

# Teeth/Palate and Interdental Segmentation Using Artificial Neural Networks

Kelwin Fernandez and Carolina Chang

Grupo de Inteligencia Artificial  
Universidad Simón Bolívar  
Caracas, Venezuela  
kelwin@gia.usb, cchang@usb.ve

**Abstract.** We present a computational system that combines Artificial Neural Networks and other image processing techniques to achieve teeth/palate segmentation and interdental segmentation in palatal view photographs of the upper jaw. We segment the images into teeth and non-teeth regions. We find the palatal arch by adjusting a curve to the teeth region, and further segment teeth from each other. Best results to detect and segment teeth were obtained with Multilayer Perceptrons trained with the error backpropagation algorithm in comparison to Support Vector Machines. Neural Networks reached up to 87.52% accuracy at the palate segmentation task, and 88.82% at the interdental segmentation task. This is an important initial step towards low-cost, automatic identification of infecto-contagious oral diseases that are related to HIV and AIDS.

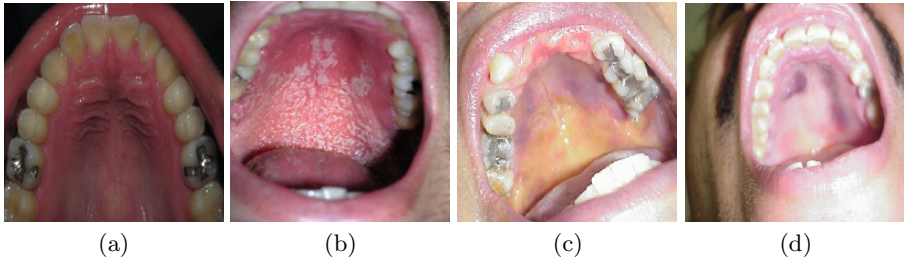
**Keywords:** teeth/palate segmentation, multilayer perceptron, support vector machines.

## 1 Introduction

Thirty-four million people were living with HIV by the end of 2010 according to the World Health Organization [1]. A 93.5% of these people resided in developing countries, and only 47% of eligible patients in this subgroup received the antiretroviral treatment they needed.

Evidence suggests that about 70% of people with HIV have oral diseases, including Leucoplasia Vellosa, Kaposi's Sarcoma and Candidiasis [2]. The diagnosis of these diseases is very important because some of them may indicate the evolution of HIV towards AIDS [3]. Sadly, the number of dental care centers that treat infecto-contagious diseases in developing countries is limited, and in many cases, insufficient. We believe that some oral diseases that indicate the presence of HIV/AIDS can be diagnosed automatically. A computational, low-cost tool for detecting oral infecto-contagious diseases can lead to health care improvement, especially in low- and middle-income countries.

We focus on the problem of teeth/palate segmentation in palatal view photographs of the upper jaw (maxilla) as an initial step towards diagnosing oral



**Fig. 1.** (a) Healthy Palate. Oral diseases: (b) Candidiasis, (c) Hepatitis, (d) Kaposi's Sarcoma.

diseases. The palate is an important and large area of the mouth where several diseases can be observed. For example, figure 1 shows cases of patients with oral diseases such as Candidiasis, Hepatitis and Kaposi's Sarcoma. In addition to segmenting the palatal region, we aim to segment each tooth of the upper jaw to give a better description of the teeth region. This information could help refine the palatal region, or detect teeth diseases at future versions of the system.

## 2 Related Work

Methods for interdental segmentation in dental radiographs for general post-mortem identification are [4] [5]. In [4] a neural network based method is proposed for a postmortem identification system by matching image features extracted from dental radiographs. This system tries to match post-mortem and ante-mortem radiographs of a person. It proposes the use of learnable inherent dental image features for tooth-to-tooth image comparisons. In [5] a dental classification and numbering system to segment, classify, and number teeth in dental bitewing radiographs is proposed. Radiographs are enhanced to isolate teeth to regions of interest using image filters. Once teeth are isolated, a support vector machine classifies each tooth to molar or premolar.

In [13] the problem of lip segmentation in color space is handled using color photographs. The proposed solution set a Gaussian model using the hue and saturation value of each pixel within the lip segment. Then, the memberships of lip and non-lip regions are calculated, and the desirable lip region is obtained based on the memberships.

