CONCEPTION OF A PHOSPHENE BASED VISUAL SYSTEM

A Functionnal Approach

Guillaume Tatur, Isabelle Marc
Équipe STINIM LGI2P, Ecole des Mines d'Alès, Site Eerie Parc Scientifique et Technique Georges Besse
69 rue Georges Besse, 30035 Nimes cedex 1, France

Michel Dumas
Équipe STINIM, Institut d'Electronique du Sud, Montpellier, France

Gerard Dupeyron
Institut ARAMAV, Nimes, France

Keywords: Visual prosthesis, Mobility, Man–machine interaction (MMI), Traversability analysis, Prosthetic vision simulation.

Abstract: This work falls within the broader framework of visual prostheses conception for blind people suffering from a retina photoreceptor cells degenerative disease. Studies on the issue of informational content of the prosthetic vision propose, in majority, a simple reduction in the resolution of greyscale images from a single camera. Our work proposes a novel approach based on functional vision. This functional vision is dedicated to specific needs in mobility, and is based on a simple analysis and less ambiguous representation of the scene layout. Emphasis is placed on the usefulness of providing access to 3D information combined with the use of an eye tracking device that should greatly improve spatial knowledge acquisition. This work presents extraction and proper formatting of data from optical sensors in order to establish a coherent vision based on the generation of a limited number of phosphenes.

1 INTRODUCTION

Prosthetic vision is currently one of the most promising fields of research in the domain of therapies for degenerative retinal diseases like Age related Macular Degeneration (AMD) or Retinitis Pigmentosa (RP). All of the works around the world are based on the same concept: bypass the destroyed outer retinal layer and use information extracted from camera images for direct stimulation of inner retinal neurons, optic nerve or visual cortex. Electrical stimulation by current injection in electrode array is known for a long time to induce discrete visual sensations like spots of light, called phosphenes. Research is underway to define the better place for the stimulation, the most efficient electrode geometry and current pulse pattern. Our work is devoted to another question: what is the minimal information which has to be sent to the stimulator?

One can define a two-step approach: information selection, which is obtained by image processing, and information representation, which is the method used to link visual objects and visual perception. The former is not so intensively studied, although the latter is widely discussed. However, even the most recently designed stimulation implants exhibit a very limited number of electrodes compared to the number of fibres involved in the natural vision process. In consequence, the amount of data that can be sent at a moment to the stimulated neurons appears to be extremely poor. Preliminary image analysis may then increase the relevance of stimulation data.

This paper presents our attempts to define the concept of "functional vision by phosphenes", which is in relation with both information selection and representation steps, as previously defined. Image processing is directly inspired by collaboration with rehabilitation specialists. The final objective is to

Tatur G., Marc I., Dumas M. and Dupeyron G. (2010).
CONCEPTION OF A PHOSPHENE BASED VISUAL SYSTEM - A Functionnal Approach.
In *Proceedings of the Third International Conference on Bio-inspired Systems and Signal Processing*, pages 339-344
DOI: 10.5220/0002758103390344

provide the subject with meaningful visual cues of the environment, based on the perception of phosphenes. This functional vision should restore autonomy in day to day tasks of mobility in known or unknown environment, in contrast with an acuity based vision (supposed to restore the ability to read) proposed by a majority of research teams in the domain.

2 PROSTHETIC VISION

When some elements of the chain of perception (the photoreceptor cells in the case of RP or AMD) are no longer operative, it is potentially possible to bypass them and stimulate the optical pathways downstream, in order to restore a kind of rough visual perception. Fundamental element that composes this perception is called phosphenes. Phosphenes are entoptic phenomena characterized by the sensation of perceiving light in the form of dots or bright areas, caused by mechanical, electrical or magnetic stimulation of the optical pathways. Up to now, visual prostheses are based on electrical stimulation through microelectrodes. Several types of implants are under test; the different kinds of implant are: Cortical implants (Fernandez, 2004), sub-retinal implants (Chow, 2002), optic nerve electrodes (Delbeke, 2001) and epi-retinal implant (Humayun, 2003).

