

KU LEUVEN

COMPUTER VISION

---

# Final Project: Incisor Segmentation

---

Joran Van de Woestijne (r0378602)  
Tim Van Den Broecke (r0296620)

August 2017

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Shape Model</b>	<b>1</b>
2.1	Shape alignment . . . . .	2
2.2	PCA . . . . .	2
<b>3</b>	<b>Radiograph pre-processing</b>	<b>2</b>
<b>4</b>	<b>Fitting model to image</b>	<b>5</b>
4.1	Initial pose estimate . . . . .	5
4.2	The ASM algorithm . . . . .	8
<b>5</b>	<b>Evaluation</b>	<b>11</b>
5.1	Euclidian distance . . . . .	11
5.2	Ground truth surface found . . . . .	12
5.3	Correctness of solution surface . . . . .	13

# 1 Introduction

This report discusses our solution for the Incisor Segmentation project for the Computer Vision course. The goal of this project was to develop a model based segmentation approach, capable of segmenting the upper and lower incisors in panoramic radiographs. We achieved this by utilizing the Active Shape Models approach, as it was first presented by Cootes et al [1]. This report is structured as follows: Section 2 starts off by discussing the deformable shape model, which is followed up by Section 3 concerning the pre-processing of the dental radiographs. Section 4 presents the actual model fitting algorithm, with the evaluation of the fitting presented in Section 5.

## 2 Shape Model

We use two models in our algorithm. One describes the mean and variation of all four upper teeth (incisors 1-4), the other describes the lower teeth (incisors 5-8).

The choice for two models containing four incisors each, as opposed to one model per tooth, comes from two main reasons that both stem from simple empirical observations about the human dental structure. The first is that the upper and lower incisors are always found next to each other and in the same order. Natural variations on the spacing between them, their height differences and sizes are all captured in our final model description. Our approach ensures that teeth will never be marked out of order within the model, even when some overlap. The second observation is that natural variations don't favor one side over the other and thus we expect a symmetric model. Including mirrored examples during training is one way to incorporate this information. Our model adds to the effect by also enforcing symmetry in the variation in spacing, height and size.

The model construction happens in two steps, as described by Cootes in [1] (ignoring the labelling as it has been performed for us). First, we align all the shapes. This means making all example sets of four incisors (either all upper or all lower) neutral with regards to translation, rotation and scale. Afterwards, we perform a principal component analysis (PCA) on the aligned shapes to acquire a robust description of the average upper/lower incisors and the way they vary in shape.

## 2.1 Shape alignment

Starting from the labelled set of images and their eight manually landmarked incisors, we group the upper and lower landmarks and combine them into one shape vector. This vector contains the coordinates of 160 ( $= 4 \times 40$  points per landmark) points. When testing the performance of our algorithm for a specific image, we remove it and its mirrored version from the training set. We are then left with 26 training examples per model (upper or lower) for an image.

We take the first example as the initial mean shape estimate and normalize it. Using Procrustes analysis, we refine the estimate using the other examples until it converges. At the start of each refinement step we align all examples with the new mean so that when the procedure has converged, we have acquired a fully aligned set of examples. We see the process illustrated in Figure 1 when training for image 4. Notice that the slight tilt of the first example (and therefore also the initial mean estimate) has cascaded into the full alignment set. This can not be prevented using our technique.

## 2.2 PCA

After aligning our training set, we want to extract statistical properties from it describing the mean shape of a model and the typical variation found within them. A principal component analysis delivers just that. We create a procedure that takes a training set of models and produces a PCA model of the set and a description of the training set in terms of said model (called the *reduced training set*). The model is used throughout the algorithm to correct new estimated fits. The terms in the reduced training set are analysed to determine a plausible range for models. We confirm in Figure 3 that (positive or negative) three times the standard deviation on each term captures the range of plausible shapes.

To determine the number of modes we need for the PCA, we provide a dynamic procedure. It first calculates all eigenvalues in descending order and then searches for the number of values needed to cover 95% of the shape variation inside the training set. We plot the coverage per number of modes used to train for image 4 in Figure 2.

# 3 Radiograph pre-processing

Dental radiographs are inherently noisy and contain too much information when just focusing on locating the upper and lower incisors. For instance, all the teeth (not just the incisors), the jaw and the nose are present in



(a) Model 1: upper incisors



(b) Model 2: lower incisors

Figure 1: Our two models before and after alignment. Examples 4 and 18 (mirrored 4) were left out. On the left we see the original landmarks simply superposed in one image, each with their own color gradient. On the right, we see the the aligned models with their same respective color.

