LANDMARK-BASED SEGMENTATION OF LUNGS WHILE HANDLING PARTIAL CORRESPONDENCES USING SPARSE GRAPH-BASED PRIORS

Ahmed Beshes\textsuperscript{(1,2)} and Nikos Paragios\textsuperscript{(1,2)}

\textsuperscript{(1)} Laboratoire MAS, Ecole Centrale Paris, Châtenay-Malabry, France \textsuperscript{(2)} Equipe GALEN, INRIA Saclay - Île-de-France, Orsay, France

ABSTRACT

In this paper, we propose a new segmentation algorithm that combines a graph-based shape model with image cues based on boosted features. The landmark-based shape model encodes prior constraints through the normalized Euclidean distances between pairs of control points, alleviating the need of a large database for the training. Moreover, the graph topology is deduced from the dataset using manifold learning and unsupervised clustering. In a graph-matching-like manner, we formulate the segmentation task as a labeling problem where we seek to match the model landmarks to image points that are extracted using the boosted classifiers. We also propose to overcome the limitation of missing correspondences by incorporating an additional label to account for outliers. Then, we repair the outlier positions to complete the segmentation. State-of-the-art discrete optimization techniques are used to provide our experimental results for the segmentation of the right lung in 2D chest radiographs, demonstrating the potentials of our method.

Index Terms— Shape Modeling, Sparse Graphs, Graph Rigidity, Segmentation, Machine Learning, Clustering, MRF, Outliers

1. INTRODUCTION

Organs segmentation is a fundamental problem in medical image analysis. As prior knowledge about the structures of interest is often available, this task can benefit from incorporating shape models in the segmentation process. Linear models like point distribution models (or active shape models) \cite{1} have been widely used in this context to constrain the solution space, through an explicit representation of the shape. Such linear models were also used in the implicit representation framework \cite{2}, or level-sets. The estimation of the shape density in a kernel space in \cite{3} overcomes the limitation of the linearity assumption. However, the training of these global models that aims to estimate a high-dimensional distribution from a small number of examples represents their main limitation.

In this paper, we adopt a local point-based representation of anatomical structures. In order to deal with the non-linear structure of the manifold and the lack of sufficient samples to estimate the corresponding density, we propose a Markov Random Field (MRF)-based model. The vertices of the graph correspond to landmarks lying on the shape boundary, whereas the edges of the graph encode the dependencies between the landmarks. Towards computational efficiency, we approximate the complete graph by the so-called \textit{k-fan graph}, which structure is determined from the training set of shapes using manifold learning and unsupervised clustering techniques. The inter-point constraints are enforced using the learned (one-dimensional) probability distributions of the normalized pairwise Euclidean distances. This approach that was previously introduced in \cite{4} and \cite{5} is further developed by linking the proposed graph structure with graph rigidity and Euclidean geometry theory.

The next step consists of combining the shape model with the image cues for organs extraction. Classically, image support can be introduced through edges, regional statistics, point correspondences or some combination of the above. Edges are simple and computationally efficient measurements, and were often considered in the active contours framework. Their limitations appear when the initial condition is far from the optimal one, or when one has to deal with low contrast images or in the presence of discontinuities, leading to erroneous segmentation results. Regional statistics are more global measurements that are relatively robust with respect to the initial conditions, but also fail in the case of highly textured images where the assumption of statistical separation between the object and the background is not satisfied. As an alternative to image intensity distributions, one can consider feature vectors extracted from the image. Such high-dimensional descriptors benefit from the neighboring data around a particular image location, and can exhibit strong discriminative power. In our approach, we seek optimal landmark positioning. Therefore, a discriminative image support that isolates highly probable control point locations in the image is of a great interest. Indeed, combined with a shape prior, feature-based detectors can yield very powerful segmentation tools. The authors of \cite{6} pursue this line of thought as they use a chain-structured shape prior to find the best landmark positions among candidates extracted from the image, based on locally orderless image (LOI) features. In the same spirit, the work of \cite{7} focuses on symmetry interest points detected using Gradient Vector Flow (GVF)-based cues, and enforce a shape prior through a MRF formulation. A possible extension of our shape model is presented in \cite{8}, where high order terms are used to enforce a similarity-invariant shape prior by considering all possible triplets of control points. This model is combined with Gabor-filter-based features to segment calf muscles in MR images. However, this recent approach does not scale well when many points are needed to model the organs. Hence, the computational complexity due to the high order energy terms represents its main limitation.

