - p(y|x_j))) \tag{4}$$

Because of these characteristics of equivalence constraints, many supervised learning algorithms, which receive canonical labelled points as input, often turn them into equivalence constraints and operate directly on them.[7] The drawbacks of distance learning from equivalence constraints, making it difficult to apply in the context of our problem, include the fact that it has been studied so far in the context of binary classification only and no applications to regression are known (although perhaps some existing approaches to distance learning from equivalence constraints might be extended to work with regression problems, which is an important direction for future research), and, second, usually the number of equivalence constraints representing a problem is big making application of sophisticated learning computationally expensive and sometimes even not feasible (for a typical classification or regression data set with a few thousand instances, the number of equivalence constraints generated may reach a few millions).

Another notable example of a non-metric distance function that can be employed both in classification and regression is the intrinsic RF dissimilarity. For a RF learnt for a certain problem, the proportion of the trees where two instances appear together in the same leaves can be used as a measure of their similarity.[5] For a given forest F the similarity between two instances x_i and x_j is calculated as follows: the instances are propagated down all T trees within F and their terminal positions $\mathbf{z}$ in each of the trees ($\mathbf{z}_1 = (z_{11}, , z_{1K})$ for x_1, similarly z_j for x_j) are recorded. The similarity between the two instances then equals to (I is the indicator function):

$$S(x_i, x_j) = \frac{1}{T} \sum_{t=1}^{T} I(z_{ik} = z_{jk}) \tag{5}$$

When dissimilarity or distance is needed and not a similarity (e.g., for clustering or multi-dimensional scaling) it is normally calculated as follows as suggested by Breiman:[5]

$$D(x_i, x_j) = \sqrt{1 - S(x_i, x_j)} \tag{6}$$

RF itself is known to possess a number of appealing properties. It was demonstrated to compare favorably with boosting and SVM in terms of predictive performance and to be often more robust with respect to overfitting noise. It is also known to provide a reliable likelihood estimate for its predictions, better than with boosting. Moreover, RF models are fast to train and apply, and both the training and application phases may be naturally parallelized. RF-based similarity (Eq. 5) has been successfully used, for instance, in bioinformatics[16] for hierarchical clustering of tissue microarray data. It is interesting that using this similarity for the most immediate task, NN classification or regression, is still rather uncommon, comparing to its use for clustering.[4]

The intrinsic RF dissimilarity (Eq. 6) is known not to be metric,[4] as the triangular inequality (Eq. 7) is often violated:

$$\forall x_i, x_j, x_k : D(x_i, x_j) \leq D(x_i, x_k) + D(x_k, x_j) \tag{7}$$

2.3 Estimating Patient-Specific Model Parameters

The abstract patient-specific model parameters proposed in Sec. 2.1 are determined using learning-based algorithms.[6] For both cardiac acquisition protocols, CT and cMRI, detectors (D^B, D^L, D^{CL} and D^S) are learned separately and applied in a four-step approach by first detecting bounding boxes at the RVOT and the Bifurcation planes, landmarks, followed by center line and full dynamic surface model estimation.

By defining the object localization task as a classification problem, the patient-specific model parameters are estimated within the Marginal Space Learning (MSL) framework.[17] For both modalities, CT and cMRI, the detectors (D^B, D^L, D^{CL} and D^S) are learned separately using the Probabilistic Boosting Tree (PBT)[18] in combination with 3D Haar-like and Steerable feature from a training dataset annotated by experts.

2.3.1 Bounding Box and Landmark Detection

Each anatomically relevant structure of the pulmonary trunk is defined by one bounding box, parameterized by an similarity transformation $\theta = (X, Y, Z, \alpha, \beta, \gamma, S_x, S_y, S_z)$. The RVOT plane is defined with the θ_{rvot} centred at L_{rvot} and orientated with the tangent of the centre line sampled at CL_0. The main bifurcation plane is included within the parameterization of the second bounding box θ_{mb} centred at the L_{mb} and orientated with the tangent of the centre line sampled at CL_n.