According to various clinical trials, the generated phosphenes have some properties that it is possible to set according to stimulation parameters (Dagnelie, 2001): the position in the visual field appears to rely on the position of excitation electrode, the size depends on the number of cells recruited by the stimulation, and the brightness, depending on the intensity and repetition frequency of stimulation.

2.1 Simulations of Prosthetic Vision

Experimental procedures with sighted people and pixelised vision simulator are the more convenient way to test methods for the delivery of information to the stimulator, even though these tests are far from the real situation of blind patients with visual prosthesis.

Studies of Cha et al (Cha, 1992) are often cited as a reference for either performance in acuity or in navigation following the characteristics of the vision provided. The conclusion is that the best compromise between FOV size, number of pixels and performances in mobility tasks is obtained for a 30 ° FOV and an image of 25x25 pixels uniformly spaced, captured by a simple camera. These results are similar to those found in (Sommerhalder, 2006).
In these previously mentioned papers, information representation is based on gray level dots, with square or circular shapes, which are mapped on square or hexagonal lattices. Information is processed from a gray level image. The camera resolution is much higher than the possible number of electrodes for implants currently in progress: 4*4 electrodes for the Argus I epiretinal implant and 60 electrodes for the second-generation “Argus II” (Dagnelie, 2007). Selection of information data is then required. The common choice is to reduce the resolution of the entire image: split it into blocks of values equal to the average gray level of pixels that they contain. Light intensity for each dot is then driven by the value of the corresponding block (see fig.1). In (Chen, 2009), different others methods are compared. Gaussian filtering of the corresponding visual field of the phosphene appears to induce slightly more efficient perception. Improvement carried on by edge detection is also studied.

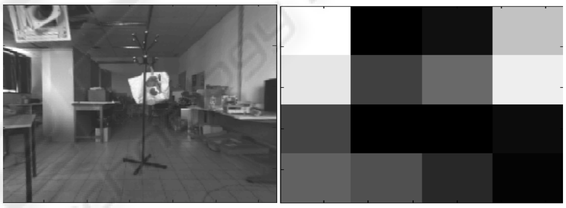

Figure 1: Left: Input greyscale image. Right: prosthetic vision simulation (4x4 matrix, regional averaging technique: mean gray values of the corresponding block, 256 gray levels, contrast: 100%).

2.2 First Clinical Test Results

First clinical trials give the possibility to directly test the methods used to generate phosphene. In (Brelen, 2005), four cuff electrodes are wrapped around the optic nerve. At the first step of experimentation, reproducible cartography of the perceived phosphenes is established in relation with the stimulation parameters. Recently, in (Margalit, 2002), (Yanai, 2007), (Humayun, 2003) are described the concepts, tests and results in some tasks of vision (discrimination of the direction of movement, forms, object detection) performed by implanted patients.

In all these experiments, implants are composed of very few electrodes. It results in a reduction of resolution that makes information highly ambiguous; however it allows the detection of light sources and of very large and contrasted objects. Furthermore, mobility tasks are not intensively studied.

3 FUNCTIONAL VISION BY PHOSPHENES

First results in prosthetic vision fields raise considerable interest and hope in the population concerned by blindness. Nevertheless, the sparse visual perceptions provided to the first implanted patients are far from being really useful in their everyday life. We believe that emphasis should be placed on works dealing with the first stage, selection of information by image analysis.

Transmission of ambiguous information of the environment as proposed by the approaches based on the reduction of the image resolution leads to two main problems: first the cognitive load of the subject increases as he has to decode his perception, and secondly these perceptions can potentially lead to misinterpretations, which can be dangerous in the context of mobility. Moreover, in all cases, the entire field of view of the camera is transmitted; it is not possible for the subject, like in natural vision, to precisely focus his attention on a particular area of the observed scene. To explore the scene, it is necessary to change the position of the camera and thus the head (in the case of an acquisition platform embedded in a pair of glasses for example). It is a long lasting and tiring action that does not necessarily provide a better understanding of the scene since each pixel captures the variations in light reflected by the surrounding surfaces and objects. Then, stable perception can hardly be generated and the implant recipient will experience difficulties to quickly extract information about the geometrical structure of the scene or its constituent entities during mobility tasks.

