Segmenting Articulated Structures by Hierarchical Statistical Modeling of Shape, Appearance, and Topology

Rok Bernard, Boštjan Likar, Franjo Pernuš

University of Ljubljana, Faculty of Electrical Engineering, Tržaška 25, Ljubljana, Slovenia
{rok.bernard, bostjan.likar, franjo.pernus}@fe.uni-lj.si

Abstract. This paper describes a general method for segmenting articulated structures. The method is based on statistical parametrical models, obtained by principal component analysis (PCA). The models, which describe shape, appearance, and topology of anatomic structures, are incorporated in a two-level hierarchical scheme. Shape and appearance models, describing plausible variations of shapes and appearances of individual structures, form the lower level, while the topological model, describing plausible topological variations of the articulated structure, forms the upper level. This novel scheme is actually a hierarchical PCA as the topological model is generated by the PCA of the parameters obtained at the lower level. In the segmentation process, we seek the configuration of the model instances that best matches the given image. For this purpose we introduce coarse and fine matching strategies for minimizing an energy function, which is a sum of a match measure and deformation energies of topology, shape, and appearance. The proposed method was evaluated on 36 X-ray images of cervical vertebrae by a leave-one-out test. The results show that the method well describes the anatomical variations of the cervical vertebrae, which confirms the feasibility of the proposed modeling and segmentation strategies.

1   Introduction

Ascertaining the detailed shape and organization of anatomic structures is important not only within diagnostic settings but also for tracking the process of disease, surgical planning, simulation, and intraoperative navigation. Accurate and efficient automated segmentation of articulated structures is difficult because of their complexity and inter-patient variability. Furthermore, the position of the patient during image acquisition, the imaging device itself, and the imaging protocol induce additional variations in shape and appearance. To deal with the variations, a segmentation method should use as much available prior information on shape, location, and appearance of the analyzed structures as possible. When segmenting articulated structures, like the spine, knee, or hand, prior knowledge on topology, i.e. organization of anatomical structures, should also be considered.

   In recent years, a great variety of shape and appearance models have been proposed as a source of prior knowledge and applied to various tasks in medical image analysis [1]. Efficient models should be general to deal with inter-patient variability and yet specific to maintain certain anatomical properties [1, 2]. Models,

which are trained on a set of labeled training images meet these requirements and have therefore received much attention. For example, point distribution models, active shape models, and active appearance models, all proposed by Cootes *et al.* [3, 4], were successfully applied to bony structures, e.g., vertebrae [5], spine [6], knee joint [4], hand [7], rib cage [8] or hip and pelvis [9], most often for segmentation purposes.

Articulated structures exhibit two kinds of shape variations, i.e. variations in shapes of individual parts and variations in spatial relationships between the parts. Such combined variations cannot be optimally described by a single linear model unless variations of spatial relationships are sufficiently small and a sufficiently large training set is used [6]. Therefore, alternative approaches are required to describe the non-linear shape variations. This can be assessed by a piecewise linear models [10] or by separately modeling the variations of spatial relationships and variations of shapes of individual parts [7]. The problem with piecewise linearization is that it can only approximate the non-linear shape variations without using prior knowledge on organization of articulated structures, while in [7] the prior knowledge is used only for model initialization and not throughout the matching process.

In this paper we propose a general statistical hierarchical modeling of shape, appearance, and topology of articulated structures, which efficiently deals with non-linear shape variations and incorporates prior knowledge on organization of articulated structures. The hierarchical scheme is comprised of two levels. Shape and appearance models, which describe individual structures form the lower level, while the topological model, which describes the organization of anatomical structures, forms the upper level and supervises spatial relations between individual models at the lower level. The proposed method is applied to the segmentation of cervical spine vertebrae.

2   Hierarchical Scheme

To build up a general scheme that can describe the shape and appearance variations of anatomical structures, such as vertebrae, and the topological variations of the articulated structures, e.g. the cervical spine, we use the principal component analysis (PCA), which is a well-known statistical tool [11]. By PCA the principal variations of average shape, appearance, and topology can be derived from a set of representative training images.

