FACE AND EYE TRACKING FOR PARAMETERIZATION OF COCHLEAR IMPLANTS

M. Cabeleira[1a], S. Ferreira[1b], L. F. Silva [2] , C. Correia[1] and J. Cardoso[1]
[1]*Instrumentation Center, Physics Department, University of Coimbra, R. Larga, Coimbra, Portugal*
[2]*Serviço de Otorrinolaringologia do Centro Hospitalar de Coimbra - (CHC), Coimbra, Portugal*

Keywords: Cochlear Implants, Eye-tacking, Colour based Trackers, Viola & Jones Face Tracker, Between-the-Eyes, Gabor Filter.

Abstract: This work presents a free head eye-tracking solution created for use as a complementary tool in the parameterization of cochlear implants. Nowadays, the parameterization of these implants is a long and cumbersome process performed by audiologists and speech therapists that throughout many periodic evaluations where audiometric tests and electrode adjustments are performed. The eye tracking system will assist this process through detection of saccades generated when a subject hears sounds produced during the audiometric test procedure. The main purpose is to ease and improve the implant re-parameterization procedure with uncooperative subjects, like children. The developed system is composed of three digital video cameras where two of the cameras are responsible of the detection of the position of the face and eyes and the third is responsible of the gaze detection. The developed face and eye detectors are also compared in order to choose the best combination of algorithms to perform robust eye detection with unpredictable subjects. The best combination of algorithms is the Viola-Jones face detector combined with an eye detector Ring Gabor filters, that correctly detected the eye-position in 76,81% of the tested videos at 18 frames per second.

1 INTRODUCTION

Cochlear implants are used in cases of profound or total deafness to restore hearing capabilities. An early implementation, particularly in children younger than three years old, allows for the exposition to sound stimuli which is essential for the development of speech and the articulation of language. A cochlear implant is composed by a microphone, a speech processor and a set of surgically implanted electrodes inside the cochlea, that stimulate the vestibulocochlear nerve responsible for hearing. After the surgery, the patient must undergo an extensive electrode parameterization procedure to achieve optimal hearing capabilities, in a procedure strongly dependent on the expertise of audiologists and speech therapists. Many periodic audiometric evaluations are performed to fine tune each electrode's gain. This slow and subjective task becomes even more complex when applied to children under twelve. Thus, innovative and objective implant parameterization techniques capable of minimizing technician's errors and allow faster and reliable procedures are mandatory (Porter, 2003).

A possible solution makes use of eye-tracking techniques, as a complement of the existing protocol, to determine the gaze point whenever a sound is heard. The system must be non-invasive, comfortable and easy to use. The proposed eye-tracking system should robustly detect and track the face region after a simple calibration procedure, as well as detect the gaze vector (even when contact is lost and without being recalibrated) at around 50 frames per second (i.e. time resolution should reach 20 ms).

The eye-tracking DAQ module is composed by two webcams and one high resolution (HR) digital camera. The webcams are for face and eyes detection along with distance assessment, while the digital HR camera is used to acquire the eye ROIs. The 3 camera solution minimizes processing time and lowers overall costs.

The purpose of this work is to select the best set of algorithms for face and eye detection. To perform the face detection, colour based algorithms were tested against feature based algorithms for robustness, detection accuracy and processing time. Similarly, eye detection is accomplished both by morphological fea-

Cabeleira M., Ferreira S., F. Silva L., Correia C. and Cardoso J..
FACE AND EYE TRACKING FOR PARAMETERIZATION OF COCHLEAR IMPLANTS.
DOI: 10.5220/0003792604290433
In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing* (BIOSIGNALS-2012), pages 429-433
ISBN: 978-989-8425-89-8

tures extraction ('between the eyes') and Gabor filtering.(Zhao, 2003)

