Computer Vision
Incisor Segmentation

Marcell Zoltán Szalontay, Matej Jeglič

June 25, 2017

Introduction

Computer vision nowadays a highly researched field that has proven its utilities in many areas. On of most important use of computer vision has been closely related to the development of medical applications. Dental radiographs usually contain complex objects whose shape can vary greatly between images. Model-based approaches have been proven to be robust and able to detect rigid objects, even in the presence of noise and other imperfections. Dental radiographs apart from teeth also contain bone structures form the jaw, cheeks and other facial features, which makes object outlining more challenging.
The purpose of this project is to detect and outline the upper and lower incisors in dental radiographs. The used approach is based on a point distributed model. Point distributed model is trained with images form a training set. Each image in training set contains points that present incisor boundary and were marked by an expert. The explored method for incisor detection consists of three main tasks: i) shape modeling; ii) image preprocessing; and iii) model fitting to image. In the first part a shape model is created for each incisors separately. The information encoded in the model presents the distribution of boundary points, which do not depend on global position, scale, and orientation. In the second part, preprocessing of images takes place in order to enhance the useful features of the objects and to remove unnecessary information which would lead to corrupted results. Preprocessing is also indispensable during initial position estimation and model fitting. Finally, in the third part, the created shape model is fitted to the image. The first important part of this task is to get a good initial estimation of incisor position examined image. After that, the iterative model fitting is done in order to align shape model to the incisor in the image.

1 Shape Modeling

Various method can be used to describe an object shape. In this project a point distribution model was used in order to obtain a shape representation of an object. Therefor objects are described as set ob points which can be also called landmarks. Each landmark embodies a particular part of an object or its boundary. As objects shape varies trough images, this methods models how landmarks tend to move. Landmarks selection is very important and must be done consistently. In order to create such point distribution model, a training set must be provided. In this project 14 dental radiographs were provided with landmarks for all 8 incisors in each radiograph. Each incisor is described with 40 landmarks.
One shape model per incisor was designed, meaning there is one shape model for first from the left upper incisor, one for second form the left upper incisor and so on. This means 8 different shape models for incisors were designed.
A shape model must only describe shape of the object or, in other words, corresponding relations of landmarks positions relative to each other. Shape should be independent of global position of the object in the image, it should be independent of object scale and it should be independent ob object orientation. Shape model must also poses means which describes how deformation to shape can be made until the shape looses similarity with a true object.
Landmarks for each object are collected in landmarks vector.

$$\mathbf{z} = ((x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N))$$

1.1 Procrustes Analysis

As mentioned before, shape model is build from training set. This section describes how landmarks from different objects can be aligned in order to present shape independently of object global position, scale and orientation. This is done with a so called Procrustes Analysis. How this is obtained is described in the following subsections.

Translation

To eliminate global position of object from landmarks, all land marks must be translated to common coordinate frame with landmarks centroid being in the origin of the coordinate frame. Translation vector can be obtained with the following expressions.

$$\overline{x} = \frac{x_1 + x_2 + x_3 \ldots + x_N}{N}$$
$$\overline{y} = \frac{y_1 + y_2 + y_3 \ldots + y_N}{N}$$

Scale

In order to have a valid alignment, objects should also be rescaled to equal size. A scale factor s is obtained for each object according to the following expression.

$$s = \sqrt{\frac{(x_1 - \overline{x})^2 + (y_1 - \overline{y})^2 + (x_2 - \overline{x})^2 + (y_2 - \overline{y})^2 + \ldots + (x_N - \overline{x})^2 + (y_N - \overline{y})^2}{N}}$$

Orientation

Last step to objects alignment is rotation. Rotation angle θ can be computed in order to align two shapes $\mathbf{z^a}$ and $\mathbf{z^b}$.