Similarly, we propose to build image classifiers that discriminate the control points from the background. These classifiers provide landmark candidates, with associated confidence values. The segmentation task boils down to finding the trade-off between the shape prior penalties, and candidate selection costs, in an integrated manner. In practice, we adopt a similar approach to the LOI for feature extraction. Then, we apply the Adaboost algorithm \cite{9} to the computed local descriptors in order to train the landmark classifiers. Unlike the aforementioned methods that considered a predefined topology of the MRF, we learn our \textit{k-fan graph} structure from the training data. Moreover, we propose a similarity invariant representation in a general framework that is applicable in 2D or in 3D.
We observed in our experiments that in some difficult cases, the classifiers may fail to provide good candidates to the segmentation algorithm, which leads to a suboptimal result. This is another limitation of the aforementioned related methods. We propose to account for the missing correspondence problem by incorporating an additional label in the segmentation formulation to account for outliers. Although this idea has already been mentioned in [10] and [11], the authors do not provide a clear method to post-process the outlier landmarks, and the impact of the procedure is not shown in their experiments. We address these limitations in this work.

The results that we obtain show that the k-fan graph is a good approximation to the complete graph, with the advantage of faster computations. They also show that accounting for the outliers in the segmentation process can lead to improved results qualitatively and quantitatively in the case of the complete graph.

2. STATISTICAL SHAPE MODELING

We recall here our statistical shape model (SSM) that was first introduced in [4] and [5]. Then we introduce its relationship with Euclidean geometry and graph rigidity.

2.1. Complete Graphs and Euclidean Distance Constraints

Knowledge-based segmentation methods rely on the definition of a model which is then combined with image support towards object extraction. Traditional approaches consist of representing the shape using a number of landmarks and learning their behavior using a training set [1].

In this work, we consider that a shape is defined by its boundary (a curve in 2D, or a surface in 3D) as well as a set of control points. Knowing the positions of these controls points, the shape boundary can be (approximately) retrieved by the use of an interpolation method. The control points can in general belong to the boundary, or not. In practice however, it is natural that such landmarks are chosen to lie on the contour of the object, which exhibits very often distinctive features. Therefore, we will assume in the following such a hypothesis, meaning that the control points are boundary points.

We represent a shape as a Markov Random Field (MRF). Let \( \mathcal{G} = (\mathcal{V}, \mathcal{E}) \) be a graph where the nodes \( \mathcal{V} = \{x_1, \ldots, x_n\} \) represent the control points positions, and where the set of edges \( \mathcal{E} \) represents the conditional dependencies between the landmarks. We assume for the time-being that this graph is complete, that is we take into account all the pairwise dependencies between the landmarks.

We would like to enforce invariance properties to our shape model, with respect to translations, rotations, and global scale changes. Such similarity-invariance characteristics enable to dismiss object changes due to the pose variations and permit to learn the intrinsic variability of the object. Discarding the transformation parameters in the inference step from the unknowns is also an advantage that makes an invariant shape model more tractable and easier to use. First, seeking a representation that does not depend on an absolute position, the probability of co-occurrence of a pair of control points can be then written as the probability of their difference vector, or:

\[
p(x_i, x_j) \approx p(x_i - x_j). \tag{1}
\]

Invariance with respect to rotation is equivalent to ignoring any orientation direction, that to consider the Euclidean norm of the difference vector, or:

\[
p(x_i, x_j) \approx p(\|x_i - x_j\|). \tag{2}
\]

