Multiresolution Approach to Biomedical Image Segmentation with Statistical Models of Appearance

Špela Iveković[1] and Aleš Leonardis[2]

[1] Dep. of Electrical, Electronic and Computer Eng., Heriot-Watt University, Edinburgh EH14 4AS, United Kingdom,
si1@hw.ac.uk

[2] Faculty of Computer and Information Science, University of Ljubljana, SI-1001 Ljubljana, Slovenia,
ales.leonardis@fri.uni-lj.si

Abstract. Structural variability present in biomedical images is known to aggravate the segmentation process. Statistical models of appearance proved successful in exploiting the structural variability information in the learning set to segment a previously unseen medical image more reliably. In this paper we show that biomedical image segmentation with statistical models of appearance can be improved in terms of accuracy and efficiency by a multiresolution approach. We outline two different multiresolution approaches. The first demonstrates a straightforward extension of the original statistical model and uses a pyramid of statistical models to segment the input image on various resolution levels. The second applies the idea of direct coefficient propagation through the Gaussian image pyramid and uses only one statistical model to perform the multiresolution segmentation in a much simpler manner. Experimental results illustrate the scale of improvement achieved by using the multiresolution approaches described. Possible further improvements are discussed at the end.

1 Introduction

Biomedical images typically contain information about human body and its health condition. Aging, illnesses, injuries, and similar factors induce the *biological variability* of human beings. In the process of biomedical image acquisition there are several variable factors such as patient position and equipment properties which result in so called *technical variability.* Both biological and technical variability reflect in the acquired biomedical image and their volatile nature aggravates the segmentation process.

Statistical model of appearance [5, 6] successfully models the structural variability present in biomedical images. The model's training process is based on a training set of images, representing the texture of the structure in question, and the training set of structure's shapes, provided by medical experts' manual annotation of training images. The variability information is then extracted from both training sets by means of PCA analysis. After the training phase, statistical model has the ability to segment a structure which potentially belongs to the family of structures seen in the training set but exercises variability different from any of the structures actually present in the training set.

The statistical model on its own successfully deals with variability of objects in the images. However, by applying a multiresolution approach to the basic statistical model image segmentation, computational complexity can be reduced and quality of the results improved.

We applied two different multiresolution approaches to the problem of medical image segmentation. The first approach uses a Gaussian pyramid of statistical models. An independent model is trained for each of the pyramid levels. During the segmentation process the resulting shape and pose estimate of the structure in question are passed down the pyramid with each next model improving the result's accuracy. The second approach uses only one statistical model of appearance. The model is trained with the original training set. After the training phase a Gaussian pyramid of the model's eigenvectors is constructed. Consequently, the same model can perform the segmentation of the input image on various resolution levels.

The paper is organized as follows: Section 2 describes the related work. In section 3, we outline the basic concepts of image segmentation with statistical models of appearance. A detailed description of each of the applied multiresolution approaches is presented in section 4. Section 5 elaborates on our experimental results and section 6 concludes with a summary and discussion.

2 Related Work

Medical image segmentation [7] presents a demanding task for segmentation methods. Structures present in medical images exercise large variability in shape and texture. Consequently, the most popular methods used in medical image segmentation area are the *semiautomatic* methods, also called *user-steered* methods [8, 9, 12, 13], where the expert-operator initializes the segmentation process and then directs the segmentation algorithm towards the correct solution.

Although the semiautomatic methods are satisfyingly accurate, the drawback is their speed and the fact that an expert should be present during the segmentation process to guide it towards the correct solution. To avoid the need for operator-guided segmentation and instead fully automatize the segmentation process, medical image segmentation with deformable models [4, 11] has been introduced. Deformable models maintain the essential characteristics of a class of structures they represent but can also deform to fit a range of examples. A particular representative of the deformable model class is the statistical model of appearance [5, 6]. It is possible to reliably segment medical images with the statistical model of appearance. However, speed and accuracy of the segmentation performance can be further improved by applying the multiresolution approach [1, 10]. Yoshimura and Kanade [14] apply the idea of multiresolution approach to eigenimages and suggest the eigenspace construction on each resolution level.

Leonardis and Bischof [3] explore the possibility of building the eigenspace on the highest resolution only and afterwards adjusting it to the necessary resolution levels. In this paper, we further explore the idea of one eigenspace for all resolution levels and apply it to the segmentation with statistical models of appearance [6].