$$\mathbf{z^a} = ((x_1^a, y_1^a), (x_2^a, y_2^a), \ldots, (x_N^a, y_N^a))$$
$$\mathbf{z^b} = ((x_1^b, y_1^b), (x_2^b, y_2^b), \ldots, (x_N^b, y_N^b))$$
$$\theta = \arctan \frac{\sum_{i=1}^{N}(x_i^b y_i^a - y_i^b x_i^a)}{\sum_{i=1}^{N}(x_i^b x_i^a - y_i^b y_i^a)}$$

Final alignment

Original landmarks can now be transformed in different coordinate frame by the following equation.

$$\begin{pmatrix} \mathrm{x} \\ \mathrm{y} \end{pmatrix} = \begin{pmatrix} \overline{x} \\ \overline{y} \end{pmatrix} + \begin{pmatrix} s\cos\theta & -s\sin\theta \\ s\sin\theta & s\cos\theta \end{pmatrix} \begin{pmatrix} x_{org} \\ y_{org} \end{pmatrix}$$

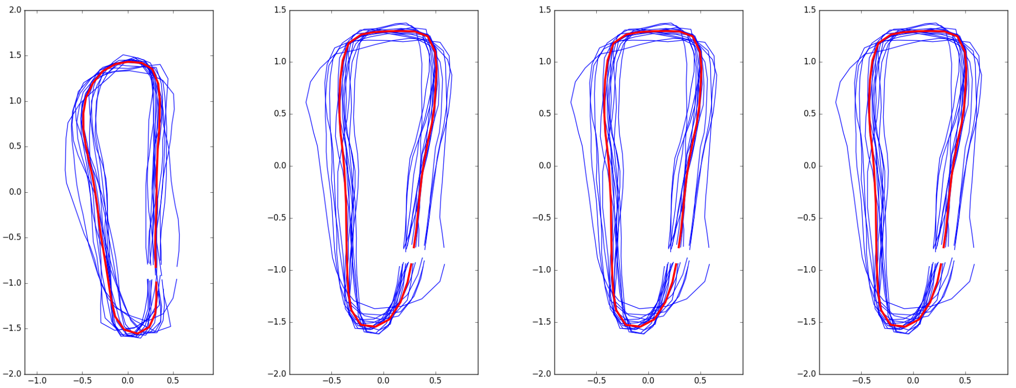
An algorithm aligns all shapes in the training shapes according to described Procrustes Analysis. The results for alignment of training set for each incisor can be seen on Figure 1.

1.2 Principal Component Analysis

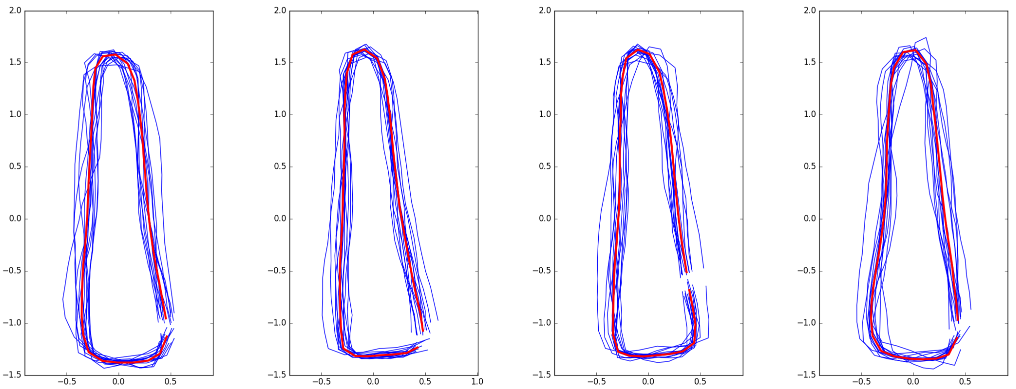
After shapes are aligned in common coordinate frame, the landmarks form a distribution over 2N dimensional space. A Principal Component Analysis (PCA) is now used to reduce the dimensionality from 2N to something more manageable. PCA produces new main axes which decrease the correlation of datapoints. Shapes differ most along this new set of axes. Any shape can be represented by the following expression.

$$\mathbf{z} = \overline{\mathbf{z}} + \mathbf{Pb}$$

Matrix $\mathbf{P}$ consist from t eigenvectors of covariance matrix and vector $\mathbf{b}$ is t dimensional vector of parameters, where t represents the number of principal components. By changing elements of vector



(a) Incisor index = 1 (b) Incisor index = 2 (c) Incisor index = 3 (d) Incisor index = 4



(e) Incisor index: 5 (f) Incisor index = 6 (g) Incisor index = 7 (h) Incisor index = 8

Figure 1: Shape alignment.

