

Analysis of Dental Images using Artificial Immune Systems

Zhou Ji, Dipankar Dasgupta, Zhiling Yang and Hongmei Teng

Abstract—This paper introduces a preliminary effort to develop an automatic image analysis method using Artificial Immune Systems for clinical dental diagnosis. To diagnose dental deformity, especially malocclusion, manual measurement of certain geometry on the X-ray images is traditionally used, which relies on subjective judgment to determine the reference points. This paper proposes a feature extraction method that is based on the brightness distribution of the image instead of the anatomical parts. A negative selection algorithm is then applied to the data represented as real-valued vectors to detect the cases of severe malocclusion. Using the same data representation, one-class SVM was also tried to compare the detection capability with the negative selection algorithm. The results show that the negative selection algorithm appears more suitable for this problem.

I. INTRODUCTION

Dental anatomy is the study of the morphology of teeth, their location, position, and relationships. Occlusion is an important topic in dental anatomy. It refers to the alignment of teeth and the way the upper and lower teeth fit together (bite). Malocclusion is the case when the teeth are not aligned properly. It is the most common reason for referral to an orthodontist. Most cases of malocclusion are mild and do not need treatment. If a misalignment is severe, especially when considered functional instead of aesthetic, orthodontics and/or oral surgery may be required. There are mainly three types of malocclusion: type I is when the bite is normal but teeth are mispositioned; type II, or retrognathism (overbite), is when the upper jaw and teeth overlap the bottom jaw and teeth; type III, or prognathism (underbite), is when the lower jaw and teeth extend forward too much with respect to the upper teeth [1]. Although some cases of severe malocclusion, e.g. prognathic lower jaw, is visible, many cases cannot be diagnosed only by practitioners' visual inspection.

To diagnose the cases of malocclusion, lateral view X-ray of the skull is commonly used [11]. Fig. 1 shows an example of such X-ray. The skull (cranium) has 23 bones. They are fit together closely except the mandible and the hyoid. The relative location of these bones may reveal malocclusion [8]. Fig. 2 shows a case of severe malocclusion. There are several existing techniques of cephalometry to diagnose malocclusion with such skull X-rays. They all use similar strategies: manually deciding the location of certain anatomical parts as reference points and measuring some distances and angles

between them, then determining the classification based on simple criterion on such measurements.

The popular Angle's classification is based on three points on the skull: N (nasion, the point where the nasal and frontal bones unite), A (subspinale, the most posterior midline point on the premaxilla between the anterior nasal spine and prosthion), and B (supramentale, the most posterior midline point, above the chin and on the mandible between the infradentale and the pogonion) [6]. The angle ANB is used as a criterion to decide severe malocclusion, or functional deformity. If angle ANB is between 0° and 5° , it is considered normal; if angle ANB is larger than 5° or less than 0° , it is considered abnormal. The difficulty exists in that the reference points are not easy to be decided on the lateral skull X-ray, and thus the angle ANB measured is not totally reliable.

Accuracy of the calculation described above is largely affected by the subjective judgment of the reference points in practice. To decide those points from the image automatically would be even harder, if not impossible. This paper proposes a totally different way to deal with the problem. Instead of identifying those anatomical parts, we try to classify the images by some real-valued features describing the brightness distribution. In other words, because the images of malocclusion profile will have different geometric characteristics from those of normal profile, we expect such difference will show up in brightness distribution as well.

In the X-rays we consider, high brightness show tissue of high density, or more exactly low X-ray penetration; low brightness shows lower density tissue.

Generally speaking, most people are considered having normal occlusion, which could in principle provide a very large training data set of normal class. That fits the assumption of one-class classification [10]. The paradigm of negative selection algorithms is one of the models of Artificial Immune Systems (AIS) [4]. It has been used in various applications, especially anomaly detection, and became a relatively new strategy to process one-class classification.

V-detector is one of the latest development of negative selection algorithms, which is efficient and more reliable in its coverage of abnormal space [7]. As any other negative selection algorithms, it involves two phases when used: (1) generate a set of so-called detectors in the complementary space; and (2) use the detector set to detect abnormal items in the incoming new data. *V-detector*'s algorithm to generate a detector set includes the following major steps:

- 1) randomly choose a point within the search space
- 2) check whether it matches any self (normal) samples available. If a match happens, it is discarded and we go back to step 1 to try another point and repeat until

Zhou Ji is with The University of Memphis, Memphis, TN 38152, and St. Jude Children's Research Hospital, Memphis, TN 38105 (email: zhou.ji@stjude.org)

Dipankar Dasgupta is with The University of Memphis, Memphis, TN 38152 (email: dasgupta@memphis.edu)

Zhiling Yang is with Yinchuan Stomatological Hospital, Yinchuan, Ningxia Hui Autonomous Region, China (email: yzl641125@163.com)

Hongmei Teng is with The University of Memphis, Memphis, TN 38152 (email: hongmei.teng@gmail.com)

we found a nonself (abnormal) point, which we call a “candidate point”.

