Citation: J. Bermudez-Cameo, A. Badias-Herbera, M. Guerrero-Viu, G. López-Nicolás, J. J. Guerrero. RGB-D computer vision techniques for simulated prosthetic vision. Pattern Recognition and Image Analysis (IbPRIA), pp. 427-436, Faro, Portugal, June 20-23, 2017. Lecture Notes in Computer Science, vol 10255. Springer, Cham.

RGB-D Computer Vision Techniques for Simulated Prosthetic Vision

Jesus Bermudez-Cameo, Alberto Badias-Herbera , Manuel Guerrero-Viu, Gonzalo Lopez-Nicolas, and Jose J. Guerrero

Instituto de Investigación en Ingeniería de Aragón, I3A, Universidad de Zaragoza

Abstract. Recent research on visual prosthesis demonstrates the possibility of providing visual perception to people with certain blindness. Bypassing the damaged part of the visual path, electrical stimulation provokes spot percepts known as phosphenes. Due to physiological and technological limitations the information received by patients has very low resolution and reduced dynamic range. In this context, the inclusion of new computer vision techniques to improve the semantic content in this information channel is an active and open key topic. In this paper, we present a system for Simulated Prosthetic Vision based on a head-mounted display with an RGB-D camera, and two tools, one focused on human interaction and the other oriented to navigation, exploring different proposals of phosphenic representations.

Keywords: simulated prosthetic vision, head-mounted displays, RGB-D vision

1 Introduction

A novel approach for treating blindness caused by retinal degenerative disorders is implanting retinal prosthesis. Cell degeneration of photoreceptors caused by retinitis pigmentosa and macular degeneration can be bypassed by artificially stimulating non-damaged cells like retinal ganglion cells and sometimes bipolar cells [36]. This effect is achieved with an implanted electrode array that provokes a set of electrical stimuli which is perceived by the blind patient as a pattern of visual spots, usually known as phosphenes [10].

The research in last years reports significant advances in the development of visual prostheses. There exist different types of visual prostheses according to the problem causing blindness: retinal prostheses, optic nerve prostheses or direct stimulation of the cortex. If the patient suffers from damaged retina receptors (cones and rods) any of the three classes is adequate. But if the damage is caused in an advanced part of the optic pathway, the visual cortex may be the most appropriate place. In particular, retinal prostheses are based in retinal ganglion cells stimulation. This may be achieved via placement of epiretinal, subretinal or suprachoroidal stimulating electrode arrays. Five representative models are Argus II, Boston Retinal Implant Project, Epi-Ret 3, Intelligent Medical Implants (IMI) and Alpha-IMS (Retina Implant AG). Currently, two

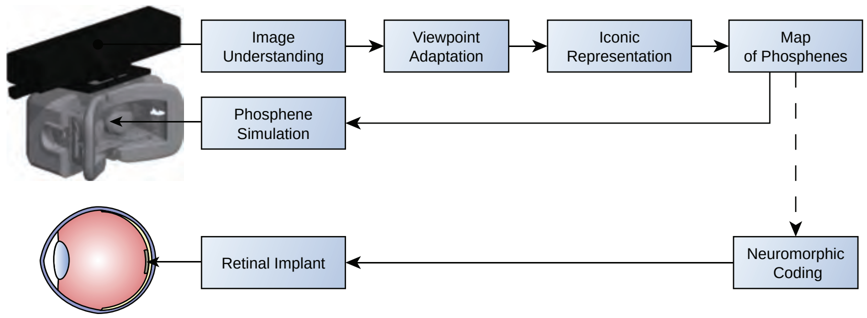


Fig. 1. Flowchart for simulated prosthetic vision. (top). SPV facilitates research on new methods to generate maps of phosphenes. Next step, would be to transfer the research results to final patients (bottom).

of them have regulatory approval (America and Europe) for the treatment of retinitis pigmentosa: the Argus II (epiretinal) and Alpha IMS (subretinal), the Boston Retinal Implant is in animal studies and all others are in clinical trials. The perception system used by most of current visual prostheses, with some exceptions, is based on the acquisition of images with an external camera. The acquired images are generally processed using basic image processing techniques to generate a phosphenes-based image which is sent to the prosthesis.