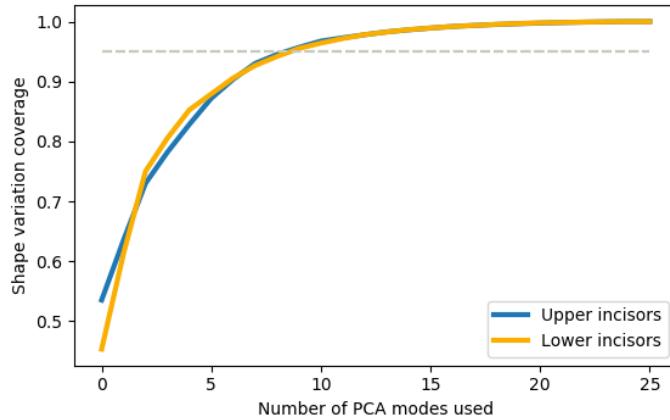


Figure 2: Coverage plot for our two models. We see that we need 10 modes of our PCA in order to cover 95% of shape variations in the training set.

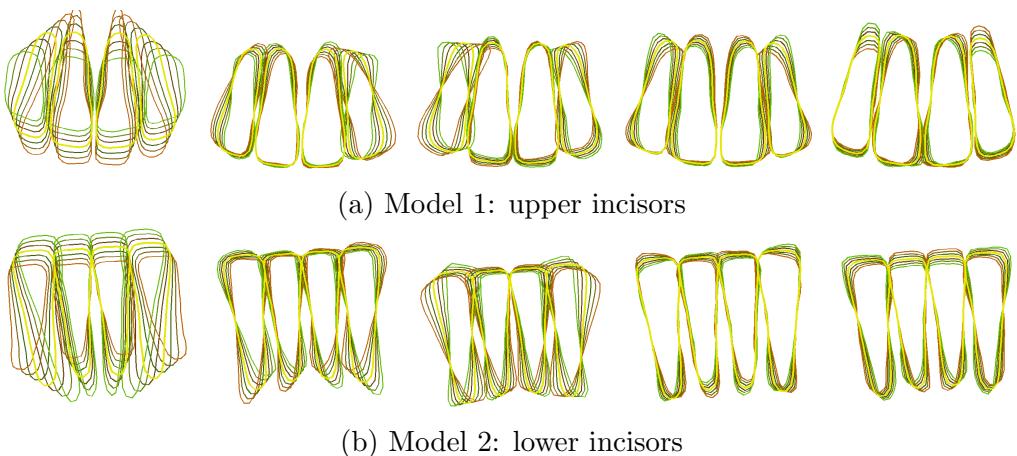


Figure 3: The five main principal components in order of our two models, excluding examples 4 and 18. The mean shape is shown in yellow. Each component is shown with a scaling between -3 and +3 times the standard deviation of the component.

the radiographs. To reduce these radiographs to a more suitable size for our application, we crop them based on domain knowledge so that only the smaller region around the four incisors remains.

After the size reduction, the radiographs still need to be de-noised so that it's easier for our algorithm to detect the edges. We achieve this by using the technique proposed in *Noise Removal and Contrast Enhancement for X-Ray Images* by Huang et al [2]. The algorithm proposed by Huang et al is visualised in Figure 4. This technique first starts off by applying an adaptive median filter to reduce the impulsive (salt-and-pepper) noise, after which a bilateral filter is used to suppress the Gaussian noise while also preserving the edges of the image. After the noise suppression step, the radiograph is sharpened by a mathematical morphology, here a top hat and bottom hat transform. The top hat transform enhances the brighter regions of the radiograph, whereas the bottom hat enhances the darker regions. Finally, the contrast of the radiograph is enhanced by applying a contrast limited adaptive histogram equalization (CLAHE). To detect the edges of the dental radiograph, a Sobel operator is used, detecting the gradients in the x and y direction. The results of applying the different pre-processing steps on a dental radiograph can be observed in Figure 5.