Radiographs have the advantage of having been taken under common standards, with low-variability machines, but provide less information about diseases compared to palatal view photographs. On the other hand, photographs are highly sensitive to phenomena such as illumination, quality of the image, angle of view, and lack of standards, among other. We address these issues on a variety of palatal view photographs, such as those shown in figure 1.

### 3 System Overview

The input of the system are palatal view photographs of the upper jaw (maxilla) such as those of figure 1. The system is able to determine if an image holds this constraint. Three stages of the system are described: teeth regions detection, palate segmentation, and interdental segmentation.

The teeth regions detection stage classifies each pixel of the image into teeth/non-teeth regions based mainly on color features (see section 4). Some pre and post-processing are required to enhance the image quality and the output. We trained Support Vector Machines and Artificial Neural Networks. Experiments show that Neural Networks reached 85.05% accuracy.

At the palate segmentation stage a curve is adjusted to the palatal arch by means of an iterative heuristic process. Outsider teeth regions are discarded after the palatal arch is found. Not only the shape but also the width of the curve is adjusted to cover teeth and leave noise outside. The palate region is considered to be the area inside the closed curve. We achieved 87.52% accuracy at the palate classification stage (see section 5).

At the final stage, Support Vector Machines as well as Artificial Neural Networks were trained to detect the starting and ending points of each tooth. To do so, the curved teeth region obtained in the previous step is flattened, and then inspected through a sliding window. This process is explained in section 6. Results show that Artificial Neural Networks reached 88.82% accuracy at this task.

For the sake of clarity, experiments and results for each stage of the system are presented at the end of its corresponding section. All experiments were carried out on an Intel Core2Duo 2.4 GHz processor, with 4GB of RAM, running Ubuntu Linux 11.04.

### 4 Teeth Detection

Color is determinant in our teeth segmentation method. A color enhancement filter improves color differentiation and normalizes the color range of images. Frequently, color images are in RGB format. When operators such as histogram equalization and contrast adjustment are applied over each component separately (red, green and blue), new undesirable colors can show up [6]. Hence, we transform original RGB images to HSV format (hue, saturation, value).

Xiao&Ohya proposed a contrast enhancement filter for color images [6]. They handle the problem of invalid new colors in the resulting image by keeping the Hue component unchanged, where the color information lays. The contrast in the V component (luminance value) is enhanced and finally, the saturation component is improved by histogram equalization.

We propose a filter based on this method. Our variation has the same approach in the Hue and Value components. However, in our variation, the Saturation component distribution is shifted such that the mean saturation is the same in all images. This variation tries to keep the difference in the saturation of each pixel while brings a saturation value suitable for distinguishing colors.



**Fig. 2.** Image preprocessing. Top: Original images. Middle: Color Enhanced images with Xiao&Ohya's filter. Bottom: Color Enhanced images with the proposed filter.

Finally, in both filters, images are restored to RGB format. These filters are useful in opaque images, where colors are almost indistinguishable. Figure 2 shows the result of applying the filters over six images.

We trained Support Vector Machines and Multilayer perceptrons to classify pixels of the images into teeth/non-teeth regions. SVM kernels include linear, polynomial and RBF kernels. Each SVM was trained varying the parameters using a logarithmic traversing over the search space. The training algorithm of the Multilayer Perceptrons is the classical random sequential back-propagation with a standard sigmoid as the activation function [11]. Neural Networks topologies had one hidden layer, which had from 5 up to 100 hidden neurons.

Experimental results are shown in section 4.1.

A post processing filter is applied to erase small blobs of the output image [10]. A blob is defined as a contiguous set of pixels of similar color. Small blobs are treated as noise. Figure 3 shows some examples of the neural network output, and the postprocessing filter output.

#### 4.1 Teeth Detection Experimental Results

A set of 100 images was used. The training and cross-validation set consisted of 10 images, and the test set of 90. Each image had a resolution of 200x200pixels, which means that the training set consisted of 400,000 pixels, and the test set of 3,600,000 pixels.

