Enhancing Facial Recognition in Visual Prostheses using Region of Interest Magnification and Caricaturing

Hesham M. Moneer
Digital Media Eng. and Technology Dept.
German University in Cairo
Cairo, Egypt
hesham.moneer@outlook.com

Reham H. Elnabawy
Digital Media Eng. and Technology Dept.
German University in Cairo
Cairo, Egypt
reham.elnabawy@guc.edu.eg

Seif Eldawlatly
Computer and Systems Eng. Dept.
Ain Shams University
Computer Science and Eng. Depr.
The American University in Cairo
Cairo, Egypt
seldawlatly@eng.asu.edu.eg

***Abstract*—Blindness is a prevalent disability with significant personal and societal consequences. While medical advancements offer treatment options, severe damage to the retina, optic nerve, or brain may remain untreated. Visual prostheses are implantable medical devices that aim at providing limited vision to such individuals. However, such prostheses offer low spatial resolution making activities like reading, facial recognition, and navigation challenging. This work aims to enhance implantees' ability to recognize faces through real-time scene preprocessing, machine learning and computer vision techniques. Virtual-reality visual models simulating prosthetic vision were tested on normally/corrected sighted subjects, investigating the use of histogram equalization for contrast enhancement, facial region magnification, and caricaturing of facial features. Results revealed that histogram equalization with magnification increases facial recognition accuracy by 60%, distinguishability accuracy by 50%, and accuracy of seeing facial details by 90%. In contrast, adding facial caricaturing improved the accuracies by 66.66%, 25%, and 75% for recognizability, distinguishability, and seeing facial details, respectively. Consequently, the combination of visual field histogram equalization, face magnification, and optional caricaturing can be considered as a promising enhancement approach that could enhance the quality of vision perceived through visual prostheses.**

***Index Terms*—Visual Prostheses, region of interest magnification, caricaturing, histogram equalization**

I. Introduction

Visual prostheses can help restore vision for patients with acquired blindness. It works by electrically stimulating functional visual pathway structures using an electrode array [1] (e.g., Second Sight's Argus II [2]). Prosthetic vision is based on the artificial generation of phosphenes, which are visual sensations caused by means other than light stimulation. Stimulation could be electrical, mechanical, magnetic, ionizing radiation, etc [3]. A single phosphene in the visual field could be thought of as a single pixel in a raster image. Engineering challenges, however, limit the maximum number of implantable electrodes and the size of the implant. Therefore, vision restored by visual prostheses suffers from low spatial resolution. Moreover, the malfunction of some electrodes may result in consistently dark phosphenes, which is known as electrode dropout [4]. Consequently, the efficiency of doing numerous tasks, like recognizing faces, for a visual prostheses implantee is very low when compared to a normally-sighted person.

Several past studies explored different processing strategies to facial regions and the impact of these strategies on improving the ability of visual prostheses implantees to recognize faces. Increasing image contrast is one of the most common image enhancement techniques. It helped in increasing the ability of visual prostheses users to recognize faces [5]. Increasing image contrast can be achieved through histogram equalization [6]. Another approach is to make the phosphenes carry as much useful information as possible. Because the visual field of the visual prostheses implantees has only a very limited number of phosphenes, these phosphenes can only carry limited information. Thus, by focusing on and magnifying a region-of-interest (ROI), it could be easier for visual prostheses users to recognize the content of the perceived image. The ROI could be identified using the Viola Jones approach or any other region of interest identification method [7], [8]. Finally, inspired by cochlear implant signal manipulations that were designed specifically to match the way humans perceive speech, manipulations designed specifically to enhance humans perception and recognition of faces were explored. One possible manipulation is face caricaturing, which is the exaggeration of the way a veridical face differs from an average face making several facial features, nose width for example, more evident [9].

The aim of this study is to utilize computer vision techniques to improve the implantees efficiency in recognizing familiar faces, distinguishing similar faces, and identifying facial expressions. Using simulated prosthetic vision [10], simulated phosphenes were presented through head-mounted virtual reality displays showing the simulated images. Also, a video camera was required to capture the real-time scene around the subject compatible with their head motion. The experiments involved three groups of subjects: one group was presented with phosphene simulations without any enhancements, one group was presented with phosphene simulations after contrast enhancement and ROI magnification, and the last group was similar to the second group but with face caricaturing applied. Our results demonstrate a performance increase in all aspects when applying the proposed enhancements.

