BIO-INSPIRED IMAGE PROCESSING FOR VISION AIDS

C. Morillas, F. Pelayo, J. P. Cobos, A. Prieto
Department of Computer Architecture and Technology, University of Granada, Spain

S. Romero
Department of Computer Science, University of Jaén, Spain

Keywords: Bioinspired real-time image processing, vision aids, neuromorphic encoding, visual neuroprostheses.

Abstract: We present in this paper a system conceived to perform a bioinspired image processing and different output encoding schemes, oriented to the development of visual aids for the blind or for visually-impaired patients. We remark some of its main features, as the possibility of combining different image processing modalities (colour, motion, depth, etc.) and different output devices (Head Mounted Displays, headphones, and microelectrode arrays), as well as its implementation on a reconfigurable chip (FPGA) or a specific VLSI chip, which allows working in real time on a portable equipment. A software design environment has been developed for the simulation and the automatic synthesis of the processing models into a hardware platform.

1 INTRODUCTION

Visual impairment is considered as one of the 4 main causes for the loss of self-sufficiency among elderly people. With different affection degrees, visual impairment affects about a 25% of persons over 65 years old, and a 15% of adults between 45 and 65 years old. In addition, the progressive ageing of the population in developed countries makes these numbers grow forth, propitiating a remarkable loss in visual acuity and a reduction of the visual field. In this context, retinal degenerations (especially the age-related macular degeneration, ARMD), cataracts, glaucoma, diabetic retinopathy, optic nerve damage, and ocular traumas, yield a relevant amount of blindness cases, often non-curable.

Visually impaired patients require optical aids (microscopes, magnifiers, telescopes, optic filters) to enhance their quality of life, exploiting their remaining functional vision. However, there is no a unique aid able to provide this enhancement under any circumstance. Electronic aids, as (LVES, 1994) V-MAX, or the recent (JORDY, 2007) provide a more efficient use of the visual functional remains of the patient by magnifying images, enhancing light/darkness and colour contrasts, but none of these systems are able to implement an efficient control of local gain to produce clear and sharp images in a variety of lighting situations. These devices also use to be relatively heavy (0.5 to 1 Kg.), quite expensive and difficult to manipulate during motion. These reasons led us to propose the system described in this paper, which is inspired by the way the biological retina works, and is fully adaptable and configurable to each patient.

The retina-like design, not only at a functional level, but also at an architectural level, is a key aspect in the development of a robust and efficient system able to apply in real time local spatio-temporal contrast processing of the visual information. The final system has been developed on a reconfigurable hardware platform in order to provide real-time and portable solutions for visual processing that fit to the particularities of the visual impairment of every person, and which can be tuned according to its evolution with time. According to diversity of affections, different output encoding modalities have been considered, including acoustic encoding, high resolution image for Head Mounted Displays and neuromorphic encoding for neuroprostheses.

The next section is devoted to explain the bioinspired image processing in the system we present, and its main architecture. In section 3, a spike event encoding method is detailed, that is able to produce trains of electrical signals intended to

Morillas C., Pelayo F., P. Cobos J., Prieto A. and Romero S. (2008).
BIO-INSPIRED IMAGE PROCESSING FOR VISION AIDS.
In *Proceedings of the First International Conference on Bio-inspired Systems and Signal Processing*, pages 63-69
DOI: 10.5220/0001063500630069

stimulate neurons of the human visual system. Section 4 describes an acoustic signal generation module, that allows to the blind to localize those objects in the visual environment that produce higher activity levels. Finally real-time hardware implementation is presented and conclusions are summarized.

2 BIOINSPIRED IMAGE PROCESSING

The development of a bioinspired system for visual processing is being pursued by several research groups, as the tuneable retinal encoder, by (Eckmiller, 1999), or the computational models of retinal functions described by (Koch, 1986). The CORTIVIS (Cortical Visual Neuroprosthesis for the Blind) consortium has also implemented a bioinspired retinal processing model as part of a system designed to transform the visual world in front of a blind individual into multiple electrical signals that could be used to stimulate, in real time, the neurons at his/her visual cortex (Cortivis, 2002; Romero, 2005).