2 ALGORITHMS

A visual representation of the algorithms used is presented in Figure 1

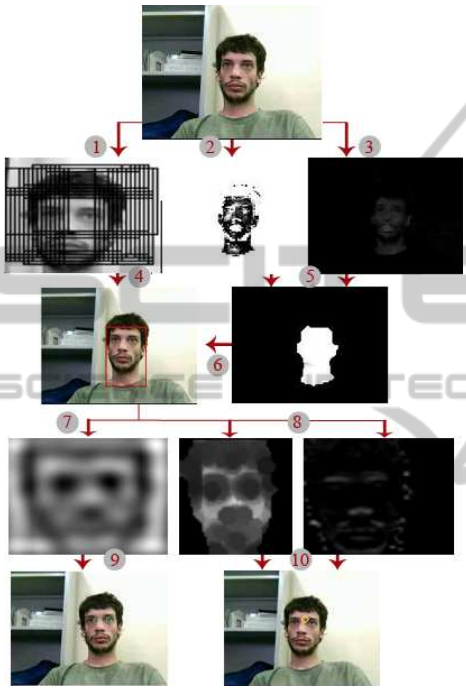

Figure 1: Schematic representation of the Face and Eye detection algorithms. (1) lighter version of the Viola-Jones algorithm, (2) Image of the Mahalanobis Distance operation, (3) Image of the Red-Green operation, (4) Bounding Box Selection, (5) Image binarization, (6) Bounding Box generation, (7) Image resultant from the Elliptical Gabor Filter, (8) Eroded image and vertical gradient image of the face, (9) Eyes marked using information obtained with the Elliptical Gabor filter, (10) Between the eyes point marked.

2.1 Face Detection and Tracking

For head or face analysis, three algorithms were studied: Red-Green (RG), Mahalanobis Distance (MD) and Viola-Jones (VJ). The first two algorithms are based on skin colour properties and the last algorithm is focused on facial features.

2.1.1 Red-Green

This algorithm was firstly developed by Saleh Al-Shehri and is the simplest amongst all the algorithms tested in this work. It uses images acquired in the RGB format and takes advantage of two basic principles; the red component is predominant in human skin and the R/G ratio is bigger than 1. Therefore if we subtract the green component to the red component, skin pixels will acquire values higher than 0 (Al-Shehri, 2004).

Non-skin pixels with values much higher than 0 can also appear and must be eliminated. This was accomplished by defining a window of values that are considered skin colour. The lower threshold used was the value 0.02 and the higher was 0.25 (considering a gray-scale ranging from -1 to 1).

This method excludes parts of the face like eyes and beard, therefore a closing operator was used to connect all the detected points in the face before calculating the Bounding Box (BBox) around the face. In order to counter several miscalculations where the generated BBox included the neck and clothes, a selection method was implemented. This method adjusts the BBox taking as premise that the face is located on the top of the body and in between of the shoulders.

2.1.2 Mahalanobis Distance (MD)

The MD allows the computation of the relative distance, or similitude, between each element of unknown data sets to known ones, being this distance measurement widely used in the segmentation of skin colours (Supriya, 2010). The colour space used for this algorithm was the YC_BC_R and it measures the MD between the colour components C_B and C_R present in the picture and the average skin colour components defined by the author as being 107.9649 for the C_B component and 140.8913 for the C_R component. The equations implemented for this algorithm may be found in (Maesschalck, 2000). The covariance matrix (C_x^{-1}) used was:

$$C_x^{-1} = \begin{bmatrix} 1.8328501 \times 10^3 & 2.2506719 \times 10^3 \\ 2.2506719 \times 10^3 & 6.8658257 \times 10^3 \end{bmatrix} \quad (1)$$

These values were obtained by the author by using a wide number of images containing skin colour samples.

2.1.3 Viola-Jones (VJ)

This method was developed by Paul Viola and Michael Jones and it differs from the previous two because instead of using pixel colour information in order to locate the face in an image, this algorithm applies Haar-like features to the image through various stages of a previously trained classifier. A more in depth explanation of the algorithm may be found in the literature (Viola, 2001).

