Katholieke
Universiteit
Leuven

**Department of
Computer Science**

Incisor segmentation

Final project

Computer Vision (B-KUL-H02A5A)

Ivan Sevcik (r0624830)
Jakub Macina (r0641432)

Academic year 2015–2016

Contents

1 Introduction

The goal of this project is to use Active Shape Model proposed by Cootes et al.[1], and develop an approach capable of segmenting the upper and lower incisors in panoramic radiographs. Our proposed approach consist of three main parts. First, the model representing the mean shape of incisor is constructed using a set of landmarks, i.e. annotated ground truth points representing shapes of incisors. The model also captures statistical variations of the incisors shapes and it is discussed in section 2. Next, estimation of initial pose of incisors in the image is essential for segmentation and section 4 is devoted to that topic. Finally, fitting the model from initial positions to the image is done by analyzing the neighbourhood pixel values, which is outlined in section 5. Both of the previous steps requires radiograph preprocessing, which is discussed in section 3. Finally, experimental results are presented in section 6.

2 Active Shape Model

The model was built for upper and lower incisors separately. For each model, training landmarks were aligned correspondingly, followed by Principal Component Analysis that was applied in order to find out main features represented by eigen-vectors and mean shape. This resulted in dimensionality reduction as from previous 80 features (x and y coordinates for each of 40 points) one could extract just 3 features that represent 90% of all variations.

2.1 Shapes alignment

In order to perform training of the model, incisor shapes need to be aligned because they come in different shapes, sizes and rotations. Shape aligning is implemented as iterative approach that was proposed in Appendix A. of [1], until the squared sum of the distance of each shape to the mean is less than 0.05. This value was found out empirically and usually the distance is already much smaller than that value after second iteration. To describe the process, lets suppose incisor shape is represented by equation 1, where j is the tooth number and n is equal to 40 (because there are 40 landmark points p_i).

$$z^j = \{p_1^j, p_2^j, \cdots, p_n^j\} = \{(x_1^j, y_1^j), (x_2^j, y_2^j), \cdots, (x_n^j, y_n^j)\} \tag{1}$$

The scale of the shape can be computed as follows:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(p_i^j - centroid^j)^2}{2n}} \tag{2}$$

Rotation angle between two shapes z^1 and z^2 is computed as:

$$\theta = tan^{-1}\left(\frac{\mathbf{x}^2\mathbf{y}^1 - \mathbf{y}^2\mathbf{x}^1}{\mathbf{x}^2\mathbf{x}^1 + \mathbf{y}^2\mathbf{y}^1}\right) \tag{3}$$

where $\mathbf{x}$ and $\mathbf{y}$ are vectors containing x and y coordinates of corresponding shapes.
In each step of iterative alignment algorithm, every incisor shape is aligned with the current mean shape by using Procrustes analysis from previously computed scale and rotation. New landmark coordinates are computed as:

$$\begin{pmatrix} x^j \\ y^j \end{pmatrix} = s^j \begin{pmatrix} cos\theta^j & sin\theta^j \\ -sin\theta^j & cosi\theta^j \end{pmatrix} \begin{pmatrix} x^j \\ y^j \end{pmatrix} \tag{4}$$

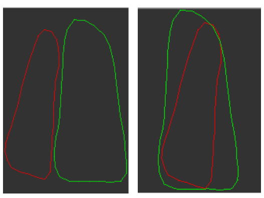

Figure 1: Visualization of aligned incisors. At the left image two incisors are unaligned and the right image presents these two teeth after applying align algorithm.

2.2 Principal components analysis (PCA)

Subsequently, a PCA is applied to the aligned training data. Every training example can be approximated as $z = \bar{z} + Pb$, where $\bar{z}$ is a mean shape, P is a matrix of eigenvectors and vector b consists of scalar eigenvalues. Together this gives a representation of shape in lower dimension. Most of the variance can be explained with the eigenvectors with the largest eigenvalues, which allows to reduce the dimensionality of the landmarks space to m, where $m << 2n$. We find m that satisfies the equation

$$\sum_{i=1}^{m} \lambda_i > 0.9 \sum_{i=1}^{2n} \lambda_i \tag{5}$$

which means that m eigenvectors with the largest eigenvalues are selected and together they explain more than 90% of the variance in the training set. It was found out that only two principal components for lower incisor and three principal components for upper incisor are required to meet this criteria and these are analyzed in figure 3. Next, each eigenvalue b_i is constrained into a range of $-3\sqrt{\lambda_i} \leq b_i \leq 3\sqrt{\lambda_i}$, where λ_i is the value computed by PCA. By setting different limits one can change the degree of freedom of the final model.