II. Methodology

A. Simulated Prosthetic Vision Environment

In order to examine the proposed approach before utilizing it for visual prostheses implantees, we developed a simulated prosthetic vision environment [11]. This environment was used to present normally/corrected sighted subjects, via a head-mounted display, with images similar to what visual prostheses implantees perceive. To generate simulated phosphenes from a raster image, the image was first transformed into grayscale, and pixels were then mapped one by one into corresponding phosphenes. One possible approach to simulate a phosphene, called the array method, is to draw multiple circles, with an increasing radius and a decreasing intensity, around the phosphene center in order to represent the phosphene [11]. In order for the phosphenes to represent pixels of different intensities, they could be either color modulated (where the inensity of the phosphene was directly proportional to intensity of the pixel it is representing), or size modulated (where the size of the phosphene was directly proportional to the intensity of the pixel it is representing). In this study, color modulation was employed in the conducted experiment. Initially, size modulation was considered. However, subjects experienced difficulty perceiving any details, even when enhancements were applied. In contrast, the utilization of color modulated phosphenes allowed subjects to discern the visual field more effectively.

B. Histogram Equalization

Before applying any face-specific enhancements, a preliminary step of applying histogram equalization to the entire visual field was undertaken. Histogram equalization is a technique used in image processing to enhance the contrast and visibility of an image. The purpose of histogram equalization is to transform the pixel intensity distribution of an image to a more uniform distribution, effectively spreading out the intensities over the entire range. This redistribution of pixel intensities results in an improved dynamic range and clearer visualization of details in the image, making it particularly useful for enhancing images with low contrast or uneven lighting conditions [6].

C. Viola Jones Haar Cascade Classifier

Because enhancements were applied to only face regions in the visual field, these faces had to be first detected. The Viola-Jones Haar Cascade Classifier is a widely used object detection algorithm in computer vision [8]. The classifier is primarily designed for real-time face detection, but it can be adapted to detect other objects as well. The algorithm's mechanism involves a two-step process: feature extraction and classification. During feature extraction, the algorithm uses Haar-like features, which are simple rectangular filters that capture intensity variations in different regions of the image. These features are computationally efficient, making the algorithm suitable for real-time applications. The classifier then employs an adaptive boosting (AdaBoost) algorithm to select and combine the most relevant features, creating a strong classifier capable of distinguishing between object and non-object regions efficiently.

D. Facial Landmarks

Facial landmarks are specific points or key locations on a human face that serve as reference points for facial analysis and enhancement [12]. These landmarks are typically represented as coordinate positions indicating their precise locations on the face. Common facial landmarks include the corners of the eyes, the tip of the nose, the edges of the mouth, and various points along the eyebrows and jawline.

E. Caricaturing Implementation

Face caricaturing can improve the perception of the differences between pairs of simultaneously seen similar faces [9], and it can be applied on the faces using the opposite of Face Morphing techniques. Face Morphing can be defined in the following sense: given two input images I_0 and I_1 of human faces, morphing is generating a fluid transformation (video clip) transitioning from I_0 to I_1 [13]. Applying the opposite of Face morphing (i.e. transitioning a veridical face *away* from an average face) is one way of doing caricaturing.

In order to caricature a face, first facial landmarks had to be detected. Moreover, an average face had to be generated, by averaging together the landmarks of multiple faces of several people, with different ethnicities and genders. The facial landmarks of an average face were computed and stored. The opposite of face morphing [13] was then applied to achieve caricaturing by straying the veridical face landmarks away from the computed average face landmarks [9]. To stray the veridical landmarks away from the average landmarks, a series of steps was followed. First, the landmarks were converted into a Delaunay Triangulation mesh. Subsequently, Affine Transformations were applied to each triangle within the veridical Delaunay Triangulation mesh, effectively transforming them into new triangles. By displacing each vertex of these triangles away from their corresponding vertices in the average face Delaunay Triangulation, with a scaling factor representing the percentage of caricaturing applied, a new triangle was formed. The resulting collective image, comprising the transformed triangles, formed the output caricatured face.