In this work the algorithm used was developed by Dirk-Jan Kroon (Mathworks, 2011), it consists of a previously trained classifier cascade with 22 stages of classification and spanned the images 11 times at different scales. To overcome processing time requirements, only the first two frames of the video used the original algorithm. The role of these two frames was to define the scale at which the face and its initial position in the image. That information was then used in the subsequent images where a lighter version of the algorithm was applied. This version used only a subwindow of the original image centred in the position detected in the previous frame and spanned this subwindow only in the scale defined in the initial frames, the classifier cascade was also simplified using only the first 5 stages. As a result of this simplification, the algorithm performed a lot faster not without losing some of its specificity generating lots of BBox around the real face. To select the correct BBox, the image was binarized using a threshold of 0.45 to 0.6 highlighting the darker pixels. The BBox with the more dark pixels was the one used and the sub-window was cropped and forced to become a 50x50 picture for further analysis.

2.2 Eye Detection

To locate the eyes in the previously detected face, two methods were studied: Between the Eyes (BTE) and Elliptical Gabor Filter (EGF).

2.2.1 Between-the-Eyes (BTE)

This algorithm was developed having as reference the work done by Peng et al. (Peng, 2005). This algorithm takes advantage of the unique illumination profile of a face and extracts the point that lies in between of the eyes. It works in two fundamental steps. The first consists in computing the vertical gradient image of the face to detect the increased shadow present in the transition between the forehead and the eyebrow region, to do this a horizontal cumulative sum is performed and the line of pixels with higher value is considered as the horizontal coordinate of the BTE point. A sub-window of the original greyscale image, centred in this line is then extracted and eroded by a circular operator with 4 pixels in diameter. Vertical cumulative sum is performed and again the column with higher value is considered the vertical coordinate of the point.

2.2.2 Gabor Elliptical Filter (GEF)

The Gabor filter used to detect the eyes was based on the work done by Yanfang Zhang, it consists in the use of an elliptical Gabor filter that is a 2-D band pass filter generated by a Gaussian function modulated by a sine function. The equation and pictures of the filter are also presented in his work (Zhang, 2005). In this work, the parameters used to compute the filter were $\sigma_x = 15$, $\sigma_y = 15$, F=64, $\theta = 0$ and the filter size was 15x15. Here σ_x and σ_y stand for variances or scale factors along x and y axis, F for spatial central frequency and θ for the rotation angle of the filter. The real part of the filter was selected and normalized for further use. The computed filter was then convolved with the gray-scale face image to highlight the eyes position. The eye region appeared in the image as holes, therefore to select the eye regions the picture was binarized highlighting the darkest regions. The centroid position of these regions was then used to mark the position of the eyes.

3 TEST SETUP

In order to compare the effectiveness of the developed algorithms, a video database was created. This database encompassed 6 types of videos of 23 different subjects performing different head movements on each video. The protocol for each video is presented in Table 1. The subjects used for this database were all voluntary members found in the investigation centre.

Table 1: Description of the head movement performed on each video.

Video	Subject's movement description
I	Upright position at rest
II	Approach the camera and stray away
III	Laterally tilt the head
IV	Tilt the head
V	Pan the head
VI	Free head movement

The videos were acquired using a Logitech C210 USB webcam at 30 frames/s and with a length of 10s with a resolution of 120x160. The recorded videos were then processed in Simulink where each video was run 6 times covering all possible combinations of face detector and eye detectors. All videos were processed in offline using a Intel Core 2 Duo CPU E6750 @ 2.66 GHz computer (RAM 2GB and 64-bit Operating System).

To evaluate the results in the terms of accuracy a score scale with three steps was created: the score "0" was attributed to algorithms that failed to detect the correct face or eye position, "1" was attributed

to algorithms that can detect the face or eyes position with minor errors and ”2” was attributed to algorithms that can correctly detect the face or eye position. The time performance of each algorithm was also measured (maximum frame rate).

4 RESULTS & DISCUSSION

The results for the time performance tests performed on each combination of algorithms is presented in Table 2.

Table 2: Results for the time performance tests.