- 3) check whether the candidate point is already covered (detectable) by the detectors that are already generated so far. Repeat from step 1 until we find a uncovered candidate point. A new detector, which centers at the candidate point and has the maximum radius that does not conflict any self samples, is then added to the detector set.
- 4) the number of consecutive candidate points that turn out to be already covered is used as a statistical basis to decide when the confidence in the coverage is enough and no more detector needs to be added.

This paper reports a series of experiments to evaluate the performance of negative selection algorithms, or more specifically *V-detector*, in such a high-dimensional one-class classification problem.

To compare, we also tried a more popular statistical learning model, SVM (Support Vector Machine), which can also deal with one-class classification [2].



Fig. 1. A typical lateral view skull X-ray for dental diagnosis

II. IMAGE DATA AND FEATURE EXTRACTION

A. Data collection

The lateral view X-rays used in this paper’s experiments were collected at Yinchuan Stomatological Hospital, China. They were gathered as part of the clinical malocclusion diagnosis process. Traditional calculations have been done



Fig. 2. An case of malocclusion

and proper diagnosis were made accordingly. The diagnoses were used to decide consequential treatment or otherwise medical decision. Classification between normal (light or none malocclusion) and abnormal (severe malocclusion that must be treated with orthodontics or surgery) is part of the diagnosis and can be used to compare with detection of negative selection algorithms or other machine learning methods.

The ultimate goal of such study is to develop an automatic tool either to replace or to expedite the process to diagnose malocclusion.

B. Data representation

To process the image data, the first step is to extract from each original image interesting geometric information as a real-valued vector. The features we try to extract is based on the distribution of the pixels’ brightness. In the RGB color space, brightness can be thought of as the arithmetic mean μ of the Red, Green, and Blue color coordinates. To simplify implementation, we will use the sum instead of the mean to describe quantitatively the brightness of each pixel as value V

$$V = R + G + B,$$

where $R, G, B \in [0, 255]$.

Although the X-Rays are gray-scale images, the above more general formula still can be used without much sacrifice

in computing performance. In this case, the only peculiarity is that R , G , and B will always be the same value.

The steps to extract real-valued features from the images are described in more detail as following:

- 1) Decide the value range of the image. The maximum value, V_{\max} , is the number describing the brightest part of the image, which shows the bony tissue with high density. The minimum value, V_{\min} , is the number describing the darkest part, which is likely to be the empty space or tissue with very low density. The surrounding area close to the border of the image has artificial highlight from the exposure process, so it is excluded from the range and considered the same as black region in the following processing. So is the upper right corner, where the image of the external artifact is sitting. The region that should be neglected is shown as black in the diagram in Fig. 3.

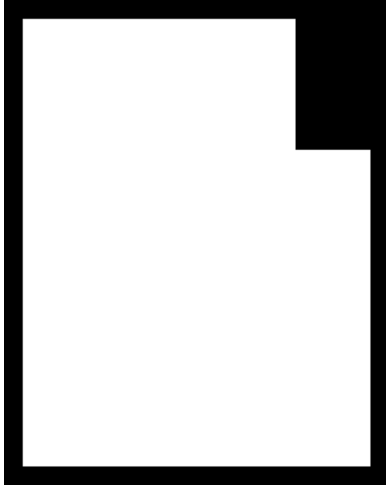


Fig. 3. The mask (region shown as black) shows what is excluded from analysis of the images.

Then, a certain number, n_T , of thresholds are chosen based on the value range. For example, we can choose evenly separated thresholds starting with the highest brightness:

$$\begin{aligned} T_0 &= V_{\max}, \\ T_1 &= V_{\max} - (V_{\max} - V_{\min})/n_T, \\ &\dots, \\ T_{n_T-1} &= V_{\max} - (n_T - 1)(V_{\max} - V_{\min})/n_T, \end{aligned}$$

The experiments described in section IV are based on this setting of thresholds.

- 2) Binarization.