F. Experiments

The experiment was conducted on 3 groups of normally/corrected sighted subjects. Each group comprised 5 subjects. The subjects' average age was 21.5 years old, and their gender distribution was 73% males and 27% females. The goal of the experiment was to assess the impact of the enhancements on the ability of the subjects to recognize familiar faces, distinguish similar faces, and recognize facial expressions. Groups names and enhancements applied to each group are presented in TABLE I. For groups where magnification of the face ROI was applied, participants were granted the freedom to bypass the magnification and observe the entire visual field. Furthermore, they were granted the capability to switch between various faces if multiple faces were present within the visual field. The subjects' visual field was made up of 32 × 32 phosphenes with 16 color modulation levels [7].

TABLE I Virtual Reality Experiment Groups

Group name	Description
DLR	Direct lowering of resolution (without any enhancements)
VJm	Visual field histogram equalization Viola Jones (VJ) face region magnification
VJc	Visual field histogram equalization Viola Jones (VJ) face region magnification Caricaturing

For an immersive experience for the subjects, the experiment was conducted using Virtual Reality (VR). In the conducted tests, subjects were shown printed images, or expressions, in the VR set through a webcam. The scene would then be captured, phosphene-simulated, and enhanced in real-time. In the following list of tests that subjects have undergone, accuracy refers to the number of a subject's correct responses divided by the total number of questions asked in a test. The tests were as follows:

1) Recognition test:
   Prior to the test, it was ensured that the participants were familiar with specific celebrities. During the testing phase, they were presented with 6 different images of these celebrities, each for 10 seconds, and tasked with recognizing them. The tester recorded subjects' accuracy, their confidence levels (rated on a scale of 1-5), and response times (the time taken to recognize a face). The goal was to assess the subjects' ability to recognize the identity of the displayed faces.
2) Expressions test:
   The experimenter would mimic a multisubset of 4 facial expressions, each for 10 seconds, from the following set: happy, sad, surprised, and frowning. The subjects were then asked to tell the expression they had seen, and their accuracy in doing so was recorded. The aim of this test was to assess the subjects' ability to perceive facial details within their visual field.
3) Distinction test:
   Participants were presented with 4 images, each containing either two faces of the same individual or of different individuals who shared general category (sex, race, and age group) and extra-face details (worn glasses, facial hair, and hair style) for 10 seconds. The accuracy of the subjects in telling whether the faces were of the same person or of two different people was recorded. The goal of this test was to assess the subjects' ability in spotting the differences between similar face.

III. Results

A. Simulations Output

We first demonstrate the output of the phosphene simulation process for each of the 3 groups (DLR, VJm and VJc). Fig. 1 shows the result of directly applying phosphene simulation to the input face image without any enhancements as viewed by the DLR group. As seen in the figure, the phosphene simulated image does not provide enough detail to identify the input face, which would make it challenging for visual prostheses users to recognize faces. This illustrates the need for the enhancements proposed in the paper.

One enhancement was applying histogram equalization and face region magnification (VJm gorup), as shown in Fig. 2. In comparison to Fig. 1, the face area is magnified with its contrast enhanced.

Finally, the last group of subjects (VJc) was presented with the face after caricaturing in addition to applying histogram equalization and face magnification. To perform face caricaturing, facial landmarks are first identified as shown in Fig. 3. Next, the caricaturing process is applied (shown in Fig. 4) as described in Subsection II-E. Finally, Fig. 5 illustrates the steps of applying VJc group enhancements (histogram equalization, ROI magnification, and caricaturing) to the same frame. The figure demonstrates that, while caricaturing changes the face appearance, the face can still be easily recognized.

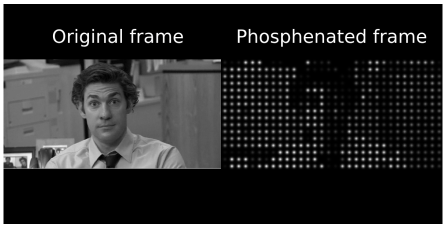

Fig. 1: Sample phosphene-simulated frame as shown to the DLR group.