## 4 Fitting model to image

To fit our model to an image, we have to find the model instance  $x = T_{(X_t, Y_t, s, \theta)}(\bar{x} + Pb)$  that most accurately aligns with the incisors in the image. This translates to finding the pose parameters  $T_{(X_t, Y_t, s, \theta)}$  and the shape parameters  $b$ . This is achieved in two steps: In the first step, an initial estimate for the pose parameters is obtained, which is discussed in Section 4.1. The second step finally fits the model to the image by applying the Active Shape Model algorithm, which is discussed in Section 4.2.

### 4.1 Initial pose estimate

To determine an initial estimate for the pose parameters  $T_{(X_t, Y_t, s, \theta)}$ , we need to find an estimate for the horizontal offset  $X_t$ , the vertical offset  $Y_t$ , the scale  $s$  and the rotation  $\theta$ .

**Scale and rotation** To find an initial estimate for the scale and rotation, we use domain knowledge. A good scale was empirically determined, and the rotation is chosen so that the teeth are positioned vertically.

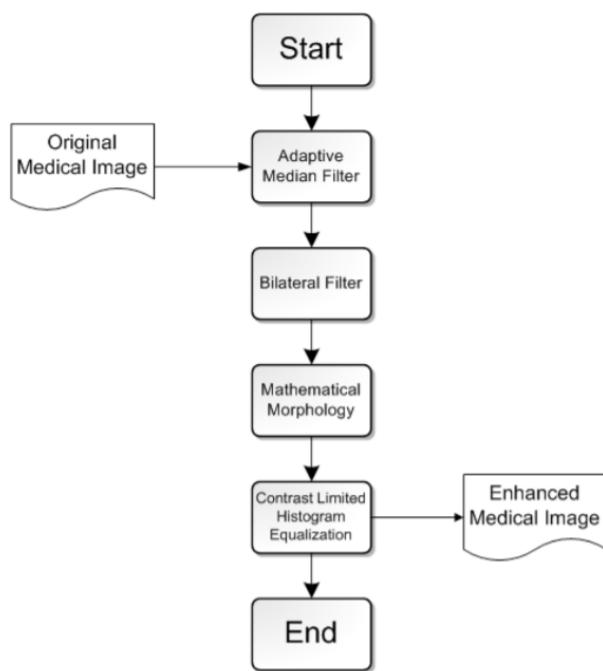


Figure 4: The algorithm proposed by Huang et al [2] for denoising dental radiographs. The algorithm has four steps, with a top hat transform and a bottom hat transform representing the mathematical morphology step.