One can question the use of this chord length representation for shape modeling. Associated with the complete graph, this representation is equivalent to considering the Euclidean distance matrix (EDM) corresponding to the control points:

**Definition 1. Euclidean Distance Matrices (EDM).**

A matrix \( D = [d_{ij}] \in \mathbb{R}^{N \times N} \) is an EDM if there exist \( N \) points \( (x_1, \ldots, x_N) \in \mathbb{R}^n \), for some \( n \geq 1 \), such that \( \forall 1 \leq i, j \leq N, d_{ij} = \|x_i - x_j\|_2 \). The set of points \( (x_1, \ldots, x_N) \) is called a realization of \( D \) in \( \mathbb{R}^n \).

EDM are invariant with respect to translation, rotation and reflection [12]. Hence, two sets of points \( (x_1, \ldots, x_N) \) and \( (p_1, \ldots, p_N) \) have the same EDM if and only if they are related by an isometry. This observation justifies the use of EDM (as being a "strong constraint") to represent shapes.

Eventually, let us consider \( \tilde{d} = \frac{2}{n(n-1)} \sum_{i<j} \|x_i - x_j\| \) the mean Euclidean distance between the control points. This mean chord length can be considered as an estimate of the shape scale. Hence, dividing the chord lengths by \( \tilde{d} \) gives a representation that is invariant to scale changes as well, or:

\[
p(x_i, x_j) \approx p\left(\frac{\|x_i - x_j\|}{\tilde{d}}\right). \tag{3}
\]

Rigorously, in equation (3), \( p(x_i, x_j) \) depends on all the control points through the scale term \( \tilde{d} \). However, we will circumvent this difficulty in the experiments by estimating \( \tilde{d} \). Moreover, this scaling factor simply extends the result of [12] to the similarity-invariant case. Upon combining all these pairwise similarity-invariant constraints, the graph energy is defined (up to an additive constant) as:

\[
E_{\text{shape}}(x_1, \ldots, x_n) = -\alpha \sum_{(i,j) \in \mathcal{E}} \log \left( p\left(\frac{\|x_i - x_j\|}{\tilde{d}}\right)\right), \tag{4}
\]

where \( \alpha \) is a weighting factor. Minimizing this energy with respect to the control point positions \( x_i \) constrains the solution w.r.t. the shape manifold. Let us consider now a set \( \mathcal{S} = \{s_1, \ldots, s_r\} \) of \( r \) instances of the object, where each example is represented using \( n \) control points, i.e. \( s_u = \{x_{u1}, \ldots, x_{un}\} \). Hence, \( \forall u \in \{1, \ldots, r\} \), the set \( X_u = \{x_{u1}, \ldots, x_{un}\} \) represents instances of the \( u \)th control point of the shape. In practice, this training set is obtained by manually labeling the landmarks for each instance of the shape, or by deducing the landmarks from the registration between a labeled shape and a set of non-labeled shapes. Note that an alignment of the shape is not needed before the learning phase since our representation is intrinsically invariant to similarity transforms.

Then, given a statistical model, we learn from the training set the probability density distributions of the relative positions of the control points \( p_{ij} \equiv p(x_i, x_j) \). These \( \frac{n(n-1)}{2} \)-d densities allow us to describe the information contained in the training set. In practice, we used a Gaussian kernel density estimation to compute the probability distributions. Interestingly, we only need to estimate one-dimensional densities, which alleviates the need of large training datasets.

However, this representation suffers from redundancy and could be computationally expensive and hard to optimize during inference, because of the complete graph structure. We describe in the following how we approximate the complete graph to remove the redundancy from our model, in order to obtain a sparser representation. But before proceeding, we discuss some geometric aspects that will enlighten the redundancy encapsulated by the complete graph.
2.2. Graph Rigidity and k-trees

This section is built on the work in [13] where the notion of global rigidity of a graph is defined in an intuitive manner:

**Definition 2.** Globally rigid graphs.
A graph $G = (V, E)$ in $\mathbb{R}^n$ is said to be globally rigid if the lengths of the edges in $E$ determine uniquely the lengths of the edges of the complement graph (the "missing" edges).

