

USE OF SUPPORT VECTOR MACHINE FOR TEETH RECOGNITION FROM
OCCLUSAL INTRAORAL DIGITAL PHOTOGRAPHIC IMAGESRAMON AUGUSTO SOUSA LINS*, KEYLLY EYGLYS ARAÚJO DOS SANTOS*, ADRIÃO DUARTE DÓRIA
NETO*, LUIS NORO[†], ANGELO GIUSEPPE RONCALLI[†], MARIA CRISTINA DOS SANTOS MEDEIROS[†],
PEDRO HENRIQUE SETTE DE SOUZA[†], SAMARA MARTINS DA SILVA[†]**Department of Computer and Automation
Federal University of Rio Grande do Norte, Natal, Rio Grande do Norte, Brazil.**[†]Department of Odontology
Federal University of Rio Grande do Norte, Natal, Rio Grande do Norte, Brazil.*Emails: ramon_asl@yahoo.com.br, eyglys.araujo@ifrn.edu.br, adriao@dca.ufrn.br,
luiz_noro@hotmail.com, roncalli@terra.com.br, mcristinamedeiros@hotmail.com,
pedro_sette@hotmail.com, samartins22@gmail.com

Abstract— In this work we propose the development of an intelligent system able to recognize teeth from occlusal intraoral digital photographic images. The system combines traditional techniques of machine learning and digital image processing. At first, we perform a color-based segmentation to find the regions of interest, teeth and non teeth, using the Support Vector Machine algorithm. After this, we use techniques based on morphological operators (erosion, watershed transform) to detect the teeth boundaries. Therewith, we calculate the Fourier descriptors to take into account their shapes and positions. Then, they are classified according to their types by the Support Vector Machine algorithm with the one-against-all method. The problem of multiclass classification is approached in two different ways. At first we consider three types of teeth classes: molar, premolar and non teeth, while in the second approach we consider five classes: molar, premolar, canine, incisor and non teeth. The results are shown through methods of performance estimation such as random sampling and confusion matrix.

Keywords— Occlusal intraoral digital photographic images, support vector machine, Fourier transform, multi-class classification.

Resumo— Neste trabalho nós propomos o desenvolvimento de um sistema inteligente capaz de reconhecer dentes a partir de imagens fotográficas digitais intraorais oclusais. O sistema faz uso combinado de técnicas de aprendizado de máquina e processamento digital de imagem. Inicialmente é realizada uma segmentação baseada em cores das regiões de interesse, dentes e não dentes, por um classificador binário usando máquina de vetores de suporte. Após a identificação dessas regiões são utilizadas técnicas baseadas em operadores morfológicos (erosão, transformada *watershed*) para detectar os limites dos dentes. Com isso é possível extrair seus descritores de forma (descritores de *Fourier*) e seus descritores de posição. Em seguida, os dentes são classificados quanto aos seus tipos por uma máquina de vetores de suporte com uso do método um-contra-todos, usado em problemas de classificação de múltiplas classes. O problema de classificação de múltiplas classes é abordado de duas maneiras diferentes. Na primeira abordagem, consideramos três tipos de classes: molares, pré-molares e não dentes, enquanto na segunda abordagem são considerados cinco tipos de classes: molares, pré-molares, caninos, incisivos e não dentes. Os resultados são mostrados através dos métodos de estimativa de desempenho amostragem aleatória e matriz de confusão.

Palavras-chave— Imagens fotográficas digitais intraorais oclusais, máquinas de vetores de suporte, transformada de Fourier, classificação multi-classe.

1 Introduction

The fast development of intelligent medical image systems is improving diagnoses and their interpretations. In dental industry, computer aided procedures such as dental implants, orthodontic planning, among others are making a widely use of digital dental images as radiographs, photographs and computerized tomographies in support to these procedures. Many efforts have been made in this area to develop methods that generates good results for different applications. Some developed techniques were based on statistics properties (Choorat et al., 2011), mathematical morphology (Mahsa Sepehrian, 2013), mathematical transforms (Gottlieb et al., 2014), among others.

