Person Recognition Based on FaceNet under Simulated Prosthetic Vision

Ying Zhao\textsuperscript{a}, AiPing Yu\textsuperscript{b}, DanTong Xu\textsuperscript{c}

School of Information Engineering, Inner Mongolia University of Science and Technology, Baotou, China, Telephone number, incl. CN
\textsuperscript{a}amengs@imust.cn, \textsuperscript{b}ariafisherwater@gmail.com, \textsuperscript{c}Danielle0xv@gmail.com

ABSTRACT. Face information is important information for identifying people, but at low resolution, face information is not well recognized. A psychological physics experiment of person recognition in daily life was designed to discover the best strategy for the limited number of stimulating electrodes, and provide useful visual information. The real-time image-processing strategy based on FaceNet were used to optimize the person information by turning complex face information into simple Chinese character information. Noted that all processed target faces were obscured by separate Chinese characters which are the target people’s last names. The psychological results showed that the image-processing strategy based on FaceNet improved recognition accuracy. The proposed strategy, to convert complex face information into simple Chinese character information, could help subjects to use their own first knowledge to identify the person who need to be identified more faster and accurately.

CCS Concepts
\begin{itemize}
  \item Human-centered computing \rightarrow Human computer interaction;
  \item Computing methodologies \rightarrow Artificial intelligence \rightarrow Computer vision \rightarrow Computer vision tasks;
  \item Social and professional topics
\end{itemize}

1. INTRODUCTION
Retinal degenerative diseases such as retinitis pigmentosa and age-related maculopathy are universal causes of visual loss [1]. To date, there is no effective treatment in medicine. However, with the development of science and technology and inspired by cochlear implant, visual prostheses could restore partial vision to individuals blinded by bypassing the damaged photoreceptors and electrically stimulating intact neurons [2]. Nowadays, many studies have been performed by implanted visual prostheses in various regions such as the retina, optic nerve, visual cortex and lateral geniculate body.

Tassicker published the preliminary report on a retinal stimulator in 1956 [3]. He described in a patent how a small, flat, light-sensitive selenium cell placed behind the retina of blind patient transiently restored the patient’s ability to perceive the sensation of light [4]. After all these years of retinal prostheses development, there are three kinds of retinal implant: subretinal, epiretinal and choroidal. The subretinal device, being investigated by Chow and Zrenner et al, is implanted between the pigment epithelial layer and the outer layer of retinal, which contains the photoreceptor cells, stimulating the retinal of bipolar cells[5]. That’s the way Retinal Implant AG subretinal used. Second Sight’s Argus II which is an epiretinal device have been more promising. Argus II has received U.S. marker approval from the Food and Drug Administration on February 14, 2013 [6]. The Argus II consists of three internal components including the coil, the electrode array, and the band which is positioned around the eye and three external components including a video camera that is mounted on a pair of glasses, a visual
processing unit worn on the patient’s belt, and external coil on the sidearm of the glasses to transmit data between the internal and external components using radiofrequency [6]. Recently, a supra-choroid transscleral retinal prosthesis has been tested in human. The surgical approach of choroidal implant involves cutting a flap in the outer layer of the eye and placing an electrode in the flap, so that the exposed electrode contacts are directed towards the choroid and retina, which lies on the other side of the choroid [7].

Compared with retinal, the optic nerve is an appealing site for the implementation of a visual prosthesis as the entire visual field is represented in a small area which can be reached surgically [8]. Veraart et.al. were the first to attempt electrical stimulation of optic nerve as a basis upon which to develop a visual prosthesis [9]. He implanted a self-sizing spiral cuff electrode around an optic nerve for a 59-year-old female volunteer with retinitis pigmentosa. In 2003, they published the results showed that the blind volunteer was able to interact with environment such as demonstrating pattern recognition and processing time and orientation discrimination [10].

The cortical prosthesis is advantageous over other approaches. It bypasses all diseased visual pathway neurons rostral to the primary visual cortex. But the disadvantage is obvious. It is not easily reproducible between various patients. The organization of visual cortical is more complex at the level of the primary cortex than at retina or optic nerve [10].