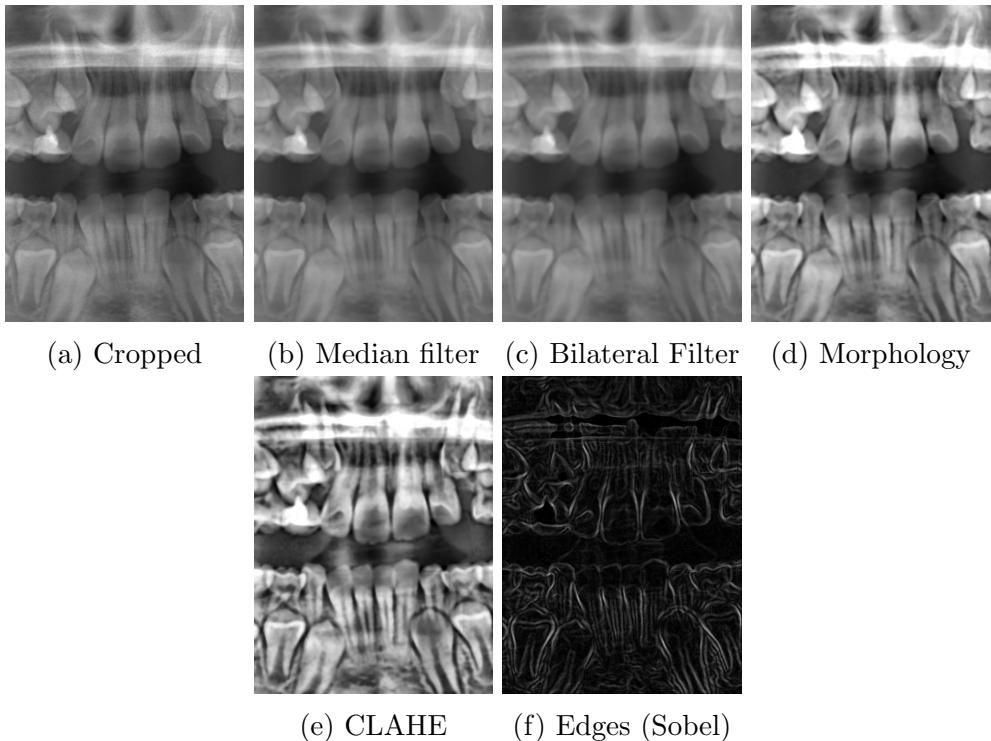


Figure 5: The results of the different pre-processing steps visualised when applied on dental radiograph 3. (a) shows the original image after cropping, (b) shows (a) after applying the adaptive median filter, (c) adds the bilateral filter onto the result of (b), (d) performs the mathematical morphology (top hat and bottom hat) on (c), and (e) finally applies CLAHE. A Sobel operator is used to detect the edges, which is visualised in (f).

**Vertical offset** To find the vertical offset  $Y_t$ , we utilise the jaw separation of the radiograph. This is accomplished by first finding the jaw separation that separates the upper incisors from the lower incisors, which corresponds to the pixel row that has the lowest total intensity value. After having found this separation, a small region above and below this separation is considered to locate the upper and lower offset where the teeth actually start. This is achieved by looking at the pixel row with the highest gradient in this region, which is caused by the transition from the black background to the white teeth. This technique was able to correctly determine the vertical offset for 12/14 images for the upper landmarks and 13/14 images for the lower landmarks. The result of locating the vertical offset can be observed in Figure 6(a).

**Horizontal offset** To find the horizontal offset  $X_t$ , we use a hough line transform to detect the boundary between the middle incisors. For the upper incisors, a small region above the vertical offset is considered, near the center of the image. A hough line transform is applied on this small region to detect the boundaries between the different teeth. The line closest to the middle of this region then represents the boundary between the middle incisors, which we use as the horizontal offset. The results of this can be observed in Figure 6(b). The method works analogously for the lower incisors, but now a small region below the vertical offset is considered. This is visualised in Figure 6(c). This method was able to correctly determine the horizontal offset for 13/14 images for the upper landmarks and 13/14 images for the lower landmarks.

## 4.2 The ASM algorithm

The Active Shape Model algorithm we use to fit the model to an image closely follows the algorithm as it is described in Cootes et al [1]. The algorithm is reproduced here in Algorithm 1.

---

### Algorithm 1: Active Shape Model Algorithm

---

- 1 Examine a region of the image around each point  $\mathbf{X}_i$  to find the best nearby match for the point  $\mathbf{X}'_i$
  - 2 Update the parameters  $(X_t, Y_t, s, \theta, \mathbf{b})$  to best fit the new found points  $\mathbf{X}$
  - 3 Apply constraints to the parameters,  $\mathbf{b}$ , to ensure plausible shapes  
(limit so  $|b_i| < 3\sqrt{\lambda_i}$ )
  - 4 Repeat until convergence
-

