

FACULTY OF ENGINEERING SCIENCE
PROGRAMME FOR EXCHANGE STUDENTS
COMPUTER SCIENCE
KASTEELPARK ARENBERG 40 B-3001 LEUVEN



Computer Vision Report

Incisor Segmentation

Orsolya Lukacs-Kisbandi, r0650042
Stefano Proto, r065080

June 10, 2017

Contents

1	Introduction	1
2	Active Shape Model	1
2.1	Mean shape	1
3	Image Processing	2
3.1	Pre-processing	2
3.2	Detecting edges	4
4	Fit model on an image	4
4.1	Hand draw boxes	4
4.2	Fine-tune boxes	5
4.3	Fitting the Active Shape Models	5
5	Evaluation and final considerations	7

1 Introduction

The aim of this project was to write an algorithm that is capable of segmenting the upper and lower incisors in panoramic radiographs using a model based approach. The chosen method is called Active Shape Models, and it is described in [1] Cootes et al. The implemented algorithm can be broken down into sub-parts, namely: building the Active Shape Model, described in section 2, pre-processing the input images, explained in section 3, and finally fitting the model on the images, as presented in section 4. In addition, our observations and experimental results will be discussed in the final part of the report.

2 Active Shape Model

The reason for building Active Shape Models for every incisor is to create a general representation of each one of them. By obtaining these models from combining different variations of the same tooth extracted from different people's radiographs, we can build a global model which can be fit on the same incisor for new images. This is a good approach for such problems because it creates a general representation of the object in interest, becoming more robust against variability.

Active Shape Models are based on a prior model of what is expected to be seen on the new image. In our case these are the incisors, represented by a set of characteristic points, the landmarks. Based on these landmarks, a general model of the expected shape is created, which can be used to find the best match of the model in the new image. This model gives a compact representation of the previous knowledge allowing some variation, but being specific enough to exclude changes which haven't been seen in the training data. The landmarks are specific coordinates on the training images which represent a distinguishable point which appears on every training example.

2.1 Mean shape

The training set contains samples from several people. For each person in the training data, every incisor is represented by a set of 40 landmark points describing the shape of the tooth. In order to create the models, the shape of each tooth had to be scaled and translated to the origin, to make the computations easier and for the comparison of landmarks to be possible. Translating the images also included rotating them by an angle θ so their x and y axes would be aligned. After the shapes were centered to the origin based on their center of gravity, the scaled mean shape for each tooth was computed. This was necessary because we wanted to build a scale and rotation invariant model which could handle any type of input.

The following procedure was done iteratively until the mean shapes converged to a value which did not change after realigning the landmarks to the new mean as proposed in [2]. After obtaining the mean shape, the landmarks were scaled and centered again around the new mean shape. This technique is called the Generalized Procrustes Analysis. The main idea is that the alignment and re-centering of the shapes is done iteratively until the model converges to a value which does not change after adding another iteration. Our models were con-

sidered stable when the squared sum of the distance of each shape to the mean was less than 10^{-10} .

The used approach is shown in Figure 1, while the final result of the computed models can be observed, for each tooth, in Figure 2.

Figure 1: Followed procedure to create mean shape models, based on [2]