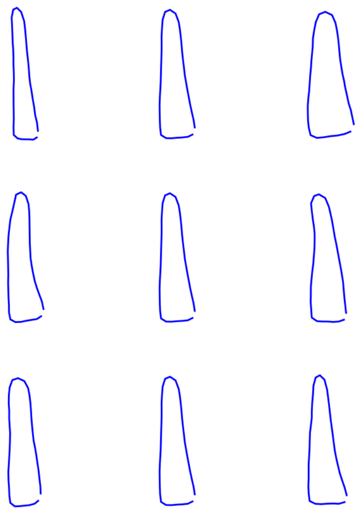


Figure 2: Principal Component Analysis.

b a shape variation can be made. To ensure that shape is representative of the object, limits should be applied on parameters in vector **b**. In this project parameters were limited to $b_i \leq \pm 2 \cdot \sqrt{\lambda_i}$, where λ_i corresponds to the i-th largest eigenvalue of the covariance matrix.
Figure 2 is presented in order to show how shape varies in relation to parameter change b_i. First row ob sub-figures shows shape variation when setting parameter b_1 to $-2 \cdot \sqrt{\lambda_i}$, 0 and $+2 \cdot \sqrt{\lambda_i}$. The second row show same results when manipulating parameter b_2 and third row show results for parameter b_3.

2 Image preprocessing

Dental radiographs are inherently noisy data, which also contain a lot of unnecessary information. The image does not include only teeth but also nose, jaw and other bones. If we work with high resolution images, we can waste a lot of computing power if while analyzing parts of the image that are irrelevant to application. Fortunately, based on some prior knowledge, we can assume that in the case of dental radiographs, the recording is made from the front of the face and the incisors are located in the middle of the picture. In that case first we can assume the roughly assume the position of the teeth and can crop out the central area of the image. Focusing on only on cropped image can save a lot of time and computation power in further work.

The next step is to examine the noise in the image. One can notice a particular type of noise, called 'salt and pepper' noise. The median filter can be used to treat this imperfection. The median filter works for removing 'salt and pepper' noise, but it cannot handle noise that is spread out in large areas. To reduce effect of this type of noise a bilateral filter is applied next in the pre-processing line. Bilinear filter is a non-linear, edge-preserving and noise-reducing smoothing filter, where the intensity value at each pixel in an image is replaced by a weighted average of intensity values from nearby pixels.Goal of the last step of preprocessing is to highlight edges ob objects. This is done with implementation of Scharr gradient operator. This operator provides a simple but efficient gradient approach where the gradient magnitude in each pixel is approximated by a finite differential. A histogram equalization was also tested, but the result overwhelmed the important parts of the image and did not contribute to better image quality.