This result suggests that *globally rigid graphs* in $\mathbb{R}^n$ are good approximations of the complete graph in $\mathbb{R}^n$. Let us now introduce the definition of *k-trees* (see Fig. 1(a)), which are particular graph structures that are of great interest in our modeling:

**Definition 3.** k-trees, base k-cliques.

The complete graph on $k$ vertices is a $k$-tree. A $k$-tree $T_n$ on $n + 1$ vertices can be generated from a $k$-tree $T_n$ on $n$ vertices, by adding a new vertex to $T_n$ and connecting it with $k$ edges to some $k$-clique (a clique with $k$ vertices) of $T_n$. This latter clique is called then a base $k$-clique.

Then, the authors of [13] show that a $k$-tree with all base $k$-cliques in general position in $\mathbb{R}^{k-1}$ (i.e. do not lie in a $k - 2$-dimensional vector subspace) is globally rigid in $\mathbb{R}^{k-1}$. We now extend this property to a larger case of graphs by invoking the following result:

**Lemma 1** ([14]). Every $k'$-tree $T'$ with $k' \geq k \geq 1$ possesses a spanning $k$-tree $T$ with the property that every $k'$-clique of $T'$ contains a $k$-clique of $T$.

Then, it follows that the property proven in [13] for the $k$-trees in $\mathbb{R}^{k-1}$ is generalizable. It is easy to show that a $k'$-tree, with $k' \geq k$, and with all subsets of size $k$ of its base $k'$-cliques being in general position in $\mathbb{R}^{k-1}$ is globally rigid in $\mathbb{R}^{k-1}$. In other words, given the above general position condition, a $k$-tree is globally rigid in $\mathbb{R}^n$, with $k > n$, and hence is equivalent to the complete graph model.

This result shows that such structures encode sufficient information and dependencies to correctly model the shape, from a geometric point of view. It also suggests that the complete graph model is redundant, and hence can be made sparser. In the next step of our modeling we tackle the redundancy problem, and we build a $k$-tree graph structure where the parameter $k$ is learned from the data, and where the particular structure of the graph is related to the training shape population.

2.3. The k-fan Shape Model

We propose a general framework to learn a $k$-tree graph structure from the training data. First, we propose to define a behavior-explaining distance between the control points. Mutual information, the Bhattacharyya coefficient or the Hellinger distance can be used to reflect their joint behavior. Alternatively, we considered in practice the shape maps distance [15] that combines the notion of compactness of models with manifold embedding using diffusion maps. Then, we propose to cluster the control points based on this distance into $k$ clusters. We used a linear-programming-based unsupervised clustering method [16] that determines automatically the number of clusters and sets the cluster centers as prototypes. Affinity propagation is an alternative clustering method that exhibits the same properties.

Hence, a cluster of landmarks is defined as a set of control points of highly correlated behavior. Thus, such a group of points carries a high amount of redundant information in terms of shape variability. This observation drives us to factor out the redundancy using the obtained clusters, through the center landmarks. The cluster centers being the best representatives of their respective groups, we make the mild assumption that for a given control point, that is not a cluster center, the joint landmark-cluster information can be encoded as landmark-cluster center information. Hence, the landmarks will be considered to be conditionally independent, knowing the cluster centers.

Then, it follows that the property proven in [13] for the complete graph model is generalizable. It is easy to show that a $k$-tree is globally rigid in $\mathbb{R}^n$. Let us now introduce the definition of *k-fans* (see Fig. 1(b)), following [17]. The novelty here consists in the method that defines automatically from the training data the number of clusters and their centers. We also presented in section 2.2 the geometric rationale behind this graph structure. To each one of these pairs $(x_i, x_j)$ we associate a probability density distribution $p_{ij}$ learned from the training set as previously stated in section 2.1.