The input data of our system are provided by cameras, as for the other research projects. But some significant differences exist, which will be discussed in the following. First, our objective is to increase quality of life for blind people by increasing their autonomy in mobility tasks. Second, selection of relevant information is performed via a 3D image analysis, using a stereoscopic pair, and it is enriched with some additional data provided by an inertial measurement unit and an eye tracking system. Finally, the complete process of information selection and representation is designed in collaboration with functional rehabilitation specialists (psychomotor therapists, orthoptists...). All previously mentioned experiment results emphasize the importance of the learning and self adaptation to the extremely impoverished information provided by an implant. Rehabilitation strategies have to be established in parallel with image analysis.

3.1 Functional Approach of Prosthetic Vision: Geometrical Scene Analysis

Acuity tasks (recognition of characters or small objects with simple shape) are clearly not in the scope of this study. Instead, the goal is to give the possibility to detect visual cues for mobility purpose (detecting free space, anticipate and avoid static or mobile obstacles, maintaining the body balance ....) and orientation (recognition of places, landmarks...). Knowledge of ego-centred distances between the subject and the obstacles is the basis for safe navigation. It is expected that this information of distance may be encoded in a simple way (stimulus intensity proportional to the obstacle proximity, for example) in order to be interpreted quickly and unambiguously. A vision based on this notion of distance would require a minimum time of interpretation (as opposed to the understanding of a gray scale image, which is a very complex task) in the frame of mobility tasks.

3.2 Gaze Direction and Construction of Mental Map

As it is pointed out in (Berthoz, 2003), perceiving is choosing, in all the available information, those that are relevant related to the planned action. A possible way to select this information is to measure the gaze direction of the subject in real-time.

There are several advantages in the use of gaze direction: interaction between the implanted patient and its prosthesis is improved, which in turn facilitates his construction of mental representation of the surroundings based on rare spots of light. The human visual system uses gaze direction and saccades to explore the scene (Henderson, 2003). Therefore, it seems more natural to take into account the location where the subject wants to focus his attention and enable the system to provide more accurate local information. Moreover, some studies suggest that for the acquisition of spatial knowledge, the eye position is used to encode the perceived space in a more abstract schematic commonly called mental map, which is independent of anatomical coordinates (Andersen, 1985). Moreover, the role of eye movements in postural balance (Schulmann, 1987) could be exploited. Finally, it offers the opportunity to use all the available points of stimulation to describe the environment in the observed area, and not in the whole camera field of view.

We would like to point out that while oculomotor movement of profoundly visually impaired or almost blind people are severely reduced, it is still possible to restore this ocular control by means of training with rehabilitation specialist. Thus, restoring recently blind ocular mobility for scene exploration is conceivable.

3.3 Information Representation

At this point of the study, no stimulation device is chosen for the implantation, but some assumptions can be made, which rely on the literature data: a hundred of points of stimulation are available, maybe less, and some phosphene characteristics may be modulated : intensity and/or size.

As described above, in the standard approach, the entire image is send through the few points of stimulation the device is able to generate. Trivial observation of human eye processing may lead to a more efficient mode of data transmission: human vision involves two different mechanisms, a foveal vision with high acuity in the central FOV, and a peripheral field vision dedicated to alert and perception of motion.

To be consistent with human visual perception, we break down the image representation into two zones: a "central" zone providing local description (narrow visual FOV, high density of phosphenes) that facilitates the perception of 3D contrasts. In addition, in a "peripheral" zone, the data will be even more specific in order to provide information from a wider field of view and to give the possibility of a faster interpretation of the perceptions. The corresponding points of stimulation will be used for obstacle or movement detection purpose.