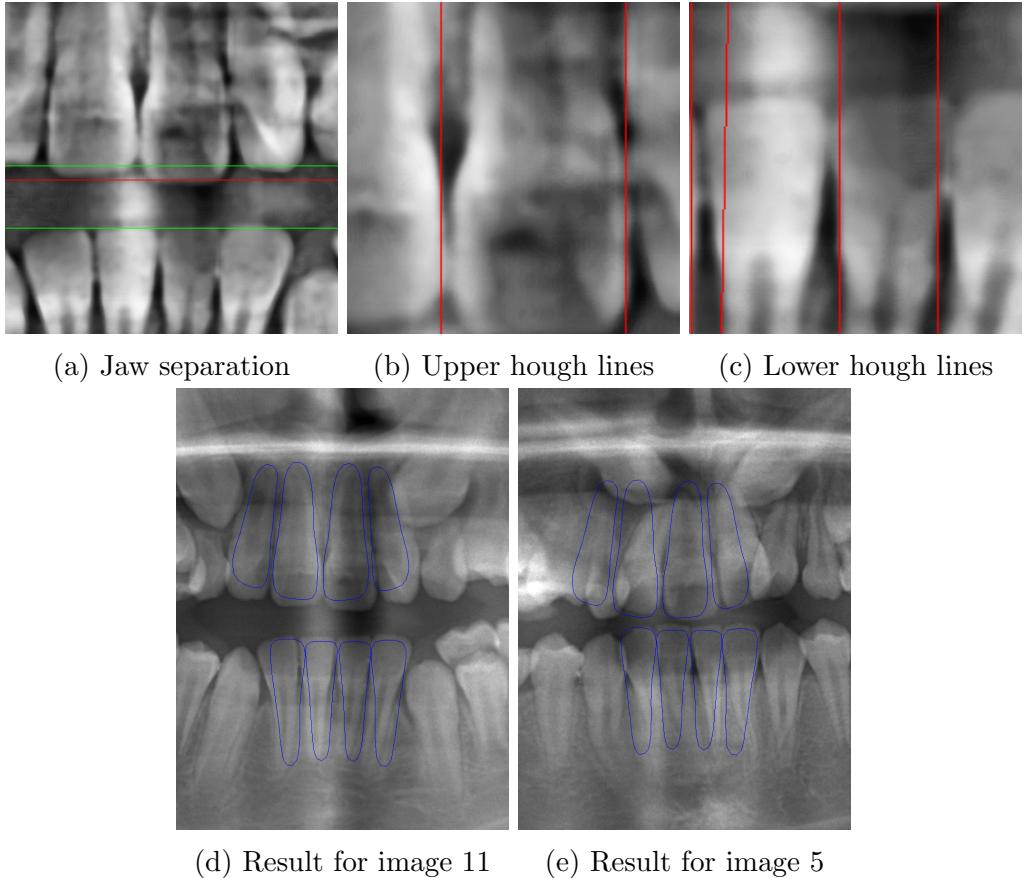
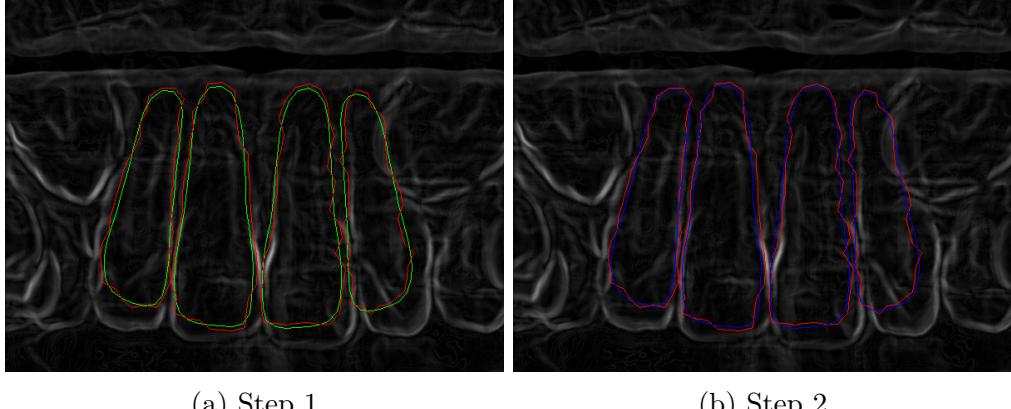


Figure 6: The process of getting the initial pose estimate. (a) shows the jaw separation lines for radiograph 11, where the red line shows the row of pixels with the lowest intensity and the green lines show the resulting upper and lower jaw separations. (b) and (c) show the hough lines for the upper and lower incisors respectively for radiograph 11. (d) shows the resulting initial estimate for radiograph 11, which is a really good fit, and (e) shows the initial estimate for radiograph 5. The estimate for radiograph 5 is a lot worse due to the off-centered position of the upper incisors.



(a) Step 1

(b) Step 2

Figure 7: The two steps of the ASM algorithm for the first iteration of radiograph 11. (a) shows the first step, where the green shape shows the initial estimate and the red shape shows the proposed new points. (b) shows the second step, where the model is fitted to the red points, resulting in the blue shape.

For the first step of the algorithm, we use a grey level model as proposed by Cootes et al [1]. This is achieved by sampling  $k$  pixels along the normal of a landmark on both sides, resulting in a slice with the size of  $2k + 1$ . This is repeated for every landmark point on every training image, after which it is combined into a statistical model for each landmark point. The result is a mean sample and a covariance matrix for each landmark. When looking for a new landmark point,  $m > k$  pixels are sampled along the normal of a landmark point of the current estimate. The Mahalanobis distance between the profile is then calculated to find the best nearby match. In our implementation, some of our covariance matrices of the grey level models happened to be singular, causing the Mahalanobis distance to fail. To alleviate this issue, we used the Moore–Penrose pseudoinverse instead of the normal inverse. For the covariance matrices that were not singular, using the inverse or the pseudoinverse resulted in the same best nearby match, thus suggesting that these can be interchanged for our application.