We wanted to know which image format was best suited for our segmentation task. Therefore, we trained the SVMs and ANNs using each of those formats. RGB, HSV and GRAY refer respectively to the RGB, HSV and gray-scale formats. eRGB and eHSV are the analogous format but with the Xiao&Ohya's enhancement filter, while eRGB', and eHSV' refer to our proposed filter. eGRAY



**Fig. 3.** Teeth Detection. Top: Original Images. Middle: Neural Network output. Bottom: Small blobs filtering.

**Table 1.** Pixel classification with SVM and ANN

Arguments	SVM				ANN			
	Kernel	Train	Test	Time	Top.	Train	Test	Time
RGB	RBF	45.34	40.40	0.63	50	82.59	80.02	0.10
eRGB'	Linear	52.07	49.86	0.80	100	<b>86.44</b>	80.71	0.13
eRGB	Linear	56.59	49.69	0.83	5	83.07	78.47	0.08
HSV	Poly	76.55	72.00	0.63	20	82.78	78.06	0.08
eHSV'	Poly	73.00	55.81	0.27	30	85.85	80.67	0.10
eHSV	RBF	73.69	71.66	0.67	10	85.60	81.44	0.08
GRAY	Linear	63.02	72.32	0.57	50	80.88	<b>81.79</b>	0.09
eGRAY	Poly	<b>82.19</b>	<b>79.02</b>	0.90	5	82.51	80.76	0.06

is the gray-scale format with histogram equalization. Each machine receives as input the format of a pixel.

Table 1 shows that Neural Networks work better in this problem than Support Vector Machines. For each color input format, the best Neural Network found outperformed the best Support Vector Machine found. Moreover, the classification time is much smaller for Neural Networks. An interesting result is that the best SVM found receives as argument the enhanced gray-scale image. Unfortunately, this is the most time consuming SVM too.

At the ANN side, it was hard to tell which was the best network found. The best training classification was achieved using the eRGB' input format. However, this network had 100 hidden neurons. Although it was the most time consuming ANN, notice that it is about 7 times faster than the best SVM found. On the other hand, the GRAY format produced the best classification of test images, yet the worst classification over training.

**Table 2.** Pixel classification on combined Input Formats for SVM and ANN

Arguments	SVM				ANN			
	Kernel	Train	Test	Time	Top.	Train	Test	Time
eRGB'+eHSV'	Poly	80.89	77.53	0.23	20	<b>89.02</b>	81.33	0.10
eRGB + eHSV	Linear	71.02	72.34	0.74	5	88.28	82.59	0.09
eRGB'+eGRAY	Linear	61.57	68.87	0.60	5	86.92	79.44	0.09
eRGB + eGRAY	Poly	80.88	81.79	0.94	30	87.68	80.85	0.10
eHSV'+eGRAY	Poly	75.05	63.30	0.27	25	86.14	80.61	0.09
eHSV + eGRAY	Linear	71.92	71.11	0.69	50	85.41	81.46	0.10
eRGB'+eHSV'+eGRAY	RBF	65.96	44.93	0.83	25	87.84	84.81	0.11
eRGB+eHSV+eGRAY	Linear	80.88	<b>81.79</b>	0.77	40	87.55	<b>85.05</b>	0.11

**Table 3.** Teeth Classification Rate After Small Blobs Filtering

Stage	Rate (%)	False Positive (%)	False Negative (%)
eRGB'+eHSV' ANN	81.3300	9.9974	8.6726
Blob Filter	86.1185	5.4553	8.4263

To explore further the impact of the image formats, we combined them, as shown in table 2. In general, significant improvements were achieved for SVMs both in classification rate and time as compared with those of table 1. However, none of these SVMs surpassed the 82.19% of classification rate over training achieved by using the eGRAY format only.

The combination of input image formats improved the classification rates of the Neural Networks as well. The best classification rate obtained over training was 89.02%. This result was achieved by a network of 20 hidden neurons, that combined the RGB and HSV images enhanced with our proposed filter. On the other hand, the best classification rate over testing was achieved by a network that combined the RGB, HSV and GRAY images enhanced by the Xiao&Ohya's filter.

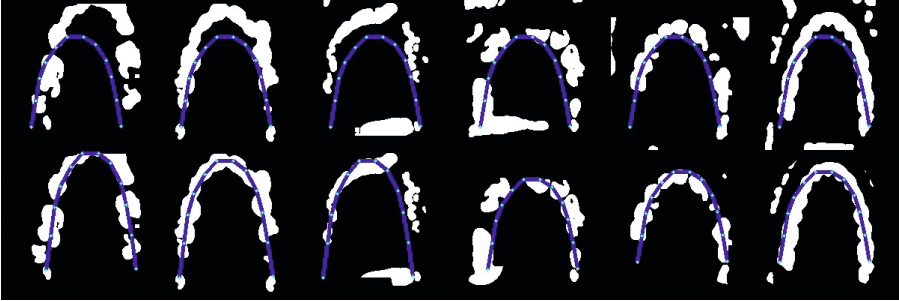
From the results we cannot conclude which is the best Neural Network configuration. However, ANNs outperformed SVMs both in classification rate and time.