4 GEOMETRICAL SCENE ANALYSIS

Restoring functional vision for blind people requires transmission of information which is both unambiguous and specific for mobility tasks. It may be noted that the ambiguity decreases as information becomes more specific. 3D contrasts and obstacle detection appear to provide data more invariant and much less ambiguous than simple mean of gray levels of pixels. This assumption has to be validated through tests using simulation of vision by phosphenes. The early stages of data selection are presented in the following.

4.1 Experimental Setup

Images are acquired from a stereoscopic device (STH- MDCS3 from Videre Design). Output data are dense disparity maps (images in which output pixel values are proportional to the depth of the scene) at real time. Axe of gravity is gained from an accelerometer (Analog Devices ADXL330). Eye tracking is performed by ViewPoint EyeTracker GigE-60 from Arrington Research, Inc.

4.2 Central and Peripheral Field Analysis

The camera FOV is decomposed into two areas, central and peripheral (Figure 2).

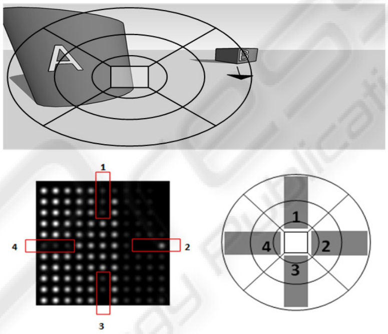

Figure 2: Schematic summary of our scene analysis paradigm and its representation. Upper input image is probed by a specific mask. Bottom left image represents simulated phosphenes. Peripheral information is analysed through angular sectors surrounding the central area. Information in each sector controls the corresponding phosphene appearance (related phosphene groups outlined in the bottom left image). The central area (square region in the bottom right and upper image) directly probes the scene depth with a high number of phosphenes.

The characteristics of these two areas, in terms of involved camera FOV and phosphene density, will be further précised by experimentation. Preliminary assumptions are the followings: in the central zone, where 3D depth data will be directly used as stimulation data, it is assumed that the spatial resolution is fine enough to let the subject capable to detect free space as wide as a doorway, at a distance of 5 meters, in his gaze direction. On the other hand, the number of phosphenes is large enough to visually cover a ground projected area of the size of a standing person at 2 meters. In the peripheral field, the choice is made to transmit not directly the 3D coordinates, but only the presence and location of obstacles. The scene is then divided into ground and objects above the ground (obstacles). Only the distance to the nearest objects will be transmitted to the stimulator. Additionally, in a more bio-inspired approach, it will be interesting to test the use of

Time-to-collision (TTC) (Bruce, 1993) cues. As the purpose of this peripheral vision is to provide alert signals, objects situated at the same distance from the user should have different visual weight depending on their relative motion.

Obstacle detection processing is commonly based on ground plan detection. “V-Disparity” (Zhao, 2007), (De Cubber, 2009) (figure 3) is one of such processing currently used for object/ ground segmentation. This algorithm analyzes the disparity image (obtained by finding the corresponding points between the two stereoscopic images) and locates the linear change in the depth along the image line axis that is characteristic of a plan, the ground in our case. In the following segmentation step, image regions that do not follow the specific depth variation of the ground plan are regarded as obstacles (i.e. depth values above or under specific thresholds for a given image line will be declared to be part of an obstacle).

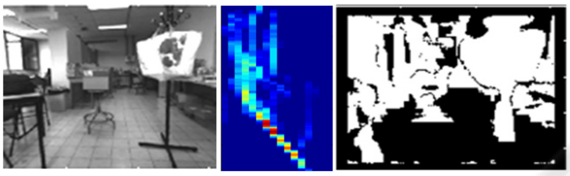

Figure 3: Left: input image. Centre: V-Disparity image. Right: Obstacle segmentation output (white=obstacle, dark=ground or no 3D data available).