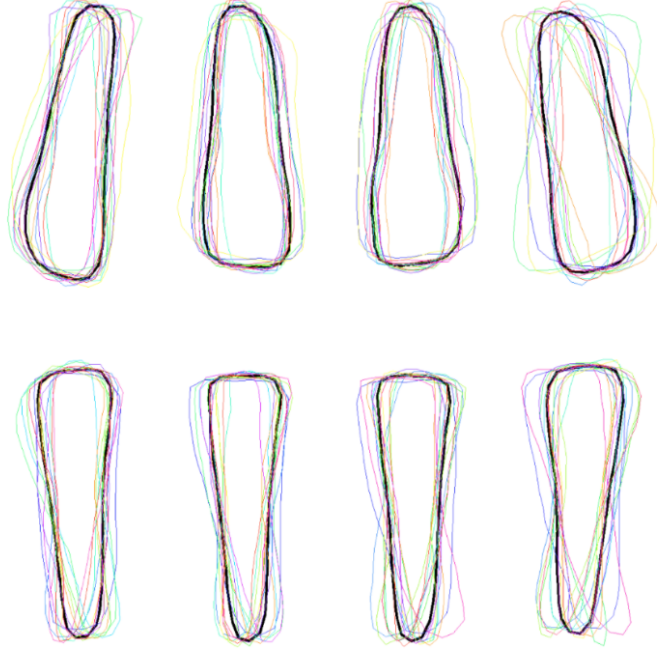


Figure 2: Final results of the Active Shape Models(black) from the given landmarks (other colors)

3 Image Processing

3.1 Pre-processing

Radiographs are usually noisy and low in contrast. These characteristics make the image features difficult to extract. In order to overcome this issue, sharpening techniques and contrast enhancement algorithms are applied to the images before they were analyzed.

In order to reduce noise, we experimented with different filters, each having its own advantages and disadvantages. A balance between these filters had to be found, since some of them removed important image features, such as edges and high contrast points. For reducing the impulsive and the Gaussian noise without affecting the edges, a median filter and a bilateral filter was applied. The median filter considers each pixel in the image in turn and looks at the

nearby pixels to decide whether or not it is representative of its surroundings. It replaces the pixel value with the median of the neighbouring pixel values. The median is calculated by first sorting all the pixel values from the surrounding neighbourhood into numerical order and then replacing the pixel being considered with the middle pixel value. Instead of the simple median filter, we tried to apply an adaptive median filter but this required too much computational power as well as time, so we decided to use the much simpler version instead. The Bilateral filter is a non-linear, edge-preserving and noise-reducing smoothing filter. The intensity value at each pixel in an image is replaced by a weighted average of intensity values from nearby pixels, which is based on a Gaussian distribution. Its formulation is simple: each pixel is replaced by an average of its neighbours. It depends only on two parameters that indicate the size and contrast of the features to preserve [5]. The results of the noise suppression can be seen in Step 1 and Step 2 of the preprocessing stages, presented in Figure 3.