From the different branches of dentistry, the public health is the area in which the dentist is responsible for the oral health diagnosis of a com-

munity. To this end, the dentist makes epidemiological surveys through oral visual inspections producing informations (e.g. identification and counting of troubled teeth) about the oral health conditions of Brazilian population, subsidizing the planning and evaluating the taken actions on different management levels of the Health Unique System (SUS, from portuguese Sistema Único de Saúde). This common process accomplished by dentists has some limiting factors, as for example a limited number of qualified professionals, different interpretations of diagnosis and so forth.

In order to assist the epidemiological survey, we propose the development of an intelligent system able to accomplish the basic visual task of recognize teeth from intraoral digital photographic images. Initially we do a color-based segmentation to find the regions of interest, teeth and non teeth, present in the images using the Support

Vector Machine (SVM) algorithm. After identifying those regions the data is used to detect their contours. At this stage, techniques based on morphological operators such as erosion and watershed transform are used to detect the teeth contours. Therewith, the Fourier descriptors are calculated taking in consideration their shapes and positions. Then they are classified according to their types by the SVM algorithm using the one-against-all method. The classification problem of multiclass is approached in two different ways. At first we consider three types of teeth classes (most of the problems happen in these teeth): molar, premolar and non teeth, while in the second approach we consider five classes: molar, premolar, canine, incisor and non teeth. The aim here is to analyse the possibility of resolving it using an approach still not explored in the literature, at least to the knowledge of the authors of this work.

This paper is organized as follows: in section II is done a brief bibliography revision of SVM. In section III the proposed system is explained. The results are described in section IV. Finally, in section V, the conclusion is discussed.

2 Support Vector Machine

The SVM is a classifier based on standard theory of statistic learning, proposed by Vapnick (Vapnick, 1995), that finds an optimal separation surface, minimizing the classification error. It can be used in linear problems solution as well as in non linear problems.

Originally the SVM was developed to solve binary problems (two classes). Extend its functionality remains current research object. Over the years, some methods were proposed to solve the problems of multiclass classification.

In the strategy one-against-all for every c classes c classifiers are built; The negative classes are all the others classes if not the positive class. The classification of each new instance is done using the methodology in which the winner-take-all, in other words, the classifier with highest output assign the corresponding class.

In this method, are used k SVM models, being k the number of classes. The i^{th} SVM model is trained with all the training data in the i class labeled positive and the others data classes labeled as negative. Thus, for a given training set $\{(\mathbf{x}_1^i, y_1^i), \dots, (\mathbf{x}_n^i, y_n^i)\}$, where $\mathbf{x}^i \in \mathbb{R}^d$ with $d = 1, \dots, n$ and $y^i \in \{1, \dots, k\}$ is the class of x^i , the i^{th} SVM model is solved as follows:

$$\min_{\mathbf{w}^i, b^i, \xi^i} \frac{1}{2} \|\mathbf{w}^i\|^2 + C \left(\sum_{j=1}^n \xi_j^i \right) \quad (1)$$

For the optimization problem there are k de-

cisions functions:

$$\begin{aligned} (w^1)^T \Phi(\mathbf{x}) + b^1 \\ \vdots \\ (w^k)^T \Phi(\mathbf{x}) + b^k \end{aligned}$$

Thus, it can be said that \mathbf{x} belongs to the class with highest value between decisions functions, being defined as:

$$\text{class of } x \equiv \arg \max_{i=1, \dots, k} ((w^i)^T \Phi(\mathbf{x}) + b^i) \quad (2)$$

For more details consult (Chih-Wei Hsu, 2002).

3 Proposed System

The process of teeth recognition from occlusal intraoral digital images is defined basically in four stages, see Figure 1. At first, we propose an image acquisition technique that fits the constraints faced by health workers in the field. Therewith, several measurements of inclination angles and distances from the acquisition camera were made. From this standardization, we created an image database that is used as input of the system. After this, the images go through a preprocessing stage to adapt their scale, color model and contrast adjustment to the problem. Next, they pass by different processes: segmentation based on colors by SVM, morphological operators, Fourier transform and multiclass classification by SVM so that the data can be represented and described adequately for the recognition process.