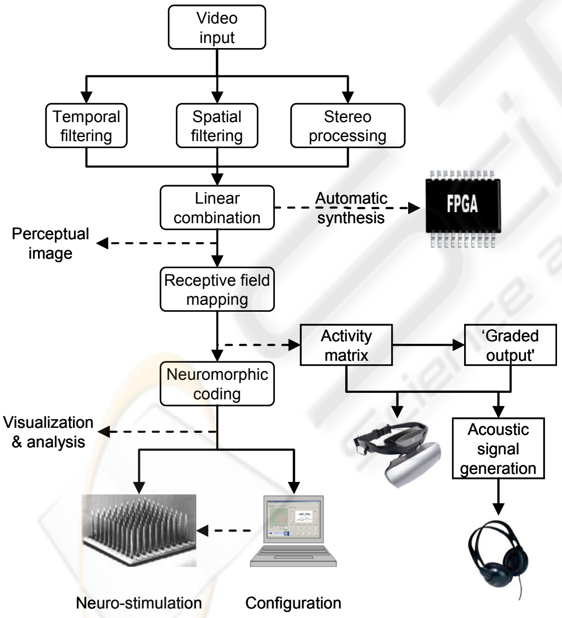

Figure 1: Reference architecture of the bioinspired image processing system for the development of visual aids. After obtaining a linear combination of spatial, temporal and depth-related features, different outputs for a variety of applications are possible. The choices include automatic synthesis for programmable devices; sending the information to the patient by means of HMD, headphones for acoustic signalling, or delivering neurostimulation to the neural tissue to evoke visual perceptions.

Even though the main objective of the CORTIVIS project, in which our research group has been involved, was the design of a complete system for neurostimulation, a part of the system is useful as a processing scheme which can be adapted to match the requirements aforementioned, and, this way, develop non-invasive aids for visualization, or sensorial transduction to translate visual information into sound patterns.

Figure 1 shows the reference architecture illustrating all the capabilities developed up to this moment. The input video signal is processed in parallel by three modules for the extraction or enhancing of image features, according to different processing modes. The first module performs a temporal filtering, as natural retinae also respond and remark temporal changes in the visual input; see for example (Victor, 1999). In our platform this temporal enhancement is implemented by remarking the differences between two or more consecutive frames, and with different strength in the periphery of the visual field (foveated model) as in natural retinae (Morillas, 2007).

For the spatial processing, an intensity and colour-contrast filtering is applied to different combinations of the three colour planes (red, green and blue) composing a frame. This spatial filtering emulates the function of bipolar cells in the retina, in the form of difference of Gaussians filters.

Our platform offers a variety of bioinspired predefined and parameterized filters, including Gaussians (1), difference of Gaussians (2) and laplacian of Gaussians (3). Even more, we can include new filters in the form of any Matlab (Mathworks, 2007) expression over the colour or intensity channels.

$$G_{\sigma 1}(x,y) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left[-\frac{x^2+y^2}{2\sigma_1^2}\right] \tag{1}$$

$$DoG = G_{\sigma 1} - G_{\sigma 2} = \frac{1}{\sqrt{2\pi}}\left[\frac{1}{\sigma_1} e^{-\left(x^2+y^2\right)/2\sigma_1^2} - \frac{1}{\sigma_2} e^{-\left(x^2+y^2\right)/2\sigma_2^2}\right] \tag{2}$$

$$LoG = \Delta G_{\sigma}(x,y) = \frac{x^2+y^2-2\sigma^2}{\sigma^4} e^{-\left(x^2+y^2\right)/2\sigma^2} \tag{3}$$

The stereo processing module obtains disparity maps at different resolutions, starting from image pairs captured by two head mounted cameras. Figure 2 shows examples of application where disparity maps are used as a weighting term for the output of a spatio-temporal filtering combination, emphasizing closer objects which produce a higher activity levels.