The original image is converted to binary images at each threshold. For example, at threshold T_0 , the binarized image shows the brightest area of the original image as white and rest of the image as black. The image binarized at the second threshold T_1 shows the area that is at least as bright as T_1 , which contains the brightest area as part of it, and so on.

Figs. 4, 5, and 6 show the first three levels of binarization of the image shown in Fig. 1, and their corresponding threshold of brightness value. The white area in fig. 4 is the brightest area (value 765) in the original image. The centroid of that white area is used as the reference point to calculate the three geometric features, Δx , Δy , and Δr , at each other level. The fourth feature, area mass A , is independent of the reference point. Fig. 5 shows the area that is brighter than the next level, 670, in the original image. It reveals more information about the configuration of the bony structure. Fig. 6's white area includes more portion of the original image. Similarly binary images can be obtained at each threshold. In our experiments, there are totally eight for every original image. However, when we actually calculate the data representation of the real-valued vectors for a group of images, it is not necessary to generate and save these binarized images as we demonstrated here.



Fig. 4. Binarization at Level 0, threshold 765



Fig. 5. Binarization at Level 1, threshold 670



Fig. 6. Binarization at Level 2, threshold 575

- 3) Calculate four real-valued features for each binary image. To describe the location and shape of the shown area (the white area) for a given threshold, we use four values with respect to the reference point, which is chosen as the centroid of the brightest area (the white area in the first level binarized image).

- a) Horizontal displacement

$$\Delta x = \overline{x_{white}} - x_0,$$

where $\overline{x_{white}}$ is the mean of the x coordinate of all the white pixels in the binarized image; x_0 is the x coordinate of the reference point.

- b) Vertical displacement

$$\Delta y = \overline{y_{white}} - y_0,$$

where $\overline{y_{white}}$ is the mean of the y coordinate of all the white pixels in the binarized image; y_0 is the y coordinate of the reference point.

- c) Displacement distance

$$\Delta r = \sqrt{(x_{white} - x_0)^2 + (y_{white} - y_0)^2}.$$

We need to note that this information cannot be derived from the first two values. For example, if the area is a ring centered at the reference point, both horizontal displacement and the vertical displacement would be zero, but the displacement distance depends on the size of the ring.

- d) Area mass

$$A = \frac{\text{total number of white pixels}}{\text{width} \cdot \text{height}}.$$

This value describes the proportion of the area whose brightness is above the threshold.

Eventually, each image is represented as a $4n_T$ -dimensional real-valued vector

$$(\Delta x_0, \Delta y_0, \Delta r_0, A_0, \Delta x_1, \Delta y_1, \Delta r_0, A_1, \dots, \Delta x_{n_T}, \Delta y_{n_T}, \Delta r_{n_T}, A_{n_T}),$$

where n_T is the number of thresholds we use.

As the last step of preprocessing, normalization is carried out before the negative selection algorithm is used. For each of the $4n_T$ elements, the value range $[\min_i, \max_i]$, $i = 1, 2, \dots, 4n_T$, is determined by all the training data

$$\begin{aligned} \{a|a \equiv (a_1, a_2, \dots, a_{4n_T}) \\ \equiv \Delta x_0, \Delta y_0, \Delta r_0, A_0, \Delta x_1, \Delta y_1, \Delta r_0, A_1, \\ \dots, \Delta x_{n_T}, \Delta y_{n_T}, \Delta r_{n_T}, A_{n_T}) \\ \text{is a normal sample.}\} \end{aligned}$$

These ranges are then used to normalize all the training data

$$\tilde{a}_i = (a_i - \min_i) / (\max_i - \min_i).$$

\tilde{a}_i will be in the range of $[0, 1]$. The test data are scaled by the same factors

$$\tilde{x}_i = (x_i - \min_i) / (\max_i - \min_i),$$

which means that \tilde{x}_i could be out of the range of $[0, 1]$. In the case that $\tilde{x}_i > 1$ or $\tilde{x}_i < 0$, we usually identify it as an abnormal item.

The resultant real vector has much lower dimensionality than the original image. Because the difference between normal occlusion and malocclusion is basically the difference in relative positions of some key anatomical parts, the four parameters described above could be adequate to catch such differences. The vector does not contain the same amount of information as the raw image by design. This strategy is only a preliminary effort to explore the idea.

III. ALGORITHMS

A. *V-detector*

V-detector is chosen because of its flexibility to process one-class classification without being limited by specific knowledge of the data to work on. In this application with the real-valued data representation described above, we don't know the distribution or other information about the data points over the high-dimensional search space. By using maximized detectors and statistical estimate of the nonself space coverage, *V-detector* could conveniently deal with high-dimensionality and utilize different distance measures, though the actual performance of *V-detector* under various scenarios still needs more experiments and analysis to verify.