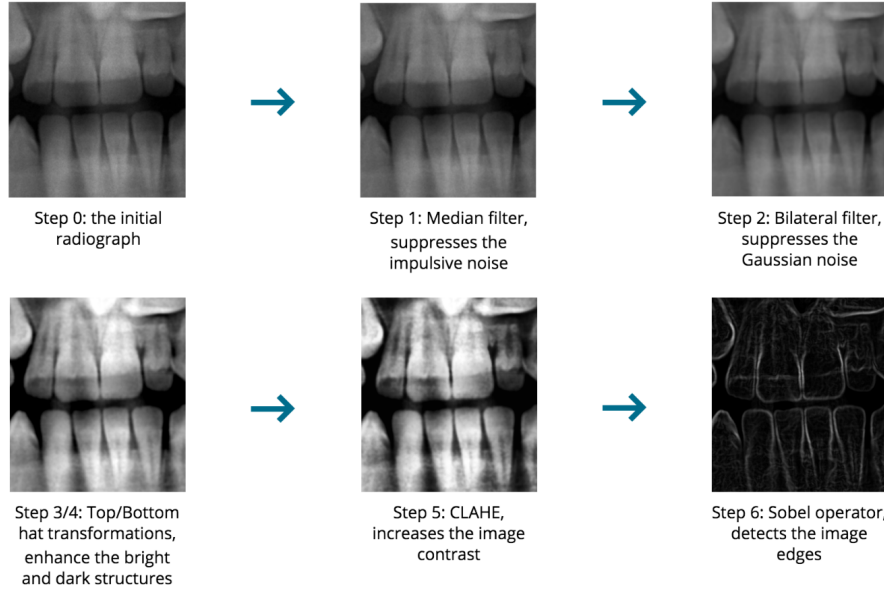


Figure 3: Preprocessing stages

After these transformations, as suggested in [6], the contrast of the image is enhanced by means of two gray-level morphology techniques (top-hat and bottom-hat transform) and a histogram equalization technique. The top-hat transform is able to enhance brighter structures, while the bottom-hat transform enhances the darker structures. The obtained picture from top hat transform is added to, from the bottom hat transform is subtracted from the initial radiograph to apply the effect of this transformations. The last step to enhance the contrast of the image is histogram equalization. For this purpose many approaches were tried: among all the candidate algorithms, in Ahmad et al. [4] it is shown that the Contrast Limited Adaptive Histogram Equalization (CLAHE) better enhances the image quality and increases the diagnostic ease, over performing both the Sharp Contrast Limited

Adaptive Histogram Equalization (SCLAHE) and the Median Adaptive Histogram Equalization (SMAHE) [4]. CLAHE is an improved version of Adaptive Histogram Equalization (AHE). The image is divided into small blocks, and then the histograms are equalized. An important improvement compared to AHE is that, instead of global histogram equalization, the algorithm adapts to local variations in an image, such as shadows and highlights. AHE could amplify the noise, to prevent this, the Contrast Limited technique is used: if any histogram bin is above a specified contrast limit, those pixels are clipped and distributed uniformly to other bins before applying histogram equalization. Steps 3, 4 and 5 presented in Figure 3 show the effects of the transforms over the images.

3.2 Detecting edges

In order to find the edges over the images, the Sobel operator is used. The Sobel operator performs a 2-D spatial gradient measurement on an image and emphasizes regions of high spatial frequency that correspond to edges. Typically it is used to find the approximate absolute gradient magnitude at each point in an input grayscale image. G_x is representing the horizontal changes G_y the vertical ones on the image.

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * I \quad G_y = \begin{bmatrix} -1 & 2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} * I$$

At each point of the image we calculate an approximation of the gradient in that point by combining both results above with the following formula.

$$G = \sqrt{G_x^2 + G_y^2} \quad (1)$$

The result of the Sobel operator can be seen at Step 6 of Figure 3.

4 Fit model on an image

Fitting the models on the image consists of multiple stages. At first hand drawn boxes have to be created to train the system, than the *estimate* function automatically fine tunes them based on the main features of the segment of interest. After returning these boxes, the Active Shape Models are fitted on the image.

4.1 Hand draw boxes

In the training part, the user is asked to draw boxes around the upper and lower incisors for every sample image in the training set.

The obtained boxes are stored in a file, and they are used as initial values for the automatic estimate. For testing, we select the coordinates which are the most extreme, the combination of these representing the largest box, in which all upper or lower incisors will fit in.

These two boxes (one for the upper and one for the lower incisors) are used

as initial estimates on the location of incisors when we are applying our model on an unseen radiograph. Choosing the largest boxes allows the system to be general and include all the teeth, even for unseen images. This assumption was made based on the fact that, when radiographs are taken, the persons' head is facing forward, having the incisors in the middle of the image.

4.2 Fine-tune boxes

Figure 4: Windows of different scales (white), sliding over the largest box, searching area, which is marked in red

For fine-tuning the boxes the function *estimate* was used. This function takes as parameter a preprocessed radiograph we want to fit the box on, a boolean variable, whether we are looking for an upper or lower incisor, the "biggest box" coordinates, which represent the biggest image segment selected from the hand-drawn boxes, the coordinates from the hand drawn boxes for the particular region and finally *b_model*, whether we want to build the model or just load the pre-trained model. Using the largest box, the radiographs are cut, resized and flattened (create a $N \times 1$ array from the image segment) based on these coordinates. We apply PCA on the cut images and take the top ten components, which capture most of the variance in the data. After these steps, we invoke the function *best_seg* which is taking as parameters the mean returned from PCA, the eigenvectors, the bounding box coordinates, the preprocessed radiograph, and the width and height of the box.

In this function, sliding windows of different scales are shifted over the region of interest on the radiograph and the error of reconstructing the segment using the eigenvectors and mean acquired from the cut shapes is calculated. Figure 4 shows the windows of different sizes and scales sliding over the searching area. When a window is able to reconstruct the image with a better error, its coordinates are saved. The segment with the smallest error is assumed to be the one better representing the incisor patterns, so it is the one chosen to be returned.