Table 3 shows the effect of filtering small blobs from the eRGB'+eHSV' Neural Network output. Blob filter leads to an increment of over 4%, decreasing considerably the false positive rate.

## 5 Palate Segmentation

Once the teeth regions are found it is easier to determine the palate location. We adjust a curve to the teeth regions by means of optimization algorithms [8]. These algorithms perform rotations, translations, scales and deformations of an original curve learned from palate curves.

When skin pixels are classified, some of them have information indistinguishable from teeth, therefore the neural network classifies them as positive examples. In these cases, the curve adjusting algorithm tries to reach a stable point between skin pixels classified as positive and teeth pixels. Cleaning iteratively the noise located outside the curve, corresponding to skin pixels, leads to better results in this stage. This process ends when fixed point is reached by the skin cleaner method.



**Fig. 4.** Curve adjusted and pixel segmentation after filtering pixels out of the curve. Top: Initial curve. Middle: Curve adjusted to the palatal arch. Bottom: Noise filtered.

As can be seen in figure 4, curves are adjusted within teeth pixels. Picture scale, differences between each individual and noise in the classified image may result in different sizes of teeth. It is needed to detect automatically the bandwidth where teeth are included.

Bandwidth detection should decide for each image the right limits where teeth are covered and noise is outside. An unsupervised and supervised combined method is proposed to solve this problem based in the algorithm exposed in [7].

At the unsupervised step, each boundary pixel classified as positive is assigned to a cluster. Assignment of the point  $p$  to a cluster  $c$  success if the distance between  $p$  and  $c$ ,  $d(p, c)$  is less than certain threshold  $W$  and  $d(p, c) \leq d(p, c')$  for all  $c'$  in the set of clusters, where  $d$  is an arbitrary function. In our case  $d$  is the squared distance between  $p$  and the mass center of  $c$ . Otherwise, a new cluster is created with the point  $p$ .

When every point is assigned, the algorithm classifies each cluster as normal or anomalous data based on the cluster size. Small clusters contain anomalous data, big clusters contain normal data. The bandwidth selected is the biggest mass center of the normal clusters.

The method described is essentially unsupervised, but there are two variables that should be set,  $W$  and normal/anomalous cluster threshold. A supervised method that tries to minimize the distance from actual bandwidth to the selected by the unsupervised algorithm is developed as an initial training method.

When the bandwidth is set, points classified positively between the curve and the allowed bandwidth are denoted as teeth. Points inside the polygon generated by the closed curve are denoted as part of palate, as shown in figure 5.

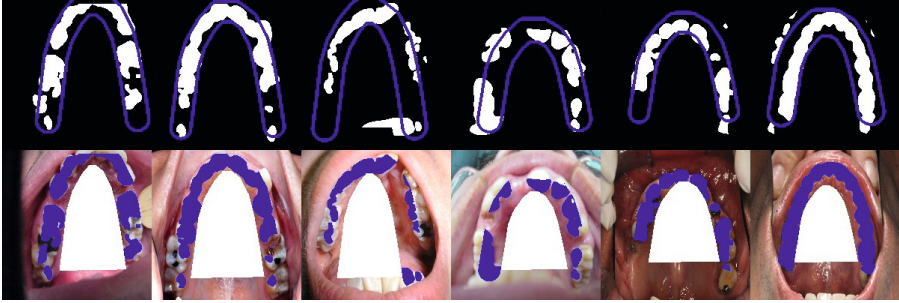


Fig. 5. Bandwidth selection and teeth/palate segmentation

### 5.1 Bandwidth Selection Experimental Results

Figure 6 shows the mean relative error of the application of bandwidth selection varying  $W$ . For each value of  $W$ , normal/anomalous threshold is selected such that minimizes the mean error.

The best results found in the training set are values of 13.3 and 0.1 for  $W$  and anomalous threshold. With this selection, the mean relative error between the bandwidth selected and the actual bandwidth is 0.09 in the training set and 0.16 in the test set.