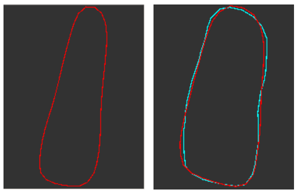

Figure 2: Left image presents trained mean shape model for upper incisors. Right image shows projection of trained model to one of the training tooth (drawn by blue color).

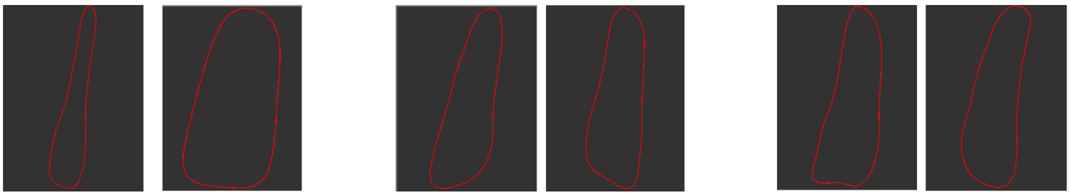

Figure 3: Analysis of three principal components for upper incisor mean shape model, which are sorted by their significance from left to right. Extreme values for principal components are presented. First principal component models the width of the teeth. Second principal component models left or right tip of the teeth. Third principal component represents left or right side curvature of the teeth.

3 Dental radiographs preprocessing

Provided dental radiographs have high resolution and contains noise at sides and capture not only teeth, but also nose and mandible. As there is some prior knowledge of how these panoramic x-ray radiographs are obtained, it can be assumed that head is centered horizontally and variation in scale is limited, therefore incisors are always approximately in the middle of the radiograph. Therefore, it is possible to only select a region of interest that actually contains the incisors by cropping and this significantly improves speed of computations.

Radiographs are inherently noisy. The kind of noise that can be found in them is usually described as quantum noise or mottle, structure noise and electronic noise. The quantum noise can be lowered to some extent by increasing the dose but a trade-off between a safe dose and good signal-to-noise ratio has to be found, which means there will be always some of this noise [2]. Therefore, there has to be some filtering stage implemented that will make the best attempt to remove the noise. By experimenting with different filters and values, it was found that best results can be obtained by applying *median filter*, which is followed by *bilateral filter*. Median filtering deals with salt-and-pepper noise, i.e. sparsely random black and white pixels, and also mottle of a small size. However, it can't remove noise that spreads large areas of image because it would also destroy other features like edges of teeth. For that reason, bilateral filter is used as it has the advantage of removing noise while preserving significant edges. In result, intensity of color is equalized throughout large areas, thus removing noise but the edgas are left intact. Finally, in order to highlight edges of teeth and remove dependence on initial intensity or luminance, *Scharr gradient operator* is used. This operator computes derivatives in x and y direction and the results can be added together to obtain a good edge representation of the original image.

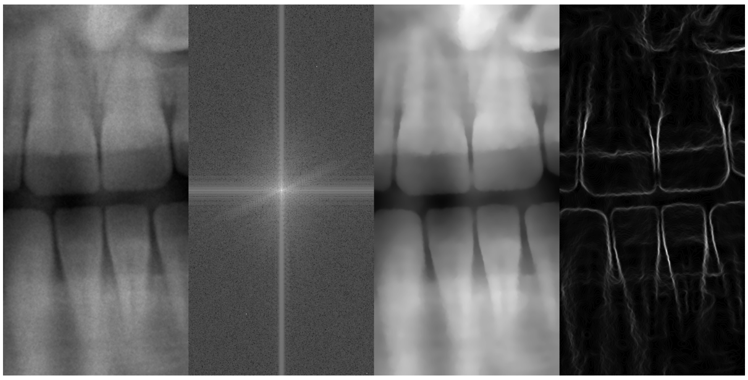

Figure 4: Result of radiograph preprocessing. Going from the left: first image shows original data, second image presents magnitude of the spectrum of filtered image after fast Fourier transform (FFT), third image shows filtered image after median and bilateral filtering, fourth image presents final filtered image after the Scharr gradient filter.