Efficient learning of the high dimensionality of θ_{rvot} and θ_{mb} is achieved by marginalizing the search space by means of MSL[17] into small subspaces which are gradually increased. Consequently, the training process is splitted into three steps: position, position-rotation and full similarity transformation learning, where classifiers, parameterized by the parameters of the current subspace, are sequentially trained.

Each estimated bounding box initialize the landmarks by its position ($L_{rvot} = \theta_{rvot}^{xyz}$, $L_{mb} = \theta_{mb}^{xyz}$) with exception of the RVOT landmark which is initialized by the scale of θ_{rvot} along the X-axis ($L_{rvot} = \theta_{rvot}^{xyz} + \theta_{rvot}^{\alpha} * \theta_{rvot}^{sx}$) and constrains the search domain Ω_i by its scale. A further accuracy improvement is achieved by learning a discriminative 2 Level PBT classifier $\vec{D}_L$ which learns the target distribution

$$p(\vec{L}_i | x_l, y_l, z_l, \boldsymbol{I}) = \vec{D}_i(x_l, y_l, z_l | \boldsymbol{I}), \ (x_l, y_l, z_l) \in \Omega_i, \ i = 1..3 \tag{8}$$

in 1mm volume resolution, where $p(L_i | x_l, y_l, z_l, I)$ is the probability of the presence of $\mathbf{L}_i$ at location (x_l, y_l, z_l).

2.3.2 Center Line Detection

In this step we first initialize the centre line as a straight line bounded by the previously detected landmarks L_{rvot} and L_{mb}. A centre line detector D^{CL} is then applied to refine the centre line which goes through the centre of the pulmonary trunk.

The detector D^{CL} is trained with PBT and 3D Haar-like features. In the learning phase positive and negatives are automatically computed from manually annotated data sets and used to train the detector D^{CL}. In the detection stage, an incremental approach is used for searching 3D centre points CL_x on a series of parallel planes. These are bounded by the corresponding landmarks, while their normal is given by the initial centre-line and updated after each detected centre point. Finally, a least-square approach is applied to fit the centre line into the set of the detected discrete 3D points.

2.3.3 Boundary Detection

In the first stage of the boundary detection a mean model, averaged over the training data set, is placed into the patient specific anatomy. Projection of the mean model is achieved by computing a piecewise rigid transformation between the detected centre line and the mean one. At equally sampled centre line points $CL_{P1}...CL_{Pn}$ a coordinate system is computed by taking the tangent at the sampled point as Z-axis and the RVOT landmarks as X-axis. Mesh points from the mean model which are close to the coordinate system CL_{Px} are then projected into the patients coordinates.