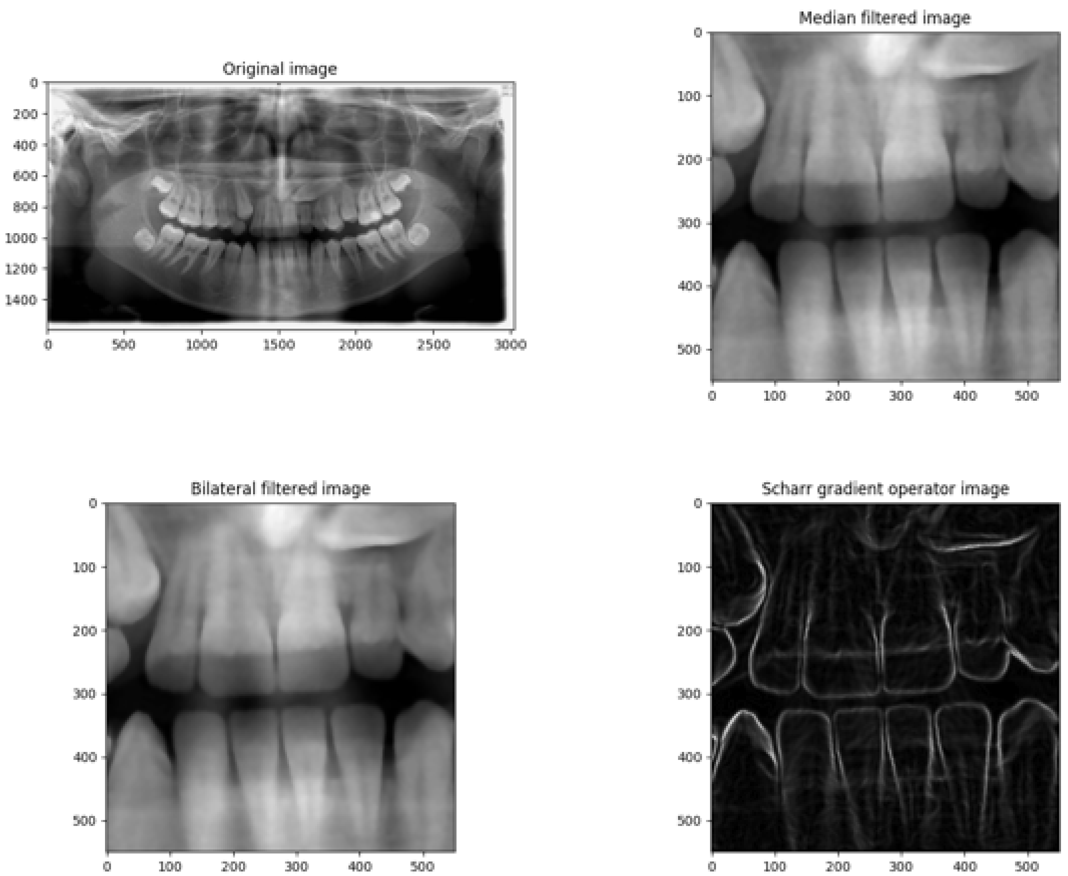


Figure 3: Filters used during pre-processing (Original image, Median filtered image, Bilateral filtered image, Scharr gradient operator filtered image) from left to right, from top to the bottom

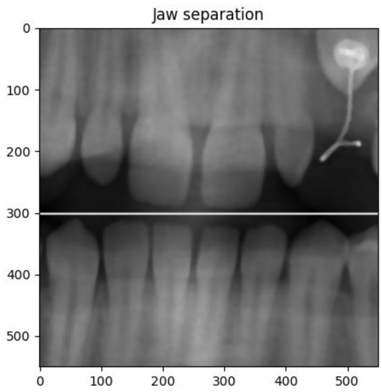


Figure 4: Separation of jaws based on a minimum search of global intensity

3 Model fitting to image

Automatic procedure that can fit shape model to a new image is roughly done in two steps. First, a good initial approximation of the object position must be determined. Second, an iterative approach of fitting a shape model to image is carried out in order to best match the model to the local surroundings.

3.1 Initial estimation

In order to fit our model as closely as possible to the new image, a good initial estimation of incisor is crucial. Approach that was chosen here is to begin with the upper and lower jaws separation. In order to separate the jaws, summation off the pixel intensities in each row is done and the line with minimal summation is selected as separation line. During this process, the method mentioned in the prepossessing - to crop the middle of the image and use just that part - is necessary, not only because we could save computing power, but we could have fake results otherwise. If we look at the original image the chin is getting narrower, we will encounter more black areas – this result that the minimum pixel intensity would be in these areas.

Once jaw separation is done the next step is to separate the teeth from each other. The teeth are bordered by intense vertical edges, so we can do better estimate their location on the x axis after detecting them in this way. For this approach first Canny edge detector was used, which gives a binary image of the shapes highlighting the edges.