Since the resolution of current prosthetic devices is far below that of natural vision [2], how to use the limited number of phosphene dots to maximize visual percepts is a question worth considering. So far, many research groups have effectively estimated various image processing strategies under simulated prosthetic vision which is an effective method of adjusting implant designs and helping researcher to find optimal strategy for prosthetic wearer. Reports on some missions including pathfinding, Chinese character recognition, and environment recognition indicated that these image processing strategies have achieved remarkable results.

Compared with environmental recognition task, pathfinding task and Chineses character recognition task, face recognition task always need higher resolution. If the resolution is low, the loss of facial information and decreased recognition accuracy would be the result. Figure 1 is a face image with two simulated grayscale level and 36×36, 48×48, and 64×64 resolution. Dagnelie et al [14], first studied face recognition under the simulated prosthetic vision. They examined four resolutions including 10×10, 16×16, 25×25, and 32×32. A face image of prosthetic vision and four different face images under normal vision were showed to the subjects who were asked to select the image of same person from the four normal face images under the prosthetic vision. In 2003, Thompson et. al. [13] experimented with 60 face images using the same experimental protocol. Li et. al. [12] studied the effects of different filtering methods, brightness modulation methods and edge extraction operators on recognition accuracy under simulates prosthetic vision. Boyle et. al. [15] proposed an image processing strategy based on image region of interest (ROI) under simulates prosthetic vision. In 2017, Irons et. al. [2] proposed that caricaturing can be used to improve the recognition accuracy under simulated prosthetic vision.

Several previous studies have tested face recognition and indicated excellent task performance. But can facial recognition get such good results in a changing environment? Face information is the important information for identifying people. However, at unsatisfactory condition such as camera motion and low resolution, face information is hard to be recognized. In this paper, a real-time processing strategy based on FaceNet were proposed and tried to solve the problem. By the trained neural network, the face information that is difficult to recognize was transformed into simple and easily identifiable Chinese characters information.
2. MATERIALS AND METHODS

2.1 Selection of subjects
Twenty-one participants enrolled from Inner Mongolia University of Science and Technology. 11 females and 10 males. They were between 22 and 25 years old. Participants had normal or corrected-to-normal acuity, defined as not lower than 20/20. Their native language is Chinese. Participants were asked to remember the name which is the person who need to be identified in the video.

2.2 Compliance statement
All of the experimental process met “the Helsinki Declaration of the Word Medical Association” with regard to ethical principles and constrains relating to biomedical research subjects. All research was conducted in accordance with national medical clinical trial provisions and with the approval of the state Department of Health Ethics Committee. All subjects gave informed written consent prior to participants.

2.3 Laboratory equipment and environment
The experimental hardware included a Logitech c920 camera and a Lenovo 30AGA29BCN tower desktop computer with windows 8.1 64-bit operating system, an Intel Xeon (Zhi Qiang) E3-1241 V3 EITH A 3.5GHz quad core processor, a Lenovo SHARK-BAY motherboard, 8 GB of memory, and a Nvidia K620 graphics card. The used experimental software included EV screen, JetBrains PyCharm Community Edition 2017.3.4 x64 and IBM SPSS Statistics. The experiment was performed in a clean and quite room. Before the experiment, subjects were assured that they were in a clam and relaxed mood without any interference. The experimental process was captured by Logitech camera and recorded by EV screen software. The experimental results were analyzed by the one-way ANOVA.

2.4 Material
Thirty-six different double videos including eighteen double sitting persons and eighteen double standing persons were filmed. The scene was designed to simulate the situation that the prosthesis wearer may encounter in life. Eight persons of the 36 video materials need to be identified were the Inner Mongolia University students, including 2 taller females (f1, f2), 2 shorter females (f3, f4), 2 taller males (m1, m2), and 2 shorter males (m3, m4). This design was to avoid too obvious height differences between male and female. Each video sequence was 7s in duration. In the videos of the sitting groups, the distance between the camera and the recognized target person was 60cm. In the videos of the standing groups, the distance between the camera and the recognized target person was 180cm.