4 Finding the initial pose of teeth

In order to fit the learned mean shape model representation of incisor to a new image, the pose in the image must be estimated. For this task, first the upper and lower jaws are separated and then individual teeth are selected by finding vertical lines with highest gradients.

4.1 Jaws separation

The method for jaws separation used in this project computes projection histogram in y axis, similarly to technique proposed in [3]. It is computed by summing the pixel intensities along each row parallel to x axis. Row with the lowest sum in the image represents a global minimum in the histogram. Afterwards, the algorithm sequentially process histogram from the global minimum to both sides and it finds local minimums within a range that fulfills threshold value. These values represents upper and lower jaws separation lines.

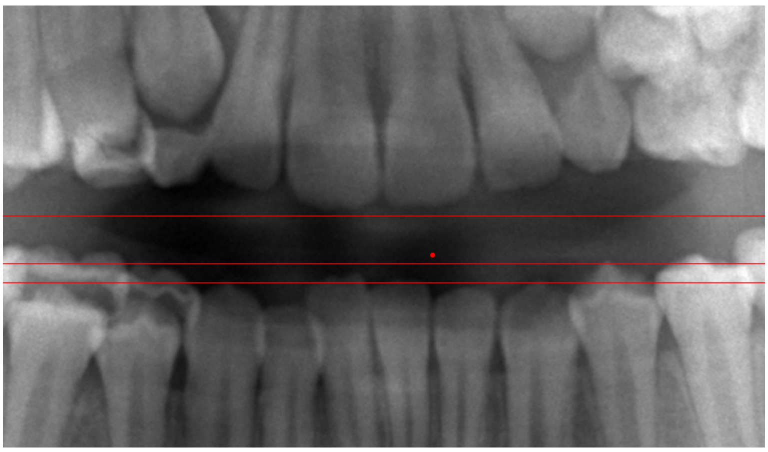

Figure 5: Separation of upper and lower jaws. Center line represents global minimum of the sum of intensities. Top and bottom lines are local minimums and these lines are used for separation of upper and lower jaws.

4.2 Tooth selection

For selecting individual teeth, it is supposed that vertical edges in the filtered image have very high intensities compared to other structures. Therefore, an approach of using Hough Transform for line detection is used. In the next step only vertical lines with maximum angle of ± 25 degrees are selected and the rest is filtered away. More lines for the tooth edge might be detected, therefore also lines that are close together are also filtered out so that only one line in specified range is retained, while preferring more vertical lines. Following this, a line closest to the middle of the separated jaw image is selected and it is assumed that it is the line representing the gap between middle incisors and which seprates the right halve of jaw from the left one. Based on this center line, one line to each side is then selected and resulting output is three lines representing the gaps between incisors.

Figure 6: Upper and lower tooth isolation using lines detection.

In the last step, an initial x position for middle incisors is computed as half of the gap between lines. Left and right incisor x position is computed as a specified gap to left and right from the edge lines. For estimating the y position, it is assumed that middle incisors are usually bigger than left and right incisors. Therefore, y position is estimated by empirically setting limit distance from the upper/lower jaw separation line.
Moreover, rotation of the model needs to be estimated. In case of rotation, it is assumed that middle incisors are usually without rotation and edge incisors are rotated away from the center. The rotation does not have to be precise but having a somewhat correct alignment beforehand helps the model converge much faster. Finally, scale is estimated similarly, based on domain knowledge and empirical experiments.

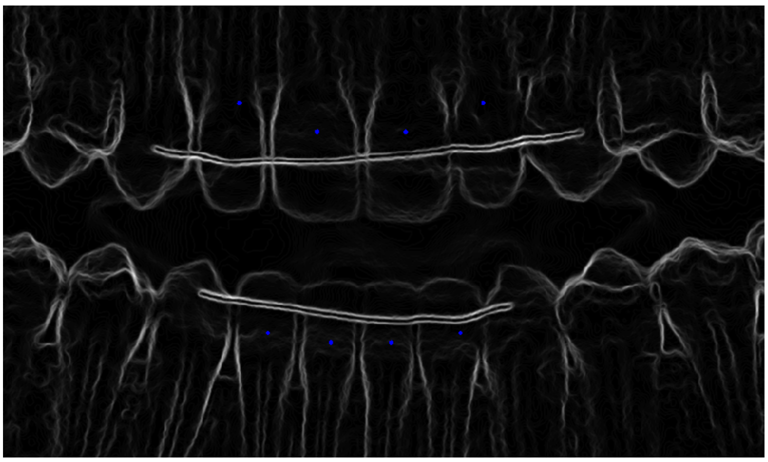

Figure 7: Result of initial pose algorithm. Initial positions are drawn by blue circles.