Table 4 shows the teeth classification rate variation after the curve adjustment method and bandwidth selection. Recall from table 3 that the Neural Networks classification rate was 81.33%. Small blob filtering combined with the curve adjustment and bandwidth selection methods, improved the teeth classification rate to 89.74%. There is an important decrease in the percentage of false positives, yet there is an increment in the percentage of false negatives because some teeth regions may not be covered by the palatal curve, as it can be seen in figure 5.

Finally, the classification rate of the pixels as palate is 87.5267%. Errors were mainly false positive (8.7598%). We will work on improving the overall performance of our system.

Table 4. Teeth Classification after Curve Adjustment and Bandwidth Selection

Stage	Rate (%)	False Positive (%)	False Negative (%)
Curve Adjustment	87.8218	3.3403	8.8378
Bandwidth Selection	88.0421	1.8916	10.0663



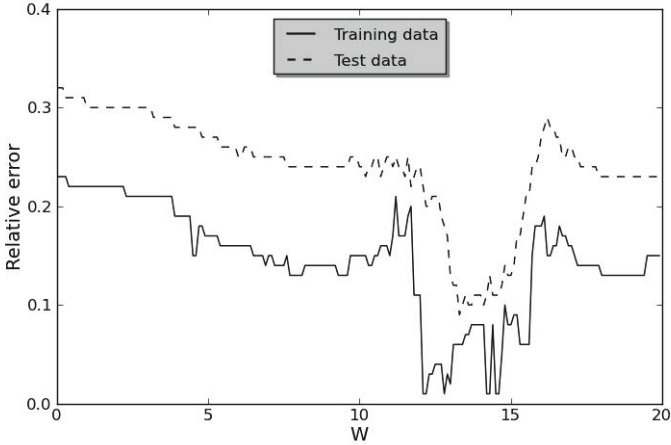


Fig. 6. Bandwidth Selection. Relative error vs.  $W$ .

### 6 Interdental Segmentation

Our first step to segment teeth is to flatten the palatal arch curve, *i.e.*, to transform the teeth band into a rectangle. Unfortunately, some pixels of the resulting rectangle are left blank because of the distortion of the image. Hence, such pixels are filled with a best first algorithm, using as priority the amount of neighboring filled pixels. Each blank pixel is filled with the average color value of its neighbors. Once the rectangle is entirely filled, the image is transformed to gray-scale, and improved by applying histogram equalization [9] [10] and edge enhancement filters [9].

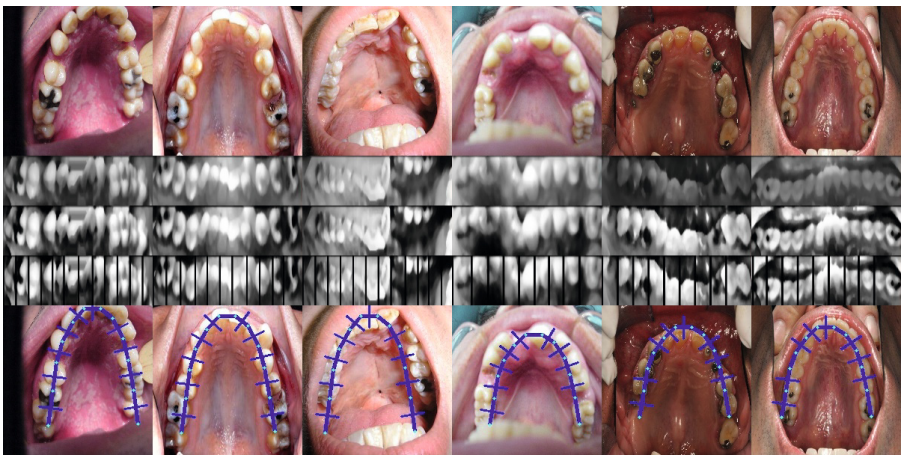


Fig. 7. Interdental Segmentation

Next, a sliding window algorithm searches for boundary between teeth. The main advantage of having the teeth region represented as a rectangle is that the sliding window can have fixed size. The algorithm complexity is linear in the number of window slides times the complexity time of the classifier.

We tested Support Vector Machine [12] and Multilayer Perceptron trained with Backpropagation [11]. Input are image slices of  $50 \times 25$  pixels. The classifiers are trained to detect gaps between teeth.

## 6.1 Interdental Segmentation Experimental Results

ANN outperform SVM at this task as shown in table 5. Best results in time and quality were reached with Neural Networks. The best Neural Network found had 50 neurons at the hidden layer and worked with the original flattened image. A smaller Neural Network of 15 hidden neurons reached the same classification rate during testing using the enhanced images as input. However, this networks did not reach perfect classification rate during training.