We will denote the $k$-fan graph by $G_k = (V, E_k)$. Without loss of generality, we can assume that the base $k$-clique corresponds to the base $k$-clique in $V$. Let $R_k$ be the subset of edges in $E_k$ that correspond to the base $k$-clique. The $k$-fan model assumes that a node $x_i$ is conditionally independent from a node $x_j$ (with $k + 1 \leq i < j \leq n$) knowing the base $k$-clique. The shape energy in equation (4) can now be approximated as:

$$E_{shape}(x_1, \ldots, x_n) \approx - \sum_{(i,j) \in E_k} \gamma_{ij} \log \left( p \left( \frac{\|x_i - x_j\|}{d} \right) \right),$$

where we introduce the relative weighting $\gamma_{ij}$ between the reference edges and the non-reference edges as:

$$\gamma_{ij} = \begin{cases} \alpha & \text{if } (i,j) \in R_k \\ \beta & \text{if } (i,j) \in E_k \setminus R_k \end{cases}.$$

Intuitively, the $k$-fan shape model is analogous to the global positioning system (GPS) where the positions of the non-reference control points are defined with respect to the positions of the control points in the base $k$-clique.

3. DETECTOR-BASED SEGMENTATION

3.1. Boosted Landmark Detectors

We will explain in the following how we learn a classifier for each control point of the object, using features that are computed from the image. These trained classifiers are then used as detectors in our algorithm, to segment unseen images. Our method consists of
extracting from the image a set of candidates for each landmark, and then choosing the best ones in terms of detection response on one hand, and global fidelity to the learned shape prior on the other hand.

We present our method by describing the features and the classifiers that we use in practice. However, the framework is flexible, and different types of detectors can be plugged in, using for instance the gradient vector flow (GVF) descriptors that were considered in [10], or the locally orderless image (LOI) features that were adopted in [6].

In our work, we first process the image $I$ using a filter bank of derivatives for different Gaussian smoothing values $\sigma$ ($\sigma \in \{0.5, 1, 2, 4, 8\}$). The output of this process is a set of feature images as showed in Fig. 2. Such a feature extraction was previously considered by [18] and [6]. Then, they obtained locally orderless image (LOI) descriptors [19] by computing the first statistical moments of the feature images in local neighborhoods. In [6], the feature vector probability function is modeled using a multivariate Gaussian distribution for each blurring level $\sigma$. Our derivation is different as we build for each control point of the object a “boosted” classifier $f^{(i)}_T$, using AdaBoost [9]. To this end, at each control point location $x_i$, we patches of size $17 \times 17$ from the feature images, amounting to a $K$-dimensional descriptor $\varphi_i$ (with $K = 8959$), as depicted in Fig. 2. Hence, given a set of $N_p$ training images, where the positions of the landmarks are known, we obtain for each control point $x_i$, a set of “positive” examples of size $N_p$. We also extract $N_n$ feature vectors from the background to form a set of “negative” examples. Then, using the training set of $N = N_p + N_n$ examples $(\varphi^{(i)}, y^{(i)})_{1 \leq i \leq N}$, with the class label $y^{(i)} \in \{-1, 1\}$ referring to the background/object, we train the classifiers $f^{(i)}_T$ in order to discriminate the control points from their neighborhoods. Hence, $f^{(i)}_T(\varphi)$ is the response of the $i^{th}$ boosted classifier w.r.t. the feature vector $\varphi$.

We define the data-related energy term $E_{\text{image}}$ thanks to the learned classifiers. Using a logistic function, we turn the strong classifier response into a probability measure as $p_i \left( f^{(i)}_T(\varphi) \right) = \frac{1}{1 + \exp \left( - f^{(i)}_T(\varphi) \right)}$. Then, by assuming independence of the feature vectors at different image locations, we can write:

$$E_{\text{image}} \left( x_1, \ldots, x_n, I \right) = \sum_{i=1}^n - \log \left( p_i \left( f^{(i)}_T(\varphi(x_i)) \right) \right),$$

where $\varphi(x_i)$ is the feature vector at the landmark $x_i$ location.