This binary image is used as an input for the Hough line transformation, which detects these edges on this basis. Since transformation will find more lines than we need, we need to filter them under certain conditions. It can be assumed that the angle of the teeth will not be greater than ± 30 degrees, so that the lines whose slopes are larger can be erased. It is also assumed that two teeth will not be within a certain threshold value. We compared the start and end point of the detected lines and if they were within a given distance we also removed them. However, during the comparison, not both lines were removed - we examined which slope was steeper and we kept that one.

3.2 Model fitting

Once a initial approximation of object is given, an iterative algorithm moves around the landmarks to best fit the object borders. It first tries to locate a new set of landmarks which would correspond to the target object and secondly it tries to align current landmarks to target landmarks with Procrustes transformations and varying principal component parameters in vector **b**. The process stops when convergence is reached or when maximum number of iterations is exceeded.

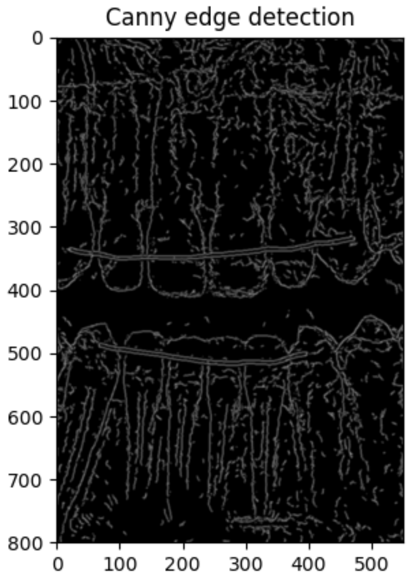


Figure 5: Separation of jaws based on a minimum search of global intensity

Locating target landmarks

New target landmarks are chosen along the normals to the object boundary. Normals are the lines perpendicular to the border at the points where model landmarks are located. To find best target landmarks the image intensity is compared to the average image intensity from the training set.
When training a model, image is sampled along the normals for each landmark to acquire a intensity profile. There are k samples made on each side, which means a vector $\mathbf{g}$ is obtained with length $2k+1$. For each landmark a mean intensity profile $\overline{\mathbf{g}}$ vector is computed.
Current intensity profile vector $\mathbf{g}_c$ samples at each iteration along the normals for each landmark. There are m samples collected on the each side of the landmark. Number m must be greater then k. A new location for the target landmark is selected at the place where intensity vectors $\mathbf{g}_c$ and $\overline{\mathbf{g}}$ match best. The matching assessment is done by the following expression.

$$f(j) = \sum_i \overline{g}_i \cdot g_{c_i}(j)$$

Number j represents which part ob vector $\mathbf{g}_c$ is taken for similarity comparison.
Figure 6 shows how sampling along normals is done for particular incisor from training set. The red dots highlight the normal along the landmark number 6. The intensity vector at landmark number 6 for different levels is presented graphically in Figure 7. More about levels is discussed in section Multi-level search.

The upper graph in Figure 8 illustrates the alignment of intensity vectors $\mathbf{g}_c(j)$ and $\overline{\mathbf{g}}$, while the lower part shows the value of function $f(j)$. Best alignment of intensity vectors is where value $f(j)$ is maximal.

Matching model landmarks to target landmarks

When new target landmarks are fund, the second step begins. Shape model landmarks are aligned to fit best to target landmarks using Procrustes analysis and varying parameter vector $\mathbf{b}$ form PCA. Values of vector $\mathbf{b}$ are limited ($b_i \leq \pm 2 \cdot \sqrt{\lambda_i}$).