3 Image Segmentation with Statistical Models of Appearance

In order to synthesize a complete appearance of an image structure, we must model both its shape and texture [6]. For this purpose, statistical model of appearance consists of two submodels - statistical model of shape maintaining the structure's shape variation behaviour, and statistical model of texture maintaining the structure's texture variation behaviour. Both submodels cooperate with each other during the segmentation process. In the sequel, we provide a general description of the statistical model, describe its training process, and present the application of the model to image segmentation. Description applies to both, statistical model of shape and statistical model of texture, providing that the training set consists of shape vectors and texture vectors respectively.

3.1 Statistical Model - General Description

Statistical model is a deformable model which in a compact way describes the information contained in the training set data. During the training process modes of variability are extracted from the training set data and presented as model's parameter space base vectors. When performing the segmentation of a previously unseen image, the mode of variation of the new image structure is expected to comply to a great extent with the modes described by the base vectors. Consequently, if the new image structure belongs to the class of structures present in the training set data, it can be accurately described as a linear combination of model's parameter space base vectors. In the opposite case the structure is assumed not to belong to the same class as those in the learning set.

3.2 Training the Statistical Model

We start with a training set of data vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_s$ which exercise as many different modes of structure variability as possible. We first calculate the mean data vector

$$\overline{\boldsymbol{x}} = \frac{1}{s} \sum_{i=1}^{s} \boldsymbol{x}_i, \tag{1}$$

then we compute the data covariance matrix C

$$C = \frac{1}{s-1} \sum_{i=1}^{s} (\boldsymbol{x_i} - \overline{\boldsymbol{x}})(\boldsymbol{x_i} - \overline{\boldsymbol{x}})^T. \tag{2}$$

In the next step we use the SVD to compute the eigenvectors $\boldsymbol{e}_i$ and corresponding eigenvalues λ_i of the matrix C (sorted so that $\lambda_i \geq \lambda_{i-1}$). When we sort the eigenvectors according to their corresponding eigenvalues we get an ordered orthogonal base of the model's parametric space. The directions of eigenvectors represent the directions of training data variability. The larger the corresponding eigenvalue, the larger the data variability along that eigenvector direction. It is sufficient to keep only the first $t, t < s$, eigenvectors to accurately describe training data variability. Number t can be calculated as

$$\sum_{i=1}^{t} \lambda_i \geq f_v V_T, \tag{3}$$

where $V_T = \sum_{i=1}^{s} \lambda_i$ is the total variance of the training data and f_v is the proportion of variance we wish to represent with t eigenvectors (e.g. $f_v = 0.98$). If the columns of the matrix Θ contain the t eigenvectors, corresponding to the t largest eigenvalues, $\Theta = (\boldsymbol{e}_1|\boldsymbol{e}_2|...|\boldsymbol{e}_t)$, we can then approximate any data vector $\boldsymbol{x}$ of the training set using

$$\boldsymbol{x} \approx \overline{\boldsymbol{x}} + \Theta \boldsymbol{p}, \tag{4}$$

where $\boldsymbol{p}$ is a t-dimensional vector given by

$$\boldsymbol{p} = \Theta^T(\boldsymbol{x} - \overline{\boldsymbol{x}}). \tag{5}$$

The approximation sign $\approx$ in the equation (4) is due to the fact that we are only using the first t eigenvectors instead of all s eigenvectors available as a result of SVD. The elements of the vector $\boldsymbol{p}$ are called *parameter space coefficients*. If we use the equation (5) for a specific input data vector $\boldsymbol{x}$, we transform the input data vector from the reference space, where it exists as a texture, shape, etc., into a point in the statistical model's t-dimensional parameter space. The elements of the vector $\boldsymbol{p}$ are the coordinate values of this point along the model's parameter space base vectors. Parameter space representation of the input vector $\boldsymbol{x}$ is useful when trying out different configurations of the input data vector. With the help of equation (4) varying the elements of $\boldsymbol{p}$ results in varying the data vector $\boldsymbol{x}$ in the reference space.

3.3 Image Segmentation with the Statistical Model

Segmentation of the input image with the statistical model of appearance is in effect an optimization procedure. Segmentation result is obtained by *minimizing* the energy function for the values of the parameter space coefficients p_i of the model. The energy function is defined as a weighted sum of *internal* and *external* energy

$$E = \alpha E_{int} + (1 - \alpha) E_{ext}. \tag{6}$$

The *internal* energy measures the deformation of the model and is defined as a weighted sum of model's parameter space coefficient values:

$$E_{int}(\boldsymbol{p}) = \sum_{i=1}^{t} u(p_i). \tag{7}$$

As mentioned before, variation of parameter space coefficient values results in the change of the reference space data vector. We refer to this change as *deformation*. The larger the absolute values of coefficients, the more expressed the deformation of data vector. The weight function $u(x)$ imposes a penalty on configurations of coefficient values which result in a large deformation and in this way limits the extent of deformation tried out during the optimization procedure.