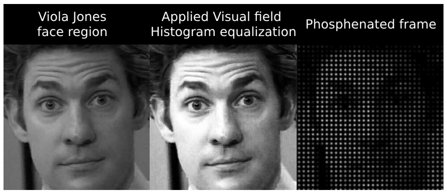

Fig. 2: Sample phosphene-simulated frame as shown to the VJm group.

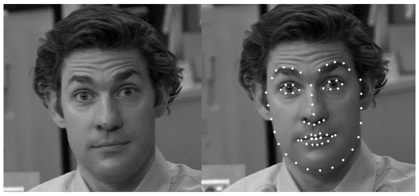

Fig. 3: Identification of 68 facial landmarks.

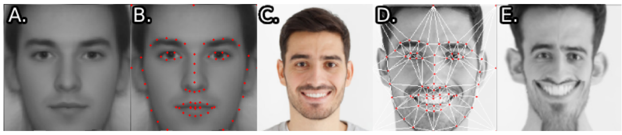

Fig. 4: The steps of generating a 60% caricature. **A.** Average face. **B.** Stored average face landmarks. **C.** Veridical face. **D.** Veridical face landmarks and Delaunay triangulation. **E.** Final caricatured face.

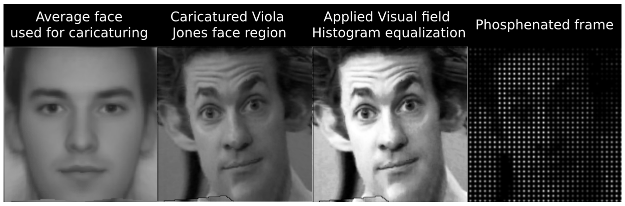
Average face used for caricaturing
Caricatured Viola Jones face region
Applied Visual field Histogram equalization
Phosphenated frame

Fig. 5: Sample phosphene-simulated frame as shown to the VJc group.

B. Experimental Results

The performance of each groups of subjects was assessed using three different tests, which were described in Subsection II-F. The first test is the face recognition test, in which the ability of the subjects in recognizing the presented faces was assessed. Fig. 6 illustrates the performance of the three groups of subjects in such test using three different performance measures, namely recognition accuracy, response time and confidence level. The figure illustrates that the highest recognition accuracy is achieved by the VJc group in the least response time compared to the other groups. However, the highest confidence level was achieved by the VJm group. These results indicate that the proposed enhancements including contrast enhancement, face region magnification and caircaturing could help visual prostheses patients in recognizing familiar faces. The slightly lower confidence achieved by the VJc group could be attributed to the use of caricaturing as it modifies the face, which could reflect on the user's confidence in recognizing the faces.

The second test assessed the ability of the subjects to recognize facial expressions. Fig. 7 shows the recognition accuracy of the three groups in this test. The figure demonstrates the significant enhancement in performance when using the proposed enhancements (both VJm and VJc) compared to not using them.

Finally, Fig. 8 illustrates the accuracy achieved by the three groups in distinguishing between similar faces. Similar to the results achieved in Fig. 7, the proposed enhancements resulted in better accuracies compared to not applying them. However, in the last two tests, caricaturing seemed to reduce the performance of the subjects. Nevertheless, these findings support the positive impact of the enhancements on face recognition capabilities, facial features perception, and similar faces distinguishability.

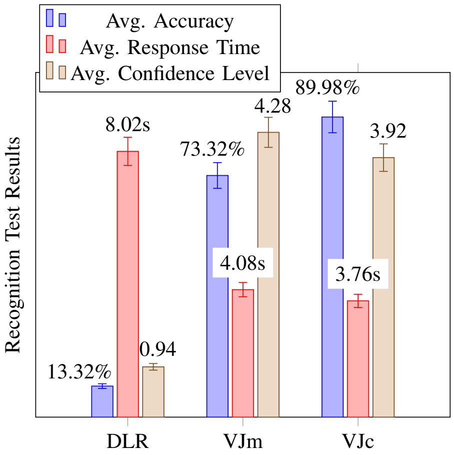

Fig. 6: Recognition Test results.

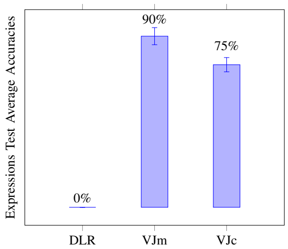

Fig. 7: Expressions Test results.