V-detector is to some extent a framework that accommodates various components. In the implementation used for the experiments in this paper, a couple of options are available to choose from. First, it has a boundary-aware version of the algorithm and a point-wise algorithm. The former uses the collection of training samples as a whole, and thus could detect anomaly close to the boundary area of self region more aggressively; the latter, on the other hand, treats each self sample as an independent point inside the self region, so it is more conservative to claim anomaly whenever it is close to a self sample. Second, *V-detector* has two versions of coverage estimate mechanism. One is the intuitive method, referred to as "naive estimate"; the other is based on hypothesis testing, which provided statistically more reliable results [7].

Fig. 7 shows the version of boundary-ware algorithm with hypothesis testing estimate, where z_α is the z score of normal distribution at the confidence level of $1 - \alpha$. It appears complicated at first look, but is in fact rather straightforward considering the fact that most of the details are already mentioned.

B. SVM

One-class classification is a learning problem with its unique issues [10]. SVM (Support Vector Machine) is probably the most used method in that area. Therefore, we choose it to compare with *V-detector*.

Beside one-class classification, SVM (Support vector Machine) is more frequently used in general classification as well as regression, ranking and other learning problems. It has the advantage of clear connection to the underlying statistical learning theory, but is relatively difficult to get reasonable results without experience to choose proper control parameters. The biggest limitation is in the choice of kernel function [2], [5].

IV. EXPERIMENTS

While more images are being collected, some preliminary experiments were carried out to evaluate the method proposed in the previous section. Considering our assumption that normal data are available in large amount, the data used in the current experiments, especially the normal samples, are not ideally adequate yet. Nevertheless, the results show encouraging potential of AIS such as negative selection algorithms in the area of clinical diagnosis.

A. Negative selection algorithm's detection ability

V-detector [7] was used to generate anomaly detector set with all the available normal images as training data, which currently include 69 images. Then all the available 132 images were tested with the detector set to be normal or abnormal. Table I shows the results compared with the classification using conventional measurement. Self threshold r_s is a key control parameter to balance between sensitivity and generalization. Experiment were done with the following values of self threshold r_s : 0.02, 0.04, 0.06, 0.08, 0.1, 0.12, 0.14, 0.16, 0.18, and 0.2. Three typical values of self threshold are shown in the table. In this case, the results were in fact not significantly affected. There may be two reasons for self threshold's insignificance here. One, for such high dimensionality, $r_s = 0.2$ may be still too small to show the difference, considering the diagonal distance of search space $[0, 1]^m$ is \sqrt{m} . Two, the number of images is so low that all the images, normal or not, distribute too far away from each other in the feature space. Another control parameter "target coverage" is set to be 99%. Although *V-detector* could use different distance measures, only Euclidean distance is used in the results reported here. The statistics are based on 100 repeated runs with the same control parameters. The performance of this method is encouraging.

Require: {Generate detector set of *V-detector*}

S : set of self samples
 r_s : self radius
 p : target coverage
 α : significant level for hypothesis testing
 T_{\max} : maximum number of detectors

- 1: $n \leftarrow \max\{5/p, 5/(1-p)\}$ {sample size required for hypothesis testing}
- 2: $D \leftarrow \emptyset$
- 3: **repeat**
- 4: $t \leftarrow 0$ {counter of "already covered" candidates}
- 5: $N \leftarrow 0$ {counter of valid candidates}
- 6: $C \leftarrow \emptyset$ {the collection of valid candidates}
- 7: $x \leftarrow$ random sample from $[0, 1]^m$
- 8: $r \leftarrow \infty$
- 9: **for all** s_i in S **do** {censored by self samples}
- 10: $d \leftarrow$ Euclidean distance between s_i and x
- 11: **if** $d < r_s$ **then**
- 12: go to 7
- 13: **else**
- 14: $r \leftarrow \min\{r, d\}$
- 15: **end if**
- 16: **end for**
- 17: $N \leftarrow N + 1$
- 18: **for all** d_i in $D = \{d_i, i = 1, 2, \dots\}$ **do** {censored by existing detectors}
- 19: $d_d \leftarrow$ Euclidean distance between d_i and x
- 20: **if** $d_d < r(d_i)$ **then** { $r(d_i)$ is the radius of detector d_i }
- 21: $t \leftarrow t + 1$
- 22: $z \leftarrow \frac{t}{\sqrt{np(1-p)}} - \sqrt{\frac{np}{1-p}}$
- 23: **if** $z > z_\alpha$ **then**
- 24: **return** D
- 25: **else**
- 26: go to 31
- 27: **end if**
- 28: **end if**
- 29: **end for**
- 30: $C \leftarrow C \cup \langle x, r \rangle$ {save a new candidate}
- 31: **if** $N = n$ **then**
- 32: $D \leftarrow D \cup C$
- 33: go to 4
- 34: **else**
- 35: go to 7
- 36: **end if**
- 37: **until** $|D| = T_{\max}$ {Exception case: too many detectors to handle}