The *external* energy measures the congruity of the texture of the currently segmented image area with the model's ability to reconstruct this same texture. It is defined as the sum of squared differences between the corresponding pixels of segmented texture and of reconstructed texture:

$$E_{ext}(M(\boldsymbol{p}), I) = \sum_{i=1}^{n} (x_i^{seg} - x_i^{rec})^2, \tag{8}$$

where n stands for the length of the vectors representing both textures. The bigger the similarity between segmented and reconstructed texture, the lower the external energy.

The idea behind the definition of the external energy is the following: statistical submodel of texture represents typical texture variation modes of the texture training set. If the texture of the currently segmented area belongs to the same family of structures as those in the training set, it should be possible to accurately represent this texture with the linear combination of the statistical submodel of texture's base vectors. In other words, if the modes of variability of the currently segmented texture comply with those of the training set, we can sequentially apply the equations (4) and (5) to the currently segmented texture input vector and the recovered texture which results from the equation (5) applied after (4) should match closely with the input texture used in equation (4). A mismatch of the two textures is an obvious sign for the segmentation procedure that it is not segmenting the right area of the image and should thus either move elsewhere around the image or change the shape of the captured area and so capture a slightly different texture.

The importance of both energies, internal and external, can be influenced through the weight parameter α. By giving a higher significance to the internal energy, the segmentation process will avoid large deformations and instead put more effort in discovering the right position in the image (capture the right texture area). On the other hand, given a higher significance to the external energy, the model will tend to look for a better solution by trying out different deformations rather than looking for a better position in the image.

4 Multiresolution approach

The general idea behind the multiresolution approach [1, 10] is to construct a Gaussian input image pyramid in order to perform coarse to fine segmentation of the input image. The original definition of the statistical model of appearance deals only with segmentation of the input image on the original resolution level. In order to perform multiresolution segmentation, some adjustments of the original model have to be made. In the next two sections we describe two different ways in which we adjusted the original statistical model of appearance to perform the segmentation on different resolution levels of the input image.

4.1 First multiresolution approach - Gaussian pyramid of statistical models

The first multiresolution approach is a straightforward extension of the original statistical model. Original statistical model is only trained on the original resolution training set and therefore performs the input image segmentation on the original resolution level only. In the multiresolution extension of this approach, we construct a Gaussian pyramid of the learning set vectors and train an independent statistical model on every level of the pyramid. The Gaussian input image pyramid also has to be constructed with the input image pyramid resolution levels corresponding to those in the learning set pyramid.

The statistical model trained on the lowest resolution training set (i.e. the highest level of the Gaussian pyramid) begins the segmentation. Since the resolution is expected to be rather low at the highest level of the pyramid, the task of the first model merely concentrates on finding the best initial guess about the possible position of the image structure that the segmentation is trying to extract from the image.

Once the first model is finished it passes the result down the Gaussian pyramid to the second model, which accepts this result as its initial estimate. The second model is performing the segmentation on the same input image but on a higher resolution level, consequently having more image details at its disposal. The greater level of detail enables the second model to improve the initial estimate received from the first model. Since the second model already has the information about the possible initial position of the image structure in question, it can concentrate on improving the quality of segmentation and doesn't spend a lot of time searching through the image for possible structure positions. This fact results in significant reductions of time needed to perform a successful segmentation.

The result obtained by the second model is then passed to the third model and so forth, eventually arriving down to the lowest level of the pyramid (i.e. the highest resolution level). The last model in the pyramid performs the segmentation of the original resolution input image and, already having a significant amount of information available about the pose and shape of the segmented structure, merely performs the possible improvements in defining the shape of the structure.

There is one more issue that we have to consider when performing the multiresolution segmentation by using a pyramid of statistical models. As we mentioned before, the models are trained independently, each on its own training set. Consequently a transformation is needed to pass the result obtained by one model as an initial estimate to the other model. By training two statistical models independently of each other with two different training sets, each of the models will have its own eigenvectors defining the base of its parameter space. Therefore, the result obtained by the first model in form of parameter space coefficients cannot be passed directly into the parameter space of the second model but must instead undergo the transformation described in the sequel.