![Figure 1. Face images with different resolution under the simulated prosthetic vision.](image_url)
Table 1. Material arrangement result of sitting and standing groups

| groups   | 1   | 2   | 3   | 4   | 5   | 6   |
|----------|-----|-----|-----|-----|-----|-----|
| female   | f1  | f2  | f1  | f4  | f3  | f4  |
| male     | m1  | m2  | m1  | m4  | m2  | m4  |
| mix      | m1f2| m2f2| m3f3| m4f4| m1f2| m2f3|

2.5 Experimental procedure

The experiment was divided into two groups: one group was 18 sitting videos, the other group was 18 standing videos. Each video processed by the strategy which based on FaceNet and at three resolutions (36×36, 48×48, 64×64, marked as f36, f48, f64). In order to avoid learning effects, the videos of each group presentation order were randomly. During the experiment, subjects were asked to accurately state the gender and name of the target person who needed to be identified. Subjects performed the recognition tasks for 5 consecutive days. The whole procedure lasted 20 minutes. Before the formal experiment, subjects were asked to remember the name of target person that needs to be identified, namely, subjects need to familiarize the targets person, so as to correctly state the information that needs to be identified in the formal experiment.

The evaluation criteria of experiment were base on the recognition accuracy of target person in the videos. The recognition accuracy was defined as the subject say the right gender and name. Each video was played only once. If subject couldn’t say the names correctly in 7s, the recognition accuracy was recorded 0. If subject could describe the gender correctly in 7s, the recognition accuracy was recorded 50%. If subject could describe the gender and one of persons’ name correctly in 7s, the recognition accuracy was recorded 75%. If subject could describe the gender and names correctly in 7s, the recognition accuracy was recorded 100%.

2.6 Image processing strategy

Firstly, the faces in the videos were identified, and assigned an occlusion picture with its last name automatically, then grayscale and binarized the processed videos, matching to the 36×36, 48×48, 64×64 three different pixelized resolution templates to get pixelized videos, the flow chart was shown in Figure 3.
3. The system used for facial recognition was the FaceNet proposed by Florian et al. [11]. The traditional face recognition method used the CNN network to extract facial features, then used SVM and other methods to classify, but the results obtained by the method were often not effective enough. The difference between the FaceNet and the previous methods is that FaceNet can learn a mapping from face images directly to compact Euclidean space where distance directly correspond to a measure of face similarity [11]. FaceNet uses a deep convolutional network trained to optimize the embedding itself directly, rather than an intermediate bottleneck layer as in previous deep learning approaches.

Compared with the traditional model structure, the biggest innovation of the FaceNet model structure is the triplet loss. A detailed explanation of the triplet loss is that an image $X^a_i$ of a specific person need be closer to all other positive images $x^p_i$ of the same person than it is to any negative image $X^n_i$ of any other person [11]. The most ideal is,

$$\|x^a_i - x^p_i\|^2 + \alpha < \|x^a_i - x^n_i\|^2, \forall (X^a_i, x^p_i, x^n_i) \in T \quad (1)$$

Where $\alpha$ is a margin enforced between positive and negative pairs. $T$ is the set of all triplets in the training set [11]. Since FaceNet in terms of face recognition is much simpler and more accurate than traditional face recognition methods, it was used in this research.

3. RESULTS
The experiments were divided into sitting identification tasks (60cm) and standing identification tasks (180cm). Figure 4 and Figure 5 showed the recognition accuracy of the standing tasks and the sitting tasks under different pixelized resolutions. The recognition accuracy was not significant represented by NS. Univariate analysis was performed on the recognition accuracy of the sitting group at the three resolution, and the LSD pair method was used to perform multiple analysis on the recognition accuracy at three resolutions. Top lines showed comparisons between resolution. * means the difference is significant at the 0.05 level. ** means the difference is significant at the 0.01 level. *** means the difference is significant at the 0.001 level.