### 3.2. Segmentation as a Labeling Problem

We formulate the task of segmenting an image $I$ as a trade-off between the shape prior and the image cues by minimizing:

$$E \left( x_1, \ldots, x_n, I \right) = E_{\text{shape}} \left( x_1, \ldots, x_n \right) + E_{\text{image}} \left( x_1, \ldots, x_n, I \right),$$

with respect to $(x_1, \ldots, x_n) \in \Omega^n$, $\Omega$ being the image domain. Let us consider now that each position in the image domain $\Omega$ corresponds to a label $l \in L$, $L$ being the labeling set. We will note $x_i(l)$ placing the landmark $x_i$ at the location labeled by $l$. Then, minimizing the energy (8) is equivalent to solving the labeling problem:

$$\min_{l_1, \ldots, l_n} \sum_{l \in L} E(l_1, \ldots, l_n) = \sum_{i=1}^n V_i(l_i) + \sum_{(i,j) \in E_k} V_{ij}(l_i, l_j),$$

where the unary and pairwise potentials are defined as:

$$V_i(l_i) = - \log \left( p_i \left( f^{(i)}_T(\varphi(x_i(l_i))) \right) \right),$$

$$V_{ij}(l_i, l_j) = - \gamma_{ij} \log \left( p \left( \frac{||x_i(l_i) - x_j(l_j)||}{d} \right) \right).$$

However, searching in the whole image domain $\Omega$ increases the complexity of the problem. Therefore, we restrict the search in practice to the image locations where the classifiers provide their highest $m$ responses (for each landmark), $m$ being a parameter that is fixed by the user. Hence, the problem boils down to finding the best configuration among the top $m$ control points candidates. For efficiency, we used an approximate optimization methods to solve (8), namely the sequential tree-reweighted message passing algorithm (TRW-S) [20]. In TRW-S, a lower bound on the optimal energy is maximized. We observed in our experiments that the lower bound value was very often equal to the graph energy at the end of the optimization, meaning that a global optimum was found. The results we obtained are presented in the next section. Considering a relatively small number of candidate points allows to solve the problem efficiently. The results can be however let down if there are no good landmarks in the candidate pool. We address this limitation in the following.

### 3.3. Handling the Outliers

The above formulation can be in some difficult cases insufficient to obtain satisfactory segmentations because of the limitation of the detected candidates. The proposed algorithm relies heavily on the quality of the detectors and the provided costs, and hence a poor generalization performance of the classifier may affect the whole segmentation process. To account for the missing correspondences problem, we consider the idea of an additional artificial candidate per landmark, as suggested in [10]. The $m+1^{th}$ candidate will refer to finding an outlier. The critical aspect is then to define the corresponding unary and pairwise costs. [10] define these costs as being proportional to the mean of the one computed for the real candidates. On one hand, this cost choice introduces an additional parameter (of proportionality) that needs to be adjusted. On the other hand, using the current test image to compute these costs may be unsuitable as these computations will be dependent on the number of candidates.

![Fig. 2. Features computation using a filter bank. Derivatives up to order 2 of the image are computed after applying Gaussian filters, to form the feature images. Then, image patches (in red) are extracted around a given position to form a feature vector.](image-url)
A different number of candidates may result in different costs, and hence different outliers detection. Moreover, the generalization performance of the classifier will have a direct impact on the outlier cost in this case. An alternative way to define these costs relies on their statistical estimation from the training set, where the energy terms of the available solutions (segmentations) can be easily evaluated. Hence, we compute from the training set the mean and standard deviation of each unary and pairwise cost, and set the outliers penalties to the mean plus two standard deviations.

The post-processing of the detected outlier landmarks is another important aspect when handling such difficult cases. Hence, one has to define a strategy to “repair” the outlier landmark positions, and obtain an allowable segmentation. It is not clear in [10] how this step is accounted for. [11] proposed to relax the positions of landmarks that have high shape energy terms to lower shape energy locations. However, the used method is not clearly explained. We provide now a way to solve for the missing outlier positions.