Institute	Image processing
Stanford University [3]	Geometric / Spatio-temporal
Dobelle Institute (University of Utah) [11]	Contour detection
Intelligent Medical Implants [16]	Spatio-temporal filtration
Second Sight (Argus II) [28]	Contrast improvement/Contour detection
University of New South Wales, Sydney [32]	Mean/Gaussian filter, zoom

Table 1. Examples of works using basic image processing techniques performed in visual prostheses.

A simulated prosthetic vision (SPV) generally consists of a vision-based system and a head-mounted display on a normal sighted subject (see Fig. 1). This system allows representing the descriptions of phosphene perception reported by visual prosthesis patients. The advantages of using SPV in prosthetic vision research have been acknowledged [1]. In particular, this approach avoids the implications derived of treating with final patients allowing an early non-invasive evaluation of advanced computer vision techniques and different representations. Most of the current approaches used in prosthetic vision and SPV are based on basic image processing techniques (see Table 1) like in [4,5]. However this configuration (camera+prosthesis) allows exploring more advanced computer vision techniques to enhance the semantics and the relevance of the information displayed to the patient. For example, visual recognition can be used for enhancing the saliency of meaningful objects [15,17], face detection can be used for human interaction [21] and going further, the recent advances in 3D visual odometry [25,13] could be used for assisting navigation with prosthetic vision.

In this paper we present a prototype consisting of a device for simulated prosthetic vision based on RGB-D vision (Section 2), a human interaction tool (Section 3) and a navigation tool (Section 4) where different representations displaying maps of phosphenes are discussed. A key issue in our proposal is to consider depth information to enhance basic image processing with advanced computer vision techniques in the context of SPVs. Note that with current technology, depth perception is transmitted by stereo displays to non-blinded people, however, since intrinsic technical limitations of prosthetic vision prevents from transmitting the stereo effect, it requires alternative strategies to be represented such as using an iconic way.

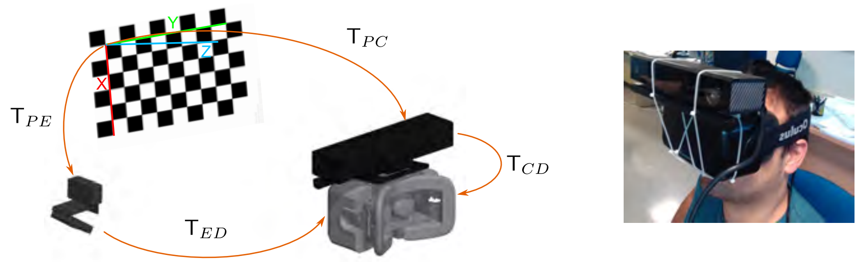
1.1 Related Work

Simulated prosthetic vision has already disclosed useful for reflecting clinical findings [1], being used for studying performance in tasks such as reading [12], finding text [9], hand-eye coordination, and mobility [29]. Most of these insights are related with the number of electrodes needed for achieving a given task. For example, according to Cha et al. [6] a pattern of 25 x 25 phosphenes allows to recognize text in a reading speed of 100 words per minute for stationary text and 170 words per minute for text moving automatically, however harder tasks like face recognition requires hundreds of phosphenes [31,33]. Despite the advances in prostheses development increase the available resolution there exist physical restrictions when miniaturizing the electrode array. In [24] the limits when increasing the number of electrodes by current spread and new strategies for improving the stimuli are discussed.