Accurate global fitting of the pulmonary trunk model into the patient's specific anatomy is achieved by projecting the mean model. However, for precise object delineation further local processing is required. A boundary detector D^S learned with PBT on 3D Haar-like and Steerable features is applied along the surface normal to test a set of hypothesis and to move the surface toward the true boundary. In case of CT, the detectors are applied in each time step to obtain the full temporal model $\{L_t, L_{rvot}, L_{mb}, CL, S\}_t$. In case of cMRI, a full model is estimated only in the end-diastolic frame S_{ED} and is used to initialize the contours LA and SA. These are refined using a trained D^S detector. A full dynamic 4D model is then estimated by using a learned regression model (Eq. 1) to predict the missing temporal information.
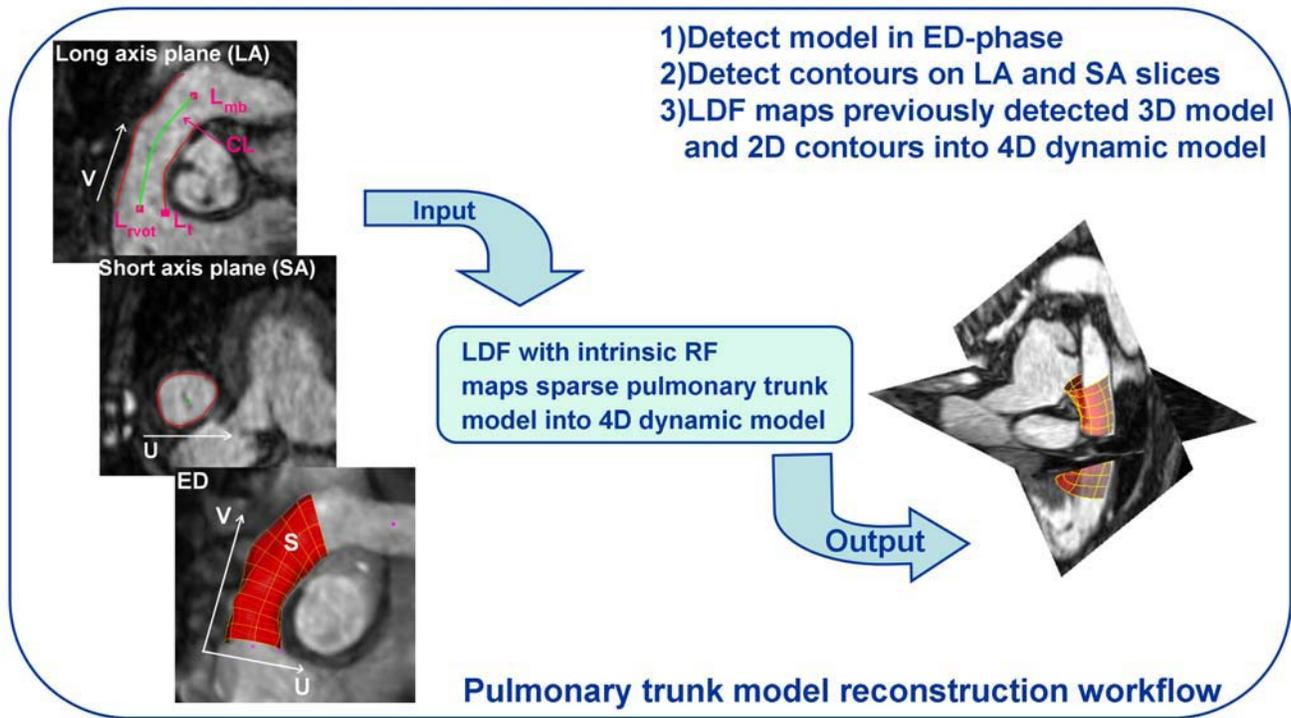
Figure 2. Workflow for pulmonary trunk reconstruction from sparse MR images

3. RESULTS

Table 1 summarizes the results for solving the regression problem with plain k-NN and Learned Distance Function (LDF) with RF. For the RF we used 25 trees, the number of features to be considered at each node was one third of the total number of features and the minimum number of instances in a leaf was set to 3. In contrary to the intrinsic RF, for the plain RF regression minimum number of instance in a leaf was one. Table 2 summaries the results for LDF with RF versus plain RF regression.

The experimental results demonstrate a clear benefit of RF-based distance learning in comparison to the use of plain Euclidean distance in the NN regression, in our context. Mean absolut error(MAE) for the intrinsic RF distance is more than 2 times less than the MAE for the Euclidean distance (Table 1).

The experimental results also demonstrate that the performance of the instrinsic RF distance-based NN regression is comparable to the performance of plain RF regression, with only insignificant differences for the best parameter settings (Table 2). This is perhaps caused by the fact that the background learning process is same in the both approaches. However, the key difference is that in RF-based NN regression, a predefined number of instances (k) only is influencing the final decision, while in the orthodox RF all instances falling into the tree leaves corresponding to the instance in question will influence the final decision.

number of neighbours	1	3	7	15	31
k-NN	0.2311	0.2312	0.2433	0.2302	0.2241
RF based LDF	0.1155	0.1123	0.1108	0.1105	0.1105