For the second step of the algorithm, we used Protocol 1 as it is described by Cootes et al [1].

These two steps are visualised in Figure 7.

**Multi-resolution framework** We implemented our ASM algorithm in a multi-resolution framework to improve the performance of our algorithm. We constructed a two level Gaussian image pyramid, since returns diminished

when going beyond two levels. This allows the algorithm to first find an estimate in a coarser low resolution image, after which it is able to further fine-tune the result in the higher resolution image.

**Stopping condition** To determine if the ASM algorithm has converged at the current resolution, we check whether  $> 90\%$  of the newly proposed landmark points fall in the central 50% of the  $2m+1$  slice among the previous landmark position. We only check this for the crowns of the teeth however, as it is a lot harder for the roots to converge due to the absence of clearly defined edges. We also allow for a maximum of 10 iterations per resolution level, as the algorithm might end up diverging when ran for a large amount of iterations. The results of this decision are elaborated upon in the evaluation.

## 5 Evaluation

We evaluate the performance of our algorithm using leave-one-out testing on all 14 training images. The results are marked **green** where the solution surface overlaps with the ground truth, **red** where the solution surface incorrectly marks an area and **yellow** where the ground truth hasn't been found.

To judge our performance, we use three measures: Euclidian distance, ground truth found rate, and percentage of solution marking the ground truth. The Euclidian distance shows trends in performance that are correlated with properties or parameters. The miss rate indicates the percentage of ground truth we have found (and missed) and shows us that certain types of radiographs are very hard to mark. Finally, the percentage of our solution that marks the ground truth teaches us that visibility of edges is crucial.

### 5.1 Euclidian distance

The Euclidian distance between a solution and its ground truth is the total absolute distance between all corresponding landmark points. When looking for trends in the results from our experiments with varying parameter settings, we notice two main properties. The first is that increasing the number of maximum iterations our fitting procedure can take has negative effects. The boxplots in Figure 8 (a) clearly show this correlation. One example is shown in (b), where the extra allowed iterations slightly improve the lower fit but send the upper fit flying off vertically.

Note that because of our stop condition a raised maximum number of iterations doesn't automatically mean more iterations done. Figure 8 (c)

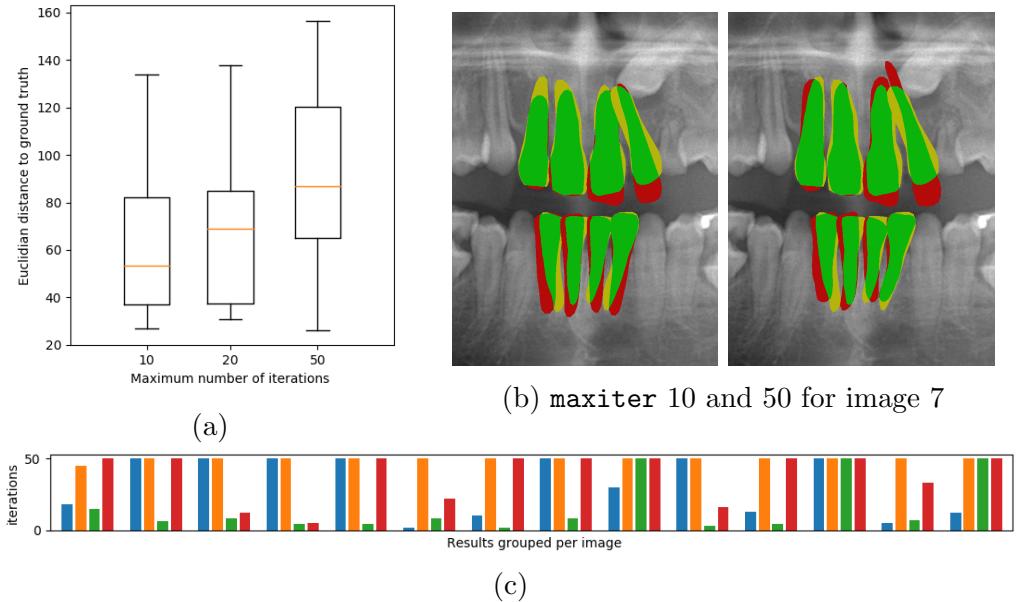


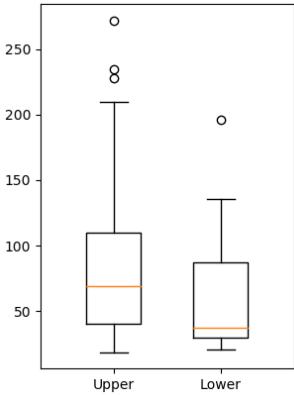
Figure 8: Euclidian distance to the ground truth per maximum number of iterations. Longer iteration has a negative effect on performance.