First addressed task in navigation for SPV is obstacle avoidance [2,27]. [26] presented a computationally model for detecting salient regions in an image frame to avoid obstacles. Another visual processing system for bionic eye with a focus on obstacle avoidance was implemented by [30]. Obstacle detection and simultaneous localization and mapping were applied to guide a user on a safe path using a stereo camera as input of the implanted prosthetic vision, and vibration motors on the shoulders [34]. In [22] a technique to find a ground-plane and the boundary with objects is presented. Objects and boundaries can be then augmented to ensure object boundary visibility [19]. More recently, [35] uses VR-based environments for evaluating the visual response for obstacle avoidance in SPVs with a simpler set-up.

Representing depth in SPV is a key concept for achieving assisted navigation, but its implementation is particularly challenging. Systems displaying depth and contrast edges in a phosphene-based display are described in [18,21] and more recently in [23]. In [14], a semantic labelling of the image provides a representation for obstacle avoidance.

2 RGB-D based prototype for simulated prosthetic vision



(a) (b)

Fig. 2. (a) Scheme for extrinsic calibration of the head-mounted display with respect to the RGB-D sensor. (b) Our prototype for simulating prosthetic vision.

A device for Simulated Prosthetic Vision (SPV) allows simulating the performance of a given vision algorithm without intervention of a blinded person. Our first prototype combines a head-mounted display (HMD) [1] with an RGB-D camera [2]. The use of a depth sensor allows representing images from the point of view of the user despite the camera is not exactly placed on the view-point of the user. The localization of the display with respect to the camera is indirectly estimated by using the external camera provided by Oculus. For calibrating the system this camera has to simultaneously perceive a led-pattern carried by the HMD and a chess pattern which is also viewed by the RGB-D sensor. The external camera is used only once to calibrate the system, since HMD and RGB-D sensor are coupled. The resulting Euclidean transformations are the transformation of the HMD from the external camera T_{ED}, the transformation of the external camera from the chess pattern T_{PE} and the transformation of the RGB-D sensor from the chess pattern T_{PC} (see Fig. 2). These transformations are linearly related by

$$\mathsf{T}_{CD} = {\mathsf{T}_{PC}}^{-1}\mathsf{T}_{PE}\mathsf{T}_{ED} \tag{1}$$

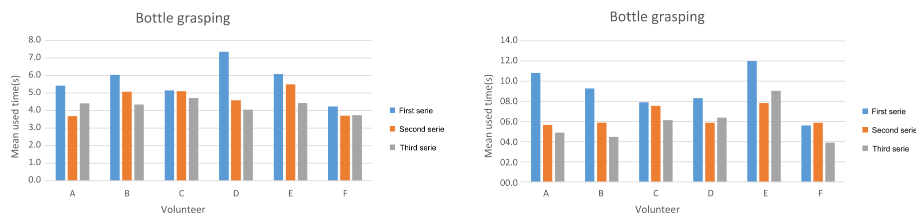
The map of phosphenes is represented using Gaussian spots with a modular configuration (e.g. [8]). This is certainly not a precise representation of the appearance of most phosphenes, but it is a standard representation used in the literature [31,7]. We also assume a circular field of view. Given a resultant processed image point, we use a look-up table for computing the corresponding phosphene. This look-up table is initialized by computing the nearest phosphene defined by the Euclidean distance. For addressing the focusing problem in very close displays and provide wider field of view, the HMD combines asferic lenses

[1] Oculus Rift www.oculus.com/rift/

[2] Microsoft Kinect V2

(a) (b) (c)

Fig. 3. Perceptual validation of calibration. (a-b) The subject must grasp a bottle and return it to its original location without blowing out the others. (c) The subject must introduce a small bottle in a basket.