The next stage in figure 1 gathers the results obtained by each of the processing modalities. Its objective is to integrate as much information as possible into a single compact representation, so it requires a maximum degree of compression to allow the integration of the most relevant features. Given a real scene, we intend to remove all the background content, so only the closest objects are remarked, which are considered to be the most relevant information for an application like the one described in this paper, conceived for basic visual exploration tasks and obstacle-avoidance navigation. After some initial experiments with a portable prototype, we considered the need for incorporating an ultrasonic range finder, which provides a measurement of the distance to the closest object that can be used to ponderate the output of later stages, based on proximity.

According to the kind of application, the resolution for the output will be different; however, we can consider a general reduction of the resolution. For a neuroprosthesis, this resolution will match the number of available electrodes in the physical interface, which is currently in the order of hundreds of channels (Fernández, 2005). If we apply this scheme for a sensorial transduction system for the translation of visual information into audible patterns, we will be restricted by the amount of different sounds that the patient is able to distinguish without interfere his/her normal perception capabilities.

The reduction of spatial resolution is based on the concept of receptive field, which can be defined as a zone of the image (set of pixels) that contribute to the calculation of the value resulting in the reduced representation, which we call “activity matrix”.

The default configuration performs a partition of the image into rectangular non-overlapping areas of equal size, however we have also developed a tool for the definition of more complex structures, allowing even different sizes and shapes, which also can be variable, depending on its localization, from the centre of the visual field to its periphery.

Once the system has computed the activity matrix, depending on the specific application we will use it in a different way. In the case of a neuroprosthesis, the next stage is the recoding of this information into a neuromorphic representation, as a sequence of stimulation events (spikes), which will be later used to drive a clinical stimulator. Another possible use is the display of this information by means of specialized portable screens as HMD (Head Mounted Displays), to assist low-vision patients suffering a visual deficiency but still holding a functional remain of his/her vision.

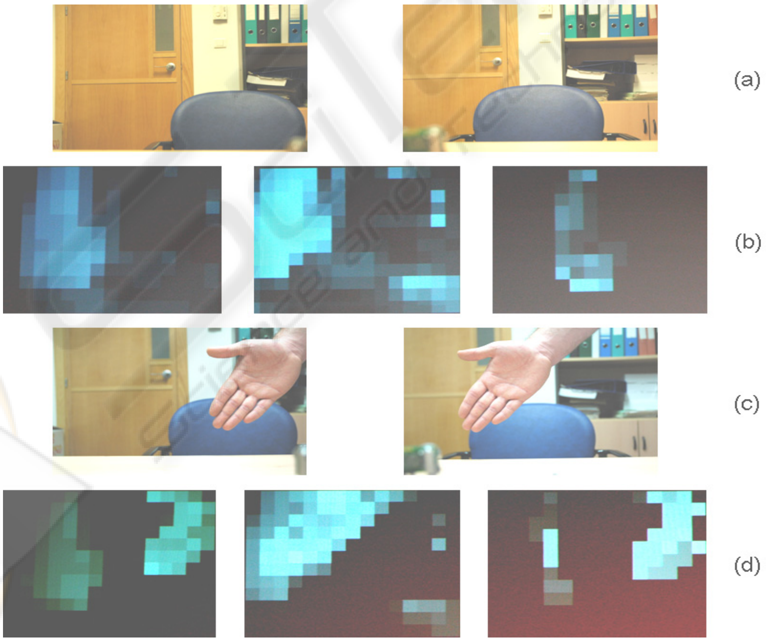

Figure 2: Stereo processing results obtained with the hardware implementation described in section 5. (a) Left and right images from the stereo cameras. (b) Activity matrices for a combination of spatial filters on the left input (left), right input (centre) and disparity weighting (right), the door is detected as relevant for spatial filters, but the chair, that is closer, is enhanced by the disparity computation. (c) A new closer object (the hand) appears in the visual field. (d) The disparity filter (on the right) detects this new object as the most relevant, due to its higher activity at the output matrix. The disparity output image is referred to the filtered left image in all cases.