Transformation The transformation description depends on the actual nature of coefficients (i.e. the result of segmentation) that are being passed down the pyramid from one model to the other. In our case the coefficients passed down the pyramid are shape coefficients complemented with the information about the position of the shape in the image. Texture coefficients are obtained during the energy evaluation part of the optimization and don't need to be explicitely passed down the pyramid. Adjustment of the position information to a higher resolution level is straightforward. We will describe the shape coefficient transformation in more detail. Shape is defined as a collection of n two-dimensional image points:

$$(x_1, y_1)(x_2, y_2)(x_3, y_3)...(x_n, y_n) \tag{9}$$

In the reference space shape can be represented as vector of $2n$ image point coordinates:

$$\boldsymbol{x}_{sh} = [x_1, x_2, ..., x_n, y_1, y_2, ..., y_n] \tag{10}$$

As we mentioned, the following equation is used to transform the data vector from the reference space into the corresponding coefficients in the parameter space:

$$\boldsymbol{p}_{sh} = \Theta^T(\boldsymbol{x}_{sh} - \overline{\boldsymbol{x}}_{sh}), \tag{11}$$

and to transform the coefficients back into the reference space data vector:

$$\boldsymbol{x}_{sh} \approx \overline{\boldsymbol{x}}_{sh} + \Theta\boldsymbol{p}_{sh}, \tag{12}$$

where Θ stands for the matrix of eigenvectors, $\overline{\boldsymbol{x}}_{sh}$ for the average shape vector, and $\boldsymbol{p}_{sh}$ for the vector of shape coefficients.

As mentioned above, the consequence of the independently trained models is the incompatibility of their parameter spaces. However, both models interpret the reference space information in the same way, the only difference being the resolution on which the model was initially trained. We base our transformation on this reference space compatibility of the models.

Let's call the eigenvector matrices of the first and the second model Ψ and Φ, respectively, and assume that the first model obtained the resulting shape coefficients $\boldsymbol{p}^1_{sh}$. In order to pass this result to the second model down the pyramid, we first transform the coefficients into the shape vector in the reference space of the first model:

$$\boldsymbol{x}_{sh}^{1} \approx \overline{\boldsymbol{x}}_{sh}^{1} + \Psi \boldsymbol{p}_{sh}^{1}. \tag{13}$$

The next step is to adjust the shape vector to the reference space of the second model by means of upsampling. If the resolution of the first model is n-times smaller than the resolution of the second model, we get:

$$x_{sh_i}^{2} = n x_{sh_i}^{1}, \tag{14}$$

$$y_{sh_i}^{2} = n y_{sh_i}^{1}. \tag{15}$$

The remaining bit of the transformation is now to transform the new shape vector

$$\boldsymbol{x}_{sh}^{2} = [x_{sh_1}^{2}, x_{sh_2}^{2}, ..., x_{sh_n}^{2}, y_{sh_1}^{2}, y_{sh_2}^{2}, ..., y_{sh_n}^{2}] \tag{16}$$

into the parameter space of the second statistical model using the equation:

$$\boldsymbol{p}_{sh}^{2} = \Phi^{T}(\boldsymbol{x}_{sh}^{2} - \overline{\boldsymbol{x}}_{sh}^{2}). \tag{17}$$

Vector of shape coefficient values $\boldsymbol{p}_{sh}^{2}$ can now be used as an initial estimate for the second statistical model.

4.2 Second Multiresolution Approach - Gaussian Pyramid within the Statistical Model

The first approach in a straightforward manner upgrades the basic statistical model. However, the straightforward upgrade involves at least two redundant steps. First redundancy is the separate training of an independent statistical model for each resolution level. Consequently, the second redundancy is the transformation needed to pass the result down the pyramid due to incompatibility of the models' parameter spaces.

Our goal in the second multiresolution approach was to replace the training of several independent statistical models with only one model handling all resolution levels of the input image and being trained only once on the original resolution training set. Training only one single statistical model simplifies the application of the model to different learning sets extensively. The training process in the second multiresolution approach is hardly any more demanding than that one of the basic statistical model. As a result, time needed to train the model is reduced and, since we are still using the multiresolution approach, the quality of the results is preserved, as we will see later.