We select “V-Disparity” considering the fact that real time processing is needed and because of its relative simplicity despite its good performance. However, the V-Disparity algorithm is valid only if the scene is observed in such a manner that image lines are approximately parallel to the lines of equal depth in the range image. In prosthetic vision application, when the cameras are head worn, this condition would restrain the user’s posture in order to achieve best performance. The algorithm has to be made invariant under rotation along the depth axis. This is done by analysing images along the projected gravity vector, through the use of an accelerometer (figure 4).

Besides, the segmentation by means of this algorithm is not controlled: i.e there is no metric that evaluates the accuracy, and if the detected plan is more likely to be the true ground plan or some artefact. A method of verification has to be defined. We chose to use the RANSAC algorithm (Fischler, 1981) to extract the parameters of the detected ground from the V-Disparity outputs. As it works on a reduced subset of 3D data points, this verification should use fewer CPU time and memory than if we try to locate the ground plane only by the RANSAC method.

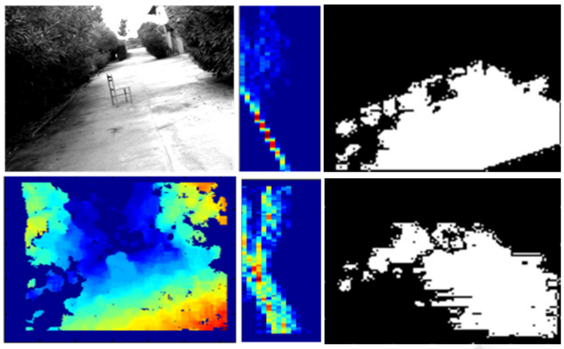

Figure 4: Upper left: Input greyscale image. Bottom left: disparity map. Upper right: raw output of the ground segmentation process with accelerometer. Bottom right: output of the ground segmentation process without accelerometer (white=ground). Corresponding V-Disparity images in the middle column.

RANSAC offers the advantage of handling noisy data. Moreover, the estimation is valid if at least 50% of the data follow the model sought. By RANSAC, we look for the plan that best fit the 3D points provides by the V-Disparity algorithm, according to the least squares method. If the parameters belong to a domain of acceptable values (given the angle formed by the camera in relation to gravity), the segmentation is accepted. Otherwise a virtual plane is used; its normal vector is updated using the inertial navigation system. Furthermore, by refining the segmentation obtained via V-Disparity by the supposedly more precise data from RANSAC, we should be able to provide a more accurate segmentation.

5 CONCLUSIONS

Methods used so far for information extraction and transmission via prosthetic devices do not respond adequately to the expectations of blind people. A novel approach based on information’s relevance, needed in a context of mobility, is described. It directly results of interactions between the world of functional rehabilitation (cooperation with rehabilitation specialist) and the world of mobile robotics, with attention to the recent advances in exploration and traversability analysis. Perception based on a geometric description of the scene coupled with an eye tracking device appears to be satisfactory. First step in the information selection is the obstacle detection. It is performed by image processing methods defined up to now for robotic applications and somehow improved for a better matching with prosthetic vision application. 3D information is then transmitted in two ways for different purposes: a central area for a more acute and local analysis of the geometrical layout of the observed scene, and a peripheral area for obstacles proximity and motion detection in a wider field of view. We will make use of an eye tracking device in order to enhance communication between the prosthesis and the wearer and to transmit more significant and useful information in relation with the user's intention and current action.

In order to validate this functional vision approach, experiments are scheduled with both sighted subjects and visually impaired subjects, provided with pixelised vision simulator. Results will be discussed elsewhere.

REFERENCES

Andersen R.A., Essick G.K., Siegel R.M, 1985. Encoding of spatial location by posterior parietal neurons, In *Science*, 230, 456-458.

Brelén M. E., Duret F., Gérard B., Delbeke J., Veraart C., 2005. Creating a meaningful visual perception in blind volunteers by optic nerve stimulation, In *J. Neural Eng.*, 22-28.