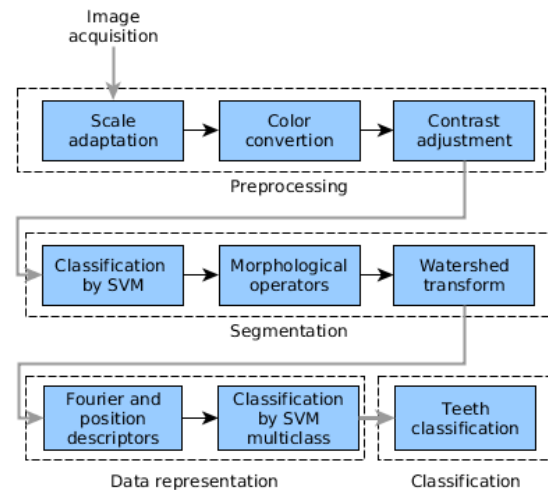


Figure 1: Block diagram of the proposed system.

3.1 Image acquisition

We use 40 images from compact cameras CANON, model A1300 with built-in flash. The images were

taken of two ways: upper and lower occlusal using a specific side mirror and lip retractor in adults.

The occlusal photograph is a two-dimensional image corresponding to the distribution of all teeth (third molar right to the third molar from left) in the dental arch (upper or lower) view from the top down or the opposite way as shown in Figure 2a.

3.2 Preprocessing

The preprocessing of images is a procedure frequently used in pattern recognition problems to adapt the input data for different purposes, to obtain the best possible solution to the problem. We use as input the image as a pixel matrix of RGB values (red, green, blue), with values ranging from 0 to 255 for 8-bit depth.

There still other models of color images, in this work is used the YCbCr color model, brief explained in section 3.2.2.

3.2.1 Scale adaptation

Due to the different scales resulting from the image acquisition process, is performed a scale adaptation, procedure that reduces or enlarges the image from a scale obtained according to the dimension of the image and an maximum dimensions of a fixed window 600 x 600 pixels, so that the fixed dimension is not exceeded.

This adjustment is accomplished through the bicubic interpolation method where the output pixels values are the average values of a 4x4 window pixels from the point neighborhood.

3.2.2 Conversion from RGB to YCbCr

The YCbCr model has redundancies that can be eliminated without prejudice the image, making them files smaller without major visual loss. Another important feature is the Y component that contains high frequencies of luminance in grayscale that facilitate the learning system process.

In this model, Y represents the luminance of an image, while the chrominance Cb is the blue (B-Y) and red chrominance Cr (R-Y), for more details consult Acharya and Tsai (2005).

3.2.3 Contrast adjustment

After the image be scaled and converted to YCbCr model, it's applied to each component a contrast enhancement to achieve better discrimination of their color pixels. The contrast adjustment is done by mapping each image pixel in a corresponding pixel to an [0,1] interval.

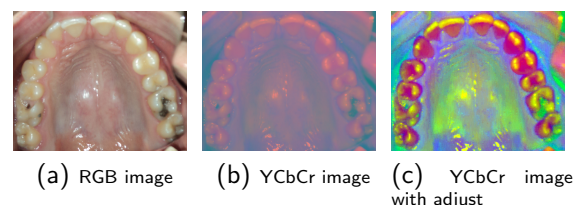


Figure 2: RGB image converted to the YCbCr model with contrast adjustment.

3.3 Segmentation

3.3.1 Based-color segmentation

From the conversion and contrast adjustment of the original image, the segmentation is done through an binary classifier (SVM). Each image pixel is classified as teeth or non teeth, setted to different sample colors for teeth and restorations in a class, and gums, tongue and other oral elements in a second class. After the creation of the training set the optimal hyperplane is calculated and the network parameters are defined. A test data set is utilized as input of the classifier and the output data forms a binary image with the regions of interest, non teeth and teeth, plus noise and unwanted regions, see sub-picture 3a.

