Electronic Imaging

JElectronicImaging.org

Comparing object recognition from binary and bipolar edge images for visual prostheses

Jae-Hyun Jung Tian Pu Eli Peli



Jae-Hyun Jung, Tian Pu, Eli Peli, "Comparing object recognition from binary and bipolar edge images for visual prostheses," *J. Electron. Imaging* **25**(6), 061619 (2016), doi: 10.1117/1.JEI.25.6.061619.

Comparing object recognition from binary and bipolar edge images for visual prostheses

Jae-Hyun Jung,^{a,*} Tian Pu,^{a,b} and Eli Peli^a

^aHarvard Medical School, Massachusetts Eye and Ear, Department of Ophthalmology, Schepens Eye Research Institute, 20 Staniford Street, Boston, Massachusetts 02114, United States

^bUniversity of Electronic Science and Technology of China, School of Optoelectronic Information, No. 4, Section 2, North Jianshe Road, Chengdu 610054, China

Abstract. Visual prostheses require an effective representation method due to the limited display condition which has only 2 or 3 levels of grayscale in low resolution. Edges derived from abrupt luminance changes in images carry essential information for object recognition. Typical binary (black and white) edge images have been used to represent features to convey essential information. However, in scenes with a complex cluttered background, the recognition rate of the binary edge images by human observers is limited and additional information is required. The polarity of edges and cusps (black or white features on a gray background) carries important additional information; the polarity may provide shape from shading information missing in the binary edge image. This depth information may be restored by using bipolar edges. We compared object recognition rates from 16 binary edge images and bipolar edge images by 26 subjects to determine the possible impact of bipolar filtering in visual prostheses with 3 or more levels of grayscale. Recognition rates were higher with bipolar edge images and the improvement was significant in scenes with complex backgrounds. The results also suggest that erroneous shape from shading interpretation of bipolar edges resulting from pigment rather than boundaries of shape may confound the recognition. *© 2016 SPIE and IS&T* [DOI: 10.1117/1.JEI.25.6.061619]

Keywords: edge detection; feature detection; object recognition; contrast polarity; bipolar edge; cusp. Paper 16370SSP received May 1, 2016; accepted for publication Nov. 17, 2016; published online Dec. 22, 2016.

1 Introduction

Human object recognition is a complex process of interpretation. Various models of object recognition have been proposed, including view-based models^{1,2} and structural description models.^{3,4} View-based models assume that objects are represented as collections of viewpoint-specific local features, while structural description models, such as the recognition by components model, assume objects are represented as configurations of simple volumes or parts ("geons" or geometric cones) and recognized using a bottom-up process.^{3,4} Whether object recognition is purely based on a view-invariant structural description (object-centered models) or on view-specific features (view-based models) is arguable; however, features that include edge lines, narrow bars, and cusps are presumed to be visual system primitives in object recognition.^{1–6}

Although the key features are essential for object recognition, the natural color image is a better representation for object recognition than only a feature-detected image.⁵ However, in limited display conditions, such as low dynamic range (the number of displayable or perceivable gray levels) or low-resolution display, or in limited observation conditions (visually impaired people), such as patients who experience reduced visual acuity and contrast sensitivity due to macular disease,^{7,8} a scene filtered to feature representation may be an effective visual descriptor for object recognition.^{4,5}

Various visual prostheses for blind people have been developed, including retinal and cortical implants ^{9,10} and visual sensory substitution devices (using inputs from other

sensory channels). Although some level of vision can be partially restored by such prostheses, extremely low spatial resolution (60,¹¹ 400,¹² and 1500^{13,14} electrodes resolution) and low dynamic range (binary or at most 3 or 4 levels¹⁵) limit the utility of the current visual prostheses. Such devices have been tested with blind subjects and demonstrated improvement of discrimination performance, but limited ability in providing the recognition of objects.^{12,16,17} Dramatic improvement in the resolution and dynamic range of the visual prostheses is unexpected in the near future. Providing effective representation methods suited for the limited resolution and dynamic range of the visual prostheses is a promising approach to improve object recognition.^{18–21}