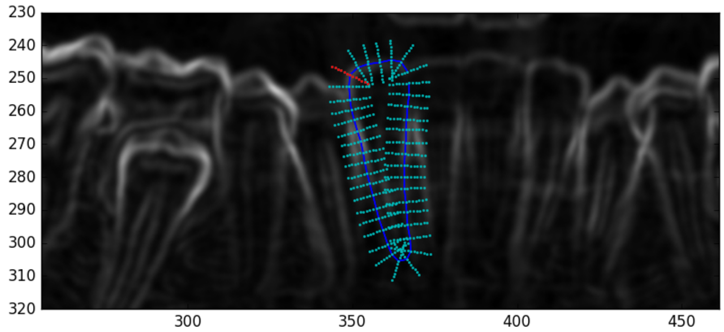


Figure 6: Profile normals along which sampling is made.

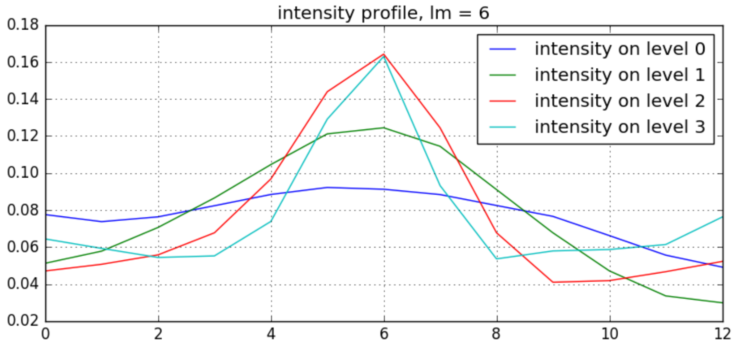


Figure 7: Profile normals along which sampling is made.

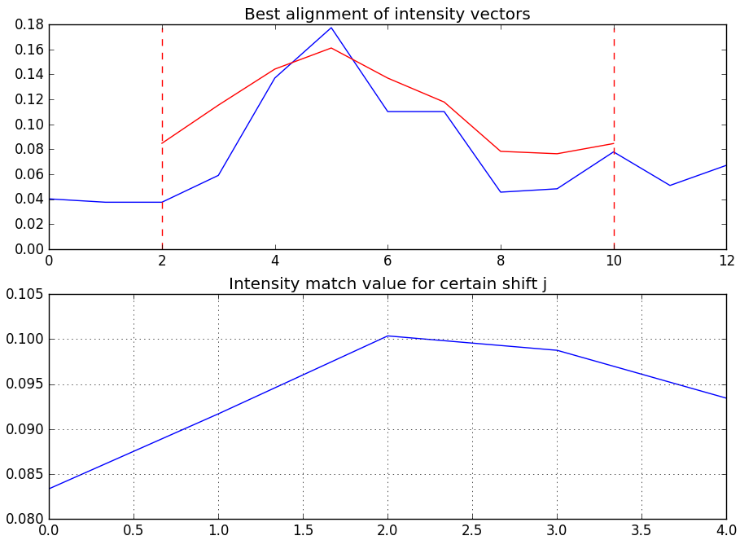
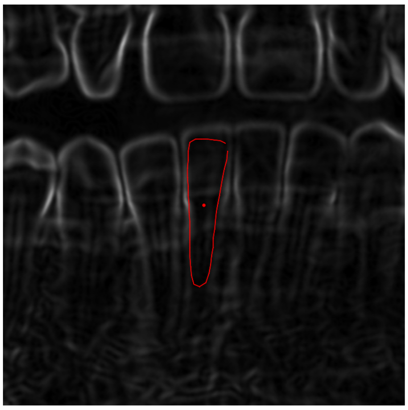
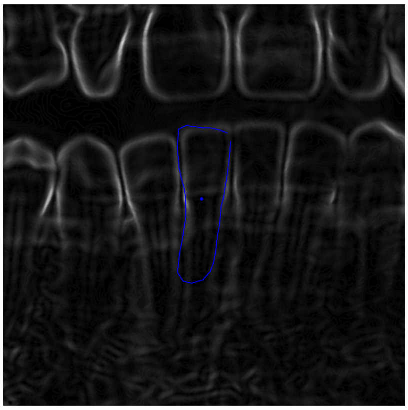


Figure 8: Best match for intensity vectors for certain landmark.



(a) Initial guess.



(b) After model fitting.

Figure 9: Results of model fitting.