Our second multiresolution approach is based on the work of Bischof and Leonardis [3]. In this work Bischof and Leonardis discuss among other things the multiresolution coefficient estimation. They point out that one can use the same coefficients to reconstruct a filtered and subsampled image from the filtered and subsampled eigenimages as well as to reconstruct the original resolution image from the original resolution eigenimages. This idea constitutes the essence of our second multiresolution approach that is why the following section describes it into some more detail.

Multiresolution Coefficient Estimation The idea we are about to present stems from the efforts to find a method of eigenimage coefficient estimation which would be robust against occlusion, varying background, and other types of non-Gaussian noise [3]. It is based on the following observation: If we take into account all s eigenvectors and if there is no noise in the data, then, in order to calculate the coefficients p_i, we only need s points $\boldsymbol{r} = (r_1, r_2, ..., r_s)$ where $\boldsymbol{r}$ is a vector of point coordinates. It is sufficient to compute the coefficients p_i by simply solving the following system of linear equations:

$$\boldsymbol{x}(r_j) = \sum_{i=1}^{s} p_i \boldsymbol{e}_i(r_j) \tag{18}$$

where $\boldsymbol{x}(r_j)$ denotes the recovered value of $\boldsymbol{x}$ at position r_j. This notation emphasizes the fact that the equation is valid pointwise.
In a more common case we assume that we have eigenvectors $\boldsymbol{e}_i$ ordered in descending order with respect to the corresponding eigenvalues λ_i. Then, depending on the correlation among the data vectors in the learning set, only t, $t < s$, eigenvectors are needed to represent the data vector $\boldsymbol{x}$ to a sufficient degree of accuracy:

$$\tilde{\boldsymbol{x}}(r_j) = \sum_{i=1}^{t} p_i \boldsymbol{e}_i(r_j). \tag{19}$$

We can now derive the following property which holds due to the linearity of the equation:

$$(f * \tilde{\boldsymbol{x}})(r_j) = \sum_{i=1}^{t} p_i (f * \boldsymbol{e}_i)(r_j), \tag{20}$$

where f denotes a filter and $*$ the convolution.
The next property holds because Eq.(19) is valid for each point r_j, therefore

$$\tilde{\boldsymbol{x}}_{\downarrow}(r_j) = \sum_{i=1}^{t} p_i \boldsymbol{e}_{i\downarrow}(r_j), \tag{21}$$

where $_{\downarrow}$ denotes the subsampling operation. Combining (20) and (21), we obtain

$$(f * \tilde{\boldsymbol{x}})_{\downarrow}(r_j) = \sum_{i=1}^{t} p_i (f * \boldsymbol{e}_i)_{\downarrow}(r_j), \tag{22}$$

which states that using the same coefficients we can reconstruct a filtered and subsampled image from the filtered and subsampled eigenimages.

Application in the Second Multiresolution approach The multiresolution coefficient estimation presents some significant advantages for the statistical model segmentation. By incorporating the results from the previous section, we can consequently perform the multiresolution segmentation of the input image using only one single statistical model trained on the original training set.

Training the statistical model essentially means extracting the data variability information from the learning set and representing it in terms of eigenvector directions. Once we have trained the original statistical model using the original training set, we can adjust the model to different resolution levels by adjusting its eigenvectors. As it follows from the previous section, the coefficients representing the input vector in the parameter space can be recovered from the original resolution image and corresponding original resolution eigenvectors as well as from subsampled input vector and corresponding subsampled eigenvectors. There is a direct relationship between the coefficients recovered using the first or the second approach - they are essentially the same, their accuracy being the only difference.

Applying this idea to the multiresolution statistical model segmentation is very straightforward. We first build the Gaussian pyramid of the input image. According to the resolution levels of this pyramid we then build the Gaussian pyramid of model's eigenvectors. The segmentation process starts on the highest level of the pyramid and performs the segmentation of the lowest resolution input image with the lowest resolution eigenvectors. The resulting coefficients are now passed straight down the pyramid to the next resolution level. The transformation from the first approach is not necessary anymore, since what we had on the previous level was only a filtered and subsampled version of the eigenvectors which will be used on the following level. According to the robust coefficient estimation idea [3] in this case coefficients are interpreted in the same way on both pyramid levels. However, the accuracy of coefficients improves as they are passed down the pyramid since every next level contains more details and thus enables the segmentation process to more reliably determine the coefficient values.