3.3.2 Morphological operators

The morphological operators are used in the extraction of image components, so they can be properly used in the representation and description of its forms. From the obtained binary image we perform an image opening process by a structure element in a disk shape with radius equal's 1 extracting the image background, in a way that the most representative pixels of teeth are identified.

Next, the image passes through a median filter with 17x17 dimension. After the filtering process the image passes through a hole fill. According to Soille (Soille, 2002), binary image holes are defined as a set of background components that are not connected to the image edges. Following this idea, we can say that holes are collections of background pixels (black) surrounded by foreground pixels (white) that do not connect the edges of objects. This process is based on morphological reconstruction where masks and markers are defined from the filtered image.

After filling the holes, the image undergoes by an erosion process to remove some noises and unwanted components.

The combined use of these techniques makes the binary image adequated to to calculate the distance mapping of the regions of interest. The distance image is utilized by the watershed algorithm to identifying the boundaries of the teeth..

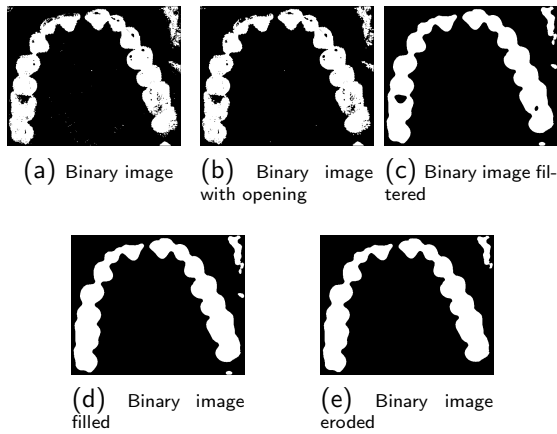


Figure 3: Identification stages of interest regions.

Meyer (1994), defines the distance image as topographical reliefs. This relief suffers a uniform flooding process from its regional minimums. From the moment the floods begin to mix, barriers are erected to prevent this from happening, these barriers are known as watershed lines.

Applying the watershed transform in the distance image it follows that the elements are labeled into integers larger or equal to zero known as watershed pixels, generating as output a label matrix that represents the regions of interest segmented. These steps can be better observed in Figure 4.

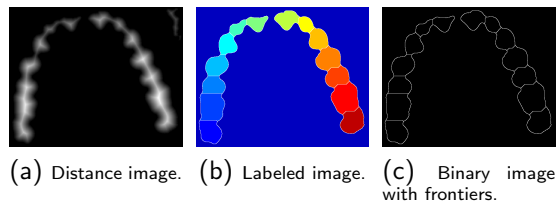


Figure 4: Frontier detection process of teeth from distance images.

3.4 Teeth type classification

At this stage the binary image with teeth borders are processed in order that the boundaries (Fourier) and position descriptors can be extracted of each segmented object. First, each object has its border decomposed in a 2-D outline as a polygonal curve of N points on the x and y plane. Next is defined an array of complex numbers through the vertices coordinates. We apply the Discrete Fourier Transform (DFT) on a complex vector obtaining a Fourier descriptors vector that represents the boundaries of each tooth. The DFT decomposes the data into a set of frequency components. We use ten Fourier coefficients for representing the data, number of descriptors that best described the teeth boundaries on the images

without significant loss.

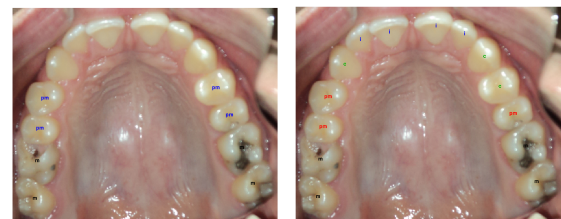
It is important that the descriptors be invariant to translation, rotation, scale and starting point given the different specific features of each patient's teeth. Make them invariant to scale is also important because of adversities found in the process of image acquisition.

According to Kaick et al. (2010), to obtain invariant descriptors to rotation of its on contour in the 2-D plane, the phase component must be discarded using only their magnitude. Discard the phase component also makes the descriptors invariant to start points contour selection. Eliminating the DC component makes the descriptors become invariant to translation. The DC component, which is the first sign of the coefficient corresponds to the average value of the signal shape. Finally, to obtain invariance to scale each coefficient is divided by the magnitude of the first frequency component, taking into account that the DC component was discarded.