shows how many iterations were passed in each step. Blue is the subsampled upper fit, orange the full resolution upper fit, green and red analogous for the lower fit. The upper model clearly runs into `maxiter` more often, implying it is fitting poorly.

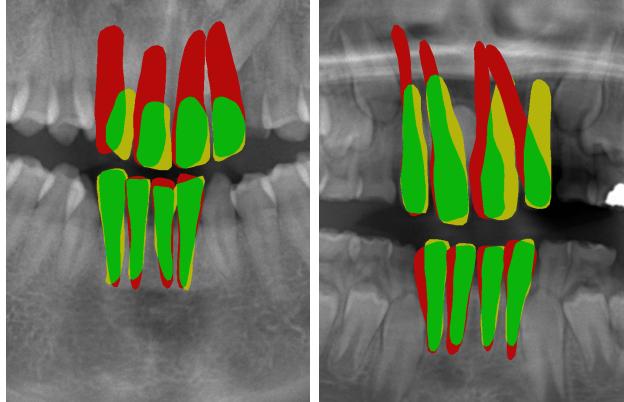
Inspecting the boxplots in Figure 9 (a) confirms that the lower teeth are fitted generally better. This is due to a more homogeneous appearance of bottom incisors. Two examples can be found in 9 (b). Although the upper incisors have very different shapes, the lower are very similar and are marked much better as a result.

## 5.2 Ground truth surface found

The bar chart in Figure 10 (a) shows the percentage of ground truth we have missed in our solutions. The blue bars are for the runs with 10 maximum iterations, orange for 20 iterations and green for 50. We have an average miss rate of about 22%. Some results stand out because of their high miss rate. There are two reasons for this. The first is the presence of highly visible edges due to orthodontic materials. This confuses the algorithm and pushes the iterated approximations away, as shown in subfigure (b). The second is that some radiographs are off-center by more than half a tooth, causing our



(a)



(b) Image 8 (left) and 4 (right)

Figure 9: Euclidian distance to the ground truth for upper and lower models. Lower incisors appear to be more homogeneous and easier to mark.

initialisation to start a whole tooth to the left or right. This is shown in subfigure (c).

### 5.3 Correctness of solution surface

In Figure 11 (a) we show the percentage of the solution that overlaps with the ground truth. The effects of Figure 10 (b) and (c) also influence this measure. However, new peaks in error rate emerge in images 8 and 12. These come from very noisy radiographs that appear almost blurred. Detecting the roots is nearly impossible. We see the result of edge detection on the enhanced images in subfigure (c) and even the manually labelled ground truth landmark indicates only the crowns in image 8. These effects accumulate to an overall error rate of about 35%.

## References

- [1] Tim Cootes, ER Baldock, and J Graham. An introduction to active shape models. *Image processing and analysis*, pages 223–248, 2000.
- [2] Ren-You Huang, Lan-Rong Dung, Cheng-Fa Chu, and Yin-Yi Wu. Noise removal and contrast enhancement for x-ray images. *Journal of Biomedical Engineering and Medical Imaging*, 3(1):56, 2016.