Bruce V., Green P., 1983. *La perception visuelle : physiologie, psychologie et écologie*, Presse Universitaire de Grenoble.

Cha K., Horch K.W., Normann R.A., 1992. Mobility performance with a pixelized vision system, In *Vision research*, vol. 32, 1367-1372.

Chen S. C., Suaning G. J., Morley J. W., Lovell N. H., 2009. Simulating prosthetic vision: I. Visual models of phosphenes, In *Vision research*, Vol. 49, 1493-1506.

Chow A. Y., Pardue M. T., Perlman J. I., Ball S. L., Chow V. Y., Hetling J. R., Peyman G. A., Liang C., Stubbs E. B. Jr , Peachey N. S., 2002. Subretinal implantation of semiconductor-based photodiodes: durability of novel implant designs. In *J. Rehabil. Res. Dev.* May-Jun; 39(3): 313-21.

Dagnelie G., 2001. Prosthetic vision: exploring its qualities and limits by means of simulations», In *Visual impairment and rehabilitation*, vol. IV. Montréal University.

Dagnelie G., Keane P., Narla V., Yang L., Weiland J., Humayun M., 2007. Real and virtual mobility performance in simulated prosthetic vision, in *J. Neural Eng.*, 92-101.

Delbeke J., Pins D., Michaux G., Wanet-Defalque MC.,. Parrini S. Veraart C., 2001. Electrical Stimulation of Anterior Visual Pathways in Retinitis Pigmentosa, In *Investigative ophthalmology & visual science*, vol 42, 291-297

De Cubber G., Doroftei D., Nalpantidis L., Sirakoulis G. Ch., Gasteratos A., 2009. Stereo-based Terrain Traversability Analysis for Robot Navigation.

Fernández E., Ahnelt P., Rabischong P., Botella C, Garcia-de Quiros F., Bonomini P., Marin C.,

Climent R., Tormos J.M., Normann R.A., 2002. Towards a cortical visual neuroprosthesis for the blind, In *EMBEC'02. IFMBE Proc* 2002 3(2).

Fischler M. A., Bolles R. C.,1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, In *Commun. ACM*, Vol. 24, No. 6. 381-395.

Henderson J. M., 2003. Human gaze control during real-world scene perception, In *Trends in Cognitive Sciences*, Volume 7, Issue 11,498-504.

Humayun M. S., Weiland J. D., Fujii G. Y., Greenberg R., Williamson R., Little J., Mech B., Cimmarusti V., Van Boemel G., Dagnelie G., De Juan E. Jr., 2003. Visual perception in a blind subject with a chronic microelectronic retinal prosthesis, In *Vision Research*, Volume 43, Issue 24, 2573-2581.

Margalit E., Maia M., Weiland J., Greenberg R., Fujii G., Torres G., Piyathaisere D., O'Hearn T., Liu W., Lazzi G. , 2002. Retinal Prosthesis for the Blind In *Survey of Ophthalmology*, Volume 47, Issue 4, 335-356.

Schulmann DL., Godfrey B., Fisher AG.,1987. Effect of eye movements on dynamic equilibrium, In *Phys. Ther.*
Vol. 67, No. 7, pp. 1054-1057.

Sommerhalder J.R., Perez Fornos A., Chanderli K., Colin L., Schaer X., Mauler F., Safran A.B. and Pelizzone M., 2006. Minimum requirements for mobility in unpredictible environments, In *Invest Ophthalmol Vis. Sci.*, vol 47.

Yanai D., Weiland J., Mahadevappa M., Greenberg R., Fine I., Humayun M., 2007. Visual performance using a retinal prosthesis in three subjects with retinitis pigmentosa, In *American Journal of Ophthalmology*, Volume 143, Issue 5, 820-827.

Zhao J., Katupitiya J., Ward J., 2007. Global correlation based ground plane estimation using v-disparity image, In *IEEE International Conference on Robotics and Automation*, Roma, Italy, 10-14.