![Figure 4](image)

Figure 4. The recognition accuracy of standing group. Top lines show comparisons between resolution (*: p <0.05, **: p < 0.01, ***: p <0.001).
As shown in Figure 4, in the standing group, there was a significant difference in the accuracy of recognition in the case of different resolutions (F=163.607, *P*<0.001). When the resolution was 36×36, the average recognition accuracy was 77.37%. The average recognition accuracy increased to 96.03% when the resolution was 48×48. While the resolution researched 64×64, the average recognition accuracy was 100%. Comparing the resolution between 36×36 and 48×48, the difference was significant (*P*=0.000). Comparison between 48×48 and 64×64 resolution showed significant difference (*P*=0.004). When the resolution was from 36×36 to 64×64 the difference was significant (*P*=0.000).

As shown in Figure 5, in the sitting group, there was no a significant difference in the recognition accuracy at different resolutions. The average recognition accuracy was all 100% when the resolutions were 36×36, 48×48 and 64×64. The results showed that in the close distance of 60cm, even at a resolution of 36×36, the resolution can achieve the ideal recognition accuracy.

4. DISCUSSION

4.1 The effect of the resolution
The experimental results of the standing group showed that the resolution has a crucial influence on the recognition accuracy. And as can be seen from Figure 4, with the resolution increasing, the recognition accuracy increased. At 64×64 resolution, the recognition accuracy can reach 100%. Analysis of recognition accuracy between various resolutions showed that when the resolution increased from 36×36 to 48×48, there was a significant difference in recognition accuracy (*P*<0.001). When the resolution increased from 48×48 to 64×64, there was a significant difference in recognition accuracy (*P*<0.01). This means that when the resolution increased from 36×36 to 48×48, the recognition accuracy increased more than that when the resolution increased from 48×48 to 64×64. The experiment in the sitting group does not reflected the effect of resolution on recognition accuracy.

4.2 The effect of the distance
The experimental results of the standing group showed, when the horizontal distance of the target person to be recognized to the lens was 180cm, the average recognition accuracy of the subject was 77.37% in the 36×36 resolution. When the resolution increased to 48×48, the average recognition accuracy of the subject reached 96.03%. With the resolution increasing to 64×64, the average recognition accuracy of the subject achieved 100%. The experimental results of the sitting group showed, when the horizontal distance of the target person to be recognized to the lens was 60cm, the average recognition accuracy of the subject could reach 100% at three resolution even 36×36 resolution. At the same time, the judgment
of the person who needs to identify the target was also correct. Comparing the two sets of data, the closer the distance to the lens, the higher the recognition accuracy.

4.3 Comprehensive analysis

Because of long-distance experimental design, the information contained in the experiment of standing group included the face and body information, the target person’s face to be identified was smaller. That is to say, the remote experiment setting of standing group reduced the face information to a certain extent. The information contained in the experiment of sitting group basically consisted of face and a part of body information. Due to the close-range experimental design, the face information was not lost, but the body information cannot be completely read by the subject. Comparing the two sets of data, it can be seen that under the image processing strategy of personal recognition based on the FaceNet, the personal recognition accuracy of the face information converted into Chinese character information in the sitting group at close distance was higher than that in the standing group at long distance.

5. CONCLUSIONS

The relationship between resolution and recognition accuracy was examined in this paper. According to the difference in resolution, the recognition accuracy change can be leaded to the following conclusions: with the increase of resolution, the recognition accuracy increased. And the recognition accuracy under three resolution was significantly different. With the increase of resolution, the recognition accuracy of the subjects was gradually concentrated. When the resolution reached 64×64, all the recognition accuracy was concentrated at 100%.

Comparing the recognition accuracy of the two sets of experiments at different resolutions can lead to the following conclusions: under the condition of close range, the image processing strategy of personal recognition based on the FaceNet can get better result. The overall difference of the subjects was not large, and the distance has a great influence on the recognition accuracy.

In the future, we hope that the image strategy of personal recognition based on the FaceNet could be better developed and applied in the field of visual prosthesis, and help blind patients who implanted the prosthesis to attain independent mobility and better communication in daily life.