Table 1. k-NN vs RF based LDF

number of trees	1	5	10	25	50	100
simple RF	0.1287	0.1146	0.1106	0.1105	0.1083	0.1074
RF based LDF	0.1284	0.1122	0.1097	0.1081	0.1076	0.1053

Table 2. simple RF regression vs RF based LDF

4. CONCLUSION

In this paper we propose an approach for reconstruction of the pulmonary trunk from sparse MR images by learning a distance function with intrinsic random forest. We also show how we can use model of the pulmonary trunk from CT images for solving the reconstruction task. The reconstructed model can be utilized to extract morphological and functional information of the pulmonary trunk over the cardiac cycle. Extensive experiments performed on 80 cardiac CT and MR sequences demonstrated the average speed of 10 seconds and accuracy of 0.1053 mm mean absolute error for the proposed approach. Using the case retrieval workflow and local nearest neighbor regression with the learnt distance function appears to be competitive with respect to "black box" regression with immediate prediction of coordinates, while providing transparency to the predictions made.

REFERENCES

[1] *Percutaneous pulmonary valve implantation for right ventricular outflow tract dysfunction: guidance* (2007).

[2] Bonhoeffer, P. e. a., "Percutaneous insertion of the pulmonary valve," *Journal of the American College of Cardiology* **39**, 1664–1669 (May 2002).

[3] Taylor, A., "Cardiac imaging: Mr or ct? which to use when," *Pediatr Radiology* **38** (2008).

[4] Tsymbal, A. Huber, M. and Zhou, S., "Discriminative distance functions and the patient neighborhood graph for clinical decision support," (2010).

[5] Breiman, L., "Random forests," in [*Machine Learning*], 5–32 (2001).

[6] Vitanovski, D., Ionasec, R., Georgescu, B., Huber, M., Taylor, A., Hornegger, J., and Comaniciu, D., "Personalized pulmonary trunk modeling for intervention planning and valve assessment estimated from CT data," in [*MICCAI*], 17–25 (2009).

[7] Bar-Hillel, A. and Weinshall, D., "Learning distance function by coding similarity," in [*ICML '07: Proceedings of the 24th international conference on Machine learning*], 65–72 (2007).

[8] Mahamud, S. and Hebert, M., "The optimal distance measure for object detection," (2003).

[9] Short, R. and Fukunaga, K. *IEEE Transactions on Information Theory* **27**, 622–627 (1981).

[10] Hertz, T., *Learning distance functions: algorithms and applications*, PhD thesis, The Hebrew University of Jerusalem (2006).

[11] Jacobs, D., Weinshall, D., and Gdalyahu, Y. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* **22**, 583–600 (2000).

[12] Hertz, T., Bar-Hillel, A., and Weinshall, D., "Learning distance functions for image retrieval," *International Conference on Computer Vision and Pattern Recongnition* **2**, 570–577 (2004).

[13] Zhou, S., Shao, J., Georgescu, B., and Comaniciu, D., "Boostmotion: Boosting a discriminative similarity function for motion estimation," *International Conference on Computer Vision and Pattern Recongnition* **2**, 1761–1768 (2006).

[14] Yu, J., Amores, J., Sebe, N., and Tian, Q., "Toward robust distance metric analysis for similarity estimation," *International Conference on Computer Vision and Pattern Recongnition* **1**, 316–322 (2006).

[15] Amores, J., Sebe, N., and Radeva, P. *International Conference on Computer Vision and Pattern Recongnition* **27**, 201–209 (2006).

[16] Shi, T. and Horvath, S., "Unsupervised learning with random forest predictors," **15**, 118–138 (2006).

[17] Zheng, Y., Barbu, A., and et al., "Fast automatic heart chamber segmentation from 3d ct data using marginal space learning and steerable features," *ICCV* (2007).

[18] Tu, Z., "Probabilistic boosting-tree: Learning discriminativemethods for classification, recognition, and clustering," in [*ICCV*], (2005).