(a) (b)

Fig. 4. Mean time used for each subject to grasp the bottle (8 executions per subject).

with a correcting barrel distortion. However, we can avoid computing the intrinsic calibration of the RGB-D sensor and the HMD since they are internally provided by the corresponding SDKs (we have separately calibrated both systems and checked that the internal calibrations provided by SDKs are correct). Finally we have performed a perceptual validation of the whole system representing the coloured point cloud captured by the RGB-D camera in the reference of the HMD. With this configuration, different subjects have tested the prototype by performing different tasks such as executing grasping tasks in short distances and launching tasks in medium distances. In the first evaluation we measure the time invested by a subject for grasping a bottle and return it to its original location without blowing out the others. The evaluation considers two cases: first (see Fig. 3 (a)) considers separated bottles and second considers the grasping of a bottle surrounded by other bottles (see Fig. 3 (b)). This evaluation is performed 8 times by 6 different subjects in three different series to study the human adaptation to the device (see Fig. 4). In the second evaluation we measure the success rate of introducing a small bottle in a basket located 2 meters far away from the subject (see Fig. 3 (c)). The obtained success rate is 57% in comparison with a success rate equal to 67% for the same evaluation with natural vision. Conclusions from these evaluations are that the prototype permits to correctly estimate 3D location and object dimensions once it has been calibrated. There is also a fast human adaptation to the given virtual reality prototype.

3 Iconic representation of humans for prosthetic vision

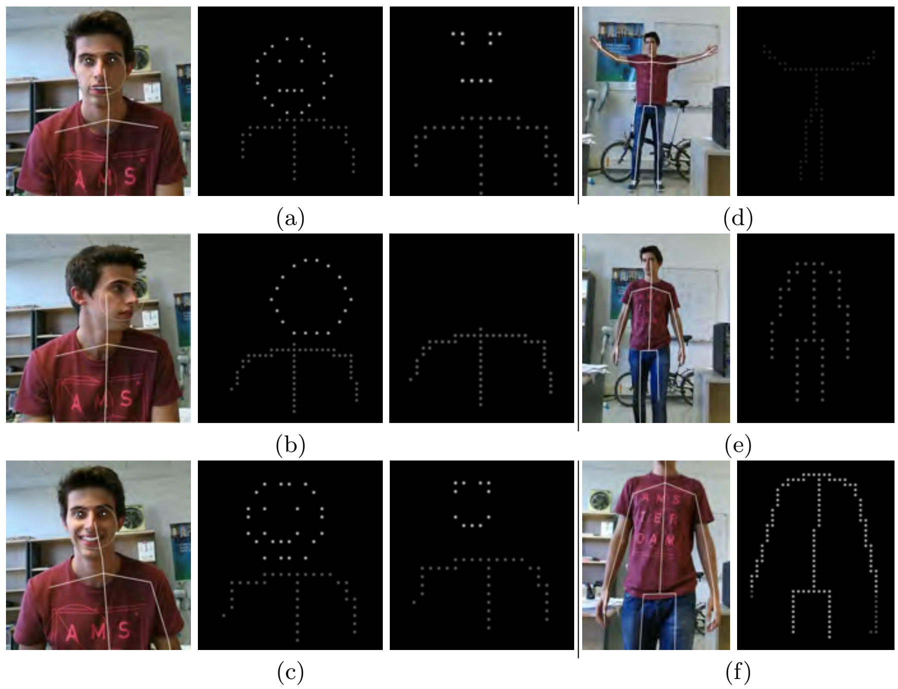
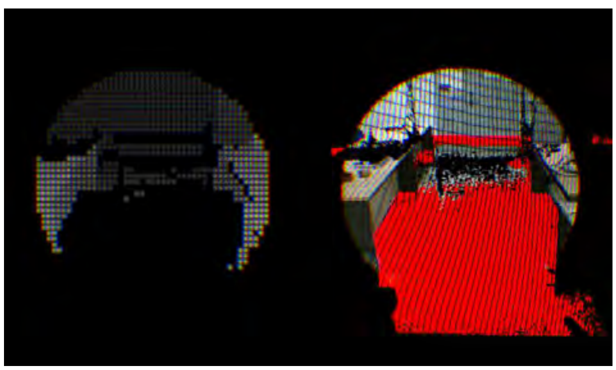


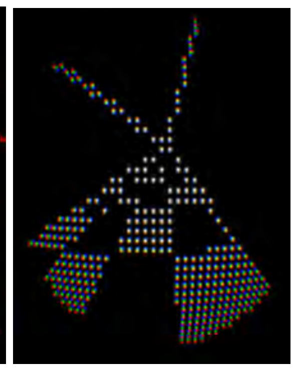
Fig. 5. Iconic representation of humans using map of phosphenes. Faces: (a) Representation for a neutral expression. (b) Representation for a face not looking at the user. (c) Representation for a happy expression. Human skeleton representation: (d) Far. (e) Middle. (f) Near.