Eye detection	Face detection	Average frames /s
BTE	R-G	43
	MD	37
	VJ	23
EGF	R-G	62
	MD	34
	VJ	18

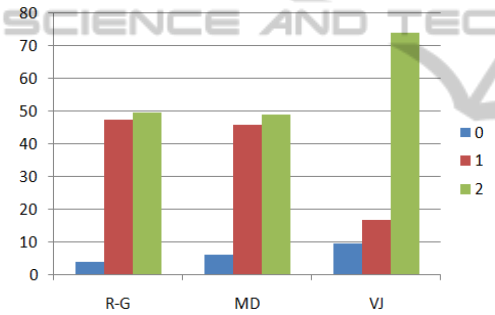

Figure 2: Graphic representation of the results obtained for the totals of each face detectors.

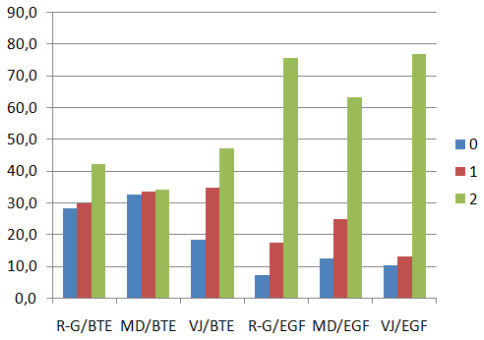

Figure 3: Graphic representation of the results obtained for the totals of each eye detectors.

As expected, the face detector algorithm that performed better in terms of processing time was the R-G, reaching an average value of 62 fps when used with the EGF. With this processing speed it should be possible to detect a saccade movement at its lower timing threshold. The VJ and the MD still need several performance improvements because, to detect a saccade movement, frame rates higher than 50 frames/s are mandatory. By analysing the results relative to the eye detectors, the EGF operated faster when combined with the R-G and slower in the other two. The exact opposite occurred with the BTE.

As can be observed in Figure 2 the results of the two colour based algorithms show similar tendencies being capable of correctly detect a face in roughly 50% of the cases. In the other cases the face was usually detected in some point, but with positional errors or with inconstant detections. The algorithm that performed the best was the VJ, being capable of correctly detecting the face in 73.91% of the studied videos. When analysing the accuracy of the face detectors, present in Table 3, it is evident that the colour based detectors are quite robust to the head movements, because they generate similar results on each type of movement. The erroneous face detection in these algorithms can be explained by illumination problems and skin-like objects present in the image that can mislead the algorithms. On the other side the VJ is more sensible to some of the head movements than to others, in particular movements that involve face scale variations mainly because, the BBox generated always have the same size.

To improve the accuracy of the eye detection algorithms, in case a face was not detected at all, the score given to the eye algorithm was also 0. As can be seen in Table 4, the EGF exhibits better eye detection results than the BTE, being able of outputting correct eye positions even when the face was segmented with errors. For this reason the EGF can compensate for the errors inherited from the face detectors and generate better results. The BTE exhibited poor eye detection results, even when it was applied to a well segmented face by the VJ in ideal conditions like in video I where the face was impeccably segmented.

The table also reveal the superior accuracy of the R-G when compared to the other colour based algorithm. In video III all the combinations of algorithms output worse results, being considered the most difficult head movement to track. In Figure 3 can be observed that the most accurate combination of algorithms to extract eye positions is the VJ and the EGF having detected the correct eye position in 76.81% of the cases. The results for the combination of R-G and EGF were also very appealing, being capable of detecting the correct eye position in 75.36% of cases. The percentage of erroneous detections was 17.39%, which was higher than the 13.04% obtained by the VJ and the EGF. In spite of this apparent superiority, the percentage of non-detections was higher with the VJ and the EGF.

Table 3: Results for the face detectors accuracy.

Face Detection Method	Videos																	
	I			II			III			IV			V			VI		
	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2
R-G	8.7	47.8	43.5	0	39.1	60.9	4.3	56.5	39.1	4.3	47.8	47.8	4.3	47.8	47.8	0	43.5	56.5
MD	8.7	43.5	47.8	0	43.5	56.5	8.7	60.9	30.4	8.7	56.5	34.8	8.7	30.4	60.8	0	39.1	60.9
VJ	13.0	0	86.9	4.3	34.8	60.9	4.3	21.7	73.9	13.0	13.0	73.9	13.0	4.3	82.6	8.7	26.1	65.2

Table 4: Results for the face and eye detectors accuracy.