Fig. 7. *V-detector*'s algorithm to generate detector set

TABLE I
DETECTION RATE - USING 8 LEVEL BINARIZATION

| Self threshold | Maximum | Minimum | Mean | Standard Deviation |
|----------------|---------|---------|-------|--------------------|
| 0.02 | 69.84 | 52.38 | 61.4 | 4.08 |
| 0.1 | 71.43 | 47.62 | 61.56 | 5.33 |
| 0.2 | 77.78 | 47.62 | 61.75 | 5.32 |

TABLE II
DETECTION RATE - USING 4 LEVEL BINARIZATION

| Self threshold | Maximum | Minimum | Mean | Standard Deviation |
|----------------|---------|---------|-------|--------------------|
| 0.02 | 76.19 | 55.56 | 67.6 | 4.34 |
| 0.1 | 76.19 | 55.56 | 66.98 | 4.49 |
| 0.2 | 80.95 | 58.73 | 67.86 | 4.55 |

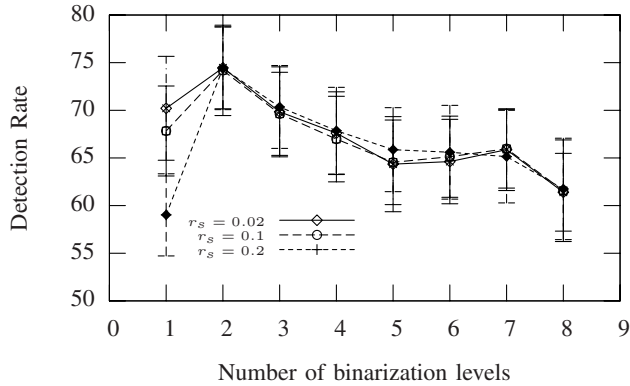


Fig. 8. Influence of number of binarization levels used in data representation

B. Influence of the number of binarization levels

It is an arbitrary design to choose 8 levels of brightness, from the brightest to the darkest, to binarize the images to extract geometric properties. Assuming we keep the same strategy, the other possible techniques are: (1) using different number of levels but still distributing the levels from the brightest to the darkest; or (2) choosing only some of levels to extract geometric features. In this paper, we tried the second alternative. When the level chosen is low and close to the darkest, most of the binarized images will be very similar to each other with large area of white and the information that can be obtained will be much less than the images binarized with higher threshold level. According to that observation, only some of the higher thresholds from the eight described above were used in the experiments. Table II shows the results of $n_T = 4$, where the four levels are from the brighter half of the value range instead of the entire value range as described in section II-B. In fact, the four binarized images are the same as the first four of the eight binarized images used earlier. The dimensionality is reduced from 32 to 16, so the algorithm runs much faster. The results didn't deteriorate at all. Instead, there is even slight improvement. That may be attributed to less noisy information introduced by including more levels of binarized images.

To further explore the effect of different data representation, experiments were done with different numbers of binarization levels chosen from the original eight. Fig. 8 shows the influence of the number, from 1 to 8, on the detection rate. As expected, two binary images gave obviously better results than one level only. However, when more levels are included, the noise introduced seemed overtaking the extra information provided.