Human interaction is very important for patients with impaired vision where identifying the presence of a person, its identity, knowing if the person is approaching or understanding their expressions is a very valuable information. A retinal prosthesis can be used to enhance the human interaction experience. However, the low resolution and reduced dynamic range of the phosphenic video stream can impede an adequate understanding of the scene. In fact, human detection and recognition are difficult tasks when having any kind of impaired vision. An interesting approach here is using computer vision techniques to perform this recognition on the original images and to present an iconic representation to the patient [18,21]. In this work, we exploit that the RGB-D camera provides human detection describing skeletons and face parameters and we use this information for an iconic representation of the person and its face. In particular, Kinect v2 API provides 3D position of human joints, eyes and mouth location, face orientation in 3D and detects other binary features like expressions. In Fig. 5 we show iconic representations of a subject using a map of phosphenes. The RGB-D sensor allows estimating the location of the person and the expression which is represented with a variation in the iconic representation. We also detect and represent the gaze direction. We have evaluated different representations taking into account the available resolution of the simulated prosthesis. In particular, we have evaluated two representations in three different resolutions. One representation involves eyes and mouth (Fig. 5 (a-c) right) and the other also includes a circle for representing the face (Fig. 5 (a-c) middle). In Fig. 5 (d-f) we show a detail of the iconic representation of the skeleton of a person. Depending on the depth we codify the range of the phosphenes. We also attach an example in the available video [1].

4 Depth for navigation using prosthetic vision



(a)



(b)

Fig. 6. (a) Ground removal from depth information for a better saliency on 3D scenario with SPV. (b) Iconic representation of a corridor using chess floor and vanishing points for showing a direction.

Assisted navigation for prosthetic vision can take advantage of the last advances in 3D localization using computer vision and RGB-D sensors [13,25]. The location of the patient in combination with a map which can be simultaneously estimated or previously stored is valuable information that can guide the patient in day to day tasks but also in emergency evacuations.

Even when the complete environment is known, transmitting depth sensation is hard due to low resolution and low dynamic range of prosthetic visual devices. Notice that stereo vision is not possible since only monocular prostheses are possible. A straight-forward approach is encoding the depth, which can be estimated with the RGB-D sensor, in a gray low-resolution image using the dynamic range

[1] `http://webdiis.unizar.es/%7Ebermudez/phosphenicRepresentation.wmv`

of the phosphene representation [20]. However, the available gray-level is considerably low making difficult the environment understanding. A first improvement of this representation is removing the background of the image (for example the ground [22]) in order to emphasize objects which are candidates to be obstacles. In Fig. 6 (a) we show an example of ground removal using a RANSAC based estimation of the ground plane.

Another of our proposals for describing depth is using an iconic description of perspective projection. Inspired by old low-resolution 3D games and exploiting the localization and mapping obtained from modern visual odometry systems, the displacement with respect the ground can be evoked by using chess patterns. Indoor scenes described by a high level map, can be represented using iconic layouts representing main directions using pairs of parallel lines. This description allows easily representing walls and corridors. In Fig. 6 (b) and in the attached video [1] we show an example of the proposed representation of a corridor using our map of phosphenes.

5 Discussion

In this paper we present a prototype for Simulated Prosthetic Vision based on a head-mounted display and a RGB-D camera. This system is the framework used to present two tools for human interaction and navigation. We have explored different alternatives for depth and human representation in SPV. We observe that there exist a tradeoff between the available resolution and the amount of information being possible to represent. Iconic representations require more performance in computer vision but they provide more understanding of the environment. As future work we consider integrating a robust visual odometry for fully exploiting the iconic representation of layouts.