The results of model fitting can be observed On Figure 9. The initial guess is presented o the left (Figure 9a) and the result after model fitting is presented on the right (Figure 9b).

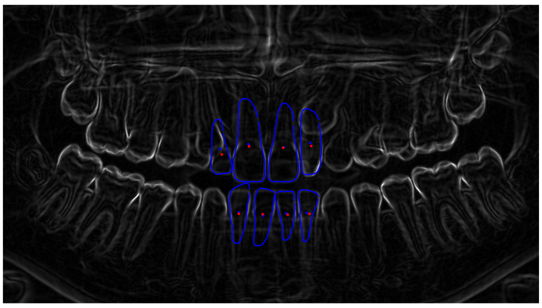
Multi-level search

For each radiograph a Gaussian pyramid is constructed in order to preform an efficient object search. Gaussian pyramid is constructed from several levels. First level or level 0 represents the original image. All the next levels in Gaussian pyramid are obtained from blurring and sub-sampling image from one level lower. Model fitting algorithm starts on the image form the highest level in the Gaussian pyramid. When model fitting converges, algorithm moves one level down and repeats the model fitting. Model fitting is done all the way to the bottom of the pyramid (level 0). This method improves efficiency and robustness ob the object search.

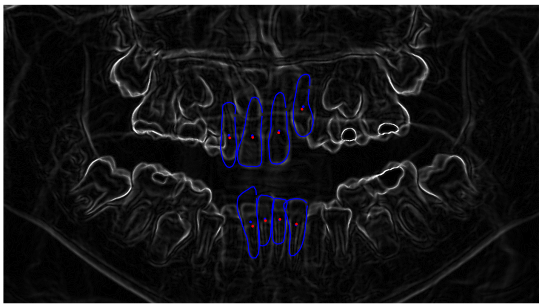
4 Evaluational

This section presents the results of described algorithm. Fourteen dental radiographs with provided landmarks were given on the beginning. This data was used to build and train the shape models for each incisor. In order to evaluate the the algorithm, another set of images are needed that were not yet seen by the algorithm. Sixteen extra dental radiographs were provided without noted landmarks. This images were used in this section to present the functioning of implemented algorithm.
In the following figure blue line represents the border of incisor that is found by the algorithm. Blue dot represents the centroid of found incisor and red dot represents the initial guess of the incisor centroid. The initial guesses was provided manually.

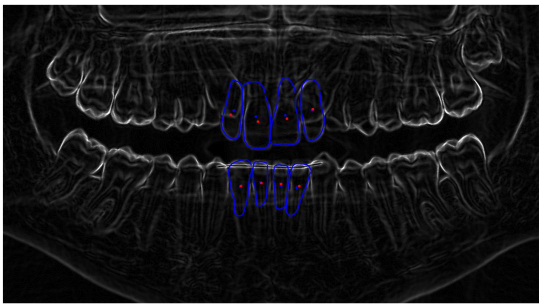
As it can be observed in Figure 10 that some incisors were located acceptably accurate, close to the real border that would be drawn by a human. But there are some case where incisor contour does not correspond well with the actual incisor. There are a few reasons why this malfunction appears. One reason is the layout of the teeth. Depending on the individual, some incisors are occluded by the neighboring teeth. An example of this scenario can be observed in Figure 10a or in Figure 10g. In bout cases, the most left upper incisor is not outlined well. Another reason for incorrect outline is the fact that our shape models were trained using a small training set, only 14 different images. In general, incisors shapes can vary greatly between people. Shape variation that was not observed in the training set cannot be detected by the shape model in later application. An example of this problem can be seen in Figure 10e, where the outlining of the most right upper incisor could have been