TABLE III
PERFORMANCE OF SVM - USING 8 LEVEL BINARIZATION

| ν | Detection Rate | False Alarm Rate |
|-------|----------------|------------------|
| 0.01 | 14.28 | 2.90 |
| 0.05 | 14.28 | 4.35 |
| 0.1 | 19.05 | 10.14 |

TABLE IV
PERFORMANCE OF SVM - USING 4 LEVEL BINARIZATION

| ν | Detection Rate | False Alarm Rate |
|-------|----------------|------------------|
| 0.01 | 11.11 | 2.90 |
| 0.05 | 11.11 | 5.80 |
| 0.1 | 19.05 | 11.59 |

C. Comparison with SVM

SVM (Support Vector Machine) is a popular statistical learning algorithm. For one-class classification problem, it may be the most tried method. It was used here on the same data representation of the images as that tested with *V-detector*. A well-known implementation of SVM, LIB-SVM [3], was used in the experiments. The default control parameters were used except that the outlier fraction ν were tested with three different values, 0.01, 0.05, and 0.1. Tables III and IV show the results, using 8 level and 4 level binarization, respectively. All the number tabulated are percentages. Compared with *V-detector*, it doesn't work as well. Fig. 9 and 10 show the results for the number of binarization levels n_T from 1 to 8. Although it was not as successful as *V-detector*, the influence of n_T is similar. Two level binarization seems the best choice. *V-detector* turned out to be more capable to solve this problem. SVM at least needs further study on the choice of kernel function and control parameters to be used in this application.

Please note that although the control parameter "self threshold" in *V-detector* has similar implication to ν in SVM, they are not comparable in terms of actual values.

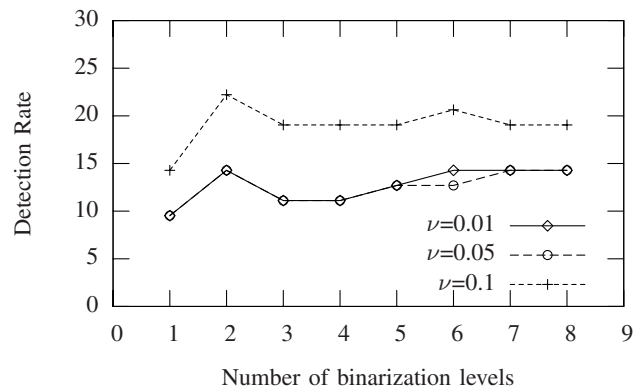


Fig. 9. Influence of number of binarization levels on SVM's detection rate

TABLE V

PERFORMANCE WHEN ONLY HALF NUMBER OF NORMAL IMAGES WERE USED AS TRAINING DATA - 8 LEVEL BINARIZATION

| Self threshold | DR_{max} | DR_{min} | DR_{mean} | DR_{SD} | FA_{max} | FA_{min} | FA_{mean} | FA_{SD} |
|----------------|------------|------------|-------------|-----------|------------|------------|-------------|-----------|
| 0.02 | 85.71 | 63.49 | 76.86 | 3.74 | 44.93 | 28.99 | 38.52 | 2.91 |
| 0.1 | 85.71 | 69.84 | 77.14 | 3.78 | 46.38 | 30.43 | 38.22 | 3.07 |
| 0.2 | 87.3 | 66.67 | 76.94 | 3.83 | 46.38 | 31.88 | 38.64 | 2.74 |

TABLE VI

PERFORMANCE WHEN ONLY HALF NUMBER OF NORMAL IMAGES WERE USED AS TRAINING DATA - 4 LEVEL BINARIZATION

| Self threshold | DR_{max} | DR_{min} | DR_{mean} | DR_{SD} | FA_{max} | FA_{min} | FA_{mean} | FA_{SD} |
|----------------|------------|------------|-------------|-----------|------------|------------|-------------|-----------|
| 0.02 | 90.48 | 73.02 | 81.08 | 3.43 | 46.38 | 36.23 | 41.83 | 2.11 |
| 0.1 | 88.89 | 71.43 | 81.7 | 3.32 | 46.38 | 34.78 | 41.94 | 2.02 |
| 0.2 | 88.89 | 73.02 | 80.95 | 3.3 | 46.38 | 36.23 | 41.94 | 2.13 |