2.1   Principal Component Analysis

Principal component analysis (PCA) is based on the statistical representation of a random variable [11]. Suppose we have a random vector population $\mathbf{x}$ and the mean of that population is denoted by $\overline{\mathbf{x}}$; $\overline{\mathbf{x}} = E(\mathbf{x})$. The covariance matrix of the same data set is $\mathbf{C}$:

$$\mathbf{C} = E\big((\mathbf{x} - \overline{\mathbf{x}})(\mathbf{x} - \overline{\mathbf{x}})^T\big) \ . \tag{1}$$

From a symmetric matrix such as **C** we can define an orthogonal basis by finding its eigenvalues and eigenvectors. By ordering the eigenvectors $\mathbf{\phi}_i$ in the order of descending eigenvalues $\lambda_i \geq \lambda_{i+1}$, one can create an ordered orthogonal basis with the first eigenvector having the direction of largest variance of the data. Data may be reconstructed by a linear combination of orthogonal basis vectors. Instead of using all the eigenvectors of the covariance matrix, we may represent the data in terms of only a few basis vectors of the orthogonal basis. Let t largest eigenvalues and corresponding eigenvectors be retained to form the matrix $\mathbf{\Phi}$; $\mathbf{\Phi} = (\mathbf{\phi}_1 | \mathbf{\phi}_2 | \ldots | \mathbf{\phi}_t)$. Knowing $\bar{\mathbf{x}}$ and matrix $\mathbf{\Phi}$, we can reconstruct the input data vector **x**:

$$\mathbf{x} \approx \bar{\mathbf{x}} + \mathbf{\Phi}\mathbf{y} \ , \tag{2}$$

from the parameters **y** of the statistical model. If the data is concentrated in a linear subspace, this provides a way to compress data without losing much information and simplifies the representation. Alternatively, the input data vector **x** can be transformed into vector **y**:

$$\mathbf{y} = \mathbf{\Phi}^T (\mathbf{x} - \bar{\mathbf{x}}) \ . \tag{3}$$

By the above statistical model we can describe shape, appearance, and topology of an articulated structure as shown below.

2.2   Shape

Each structure is described by a statistical shape model as proposed by Cootes *et al.* [3]. The model is derived from a set of training shapes. Each training shape is composed of anatomical points defined in training images. Prior to defining the mean shape of a structure, the training shapes are rigidly aligned [3]. Shape variations are found by the PCA of training sets of anatomical points and represented by the most significant eigenshapes.

2.3   Appearance

The appearance, i.e., the texture of each structure is modeled on shape-free training images, obtained by elastic registration of training shapes and mean shape. Thin-plate splines interpolation between corresponding anatomical points is used for this purpose [12]. By applying PCA to the set of shape-free training images, defined on a region of interest covering a structure, the mean image and the most significant eigenimages are extracted.

2.4   Topology

To describe topological variations of an articulated structure we need to correlate variations in pose and shape of all structures. We propose to apply the PCA on pose and shape parameters of all structures, which were obtained in the shape model generation step. In this way, the most significant eigentopologies describe the

anatomically plausible topological variations. This novel strategy can be viewed upon as a hierarchical PCA. The topological PCA (upper level of hierarchy), describing plausible topological variations of an articulated structure, is constructed from sets of parameters generated by shape PCAs and corresponding pose parameters (lower level of hierarchy) that describe plausible variations of shapes and poses of individual structures. In this way, the topological PCA enables the supervision of the spatial relations between shapes of structures, which form the articulated structure.

3   Segmentation

The above hierarchical scheme consists of parametrical models that describe shape, appearance, and topology of an articulated structure. Once extensively trained, it incorporates a valuable prior knowledge that can be used efficiently for describing the image of the articulated structure. We consider model-based image segmentation by searching the configuration $\mathbf{L}$ of the model instances that best match the given image I. The best configuration $\mathbf{L}^*$ may be found by the maximum a posteriori (MAP) estimation:

$$\mathbf{L}^* = \arg\max_{\mathbf{L}} P(\mathbf{L} \mid I) \; . \tag{4}$$

Bayes rule then implies:

$$\mathbf{L}^* = \arg\max_{\mathbf{L}} P(\mathbf{L})P(I \mid \mathbf{L}) \; . \tag{5}$$