5 Fitting the active shape model to the image

Given a mean shape model represented as $z = \bar{z} + Pb$, the task is to choose a set of parameters b to fit a model to a new image. Method which was used for this is the Active Shape Model algorithm proposed by [1] (part Protocol 2).

5.1 Landmark updates

In order to find a set of parameters b, $(2m + 1)$ pixels are sampled (m to each side) along normal vector for each point of the shape as shown in figure 8. Parameter m was set to 14 by empirical observations. Next step is to search in these sampled pixels for the best position where the landmark point will be relocated. Three different strategies to search for the best point were implemented:

- Mean model: A statistical model is built for each point by sampling training images along annotated points and creating a mean vector. During training, $(2k+1)$ pixels are sampled, where $k < m$, e.g. $k = 5$. Based on the mean vector for every point, the sampled profile obtained during update is searched for best alignment. This is done by iteratively selecting $2k + 1$ values from sampled profile and calculating sum of squared means. The alignment where the sum is minimal is considered a best match and the position of landmark is updated to this position.
- Appearance model: Assuming that intensities along normals are distributed as a multivariate Gaussian distribution, statistical model containing mean and covariance matrix can be created. Number of sampled pixels is the same as in mean model. Search is similar to previous model, however as a similarity measure it uses Mahalonobis distance. This model was proposed by Cootes [1].
- Intensity model: This model simply searches for maximum value in the sampled vector and moves landmark into that position. However, this would result in the model always searching for the maximum intensity in the image. To lessen this effect, the sampled vector is first multiplied by a triangular window. The result of this is that there is a trade-off between intensity and distance. If the landmark is already located on relatively bright pixel (edge), it will not have tendency to move from there just because there is a very bright pixel at the end of sampled vector. The changes to position are then more local and gradual.

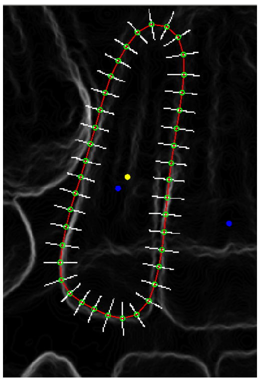

Figure 8: Visualization of sampled points for each point of the shape.

It was found out during experiments that it is not always possible to obtain the Appearance model which was suggested by original paper. This is because it assumes that an inverse covariance matrix can be created, however the training data had sometimes linearly dependent vectors. The matrix determinant is then 0 and inverse matrix can not be obtained. Therefore, the mean model was developed which was supposed to simulate previous approach only without the covariance matrix, but it failed to give good results and also required long training times. Finally, the intensity model was developed. While it is simple, it produced relatively good results during experiments. However, it is prone to favoring the strongest edges in the image which are not always the edges of teeth.

5.2 Multi-resolution search

In order to develop efficient algorithm and to decrease the probability of the algorithm getting into local optimum, a multi-resolution framework was implemented as a Gaussian image pyramid with two levels. Stopping conditions for search on every level of the image pyramid is either maximum number of steps (empirically set to 100) or when the fitting algorithm stops to converge.

6 Evaluation

Evaluation results are calculated using leave one out cross validation on the data set of 14 annotated images. The error is calculated as a normalized magnitude difference between found shape $\mathbf{z}$ and ground truth shape

t as:

$$error_z = \frac{\|z - t\|}{\|t\|}$$

The average total error across all of 14 examples after cross validation is shown in table 1 for upper incisors and in table 2 for lower incisors. Figure 9 and figure 10 presents results of two cross validations.

Table 1: Total average error for upper incisors after running full leave one out cross validation. Incisors are labelled by their number from left to the right.

First	Second	Third	Fourth
0.432397	0.321746	0.316382	0.33416

Table 2: Total average error for lower incisors after running full leave one out cross validation. Incisors are labelled by their number from left to the right.