The descriptors area and position are obtained from its area and angle. The area is normalized in function of the total area of segmented regions. The angle is obtained in function of each object relative to the central vertical axis and the center point of each image, normalized in function of the angles sum.

For the teeth type classification problem we use two approaches in trying to handle in the best possible way for this problem, since most of the problems occur into the molars and premolars. In first approach the objects are arranged in three classes for the training process and classification. To the first class are considered the molar teeth, in second class the premolars and in the third class the canines, the incisors and the objects that we considered non teeth.

In second approach we considered five classes. To the first class are considered the molar teeth, in second class the premolars, in third class canines, the incisors to fourth class and to the fifth class the objects that we considered non teeth. The result of both approaches can be seen on Figure 5.



(a) Classification of the teeth types for three classes. (b) Classification of the teeth types for five classes.

Figure 5: Classification process of the teeth types.

4 Results

4.1 Classification algorithm assessment for the based-color segmentation

4.1.1 Random Sampling

In the segmentation by colors we use a data set consisting of 22353 points with three dimensions (YCbCr), representing the pixels belonging to classes teeth and non teeth. The data set is divided in: half training, half testing. The training and test processes are carried out repeatedly for different kernel functions. By the end of the repeating process we obtained the statistical information results of average and variance shown in Table 1.

Table 1: Segmentation by teeth colors.

Function	Kernel	Accuracy(%)	Variance
Quadratic		96.59	0.02
Radial Base Function(RBF)		98.13	0.01
Sigmoid		72.83	4.46

4.2 Watershed segmentation performance

From the acquired images we recorded a total of 463 teeth, in which 117 of them are molars, 118 premolars, 78 canines and 150 incisors. Through the border detection method is possible by an expert count obtain a rate of 83.37% in the correct segmentation of the teeth present in images. The Table 2 shows concisely the results obtained during the segmentation process.

Table 2: Watershed segmentation of the teeth.

Teeth	Present	Segmented	Accuracy(%)
Molars	117	94	80.34
Premolars	118	102	86.44
Canines	78	68	87.18
Incisors	150	122	81.33
Total	463	386	83.37

4.3 Evaluation of the classification algorithm for the three classes approach

4.3.1 Random sampling

In the classification of teeth we use a data set consisting of 470 objects that represents the essence of the classes: molars, premolars and non teeth. The dataset is divided into 60% for training and 40% for testing. The training and testing processes is carried out repeatedly for different kernel functions. By the end of the repeating process we obtained the statistical information results of average and variance shown in the Table 3:

We can inferred from the classification with RBF kernel that the count of molar and premolars has an average hit rate of 88.13 % given the

Table 3: Classification of teeth type for the three classes approach.

Kernel Function	Accuracy(%)	Variance
Quadratic	85.87	10.16
Polynomial	84.16	7.55
RBF	88.13	3.01

amount of test data used. The others teeth (incisors and canines) can not be distinguished in this approach because they are considered in the same objects class labeled as non teeth.

4.3.2 Optimal classifier performance

For choosing the SVM parameters we use the grid search technique for a full search of parameters. The optimal classifier is set to RBF kernel, with a $\sigma = 0.9$ and a regularization parameter $C = 2$. The optimal classifier is assessed from their error metrics and confusion matrix.

Based on the confusion matrix the performance of the optimal classifier is shown in Table 4.

Table 4: Confusion matrix for the three classes approach

		True			
		molar	premolar	non teeth	
Predict	molar	37	3	0	
	premolar	1	36	2	
	non teeth	1	3	106	
accuracy(%)					94.7

4.4 Evaluation of the classification algorithm for the five classes approach

4.4.1 Random sampling

The five class approach evaluation is similar to the process described in 4.3.1, however two classes are inserted, canines and incisors. The statistical information results are shown in the following Table 5:

Table 5: Teeth type classification for five classes approach.