The prior $P(\mathbf{L})$ is given by the probability distributions of shapes, appearances, and topology. The likelihood function $P(I|\mathbf{L})$, measures the probability of observing image I given a particular configuration $\mathbf{L}$. The standard approach to finding the MAP estimation is to minimize the energy function $F(I,\mathbf{L})$ obtained by taking the negative logarithm of a posteriori probability:

$$\mathbf{L}^* = \arg\min_{\mathbf{L}} F(I,\mathbf{L}) \; . \tag{6}$$

The required matching strategy and the energy function are given in the following sub-sections.

3.1   Matching Strategy

Consider the configuration $\mathbf{L}$ describing an articulated structure composed of N structures where each structure is described by t_s shape, t_a appearance, and t_p pose parameters. The number of all parameters is $N^*(t_s + t_a + t_p)$, possibly causing a demanding optimization problem. To overcome this problem, we can elegantly omit the appearance parameters as they may be estimated from the image patch defined by the shape model. This reduces the number of parameters to $N^*(t_s + t_p)$. We name this optimization strategy a fine matching strategy. We consider also a coarse matching strategy by which the number of parameters can be further significantly reduced to $t_P + t_T$ by tuning only t_P global pose and t_T topological parameters of the articulated

structure at the upper level in the hierarchy. Global pose and topological parameters then drive pose and shape parameters of individual structures at the lower level of hierarchy. The coarse and fine matching strategies are considered in the following.

3.1.1   Coarse Matching

In the coarse matching step, illustrated in Fig. 1a, we tune only global pose and topological parameters that in turn drive pose and shape parameters of the structures. According to these parameters, each shape model, describing a corresponding structure, generates a shape that defines a patch on the underlying image. The patch is then elastically transformed to the shape-free form, which is fed into the appearance model that yields appearance parameters and approximates the given shape-free image patch. Finally, the match measure between the shape-free image patch and its approximation is calculated. The obtained match measure is part of the energy function that is used for selecting the global pose and topological parameters for the next iteration in the optimization process. The energy function, which considers also topology, shape, and appearance deformation energies, is described latter.
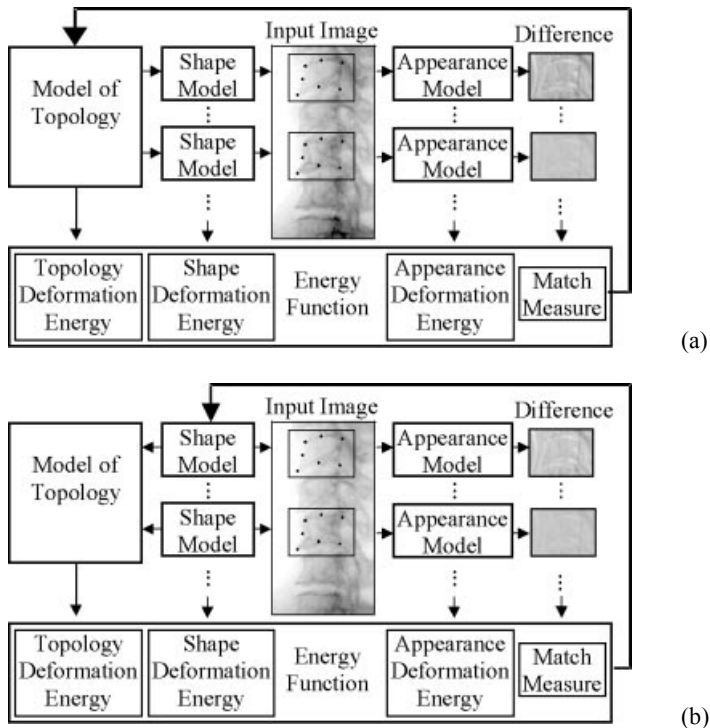

Fig. 1. Coarse matching strategy (a): global pose and topological parameters are optimized and fine matching strategy (b): pose and shape parameters of all structures are optimized simultaneously

3.1.2   Fine Matching

In a fine matching strategy, illustrated in Fig. 1b, the pose and shape parameters of all structures are optimized simultaneously, whereas the model of topology only supervises the spatial relations between shapes of structures via the topology deformation energy in the energy function.