The information provided by the activity matrix has been also employed in our system to locate the most relevant zones of the scene and translate them into sound patterns that include 3D spatial information. This way, the system can point out the location of the highest activity levels in the scene for the patient.

3 NEUROMORPHIC ENCODING

Features extracted by the image processing stage can be used in a complete neural stimulation system, being transformed by a spiking neuron model that is able to translate numerical activity levels into spike trains that the stimulation device can handle.

Different neuron models can be found in the literature (Gerstner, 2002), and we decided to implement an integrate-and-fire spiking neuron model, including a leakage factor, because of the simplicity to be implemented in a discrete system.

The selected spiking neuron model, depicted in figure 3, needs a set of accumulators which gather activity levels resulting of the current frame processing. When a value is integrated, the result is compared to a previously defined threshold, and if reached, the accumulator is initialized and a spike event is raised. The leakage factor avoids unexpected events due to ambient noise or residual activity from previously processed frames.

Each spike event generated is delivered to the stimulation device which has to form the corresponding electric waveform to be applied to the neural tissue.

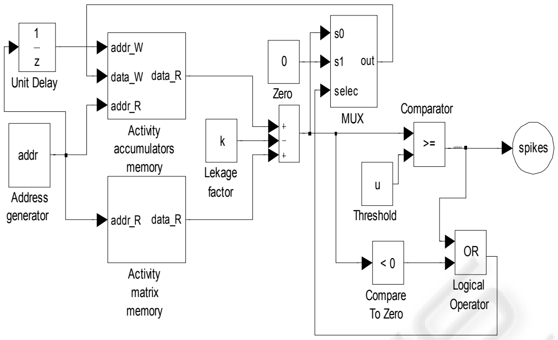

Figure 3: Block diagram of the neuromorphic coding subsystem for a sequential implementation.

All the events generated during a stimulation session can be stored for analysis. Figure 4 shows a graphic representation of all the events produced by a white horizontal bar moving from bottom to top on a black background, considering a 10 by 10 channels stimulation device. In this example the retina function is approximated by a simple model described by the expression (4):

$$retina = \frac{1}{5} \cdot I + F_{temp} \qquad (4)$$

where I is the input pixel intensity and the temporal filter F_{temp} compares, for each pixel, the current intensity value with the average of the five previous frames.

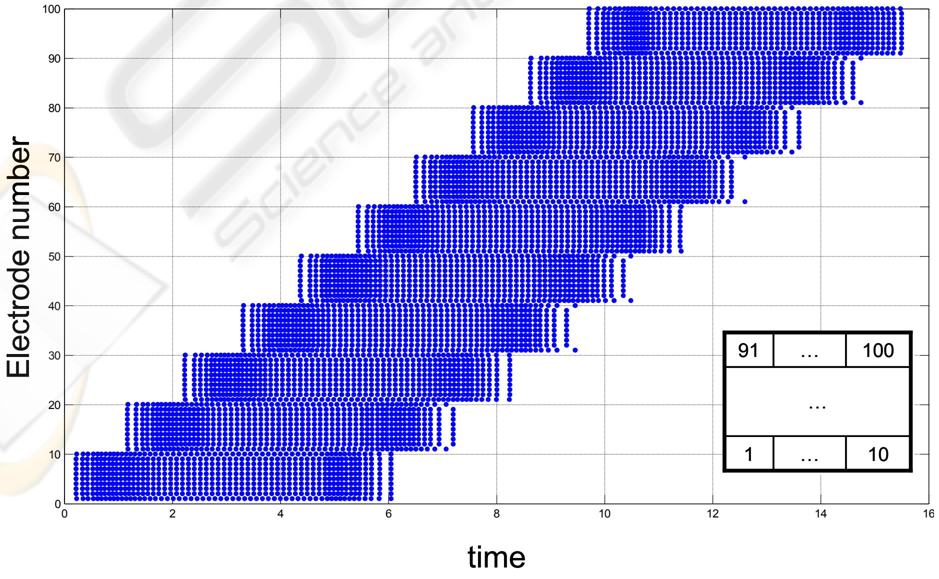

Figure 4: Spike event trains produced by a horizontal bar pattern moving from bottom to top of the image, and illustration of stimulation channels numbering (see text for details).