First	Second	Third	Fourth
0.420079	0.340895	0.31304	0.336813

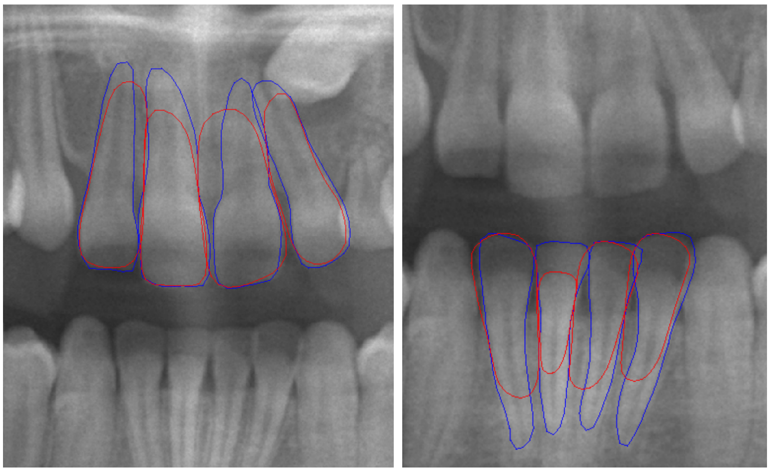

Figure 9: Resulting prediction for radiograph number 7. Ground truth labels are drawn in blue, incisor segmentation of our approach are drawn in red.

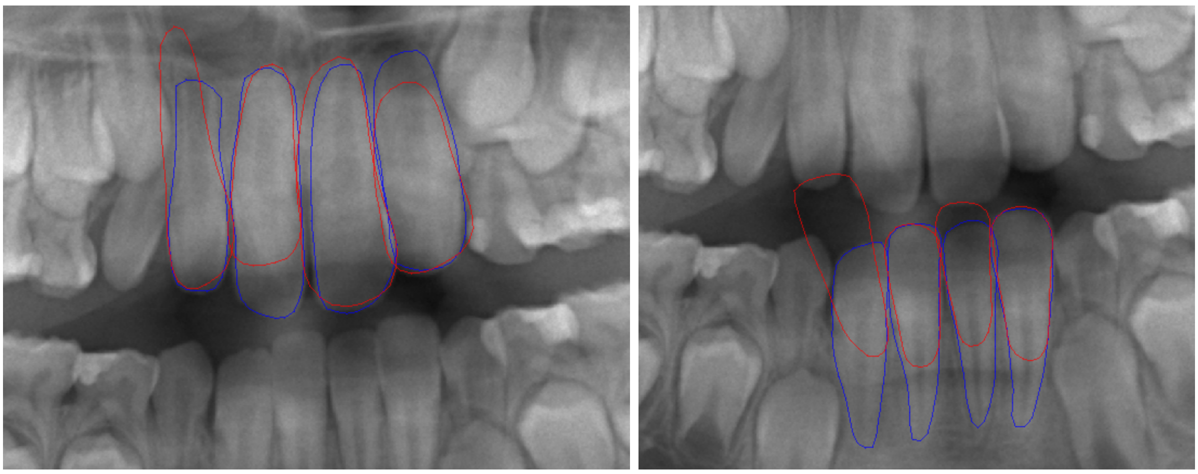

Figure 10: Resulting prediction for radiograph number 14. Ground truth labels are drawn in blue, incisor segmentation of our approach are drawn in red.

7 Conclusion

Our approach for incisor segmentation achieved promising results. Experiments show that estimation of initial position, scale and rotation of active shape model are essential to obtain good results. The approach for finding these variables can be further improved by using tooth isolation algorithm proposed in [3], where projection along x axis is also computed. Moreover, if more training data would be available, Haar cascade could be trained for detecting incisors in the image.
Also, the landmark model that was used in final solution is rather simple and with more data available, it might be possible to employ the more advanced appearance model which has higher chance of converging to correct edges. Finally, there are more advanced filtering techniques which would give better results but were not explored due to their complexity.

References

[1] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models—their training and application. *Comput. Vis. Image Underst.*, 61(1):38–59, January 1995.

[2] M Cohnen, J Kemper, O Möbes, J Pawelzik, and U Mödder. Radiation dose in dental radiology. *European radiology*, 12(3):634–637, 2002.

[3] Anil K Jain and Hong Chen. Matching of dental x-ray images for human identification. *Pattern recognition*, 37(7):1519–1532, 2004.