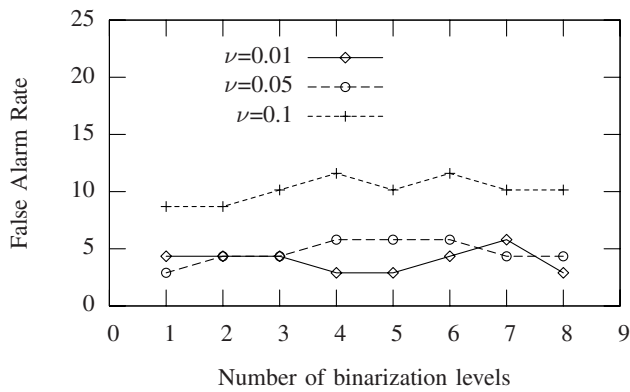


Fig. 10. Influence of number of binarization levels on SVM's false alarm

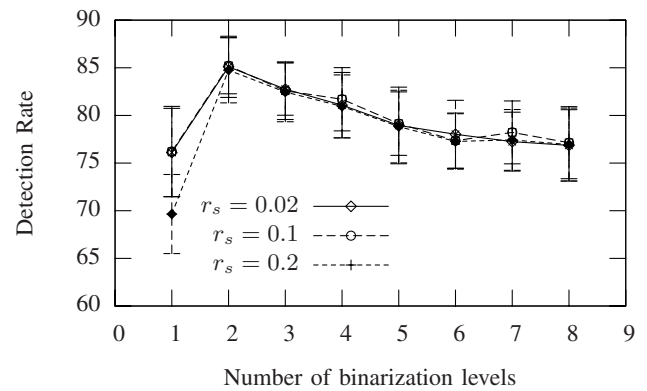


Fig. 11. Detection rate when half number of normal samples are used as training data

D. Influence of unseen normal data

Because of the current lacking in large number of normal images, we used all the available normal images (69) to train *V-detector* and SVM. The false alarm rate could be deceptive because the resulted model may overfit the training data so the false alarm does not show up. To illustrate the issue, we used only half of the available normal images (35) to train the systems and tested again. Although this smaller number will further compromise any methods' performance, we need such experiments to demonstrate the zero false alarm in the previous experiments are mainly because we don't have unseen normal data. Table V, in which DR denotes Detection Rate and FA denotes False Alarm Rate, shows very high false alarm. Basically unseen normal images were mostly decided as abnormal. This problem could be lessened when we have larger number of normal images, but further experiments are needed to verify the expected behavior.

Table VI shows similar results when 4 levels of binarization is used. The detection rate is slightly better, but the false alarm is higher too.

Figs. 11 and 12 show the results for all different numbers of binarization levels. The high false alarm appears consistently.

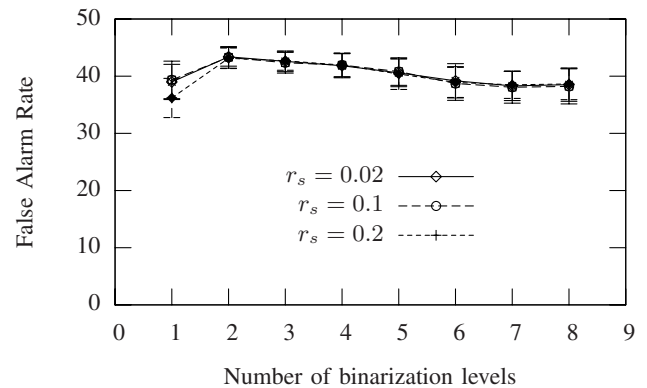


Fig. 12. False alarm rate when half number of normal samples are used as training data

TABLE VII

PERFORMANCE OF SVM WITH HALF NUMBER OF TRAINING IMAGES (8 LEVEL BINARIZATION)

| ν | Detection Rate | False Alarm Rate |
|-------|----------------|------------------|
| 0.01 | 25.40 | 20.29 |
| 0.05 | 23.81 | 20.29 |
| 0.1 | 25.40 | 21.74 |

TABLE VIII
PERFORMANCE OF SVM WITH HALF NUMBER OF TRAINING IMAGES (4
LEVEL BINARIZATION)

| ν | Detection Rate | False Alarm Rate |
|-------|----------------|------------------|
| 0.01 | 19.05 | 11.59 |
| 0.05 | 19.05 | 10.14 |
| 0.1 | 19.05 | 10.14 |

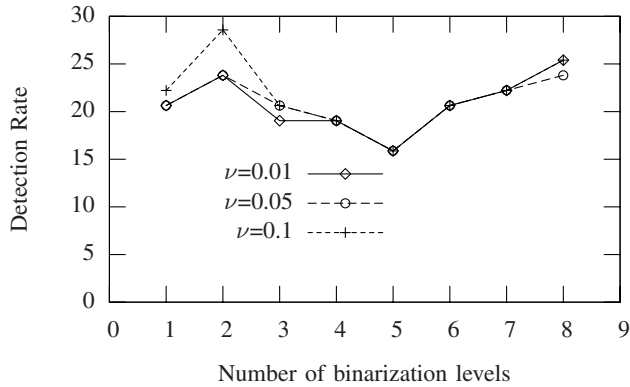


Fig. 13. SVM's detection rate when half number of normal samples are used as training data

The behavior of SVM was also tested under such condition. Table VII shows the results of 8 level binarization and table VIII shows the results of 4 level binarization. Figs. 13 and 14 show the results for more values of the number of binarization levels. The false alarm rate did not become too bad for SVM, especially when $n_T = 3$. That indicates that SVM doesn't have as strong tendency to outfit the training data as *V-detector*. However, considering its very low detection rate, such advantage is not very helpful.