3.2   Energy Function

To suppress anatomically implausible configurations we define the energy function $F(I,\mathbf{L})$ as a weighted sum of match measure $M(I,\mathbf{L})$ and topology $F_T(\mathbf{L})$, shape $F_S(\mathbf{L})$, and appearance $F_A(\mathbf{L})$ deformation energies:

$$F(I,\mathbf{L}) = \alpha \cdot M(I,\mathbf{L}) + F_T(\mathbf{L}) + F_S(\mathbf{L}) + F_A(\mathbf{L}) \quad, \tag{7}$$

where α is a regularization parameter weighting the match measure against deformation energies.

The mean square of intensity differences between shape-free image patches and their corresponding approximations was chosen as the match measure:

$$M(I,\mathbf{L}) = \sum_{i=1}^{N} \frac{1}{\Omega_i} \frac{1}{V_i} \sum_{j \in \Omega_i} r_{i,j}^2 \quad, \tag{8}$$

where $r_{i,j}$ is the intensity difference of j-th pixel in an image patch i defined on a region of interest Ω_i and V_i is the variance of the sum of the squares of intensity differences [13].

The deformation energies are calculated as a weighted sum of corresponding PCA parameters. The weights correspond to the probability density functions of PCA parameters. In this way, a configuration $\mathbf{L}$ that is not anatomically plausible is penalized.

4   Results

The proposed method was evaluated on 36 X-ray images of cervical vertebrae by a leave-one-out test. The annotated images were taken from the NHANES II X-ray database [14]. Vertebrae 3, 4, 5, and 6 were modeled by placing seven landmarks on each of them. The number of shape parameters t_s was set to 4 (capturing ~72% of all shape variations), the number of appearance parameters was set to 3 (capturing ~85% of all appearance variations), and the number of topological parameters t_T was set to 2 (capturing ~40% of all topological variations). The regularization parameter α was set to 100. The weighting function of PCA parameters was defined as $W(y_k)=sign(|y_k|-b)*(|y_k|-b)$, so that the parameters y_k had no influence on the corresponding deformation energy if lying inside the interval $[-b, b]$. The values of b were 1, 0, and 1 for shape, appearance and topological parameters, respectively. The simulated annealing global optimization method was used for energy minimization [14].

In the leave-one-out test the method was trained on 35 images and then tested on the remaining image. The initialization of the method, which provided global pose

parameters of the cervical spine model, was performed by selecting two points, one on vertebra 3 and one on vertebra 6. Points were selected in centers of the vertebrae and then perturbed by a constant distance, which was quarter of the vertebra size, in 10 different directions. After applying the method, the resulting landmark positions were compared to the manually defined gold-standard positions by calculating root mean square (RMS) error separately for each of the vertebrae. The RMS errors were calculated for the initial landmark positions and after the coarse and fine matching steps. In 80% of the cases the initial RMS error was reduced. Fig. 2 shows two cervical spine X-ray images. In Fig. 2a the landmarks were accurately placed by the proposed method, while in Fig. 2b the method failed because of the poor contrast on vertebrae 3 and 4, large osteophyte on vertebra 5, and partial overlapping of vertebra 6 and shoulder. Also a well-trained human operator hardly identifies vertebrae landmarks on such images.

   The initial RMS errors and RMS errors after coarse and fine matching steps are shown in Fig. 3. The resulting RMS errors were on the average 2.2 pixels, while the initial RMS error ranged on the average from 5.5-7.5 pixels. The coarse matching step succeeded to locate the landmarks to pixel accuracy, while the fine matching step only slightly improved the localisation.
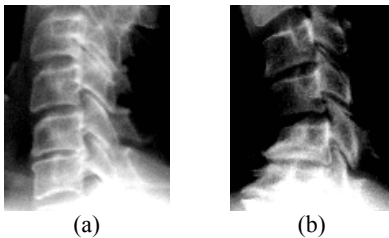

(a)   (b)