Simulated prosthetic vision open new possibilities for understanding the performance of computer vision algorithms for prosthetic vision. On the one hand we have flexibility for testing different algorithms in simulations of different prostheses. On the other hand using SVPs allows validating the proposals in testing groups crowded enough for obtaining design parameters with statistical meaning. We believe that these first steps presented here pave the way for improving SPV state-of-the-art techniques by including depth information inside the perception loop. The main challenging issue here is to encapsulate depth information in the constrained representation of the prosthetic vision system, and our proposals aim towards these goals.

6 Acknowledgements

This work was supported by Spanish Government/European Union projects DPI2014-61792-EXP and DPI2015-65962-R (MINECO/FEDER). The final publication is available at Springer via `http://doi.org/10.1007/978-3-319-58838-4_47`.

References

1. Ahuja, A.K., Behrend, M.R.: The Argus^{TM} II retinal prosthesis: factors affecting patient selection for implantation. Progress in retinal and eye research 36, 1–23 (2013)
2. Aladren, A., Lopez-Nicolas, G., Puig, L., Guerrero, J.: Navigation assistance for the visually impaired using RGB-D sensor with range expansion. IEEE Systems Journal, Special Issue on Robotics & Automation for Human Health 10(3), 922–932 (2014)
3. Asher, A., Segal, W.A., Baccus, S.A., Yaroslavsky, L.P., Palanker, D.V.: Image processing for a high-resolution optoelectronic retinal prosthesis. IEEE transactions on Biomedical Engineering 54(6), 993–1004 (2007)
4. Ayton, L.N., Luu, C.D., Bentley, S.A., Allen, P.J., Guymer, R.H.: Image processing for visual prostheses: A clinical perspective. In: IEEE International Conference on Image Processing. pp. 1540–1544 (2013)
5. Barnes, N.: An overview of vision processing in implantable prosthetic vision. In: IEEE International Conference on Image Processing. pp. 1532–1535 (2013)
6. Cha, K., Boman, D.K., Horch, K.W., Normann, R.A.: Reading speed with a pixelized vision system. JOSA A 9(5), 673–677 (1992)
7. Chang, M., Kim, H., Shin, J., Park, K.: Facial identification in very low-resolution images simulating prosthetic vision. Journal of neural engineering 9(4), 046012 (2012)
8. Chen, S.C., Suaning, G.J., Morley, J.W., Lovell, N.H.: Simulating prosthetic vision: Visual models of phosphenes. Vision Research 49(12), 1493–1506 (2009)
9. Denis, G., Jouffrais, C., Mailhes, C., Macé, M.J.: Simulated prosthetic vision: Improving text accessibility with retinal prostheses. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 1719–1722 (2014)
10. Dobelle, W.H., Mladejovsky, M., Girvin, J.: Artificial vision for the blind: electrical stimulation of visual cortex offers hope for a functional prosthesis. Science 183(4123), 440–444 (1974)
11. Dobelle, W.H.: Artificial vision for the blind by connecting a television camera to the visual cortex. ASAIO journal 46(1), 3–9 (2000)
12. Fornos, A.P., Sommerhalder, J., Pelizzone, M.: Reading with a simulated 60-channel implant. Frontiers in neuroscience 5 (2011)
13. Gutiérrez-Gómez, D., Mayol-Cuevas, W., Guerrero, J.: Inverse depth for accurate photometric and geometric error minimisation in RGB-D dense visual odometry. In: IEEE International Conference on Robotics and Automation. pp. 83–89 (2015)
14. Horne, L., Alvarez, J., McCarthy, C., Salzmann, M., Barnes, N.: Semantic labeling for prosthetic vision. Computer Vision and Image Understanding 149, 113–125 (2016)
15. Horne, L., Barnes, N., McCarthy, C., He, X.: Image segmentation for enhancing symbol recognition in prosthetic vision. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 2792–2795 (2012)
16. Hornig, R., Zehnder, T., Velikay-Parel, M., Laube, T., Feucht, M., Richard, G.: The imi retinal implant system. In: Artificial Sight, pp. 111–128. Springer (2007)
17. Jung, J.H., Aloni, D., Yitzhaky, Y., Peli, E.: Active confocal imaging for visual prostheses. Vision research 111, 182–196 (2015)
18. Li, W.: Wearable computer vision systems for a cortical visual prosthesis. In: Proceedings of the IEEE International Conference on Computer Vision Workshops. pp. 428–435 (2013)