V. CONCLUSIONS

The preliminary experiments reported in this paper show the potential usage of a negative selection algorithm as an efficient technique to diagnose malocclusion. It also eliminates

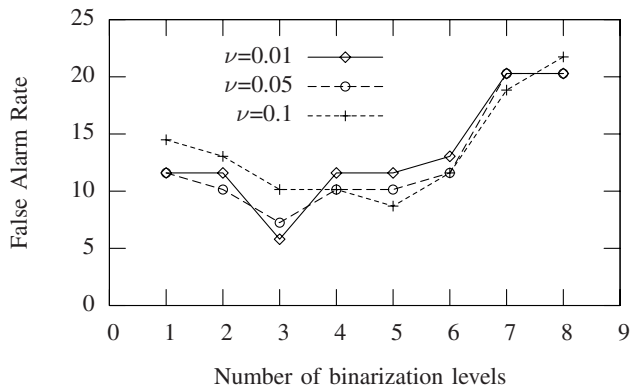


Fig. 14. SVM's false alarm rate when half number of normal samples are used as training data

unnecessary subjective influence from the diagnosis process, which could significantly affect the decision in this case.

The proposed design to represent each image with a real-valued vector describing the location and shape of areas of different brightness works well in these experiments. Some factors in such representation, for example, the number of levels to binarize, are still arbitrary. When fewer levels of binarization were used, the results did not deteriorate. Besides more efficient computation, the accuracy could in fact be a little better with fewer levels of binarization.

The data representation plays a key role to separate the two classes to be differentiated in the feature space. The current design is only a tentative effort and may be improved or replaced in many ways, but the learning methods, *V-detector* or otherwise, are still applicable.

A negative selection algorithm, *V-detector*, shows potential to detect images of malocclusion. It is generally more sensitive than SVM to detect abnormal cases in the experiments. SVM appears to have provided more generalization, but it may take specialized study, especially searching for suitable kernel functions, to make it usable in this application.

While the "curse of dimensionality" has effects [9] not only in negative selection algorithms but also in other machine learning algorithms, we observed that the *V-detector's* approach is more reliable because of the target coverage (instead of predefined detector number) that can be achieved by the algorithm.

From either the application's or the method's point of view, the current dataset is not adequate yet. A large number of normal images are important for the detector set based on them to yield more reliable results.

VI. ACKNOWLEDGMENTS

This work was supported in part by NIH Cancer Center Support Core Grant CA-21765 and the American Lebanese Syrian Associated Charities (ALSAC).

REFERENCES

- [1] Malocclusion of teeth. Web page, <http://health.allrefer.com/health/malocclusion-of-teeth-info.html>.
- [2] C. J. C. Burgers. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2:121–167, 1998.
- [3] C.-C. Chang and C.-J. Lin. LIBSVM: a library for support vector machines. World Wide Web, <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 2001.
- [4] S. Forrest, A. Perelson, L. Allen, R., and Cherkuri. Self-nonsel self discrimination in a computer. In *Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy*, Los Alamitos, CA, 1994. IEEE Computer Society Press.
- [5] C.-W. Hsu, C.-C. Chang, and C.-J. Lin. A practical guide to support vector classification. World Wide Web, <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 2001.
- [6] W. Hussels and R. S. Nanda. Analysis of factors affecting angle ANB. *American Journal of Orthodontics*, 85(5):411–23, May 1984.
- [7] Z. Ji and D. Dasgupta. Estimating the detector coverage in a negative selection algorithm. In H.-G. B. et al, editor, *GECCO 2005*, volume 1, pages 281–288, Washington DC, USA, 25–29 June 2005. ACM Press.
- [8] Z. Lu and et al. *Systemstic Anatomy*. Science Press (China), 2005.
- [9] T. Stibor, J. Timmis, and C. Eckert. A comparative study of real-valued negative selection to statistical anomaly detection techniques. In *ICARIS*, pages 262–275, 2005.
- [10] D. M. J. Tax. *One-class classification*. PhD thesis, Technische Universiteit Delft, 2001.
- [11] Z. Zou. *Dento-Maxillo-Facial Diagnosis with X-Ray*. People's Hygiene Press (China), 1993.