ACKNOWLEDGMENTS

The authors thanks all participants. This work is supported by the National Natural Science Foundation of China (81460279, 61841204); the Inner Mongolia Natural Science Foundation (2018LH08066,2015MS0604); the Inner Mongolia Autonomous Region High School Science Research Foundation (NJZY145); and the Baotou Science and Technology Project (2015C2006-14, 2017C1002).

REFERENCES

[1] RR Bourne, SR Flaxman, T Braithwaite. 2017. Magnitude, temporal trends, and projections of global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. Lancet Glob Health, 5 (Aug. 2017), 888-887. DOI= https://doi.org/10.1016/S2214-109X(17)30293-0.

[2] JL Irons, T Gradden, A Zhang. 2017. Face identify recognition in simulated prosthetic vision is poorer than preciously reported and can be improved by caricaturing. Vision Res. 137 (Aug. 2017), 61-79. DOI= https://doi.org/10.1016/j.visres.2017.06.002.

[3] GE Tassicker. 1956. Preliminary report on a retinal stimulator. The British journal of physiological optics, 13 (Apr. 1965),102.

[4] Zrenner, Eberhart. 2002. Will retinal implants restore vision?. Science. 295 (Feb. 2002), 1022-1025. DOI= https://doi.org/10.1126/science.1067996.

[5] Mark S, Humayun MD. 2016. The bionic eye: a quarter century of retinal prosthesis research and development. Ophthalmology. 123 (Oct. 2016), 89-97. DOI= https://doi.org/10.1016/j.ophtha.2016.06.044.
[6] Finn AP, Grewal DS, Vajzovic L. 2018. Argus II retinal prosthesis system: a review of patient selection criteria, surgical considerations, and post-operative outcomes. Clinical ophthalmology, (Jun. 2018), 1089-1097. DOI= https://doi.org/10.2147/OPTH.S137525.

[7] Weiland JD, Liu W, Humayun MS. Retinal prosthesis. Annual Review of Biomedical Engineering, 7 (Aug, 2005), 361-401. DOI= https://doi.org/10.1146/annurev.bioeng.7.060804.100435.

[8] Banarji A, Gurunadh VS. 2009. Visual prosthesis: Artificial vision. Medical Journal, Armed Forces India, 65 (Oct. 2009), 348. DOI= http://dx.doi.org/10.1016/S0377-1237(09)80098-1.

[9] Veraart C, Raftopoulos C. 1998. Visual sensations produced by optic nerve stimulation using an implanted self-sizing spiral cuff electrode. Brain research. 813, 1 (Nov. 1998), 181-186. DOI= https://doi.org/10.1016/S0006-8993(98)00977-9.

[10] Veraart C, et al. 2003. Pattern recognition with the Opti nerve visual prosthesis. Artificial organs. 27, 11 (Oct. 2003), 996-1004. DOI= https://doi.org/10.1046/j.1525-1594.2003.07305.x.

[11] Schroff F, Kalenichenko D, Philbin J. 2015. FaceNet: A unified embedding for face recognition and clustering. In proceedings of the IEEE conference on computer vision and pattern recognition. (2015), 815-823.

[12] Li R, Zhang X, Zhang H. 2006. Facial recognition using enhanced pixelized image for simulated visual prosthesis. In 2005 IEEE Engineering in Medicine and Biology 27th Annual conference. (April. 2006), 5219-5222.

[13] Thompson RW, Barnett GD, Humayun MS, Dagnelie G. 2003. Facial recognition using simulated prosthetic pixelized vision. Investigative ophthalmology & visual science. 11, 44 (Nov. 2006), 5035-5042. DOI= http://dx.doi.org/10.1167/iovs.03-0341.

[14] Dagnelie G, Thompson RW. 2001. Simulated prosthetic vision: Perceptual and performance measures. In Vision Science and its Applications, OSA Technical Digest. Washington DC: Optical Society of America, (2001), 43-46.

[15] Boyle JR, Maerder AJ, Boles WW. 2008. Region-of-interest processing for electronic visual prostheses. Journal of Electronic Imaging, (January, 2008). DOI= https://doi.org/10.1117/1.2841708.