Fig. 2. Cervical spine X-ray images on which the method was successful (a) and on which the method failed (b)
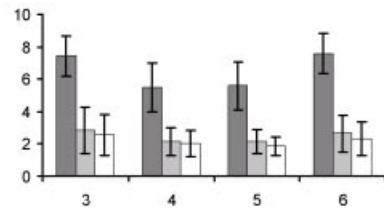
Fig. 3. Mean RMS errors and corresponding standard deviations (in pixels) for initial landmark positions (dark gray) and after coarse (light gray) and fine (white) matching steps

5   Conclusion

In this paper we presented a general method for segmenting articulated structures exhibiting variations in shape, appearance and topology. The method is based on statistical parametrical models that are incorporated in a two-level hierarchical scheme. The lower level describes the shape and appearance of individual structures, while the upper level controls the topology of the articulated structure. When segmenting a given image, the anatomically plausible configuration of the models is searched for in coarse and fine matching steps. The segmentation results on 36 X-ray images confirmed the applicability of the proposed modeling and segmentation strategies. We will focus our future efforts on extensive evaluation of the method on a larger number of spine X-ray images. The proposed hierarchical statistical modeling of shape, appearance, and topology is an important breakthrough for describing non-

linear shape variations of articulated structures. Further development and refinement of this methodology should remain an important area of research in the near future.

Acknowledgements

References

1. McInerney, T., Terzopoulos, D.: Deformable models in medical image analysis: A survey. Medical Image Analysis 1 (1996) 91-108.
2. Jain, A.K., Zhong, Y., Dubuisson-Jolly, M.P.: Deformable template models: A review. Signal process 71 (1998) 109-129
3. Cootes, T.F., Hill, A., Taylor, C.J., Haslam, J.: Use of active shape models for locating structures in medical images. Image Vision Comput 12 (1994) 355-365
4. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. In: Burkhardt, H., Neumann, B. (eds.): European conference on computer vision, Vol. 2. Springer (1998) 484-498
5. Hill, A., Cootes, T.F., Taylor, C.J.: Active shape models and the shape approximation problem. Image Vision Comput 14 (1996) 601-607
6. Smyth, P.P., Taylor, C.J., Adams, J.E.: Automatic measurement of vertebral shape using active shape models. Image Vision Comput 15 (1997) pp. 575-581
7. Mahmoodi, S., Sharif, B.S., Chester, E.G., Owen, J.P., Lee, R.: Skeletal growth estimation using radiographic image processing and analysis. IEEE T Inf Technol B 4 (2000) 292-297
8. van Ginneken, B., Haar Romeny, B.M.: Automatic delineation of ribs in frontal chest radiographs. In: Hanson, K.M. (ed.): Image processing. Medical Imaging Vol. 3979. SPIE San Diego (2000) 825-836
9. Bernard, R., Pernuš, F.: Statistical approach to anatomical landmark extraction in AP radiographs. In: Sonka, M., Hanson, K.M. (eds.): Image processing. Medical Imaging Vol. 4322. SPIE San Diego (2001) in press
10. Heap, T., Hogg, D.: Improving specifity in PDMs using hierarchical approach. In: Clark, A.F. (ed): British Machine Vision Conference. Essex, UK (1997) 80-89
11. Gonzalez, R.C., Woods, R.E.: Digital image processing. Addison Wessley (1992)
12. Bookstein, F.L.: Principal warps: thin-plate splines and the decomposition of deformations. IEEE T Pattern Anal 11 (1989) 567-585
13. Cootes, T.F., Page, G.J., Jackson, C.B., Taylor, C.J.: Statistical grey-level models for object location and identification. Image Vision Comput 14 (1996) 533-540
14. Long, L.R., Pillemer, S.R., Lawrence, R.C., Goh, G-H., Neve, L., Thoma, G.R.: World Wide Web platform-independent access to biomedical text/image databases. In: Horii, S.C., Blaine, G. (eds.): PACS design and evaluation: Engineering and clinical issues. Medical Imaging Vol. 3339. SPIE San Diego (1998) 52-63
15. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B. P.: Numerical recipes in C, The art of scientific computing. University Press, Cambridge (1992)