Augmenting the candidate set by an outlier label provides a partial segmentation of the object. In practice, after minimizing (9), most of the landmarks positions are determined, and only a few of them are set to be outliers. Then, the next step consists to locate these outliers according to the shape prior, and to the set of fixed (already solved) landmarks. Let \( O \) be the set of outliers, and let \( X \) be the set of fixed landmarks. Then iteratively, we search the optimal set of displacements \( \mathbf{d}^* \), \( i \in O \) of the outlier landmarks such that a shape energy is minimized. Let \( D_i = \{d_{i1}, \ldots, d_{ik}\}, i \in O \) be a quantization of the possible displacements \( d_{i\ell} \). Hence, we iteratively look for:

\[
\{d_{ik}\}_{k \in O} = \arg\min_{d_{ik} \in D_k} E_{\text{outliers}}(d_{ik}'),
\]

where:

\[
E_{\text{outliers}}(d_{ik}') = \sum_{(i,j) \in O \times O} - \log \left( p_{ij} (x_i + d_{ik}', x_j + d_{jk}') \right)
+ \sum_{(i,j) \in O \times X} - \log \left( p_{ij} (x_i + d_{ik}', x_j) \right).
\]

This final step is hence also expressed as a labeling problem, where each quantized candidate displacement is mapped to a label. In this iterative process, we chose in practice to start from initial outliers positions that correspond to the top candidate positions. We use the FastPD algorithm [21] to solve the energy (11).

### 4. EXPERIMENTAL VALIDATION

We applied our method to the segmentation of the right lung in 2D radiographs. We used an available public dataset [22][23] of 247 images of healthy and non healthy subjects (presenting nodules). The database contains gold standard segmentations from radiologists, that provided a delineation of the organ. Gold standard segmentation masks are hence available as well as corresponding landmark positions lying on the contour. We tested our algorithm on rescaled images of size 256 × 256. The database was split in two subsets (containing both healthy and non healthy cases). The first subset (corresponding to the odd indexed images) was used for training, and the second subset was used for testing. As explained previously in section 2, we used the annotated data in the training set to learn the structure of the graph and the prior model. We used the 44 available landmarks, and the clustering lead to the result presented in Fig. 3(a) which shows the 7 automatically determined clusters. The clustering was used to define the \( k \)-fan structure of the graph, and the pairwise normalized distance distributions were learned using Gaussian kernel density estimation. The boosted classifiers \( f_i^{(k)} \) were also learned using the 124 images of the training set according to the described scheme in section 3.1. The Fig. 3(b) shows also that we used for each node regions of interests (ROIs) to speed up the candidate search step. The limits of these ROIs were learned from the training data, as done in [6]. We estimated the value of the scaling factor \( d \) using the top candidate positions. In order to evaluate the performance of our method, we compared the segmentation results that we obtained using a complete graph against using the \( k \)-fan learned graph (without accounting for the outliers). In both cases, the maximum number of candidates \( m \) was set to 20. As in [22] and [6], we used the overlap coefficient \( \Gamma = \frac{TP}{TP+FP+FN} \) (where TP stands for true positive, FP for false positive, and FN for false negative) and the mean curve distance error (MDE) to assess the results quantitatively. They are summarized in Table 1. We also report the results in [6], where 40 control points were used. The distributions of these quantities in the testing set are also represented by the box-plots in Fig. 7. Examples of the obtained segmentations are shown in Fig. 4. The quantitative and qualitative results back our claims regarding the learned graph structure: we obtain a sparse model, and we maintain a performance that is equivalent to the complete graph. We also slightly improve the quantitative results w.r.t. [6]. From the computational point of view, after pre-computing the unary and pairwise terms, the computational cost is linear with the cardinality of the landmarks, and constant with the number of training data, and is very low compared to the training cost of the learning algorithm.