The result of the box tuning is shown in Figure 5: the upper and lower incisors are identified respectively by the two boxes.

4.3 Fitting the Active Shape Models

To fit the models, first the upper and lower incisor boxes are divided in four equal parts. This decision is based on the assumption that incisors in the same jaw generally have the same width. Following this, each tooth model is shaped in the corresponding box: the models are centred in the respective tooth box, scaled to fit in it and then centered again.

After this initial positioning and scaling of the models, they are parsed in order to look for a better estimation of the tooth in their neighbourhood. For every point in the model, the normal line is computed. In case another point is a better fit than the current one, we update the model. The criteria for a point to be better than another is to have a higher value, since edges are represented by light regions. For this reason, brightness along the norm to the model at the

(a)
Best
box
for
up-
per
in-
cisor

(b)
Best
box
for
lower
in-
cisor

Figure 5: Fitted boxes for Radiograph 0

current point is compared with the current value of the model on this line. In order to have a greater resistance to noise not only the points on the line are considered, but for each point all its neighbours are evaluated. An average of the region is computed, and then compared with the current brightest point for the line: if the value of the region is greater than a certain threshold (directly depending on the current brightest point), then the centre of the evaluated region will become part of the model. The parameters that allow flexibility in the model are the length of the norm we calculate (this can decide on the region we are looking for points which are more likely to be edges) the step we traverse the shape of the incisor, larger step is more resistant to noise, while smaller steps are following the shape more precisely. The chosen values for the parameters are based on domain knowledge, empirical experiments and evaluations.

The shape of the model is affected by the fitting, and it has to be smoothed. After the final model is found, the *smooth_model* function is invoked. This function smooths the model, computing an average of the shape over three consecutive points, reducing the effect of outliers in the model and creating a better representation of the incisor. Figure 6 shows some of the steps of the fitting process, namely the fitting of the model into the tooth box, Figure 6a, and the parsing over the normal lines of the model points, Figure 6b.

(a) Fitting of the midel into the tooth box (b) Parsing the model to improve accuracy

Figure 6: Steps while fitting the model over the image

5 Evaluation and final considerations

In order to evaluate the obtained result, two error estimations are considered.

The first value we are computing is the Mean Squared Error. This represents the quadratic distance from the obtained model and the original landmarks. It is computed both for the x and y axis for every sample of which we have the corresponding landmarks.

The Mean Squared Error for each incisor is computed as:

$$MSEx = \frac{1}{n_{teeth}} \sum_{i=1}^{n_{teeth}} \frac{1}{n_{lm}} \sum_{i=1}^{n_{lm}} (Xe_i - Xl_i)^2 \quad (2)$$

$$MSEy = \frac{1}{n_{teeth}} \sum_{i=1}^{n_{teeth}} \frac{1}{n_{lm}} \sum_{i=1}^{n_{lm}} (Ye_i - Yl_i)^2 \quad (3)$$

The second value we are computing is the Mean Absolute Error. It represents the absolute value of the difference between the estimated landmarks and the original ones. It is computed both for the x and y axis for every sample of which we have the corresponding landmarks.

The Mean Absolute Error is computed as:

$$MAEx = \frac{1}{n_{teeth}} \sum_{i=1}^{n_{teeth}} \frac{1}{n_{lm}} \sum_{i=1}^{n_{lm}} |Xe_i - Xl_i| \quad (4)$$

$$MAEy = \frac{1}{n_{teeth}} \sum_{i=1}^{n_{teeth}} \frac{1}{n_{lm}} \sum_{i=1}^{n_{lm}} |Ye_i - Yl_i| \quad (5)$$

From the Mean Absolute Error it is possible to compute an estimation of the percentage error made with respect to the image dimensions. This can give an idea of how big the error is compared to the image resolution. To compute this value, the Mean Absolute Error needs to be divided consecutively by the image width, and the image height, and then multiplied by 100.