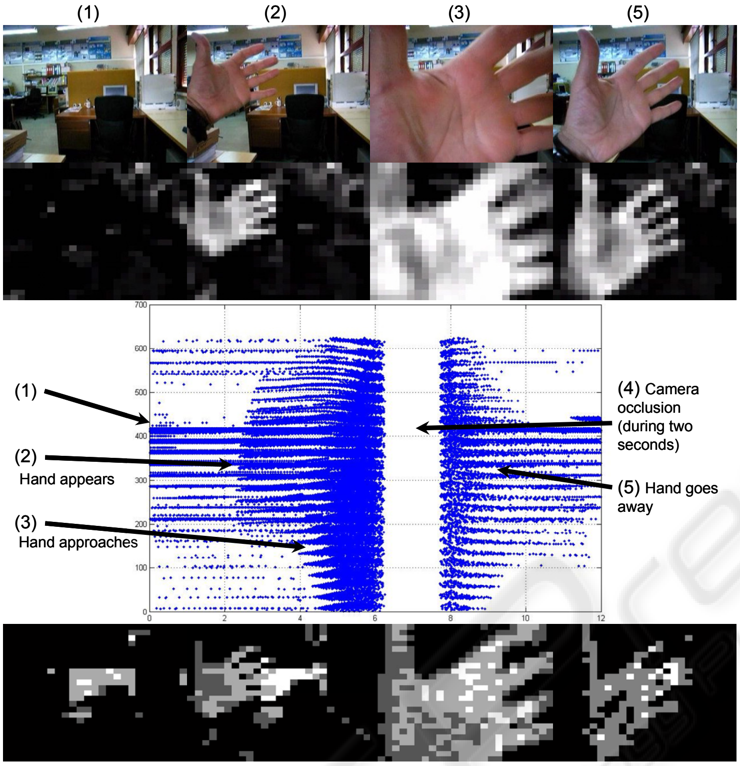

Figure 5: Image processing and coding example, including the inverse activity restoration stage. First row shows four instants of a video sequence. In the second row we can see the corresponding activity matrix obtained with a certain spatial filtering combination. Graphics shows the representation of the spike events produced by the image sequence, and finally, the bottom row represents the reconstruction of each activity matrix.

In order to test the effectiveness of this information coding method, we have developed a procedure for restoring activity matrix values from the temporal sequence of spike events (i.e. an inverse spike to activity conversion). Each spike produces an increment of the accumulated value of the corresponding activity matrix component, while a leakage factor is applied every simulation time step. A visual comparison between original and restored activity matrices was done, reporting successful results. However, a better evaluation method was implemented. Restored activity matrix was applied again to the neuromorphic pulse coding stage, producing very precise results with almost imperceptible differences. Results obtained from the restoring stage are illustrated in figure 5

4 GENERATION OF ACOUSTIC SIGNALS

As we have mentioned above, an object detected by the image processing stage can be encoded by a sound that will represent the position in which it has been detected (see figure 1).

We will represent the position of an element in the visual space by means of a sound pattern coming (apparently) from the actual spatial location of that element. This location is determined by three parameters (see figure 6): straight-line distance between the observer and the object, *d*, elevation of the object over the horizontal plane containing the head, *e*, and azimuth or horizontal angle between the front and the sides of the head, *a*.

The mechanism of spatial sound location carried out by the binaural biological system is highly dependent on the individual, making difficult the set up of an artificial system for universal filtering (Algazi, 2001). As a first approach, we have made use of the results obtained by (Gardner, 1994) in order to create sounds including spatial localization information. In their study, they placed a KEMAR (Knowles Electronics Manikin for Acoustic Research) model inside a soundproof cabin. Then, the authors played pseudo-random sound stimuli, and measured the response at the input of each of the pinnae.