The fact that we are using the filtered and subsampled eigenvectors raises another issue. In the first multiresolution approach we had an independent model trained on each resolution level. As a consequence each model had its own orthogonal set of parameter space base vectors. As a result we could use the projection equations (4) and (5) to transform the data from the parameter space representation into the reference space representation and vice versa. However, in the second multiresolution approach this is not anymore the case. In the training process of the original model we do get an orthogonal set of eigenvectors composing the parameter space base. As soon as we filter and subsample these eigenvectors, however, we lose the orthogonal property. Consequently, the projection equations (4) and (5) are no longer valid. Following from equation (19), we solve a predetermined system of m linear equations with t unknowns instead. To illustrate the approach let Ψ be a matrix of eigenvector values at m points

$$\Psi = \begin{pmatrix} \boldsymbol{e}_1(r_1) & \boldsymbol{e}_2(r_1) & ... & \boldsymbol{e}_t(r_1) \\ \boldsymbol{e}_1(r_2) & \boldsymbol{e}_2(r_2) & ... & \boldsymbol{e}_t(r_2) \\ . & . & & . \\ . & . & . & . \\ . & . & & . \\ \boldsymbol{e}_1(r_m) & \boldsymbol{e}_2(r_m) & ... & \boldsymbol{e}_t(r_m) \end{pmatrix} , \tag{23}$$

$\boldsymbol{x}_m$ be the input data vector of length m (in principle t points from the input data vector suffice to calculate the value of the coefficients, however, for the purpose of simplicity we rather take all m elements of the input vector and solve an overdetermined linear system of equations instead), and $\boldsymbol{p}$ the vector of t parameter space coefficients

$$\boldsymbol{x}_m = \begin{pmatrix} \boldsymbol{x}(r_1) \\ \boldsymbol{x}(r_2) \\ . \\ . \\ . \\ \boldsymbol{x}(r_m) \end{pmatrix} , \ \boldsymbol{p} = \begin{pmatrix} p_1 \\ p_2 \\ . \\ . \\ . \\ p_t \end{pmatrix} . \tag{24}$$

Now we can, for the purpose of the second multiresolution approach, replace the equation (4) by

$$\boldsymbol{x}_m = \Psi \boldsymbol{p} \tag{25}$$

and the equation (5) by

$$\boldsymbol{p} = \Psi^+ \boldsymbol{x}_m . \tag{26}$$

Note that Ψ represents the matrix of an overdetermined system of linear equations and thus has to be inverted using pseudoinverse.

In this section we described the idea behind the second multiresolution approach and explained the mathematical background to it. In the following section we proceed to tests and test results which show the scale and mode of improvement possible when using the described multiresolution approaches.

5 Experimental Results

The statistical model was trained on the training set consisting of 36 x-ray images of 366×499 original resolution. The images were x-ray scans of the human cervical vertebrae, courtesy of Jesenice hospital, Slovenia. An example of the training set image is given in Figure 1. Medical experts annotated every image in the training set with 7 points describing the shape of a single vertebra. These points were copied to shape vectors and assembled in the shape training set.

Database size is a distinctive problem when it comes to medical image processing. It is difficult to obtain a large database of quality medical images, especially when dealing with x-ray images. The primary aim of a doctor taking an x-ray scan of a patient is not to expose the patient to the radiation for longer than

Fig. 1. Example of a training set image.

necessary. It often happens that the image which is incomplete from computer processing point of view still has enough information for doctor's interpretation. There are severe health reasons against additional x-ray scans so one has to compromise and use the database available.

Because of the relatively small database size we conducted the leave-one-out test on the database of 36 images. The criteria used as a quality measure of the segmentation result was the Euclidean distance in pixels between the shape estimate given by the model and the shape annotated by the medical expert.

We first conducted the test with the original statistical model. The results were used as a reference in terms of segmentation quality and number of iterations. When segmenting the images with the original model, on average 35 iterations were necessary for the model to converge to a solution. In many cases the solutions were in the right position but not properly shaped.

When we forced the original model to initialize the segmentation process far from the actual solution, it regularly got lost and terminated the optimization due to exceeded maximum number of iterations. Example of a case where the model got lost is shown in Figure 2. The position of the "butterfly" in the image shows where the model terminated the segmentation. The reason behind the poor performance on the shifted structures or bad initialization was on one hand the restriction of the model to only segment the area which was covered in the training set and on the other hand the large amount of detail present in the test image which got the model confused before it arrived at the correct solution. It is always possible to perform several segmentations of the same image with the original model, every time initializing the model in a different image area and therefore increasing the chances of finding the right solution even when it is shifted from its usual position, but it is computationally prohibitive. Our goal was to explore how much the segmentation performance could improve by using the multiresolution approach.