### Table 1. Overlap coefficient and Mean Distance Error between the manual gold standard and the obtained right lung segmentation.

<table>
<thead>
<tr>
<th></th>
<th>( \Gamma ) (%)</th>
<th>MDE (pixels)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete Graph</td>
<td>94.77(±2.37)</td>
<td>1.39(±0.64)</td>
</tr>
<tr>
<td>( k )-fan Graph</td>
<td>94.74(±2.31)</td>
<td>1.40(±0.62)</td>
</tr>
<tr>
<td>Complete Graph + Outliers</td>
<td>94.95(±2.48)</td>
<td>1.34(±0.65)</td>
</tr>
<tr>
<td>( k )-fan Graph + Outliers</td>
<td>93.87(±3.74)</td>
<td>1.62(±0.87)</td>
</tr>
<tr>
<td>MISC algorithm [6]</td>
<td>93.9(±3.1)</td>
<td>1.49(±0.63)</td>
</tr>
</tbody>
</table>

Fig. 3. Modeling the right lung using the training set. (a) The unsupervised linear programming-based clustering lead to 7 clusters. Cluster centers are represented using squares. (b) A region of interest is used for each landmark to speed up the segmentation algorithm.
wise terms, the optimization of the complete graph lasts about 30s, whereas optimizing the \(k\)-fan graph approximately lasts 10s. The use of the \(k\)-fan graph also decreases the memory needs to store the pairwise terms (this storage space is linear in the number of graph edges). These two aspects can be critical in applications with a high number of control points.

Moreover, as described earlier, we learned from the training set unary and pairwise costs for the outlier label, and evaluated its impact on the segmentation results. Table 1 and the boxplots in Fig. 7 show that the quantitative performance of the segmentation algorithm is improved when using the complete graph and handling outliers. We show in Fig. 5 examples of images where outliers were detected, and where the segmentation using the complete graph was improved. The first example in Fig. 6 also suggests that the performance of the \(k\)-fan graph can be improved by accounting for the outliers. However, as we can see in the two last examples of Fig. 6, the method performs better in the case of the complete graph. This observation is confirmed by the quantitative results in Table 1 and the boxplots in Fig. 7. They show that the computed performance indicators are slightly impaired when using outliers with the \(k\)-fan graph. This difference in the performance of the \(k\)-fan graph and the complete graph may be explained as follows: selecting a landmark as being an outlier can be thought as eliminating it from the graph, and destroying all its connections. Suppressing these geometric constraints is clearly more critical for the \(k\)-fan graph than for the complete graph, especially if landmarks of the base \(k\)-cliques are labeled as outliers (recall the geometric rationale behind the \(k\)-fan graph structure, presented in section 2.2).

5. DISCUSSION

In this paper, we have developed our previously introduced landmark-based discrete shape model that encodes the prior using pairwise normalized Euclidean distance distributions. We consider two graph structures along with this representation: the complete graph structure that is justified by Euclidean distance geometry, and the \(k\)-fan graph structure which rationale is related to the notion of global graph rigidity. A particular \(k\)-fan graph instance is learned from the data. Then, we have combined this shape prior with boosted classifier responses to segment the right lung in chest radiographs. This new algorithm is robust w.r.t. missing correspondences because...
it accounts for outliers, and repairs their positions.

Despite the promising results, there are possible improvements to this method. In our approach the selection of the model was based on purely geometric properties and characteristics. Then, during inference, the model was associated with image cues. A future research direction would bridge the gap between these two aspects by combining shape and appearance in the learning step.

6. REFERENCES


Fig. 7. Quantitative evaluation of the segmentation of the right lung: boxplot representation of the distributions of (a) the overlap coefficient $\Gamma$ and (b) the mean curve distance error (MDE). The box lines represent the lower quartile, the median, and upper quartile values. The whiskers extend to a range of 1.5 times the interquartile range. The values beyond this range are drawn with a “+” sign.