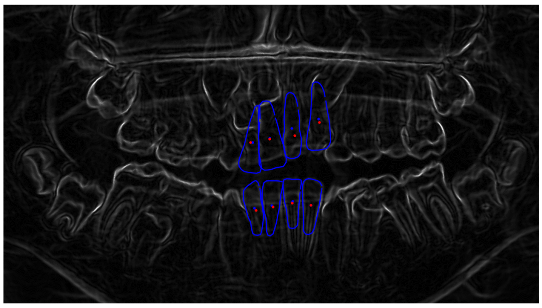
(a) Radiograph 15.tif



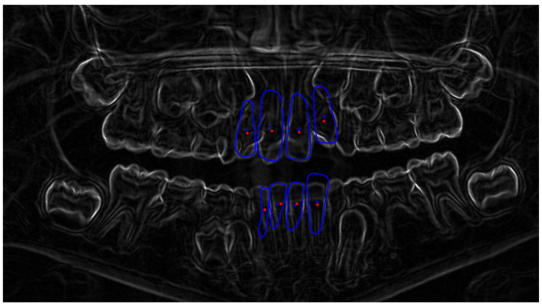
(b) Radiograph 16.tif



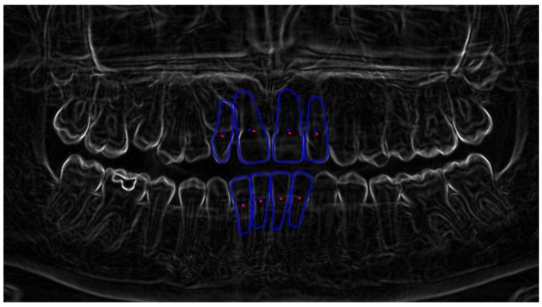
(c) Radiograph 17.tif



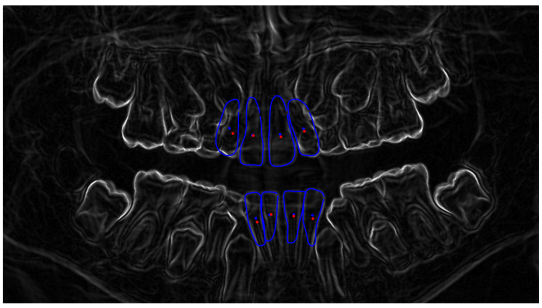
(d) Radiograph 18.tif



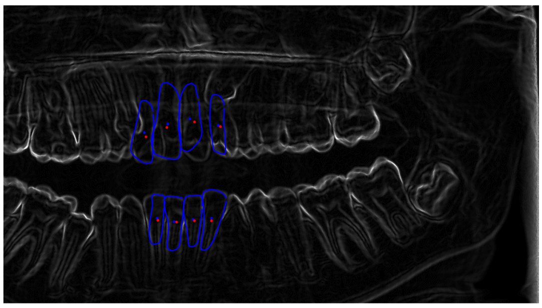
(e) Radiograph 19.tif



(f) Radiograph 20.tif



(g) Radiograph 21.tif



(h) Radiograph 22.tif

Figure 10: Final results of the algorithm.

done better. But the most important reason for incorrect outlining is the fact that dental radiographs are complex images with many similar object, laying next to each other. The edges of the objects are close by or even overlap. Many other objects or structures are present in the image that are not a matter of interest for this application. Although preprocessing of the dental radiographs does improve the quality of the image, enhances the meaningful structures (such as incisor's boundary) while reducing the noise, it cannot provide an ideal image. Improvements of the image preprocessing would definitely increase the accuracy of incisor outlining.
It is also very important to note that the final result highly depends on the initial guess. Initial guess of incisor centroid must be done well and must be close to real position of incisor centroid. If initial guess is done poorly, algorithm will converge to some local maximum that is far away from real incisor boundary.

Conclusion

The goal of the project was to automatize the incisor segmentation from a given dental radiograph. The end result of the project does not fully satisfy the chosen goal. The part where initial estimation is made encounters several problems and does not provide a suitable initial position of the incisor. This is why some changes were made in the final program that allows user to manually select the initial position of incisor. Form that point on the incisor contour automatically converges to a decent result. Therefor the final version of program can be called semi-automatic incisor segmentation. Given more time, other approaches for initial estimation could be tested. However the program allows to speed up the process of incisor segmentation in comparison to fully manual approach.