The multiresolution approach turned out to be significantly more successful than its original model counterpart. The pyramid consisted of 5 resolution levels. While the original model was forced to work with the global optimization method due to the multimodality of the problem, the multiresolution approach arrived at the same or even better quality solution using only the local optimization

Fig. 2. The original model tends to get lost when initialized too far away from the structure's expected position.

technique. Due to the multiresolution approach the number of iterations necessary to converge to a solution got much smaller. On average two iterations per resolution level were necessary. By contrast with the original model the multiresolution approach never segmented a shape that would be too small or wrongly oriented, which was often the case with the original model.

Another advantage of the multiresolution approach is the segmentation of shifted structures. It is relatively easy to perform an exhaustive search through the lowest resolution level image in order to determine the location of the shifted structure when using the multiresolution approach (see Figure 3).

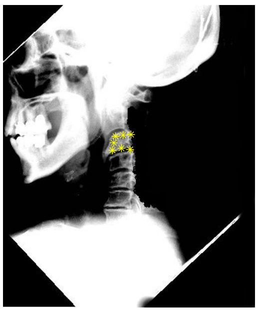

Fig. 3. The multiresolution approach is far more successful when dealing with bad initialization.

This can be a problem, however, when the image contains several instances of the structure that is being segmented, such as in the case of cervical vertebrae. However, to arrive at the correct solution in this case, instead of performing the exhaustive search of the entire image on the lowest resolution level, only the appropriate surroundings of the expected solution position are examined. In a

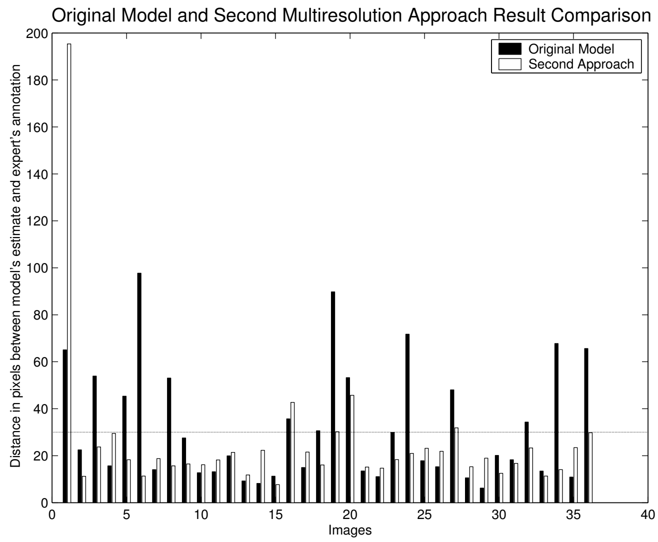

Fig. 4. Comparison of original model and second multiresolution approach test results.

case like ours where there are several vertebra structures exercising very similar modes of variability, it is helpful to exploit the spatial relationship between the individual structures in order to segment the correct one [2]. Speaking about automatizing the segmentation process, it is also good to keep in mind that when exploiting the spatial relationship and trying to locate a shifted structure, we should not strive to segment the structure no matter what its position in the image. If the structure is significantly shifted from its expected position and therefore the segmentation process fails to locate it, this could well be the indicator that there's something wrong with the structure's position in the first place and that it thus needs, for example, doctor's special attention.

Both the first and the second approach produced better results than the original approach. Figure 4 shows the comparison of the original model results and results obtained by the second multiresolution approach.

The correct solutions in Figure 4 are those which are no more than 25-30 pixels high. One would expect the correct solution to be close to 0. The reason for such a large number lies in the distance criteria. The quality of the solution is defined as a distance between the expert annotated shape and the segmented shape estimated by the statistical model. The expert annotated shape cannot be taken as 100% correct. It is the expert's decision where the distinctive anatomic points lie in the image. When the model suggests its own solution, this can differ up to as much as 30 pixels from the expert annotated solution and still make sense. The upper boundary for the correct solutions was estimated by manually examining all 36 segmentation results.

6 Summary and Discussion

We suggested the upgrade of the original statistical model with the multiresolution approach and described two different multiresolution approaches. We conducted segmentation tests on the medical image database available and showed that the complexity and quality of segmentation can be improved by using the multiresolution approach.