Eye Detection Method	Face Detection Method	Videos																	
		I			II			III			IV			V			VI		
		0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2
BTE	R-G	30.4	17.4	52.2	17.4	17.4	65.2	39.1	43.5	17.4	30.4	17.4	52.2	30.4	34.8	34.8	21.7	47.8	30.4
	MD	30.4	26.1	43.5	21.7	34.8	43.5	39.1	34.8	26.1	30.4	39.1	30.4	43.5	21.7	34.8	30.4	43.5	26.1
	VJ	17.4	34.8	47.8	13.0	43.5	43.5	26.1	30.4	43.5	17.4	26.1	56.5	17.4	30.4	52.2	17.4	43.5	39.1
EGF	R-G	13.0	8.7	78.3	4.3	34.8	60.9	8.9	26.1	65.2	8.7	4.3	86.9	4.3	13.0	82.6	4.3	17.4	78.3
	MD	17.4	4.3	78.3	8.7	34.8	56.5	13.0	43.5	43.5	13.0	4.3	82.6	13.0	21.7	65.2	8.7	39.1	52.2
	VJ	13.0	0.0	86.9	4.3	17.4	78.3	4.3	26.1	69.6	13.0	13.0	73.9	13.1	8.7	78.3	13.0	13.0	73.9

5 CONCLUSIONS AND FUTURE WORK

Both the R-G and VJ performed well in the tests. Although the VJ has a higher percentage of correct detections than the R-G, it requires a much longer processing time. The best eye detector is the EGF therefore the best possible combinations are the R-G with EGF or the VJ with the EGF. The colour based algorithms are not robust enough to the task of detecting faces, but could be used as a first step of image processing in order to isolate regions where a face could be present in order to diminish the processing time required by the VJ to segment the face.

In the future the face detectors and the eye detectors will be improved in terms of processing time and accuracy and will be implemented in C++ using the OpenCv library. The cameras used to acquire the image will also be improved to cameras capable of acquiring images at frame rates of the order of the 50-60 frames per second in order to be able to detect a saccade.

The project will enter in a second phase where algorithms of stereo vision will be implemented in order to calculate the distance between the head and the system, in order to estimate the head pose and to estimate the position of the eye under study in the third camera. Eye processing techniques will also be implemented in order to estimate the gaze point and complete in this way the eye tracking system.

ACKNOWLEDGEMENTS

We would like to thank *Fundação para a Ciência e a Tecnologia* (FCT) for the financial support of this project (PTDC/SAU-BEB/100866/2008). We would also like to thank the collaboration of the patients and their families and for all support received from *Serviço de Otorrinolaringologia do Centro Hospitalar de Coimbra* (CHC).

REFERENCES

Al-Shehri, S. (2004). *A simple and novel method for skin detection and face locating and tracking.* Lecture Notes in Computer Science, Volume 3101/2004, 1-8.

Maesschalck, R. (2000). *The Mahalanobis distance.* Elsevier Science B.V., Chemometrics and Intelligent Laboratory Systems, 50, 1-18.

Mathworks (2011). *http://www.mathworks.com/matlabcentral/fileexchange/authors/29180*.

Peng, K. (2005). *A robust algorithm for eye detection on gray intensity face without spectacles.* JCST, Vol.5, No.3.

Porter, G. T. (2003). Cochlear implants. In *Grand Rounds Presentation*. UTMB, Dept. Of Otolaryngology.

Supriya, K. (2010). *Facial Gesture Recognition using Correlation and Mahalanobis Distance.* IJCSIS, Vol. 7, No.2.

Viola, P., J. M. (2001). *Rapid object detection using a boosted cascade of simple features.* IEEE, Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp.511.

Zhang, Z. (2005). *A robust method for eye features extraction on color image.* Elsiever B.V., Pattern Recognition Letters, 26, 2252-2261.

Zhao, W. (2003). *Face Recognition: A Literature Survey*. ACM Computing Surveys, vol. 35, no. 4, pp. 399-458.