**Table 5.** Interdental Segmentation

Arguments	SVM				ANN			
	Kernel	Train	Test	Time (ms)	Top.	Train	Test	Time (ms)
Original	Linear	86.82	81.14	149	50	<b>100.00</b>	<b>88.82</b>	85
Enhanced	Poly	<b>92.27</b>	<b>83.99</b>	154	15	99.55	88.82	92

## 7 Conclusions and Future Work

We presented our results on automatic teeth/palate and interdental segmentation in palate view photographs of the upper jaw. Classic Multilayer Perceptrons trained with the error backpropagation algorithm outperformed SVMs for both of these tasks.

Image enhancement, filtering and adaptive curve fitting helped differentiate teeth and palates. Our system was able to cope with variations in patient anatomy, mouth scale in pictures, and view angle. Likewise, it was able to overcome standard problems for image processing systems, such as differences in illumination and image quality.

We wish to make our system more robust, so we are extending its capabilities to detect whether or not the input image depicts the upper jaw. Our final goal is to effectively detect common oral infecto-contagious diseases by looking at the palate, which could help diagnose HIV and AIDS. We believe this is an important initial step in that direction.

**Acknowledgment.** We would like to thank Dr. Vilma Tovar from the “Centro de Atención a Personas con Enfermedades Infectocontagiosas Dra Elsa La Corte”, Universidad Central de Venezuela. Dr. Tovar provided us with photographs of oral diseases and their diagnose.

## References

1. World Health Organization: Progress report 2011: Global HIV/AIDS response Epidemic update and health sector progress towards universal access. WHO Press, ISBN 978 92 4 150298 6,  
[http://www.who.int/hiv/pub/progress\\_report2011/en/index.html](http://www.who.int/hiv/pub/progress_report2011/en/index.html)
2. Barr, C.E.: Dental management of HIV-associated oral mucosal lesions: current and experimental techniques. In: Robertsson, P.B., Greenspan, J.S. (eds.) *Perspectives on Oral Manifestation of AIDS: Diagnosis and Management of HIV-Associated Infections*, pp. 77–95. PSG Publishing Co., Inc., Littleton (1988)
3. Arendorf, T.M., Bredekamp, B., Cloete, C., Sauer, G.: Oral manifestation of HIV infection in 600 South African patients. *J. Oral Pathol. Med.* 27, 176–179 (1998)
4. Nassar, D., Ammar, H.: A neural network system for matching dental radiographs. *Pattern Recognition* 40, 65–79 (2006)
5. Lin, P., Lai, Y., Huang, P.: An effective classification and numbering system for dental bitewing radiographs using teeth region and contour information. *Pattern Recognition* (2009)
6. Xiao, D., Ohya, J.: Contrast Enhancement of Color Images Based on Wavelet Transform and Human Visual System. In: *Proceedings of the IASTED International Conference Graphics and Visualization in Engineering, Florida* (2007)
7. Chimphee, C., Hanan, A., Noor, M.: Unsupervised Anomaly Detection with Unlabeled Data Using Clustering. In: *Proceedings of the Postgraduate Annual Research Seminar* (2005)
8. Gendreau, M., Potvin, J.: *Handbook of Metaheuristics*. Springer (2010)
9. Gonzalez, R., Woods, R.: *Digital Image Processing*, 3rd edn. Prentice Hall (2007)
10. Shapiro, L., Stockman, G.: *Computer Vision*. Prentice Hall (2001)
11. LeCun, Y., Bottou, L., Orr, G.B., Müller, K.-R.: Efficient BackProp. In: Orr, G.B., Müller, K.-R. (eds.) *Neural Networks: Tricks of the Trade*. LNCS, vol. 1524, pp. 9–50. Springer, Heidelberg (1998)
12. Burges, C.: A tutorial on support vector machines for pattern recognition. *Knowledge Discovery and Data Mining* 2(2) (1998)
13. Li, M., Cheung, Y.-M.: Automatic Segmentation of Color Lip Images Based on Morphological Filter. In: Diamantaras, K., Duch, W., Iliadis, L.S. (eds.) *ICANN 2010, Part I*. LNCS, vol. 6352, pp. 384–387. Springer, Heidelberg (2010)