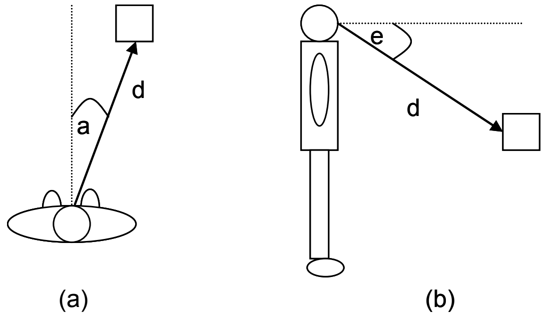

Figure 6: Basic parameters for the 3D location of sounds. See text for more details.

This way, they obtained an ample set of measurements of the HRTF (Head-Related Transfer Functions) which model the physical and mechanical features of the head acoustic system. These functions are described as a set of multiple pairs of coefficient sets of a FIR filter, one for each spatial location and auditory channel.

This technique is expressed in (5):

$$y(t)=\sum_{i=1}^{N} h(i)\cdot x(t-i) \qquad (5)$$

where h is the set of coefficients of the FIR filter, x represents the samples of base sound to which we want to add spatial information, and y is the resulting sound.

5 REAL-TIME HARDWARE IMPLEMENTATION

This section describes the implementation of our system into a programmable circuit. This choice is based on the customization needs of the application. The kind of processing to be carried out depends strongly on the specific features of the disability of the patient, which also can vary with the evolution of the illness, so the system requires being able to adapt its configuration to those changes.

Furthermore, the systems that are based on reconfigurable logic chips (FPGA) present some other features that make them suitable for this field of development, as the short time required to obtain a working prototype, its small size, allowing for portability and the integration of some other interfacing circuitry.

The description of the different modules has been written in Handel-C language, from (Celoxica, 2007), within its DK synthesis environment.

The prototyping platform selected for our tests is the Celoxica's RC300 board, which incorporates a 6 million gates FPGA and all the peripherals required for our application, as a dual video capture system to grab input images for each visual channel, VGA video outputs, and specific circuitry to obtain stereo audio. Figure 7 shows the experimental setup based on the RC300.

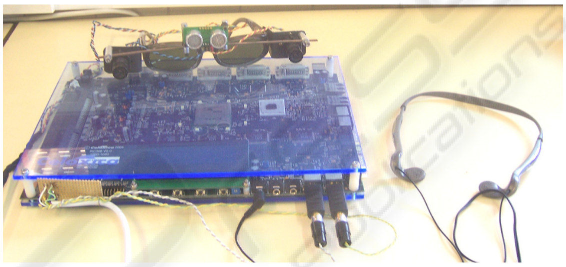

Figure 7: Experimental hardware prototype composed by two cameras, an ultrasonic range finder, headphones and a RC300 board.

This kind of devices let us implement a high degree of parallelism, so that most of the modules can process in parallel. The design has been made to exploit this capability, and a pipelined architecture has been implemented, with a high number of stages that operate concurrently.

Figure 8 shows the schematic organization of the building blocks for the image processing subsystem.

The combination of an image grabbing process to store the frame, and another process to read this information feeds constantly the computational pipeline, and achieves an uncoupling between the image capture rate and the processing carried out by the rest of the system. The information read from the memory banks is delivered to the spatial filtering module, which performs the convolution over the input images with different masks. The outputs from this stage are put together by a weighting module. The results obtained with this module will be used to perform the receptive field based mapping, in which the mean value of all the pixels in the contributing zone is stored for every point of the activity matrix.

Once the activity levels are computed for every zone of the image, the maximum values are identified, as they indicate the presence of the most relevant objects in the scene, which need to be reported to the patient. The user can select the amount of information that he/she is receiving through this system by varying the number of different zones (K) of the image that he/she prefers to be reported about. This means that we will generate K audible patterns, modulating each of them to include information regarding the location of the image from which it has been extracted, so the patient can perceive their origin.