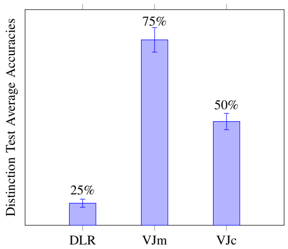

Fig. 8: Distinction Test results.

IV. Conclusion

While the recent developments in the visual prostheses domain could help the blind restore, at least partially, their vision, perceived vision using such devices has a very low resolution. In this work, we proposed using a combination of a number of enhancements; namely, contrast enhancement, face region magnification and face caricaturing. Histogram equalization effectively alleviated any visual unclarities arising from poor lighting conditions within the scene. Magnification maximized the utilization of the limited number of phosphenes to capture and convey as much information as possible from the facial details. Finally, caricaturing techniques were employed to exaggerate facial features enhancing their perceptibility for the subjects. Applying these enhancements resulted in the best performance by the subjects in three tests: face recognition, facial expression recognition and similar faces distinction test. However, caricaturing a magnified face did not significantly improve the subjects' performance when compared to magnification and contrast enhancement alone. Therefore, the combination of visual field histogram equalization with magnification, along with the option for caricaturing (where subjects can control whether caricaturing should be applied or not), could be considered the best option for enhancing faces without violating real-time requirements. The work in this study could be extended by conducting experiments on actual implanted patients and incorporating more advanced face caricaturing techniques, such as Generative Adversarial Networks (GANs).

REFERENCES

[1] E. Fernandez, “Development of visual neuroprostheses: trends and challenges,” *Bioelectronic medicine*, vol. 4, no. 1, pp. 1–8, 2018.

[2] M. S. Humayun, J. D. Dorn, L. Da Cruz, G. Dagnelie, J.-A. Sahel, P. E. Stanga, A. V. Cideciyan, J. L. Duncan, D. Eliott, E. Filley, *et al.*, “Interim results from the international trial of second sight’s visual prosthesis,” *Ophthalmology*, vol. 119, no. 4, pp. 779–788, 2012.

[3] I. Bókkon, “Phosphene phenomenon: a new concept,” *BioSystems*, vol. 92, no. 2, pp. 168–174, 2008.

[4] R. H. Elnabawy, S. Abdennadher, O. Hellwich, and S. Eldawlatly, “Electrode dropout compensation in visual prostheses: an optimal object placement approach,” in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 6515–6518, IEEE, 2021.

[5] R. W. Thompson, G. D. Barnett, M. S. Humayun, and G. Dagnelie, “Facial recognition using simulated prosthetic pixelized vision,” *Investigative ophthalmology & visual science*, vol. 44, no. 11, pp. 5035–5042, 2003.

[6] S. S. Agaian, B. Silver, and K. A. Panetta, “Transform coefficient histogram-based image enhancement algorithms using contrast entropy,” *IEEE transactions on image processing*, vol. 16, no. 3, pp. 741–758, 2007.

[7] J. Wang, X. Wu, Y. Lu, H. Wu, H. Kan, and X. Chai, “Face recognition in simulated prosthetic vision: face detection-based image processing strategies,” *Journal of neural engineering*, vol. 11, no. 4, p. 046009, 2014.

[8] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, vol. 1, pp. I–I, Ieee, 2001.

[9] J. L. Irons, T. Gradden, A. Zhang, X. He, N. Barnes, A. F. Scott, and E. McKone, “Face identity recognition in simulated prosthetic vision is poorer than previously reported and can be improved by caricaturing,” *Vision research*, vol. 137, pp. 61–79, 2017.

[10] S. C. Chen, G. J. Suaning, J. W. Morley, and N. H. Lovell, “Simulating prosthetic vision: I. visual models of phosphenes,” *Vision research*, vol. 49, no. 12, pp. 1493–1506, 2009.

[11] H. Moneer, “Enhanced Simulation of Prosthetic Vision.” https://github.com/HeshamMoneer/Phosphenes-Simulation.

[12] V. Kazemi and J. Sullivan, “One millisecond face alignment with an ensemble of regression trees,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1867–1874, 2014.

[13] A. Wang, “Face Morphing.” https://github.com/Azmarie/Face-Morphing.