Kernel function	Accuracy(%)	Variance
Quadratic	77.72	11.05
Polynomial	74.51	12.81
RBF	80.95	5.09

We can inferred from the classification with RBF kernel that the teeth count has an average hit rate of 80.95 % given the amount of test data used. All teeth can be distinguished in this approach because the canines and incisors are now distinct classes among themselves and objects labeled as non teeth.

4.4.2 Optimal classifier performance

For choosing the SVM parameters we use the grid search technique for a full search of parameters. The optimal classifier is set to RBF kernel with a $\sigma = 0.9$ and a regularization parameter $C = 2$. The optimal classifier is assessed from their error metrics and confusion matrix. Based on the confusion matrix the classifier performance is shown in Table 6.

Table 6: Confusion matrix for the five classes approach

Predict	True					
	molar	premolar	canine	incisor	non teeth	
molar	36	3	0	0	0	
premolar	1	38	4	0	0	
canine	0	0	20	3	0	
incisor	0	0	3	47	3	
non teeth	2	1	1	1	26	
accuracy(%)						88.4

5 Conclusion and Future Work

This work proposes to use intelligent image processing techniques for interpretation of digital intraoral occlusal photos facing the public oral health.

As usual, working with appropriate digital images is always a limiting factor in intelligent systems development. Using photographs taken through common low cost cameras is even more challenging, comparing with the use of X-ray or computed tomography images. External factors such as brightness, noise, resolution might even impair the system use.

With these challenges in mind, we tried to use simpler classical techniques already established in intelligent image processing area. We opted for the use of SVM that jointly with the morphological operators techniques, showed to be efficient and robust techniques for boundaries detection. Considering also the use of the SVM in classification problems of multiple classes from method one-versus-all, together with the position descriptors and shape, we observed that this approach also shown satisfactory results to teeth recognition. The system demonstrates that machine learning with digital image processing techniques can assist the dentist in carrying out epidemiological surveys in public oral health and can contribute to the collection of oral health information from a population effectively.

In the future the intelligent system can be improved and modified by inserting new restrictive elements, such as information as caries, periodontal disease, malocclusion, prosthesis, fluorosis, and oral lesions. Therewith more information about the oral situation of the population can be obtained allowing a local planning increasingly optimized and efficient facing the principle of fairness.

An intuitive next step will be to use the deep convolutional neural networks to attack this problem and possibly observe better outcomes.

Acknowledgment

This work was funded by the National Council for Scientific and Technological Development (CNPq).

The authors would like to thank the odontologic department of the Federal University of Rio Grande do Norte - UFRN for the idealization of this work and all assistance provided.

References

- Acharya, T. and Tsai, P.-S. (2005). *JPEG2000 Standard for Image Compression: Concepts, Algorithms and VLSI Architectures*, 1 edn, Wiley.
- Chih-Wei Hsu, C.-J. L. (2002). A comparison of methods for multiclass support vector machines, pp. 415–425.
- Choorat, P., Chiracharit, W. and Chamnongthai, K. (2011). A single tooth segmentation using structural orientations and statistical textures, *4th Biomedical Engineering International Conference* pp. 294–297.
- Gottlieb, R., Sarrett, D., Belle, A., Najarian, K. and Hobson, R. (2014). An Automated Dental Caries Detection and Scoring System for Optical Images of Tooth Occlusal Surface, pp. 1925–1928.
- Kaick, O. V., Hamarneh, G., Ward, A. D., Schweitzer, M. and Zhang, H. (2010). Learning Fourier Descriptors for Computer-Aided Diagnosis of the Supraspinatus 1, **1**: 1–33.
- Mahsa Sepehrian, Ali M. Deylami, R. A. Z. (2013). Individual Teeth Segmentation in CBCT and MSCT Dental Images Using Watershed, *20th Iranian Conference on Biomedical Engineering* pp. 27–30.
- Meyer, F. (1994). Topographic distance and watershed lines, *Signal Processing* **38**: 113–225.
- Soille, P. (2002). *Morphological Image Analysis*, 2 edn, Springer.
- Vapnick, V. (1995). The nature of statistical learning theory, *Springer-Verlag*.