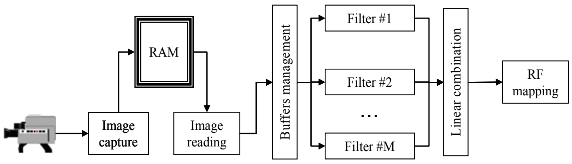

Figure 8: Architecture for the implementation of the image processing system in a Celoxica RC300 Board. The output of this common sub-system can be used for sensorial transduction, low-vision enhancing, or neurostimulation, as depicted in figure 1.

Although the working frequency for the global system is not very high, about 40 MHz, the performance-oriented design architecture allows reaching a 60 fps rate of processing, more than enough to consider the system is working in real time.

6 CONCLUSIONS

Image processing is a key stage for any device conceived to provide an aid to visually-impaired persons. We present a system that incorporates a bioinspired vision preprocessing stage which selects the most relevant objects in a visual scene to perform later processing that can be applied to different impairments. When this later translation is encoded into a stream of events for electrode addresses, the system can be applied for a visual neuroprosthesis. If we perform a sensorial transduction, the results can be translated into sound patterns, providing 3D binaural information related to the location of obstacles in the visual field. In any case, the system is highly flexible and parametric, and can be synthesized to fit into a portable, restricted power consumption board, which is suitable for a wearable aid. Our system is able, as described, of integrating different aspects of the image, as depth, colour and luminance contrast, and temporal changes detection.

We show some results on how the image analysis is performed for a variety of tuneable aspects, and specific data related to the synthesis of the processing scheme on a FPGA.

ACKNOWLEDGEMENTS

This work has been supported by the National Spanish Grants DPI-2004-07032 and IMSERSO-150/06, and by the Junta de Andalucía Project: P06-TIC-02007.

REFERENCES

Algazi, V.R., Duda, R.O., Thompson, D.M., Avedano, C., 2001. 'The CIPIC HRTF Database', *2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pp. 99-102.

Celoxica. http://www.celoxica.com [2007].

Cortivis website. http://cortivis.umh.es [2002].

Eckmiller, R., Hünermann, R., Becker, M., 1999. Exploration of a dialog-based tunable retina encoder for retina implants. *Neurocomputing* 26-27: 1005-1011.

Fernández, E., Pelayo, F., Romero, S., Bongard, M., Marin, C., Alfaro, A., Merabet, L., 2005. Development of a cortical visual neuroprosthesis for the blind: the relevance of neuroplasticity. *J. Neural Eng.* 2: R1–R12.

Gardner, B., Martin, K., 1994. HRTF Measurements of a KEMAR Dummy-Head Microphone, *Media Lab Perceptual Computing Technical Report* #280.

Gerstner, W. and Kistler, W., 2002. *Spiking Neuron Models*, Cambridge: Cambridge University Press.

JORDY, Enhanced Vision. [2007]. http://www.enhancedvision.com

Koch, C., Torre, V. and Poggio, T., 1986. Computations in the vertebrate retina: motion discrimination, gain enhancement and differentiation. *Trends in Neuroscience* 9: 204-211.

LVES, University John Hopkins, Baltimore in collaboration with NASA. [1994] http://www.hopkinsmedicine.org/press/1994/JUNE/199421.HTM

Mathworks website, The. [2007] http://www.mathworks.com.

Morillas, C., Romero, S., Martínez, A., Pelayo, F., Reyneri, L., Bongard M., Fernández, E., 2007. A Neuroengineering suite of Computational Tools for Visual Prostheses. *Neurocomputing* 70(16-18): 2817-2827.

Romero, S., Morillas, C., Martínez, A., Pelayo, F., Fernández, E., 2005. A Research Platform for Visual Neuroprostheses. In *SICO 2005, Simposio de Inteligencia Computacional*, pp. 357-362.

Victor, J., 1999. Temporal aspects of neural coding in the retina and lateral geniculate. *Network* 10(4): 1-66.