The algorithm has been tested using Leave one out Cross Validation on the data set containing 14 annotated radiographs. Therefore, the teeth models are build over the 13 images selected for testing while one is kept for testing.

We can see, observing the results presented in Table 1, that the errors have similar values. This proves that we have a stable model, which gives reliable results, having similar performance for every image. When the teeth on the radiographs are shifted to the side or have unusual orientations, the model can not fit them perfectly since the edges are not at the positions where they are expected to be.

We can observe that the accuracy of the hand drawn boxes in the training has a significant affect on the outcome of the estimation. Based on these error indexes and intermediate results, we can infer that the estimated position of the incisors in the radiographs is mainly affected by the box fitting.

In addition, unusual teeth positions, caused for example by a bad accuracy when the radiograph itself was take, might affect the model as well.

Finally, we think that a larger training set could further improve the accuracy of the results.

	MSE_x	MSE_y	MAE_x	MAE_y
1	2052.93	2255.69	10.04	6.10
2	309.05	392.54	0.97	1.12
3	335.12	1037.68	0.45	1.74
4	259.79	409.60	1.10	0.83
5	3475.37	1712.84	10.31	2.31
6	3042.80	1004.48	7.76	4.18
7	244.44	405.18	2.71	0.49
8	286.43	10644.19	0.71	10.59
9	4164.56	1980.69	11.35	2.51
10	4063.35	206.04	0.77	0.38
11	2155.47	512.22	2.19	2.34
12	2956.24	4429.92	10.85	7.74
13	186.61	205.67	1.85	2.24
14	150.68	425.43	0.71	1.30

Table 1: Results from the leave-one-out-testing. On the left the left out sample, in the first two columns the Mean Squared Error, in the last ones the Mean Absolute Error

(a) Best results - Radiograph 14 (b) Worst results - Radiograph 9

Figure 7

The cause of the bad performance on the radiograph on Figure 7b is the presence of the line with high intensity (probably braces). This is because the ground truth for the best fitting boxes is based on the pixel intensity inside the selected window. This leads to the wrong estimation of the best fitting box, returning the wrong image segment. Since the box is not the right dimension (too big) and we split it in four even parts to find the incisors, it can not detect them accurately.

When comparing the results from the hand drawn boxes and the estimated boxes, we can draw the same conclusion. There is a strong correlation between the image quality, estimated boxes and the performance of the algorithm. (Figure 8a and Figure 8b)

(a)
Ra-
dio-
graph
14

(b)
Ra-
dio-
graph
9

Figure 8: Comparing results from hand drawn boxes and estimated boxes

References

- [1] Cootes, Tim, E. Baldock, and J. Graham. "An introduction to active shape models." *Image processing and analysis* (2000): 223-248.
- [2] Cootes, Timothy F., et al. "Active shape models-their training and application." *Computer vision and image understanding* 61.1 (1995): 38-59.
- [3] Blanz, Volker, et al. "A statistical method for robust 3D surface reconstruction from sparse data." *3D Data Processing, Visualization and Transmission, 2004. 3DPVT 2004. Proceedings. 2nd International Symposium on*. IEEE, 2004.
- [4] Ahmad, Siti Arpah Bt, et al. "Utilizing contrast enhancement algorithms (CEAs) in identification of dental abnormalities." *Technological Advances in Electrical, Electronics and Computer Engineering (TAECE), 2013 International Conference on*. IEEE, 2013.
- [5] Paris, S., Kornprobst, P., Tumblin, J. and Durand, F., 2007, August. "A gentle introduction to bilateral filtering and its applications." In *ACM SIGGRAPH 2007 courses* (p. 1). ACM.
- [6] R.-Y. Huang, L.-R. Dung, C.-F. Chu, and Y.-Y. Wu. "Noise removal and contrast enhancement for x-ray images." *Journal of Biomedical Engineering and Medical Imaging*, 3(1):56, 2016. 11
- [7] "Model reconstruction". Computer Vision [H02K5a], KU Leuven