19. Li, Y., McCarthy, C., Barnes, N.: On just noticeable difference for bionic eye. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 2961–2964 (2012)
20. Lieby, P., Barnes, N., McCarthy, C., Liu, N., Dennett, H., Walker, J.G., Botea, V., Scott, A.F.: Substituting depth for intensity and real-time phosphene rendering: Visual navigation under low vision conditions. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 8017–8020 (2011)
21. Lui, W.L.D., Browne, D., Kleeman, L., Drummond, T., Li, W.H.: Transformative reality: improving bionic vision with robotic sensing. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 304–307 (2012)
22. McCarthy, C., Barnes, N., Lieby, P.: Ground surface segmentation for navigation with a low resolution visual prosthesis. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 4457–4460 (2011)
23. McCarthy, C., Walker, J.G., Lieby, P., Scott, A., Barnes, N.: Mobility and low contrast trip hazard avoidance using augmented depth. Journal of neural engineering 12(1), 016003 (2014)
24. Meffin, H.: What limits spatial perception with retinal implants? In: IEEE International Conference on Image Processing. pp. 1545–1549 (2013)
25. Mur-Artal, R., Montiel, J., Tardós, J.D.: Orb-slam: a versatile and accurate monocular slam system. IEEE Transactions on Robotics 31(5), 1147–1163 (2015)
26. Parikh, N., Itti, L., Weiland, J.: Saliency-based image processing for retinal prostheses. Journal of neural engineering 7(1), 016006 (2010)
27. Perez-Yus, A., Gutierrez-Gomez, D., Lopez-Nicolas, G., Guerrero, J.: Stairs detection with odometry-aided traversal from a wearable RGB-D camera. Computer Vision and Image Understanding 120, 1–14 (2016)
28. Products, S.S.M.: Argus ii retinal prosthesis system surgeon manual (2013)
29. Srivastava, N.R., Troyk, P.R., Dagnelie, G.: Detection, eye–hand coordination and virtual mobility performance in simulated vision for a cortical visual prosthesis device. Journal of neural engineering 6(3), 035008 (2009)
30. Stacey, A., Li, Y., Barnes, N.: A salient information processing system for bionic eye with application to obstacle avoidance. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 5116–5119 (2011)
31. Thompson, R.W., Barnett, G.D., Humayun, M.S., Dagnelie, G.: Facial recognition using simulated prosthetic pixelized vision. Investigative ophthalmology & visual science 44(11), 5035–5042 (2003)
32. Tsai, D., Morley, J., Suaning, G.J., Lovell, N.H.: A wearable real-time image processor for a vision prosthesis. Computer methods and programs in biomedicine 95(3), 258–269 (2009)
33. Wang, J., Wu, X., Lu, Y., Wu, H., Kan, H., Chai, X.: Face recognition in simulated prosthetic vision: face detection-based image processing strategies. Journal of neural engineering 11(4), 046009 (2014)
34. Weiland, J.D., Parikh, N., Pradeep, V., Medioni, G.: Smart image processing system for retinal prosthesis. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 300–303 (2012)
35. Zapf, M.P.H., Boon, M.Y., Lovell, N.H., Suaning, G.J.: Assistive peripheral phosphene arrays deliver advantages in obstacle avoidance in simulated end-stage retinitis pigmentosa: a virtual-reality study. Journal of neural engineering 13(2), 026022 (2016)
36. Zrenner, E.: Will retinal implants restore vision? Science 295(5557), 1022–1025 (2002)