The computational complexity got reduced due to the multiresolutional approach since the segmentation process was reliably performed with the local optimization technique. In more demanding cases, e.g. shifted structures, the lowest resolution level segmentation process could be initialized in several different areas of image in order to enhance the segmentation robustness. In spite of several repetitions of the segmentation on the lowest resolution level, the multiresolution approach is still far less computationally prohibitive than the original statistical model and at the same time producing better quality solutions.

The multiresolution approach also proved to be much more flexible when segmenting shifted structures. The only drawback was the quality of the images which often contained shadows and missing information. These images were already very unclear on the original resolution level. After downsampling they lost the important distinctive information about the structure since through the downsampling process the shadows mixed with the actual vertebrae. In such cases, in order to reliably segment the images with the multiresolution approach, some extra image preprocessing, e.g. removing the shadows, enhancing the edges, and similar transformations are required that would preserve their effects all the way down to the lowest resolution level and so make it easier for the model to discover the initial structure position. However, the multiresolution approach has proved to be very reliable when segmenting the images which didn't contain strong shadow problems and large portions of missing information.

When comparing both multiresolution approaches applied, the second proves to be particularly useful in spite of the results being pretty similar for both described approaches. Apart from subsampling the input image, the second approach appears to be the same on the outside as the original statistical model of appearance. The end user does not feel the difference when using the second approach adjusted model, since it does not require the downsampling of the training set. The second approach also allows for arbitrary choice of number of resolution levels actually used during the segmentation process. The coefficients obtained on the penultimate level of resolution can in most cases be directly propagated onto the last resolution level without actually performing the last level segmentation. This is due to the same interpretation of the coefficients on all resolution levels.

The elegant implementation of the second multiresolution approach and its simple application to various training sets is a significant advantage. Using this idea together with the information about the spatial relationship between the structures in the image (called *topology model* and described in [2]) could significantly improve the medical image segmentation's reliability as well as speed and remains an area to be explored.

References

1. R. Bajcsy, S. Kovacic, "Multiresolution Elastic Matching", *Computer Vision, Graphics and Image Processing*, 46, 1-21, April 1989.
2. R. Bernard, B. Likar, F. Pernus, "Segmenting Articulated Structures by Hierarchical Statistical Modeling of Shape, Appearance, and Topology", *Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention–MICCAI 2001*, Utrecht, The Netherlands, October 2001, 499-506, Springer 2001.
3. H. Bischof, A. Leonardis, "Robust Recognition of Scaled Eigenimages Through a Hierarchical Approach", *Proceedings of Conference on Computer Vision and Pattern Recognition - CVPR'98*, 664-670, IEEE 1998.
4. A. Blake, M. Isard, *Active Contours*, Springer, 1998
5. T. F. Cootes, G. J. Edwards, C. J. Taylor "Active Appearance Models" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 6, June 2001.
6. T. F. Cootes, C. J. Taylor, "Statistical Models of Appearance for Medical Image Analysis and Computer Vision", *Proceedings of SPIE Medical Imaging 2001*, Vol. 4322.
7. J. Duncan, N. Ayache, "Medical Image Analysis: Progress Over Two Decades and the Challenges Ahead", *IEEE Trans. PAMI 22(1)*, 85-106, 2000
8. A. X. Falcao, K. Jayaram, J. K. Udupa, F. K. Miyazava "An ultra-fast user-steered image segmentation paradigm: Live-wire-on-the-fly", *SPIE on Medical Imaging*, vol. 3661, 184-191, 1999
9. M. Kass, A. Witkin, D. Terzopoulus, "Snakes: active contour models", *International Journal of Computer Vision* 1(4), 321-331, 1988
10. T. Lindeberg, *Scale-Space Theory in Computer Vision*, Kluwer Academic Publishers, Dordrecht, Netherlands, 1994
11. T. McInerney, D. Terzopoulos, "Deformable models in medical image analysis: a survey", *Medical Image Analysis*, 1 (2) 91-108, 1996.
12. A. Schenk, G. P . M. Prause, H. -O. Peitgen, "Efficient semiautomatic segmentation of 3D objects in medical images", *Lecture Notes in Computer Science*, Vol. 1935, 186-195, Springer, New York, 2000.
13. D. Stalling, H. -C. Hege, "Intelligent scissors for medical image segmentation", *Digitale Bildverarbeitung fuer die Medizin*, Freiburg, 32-36, 1996
14. S. Yoshimura, T. Kanade, "Fast template matching based on the normalized correlation by using multiresolution eigenimages", *Proceedings of IROS '94*, 2086-2093, 1994