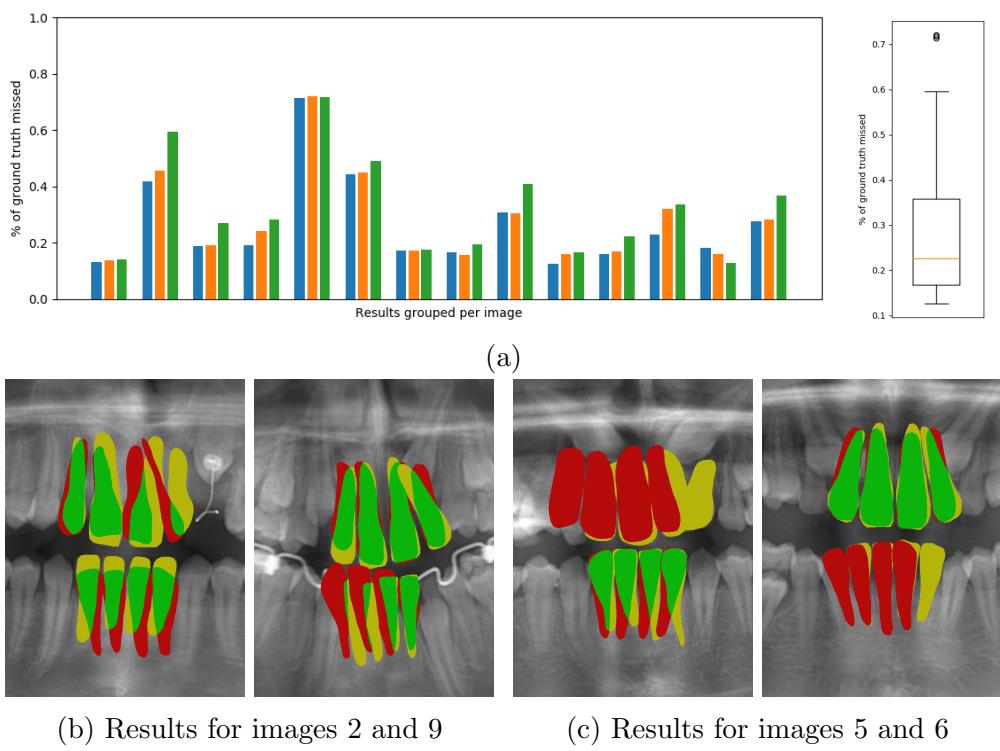


Figure 10: Ground truth surface missed per image. Orthodontic materials and off-center initialisation causes great errors.

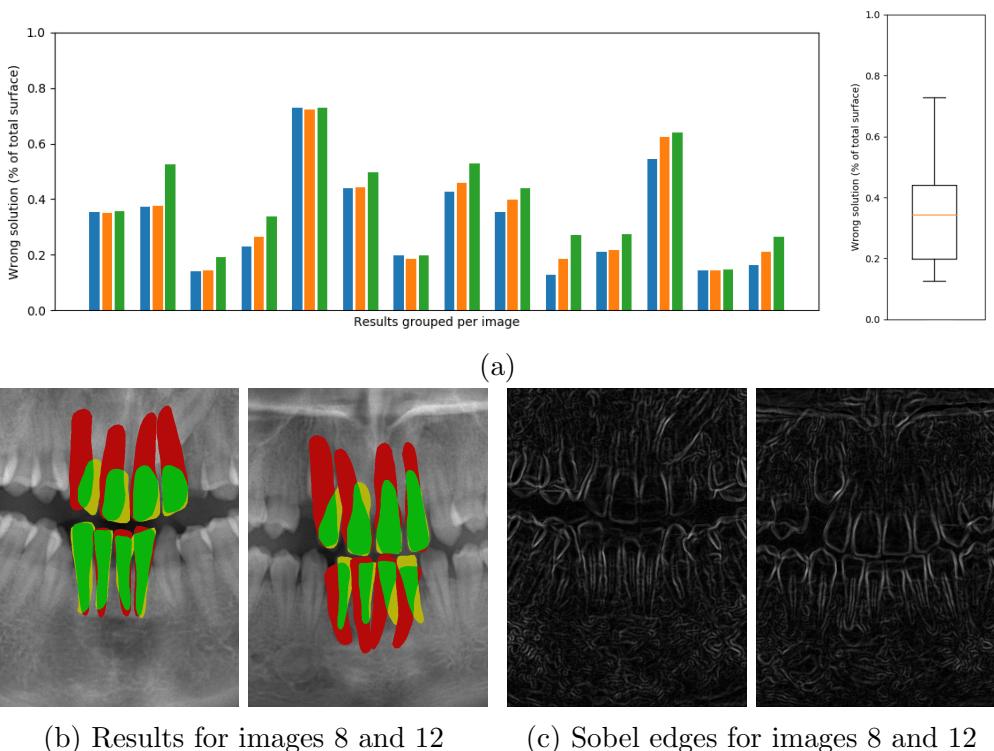


Figure 11: Erroneous solution surface per image. Poorly visible incisor roots cause the approximation to fly off.