Various feature detection (mostly edges) and representation methods have been proposed $^{6,22-24}$ to effectively convey information from scenes for object recognition in the human visual system.^{5,21,22,25} Since edge images are thought to provide useful representations for object recognition, adding high contrast edge information has been proposed as a way of enhancing image visibility for visually impaired people.^{7,26} For the same reasoning and due to the limited dynamic range of visual prostheses, many have proposed using binary edge representations for these systems.^{18,21,27} However, edge representations in human vision were fundamentally different from those used in computational algorithms.⁵ Generally, edge detection algorithms merely locate edge pixels defined by luminance or color differences within a small region of the image.^{5,28} Differences exceeding a threshold are represented in black or white pixels on a contrasting background (i.e., binary edges). The thresholds are

^{*}Address all correspondence to: Jae-Hyun Jung, E-mail: jaehyun_jung@meei. harvard.edu

^{1017-9909/2016/\$25.00 © 2016} SPIE and IS&T

typically selected manually, though a statistical method was proposed.²⁹ In comparison, edge extraction in human vision is thought to be an abstraction of the scene using global information (e.g., shading includes shadow created by illumination condition, perceived depth, and overcoming occlusion) to combine features that form regions, volumes, or some other intermediate-level representation (e.g., grouping and segmentation).^{1–5}

Using cluttered scenes as stimuli for object recognition studies may illustrate the difference between the efficacy of object recognition from original and feature-extracted images (e.g., edge image). The human visual system is thought to convert the scene to feature components by segregating the key objects and suppressing background clutter using global information and intermediate level processing.⁵ However, if an observer is presented with the binary edge image rather than the original image, segregating the target object in cluttered scenes may be more challenging because all edges from the foreground and background are represented equally without useful global information. While edges from background clutter do not contribute to the target object recognition, they frequently interfere with the edges of the target object.^{5,21}

Sanocki et al.⁵ compared object recognition between fullcolor images and binary edge images, with and without manual removal of background clutter following the 1-s presentation. The average recognition rate of binary edge images was only 41.2% with the background (70% without background), significantly lower than full-color images (90.6% and 89.8% with and without background, respectively). Manual removal of background clutter significantly improved object recognition of binary edge images because segregation requires considering different groupings of edges as figure and ground; the increased number of edges may greatly increase complexity. However, the effect was minimal in the color images, presumably because of a ceiling effect.⁵ Processing that facilitates background clutter suppression in binary edge images may improve object recognition. Jung et al.²¹ also compared object recognition in binary edge images with and without background clutter. The binary edge images without the background clutter are 5.6 times more likely to be recognized than binary edge images with background clutter.

Although removing background clutter from binary edge image resulted in significant improvement in object recognition, it is not solely based on image information and requires additional captured information, such as depth map, focus difference^{21,30–32} or high-level image processing for computational object recognition.³³ In edge images, improving the representation method that may convey any global information to suppress background clutter with minimal cost rather than an additional process for background clutter removal is promising for visual prostheses. The bipolar edge filtering represents contrast polarity of the edge line in addition to the location of the binary edge line representation.^{6,7,26} The contrast polarity is one of possible promising information^{24,34} to be represented with low additional computational cost, requiring only one more gray level for the visual prostheses. We have suggested that the bipolar edge representation may provide an advantage for prosthetic systems that can provide more than two levels of stimulation.^{18,21,27}

2 Bipolar Edge Images

2.1 Peli's Bipolar Edge Detector

Peli⁶ proposed a bipolar edge detection algorithm motivated by a model of the human visual system. Human visual system models perform multiscale bandpass filtering, and the measured contrast sensitivity function is a measure of the system detection threshold in each band. Peli's bipolar edge detector uses one-octave wide bandpass filters separated by one octave. Binary phase congruence across a range of scales results in three gray levels to represent two different polarities of contrast (black and white lines over gray background), as shown in Fig. 1.

Whereas binary edge image represents only the location of edges and cusps [Fig. 1(b)], bipolar edge image [Fig. 1(c)] can represent the luminance transition (brighter and darker side) of features as well as the location. Note that the binary edge image converted from the bipolar edge image [Fig. 1(d)] verifies the similarity of the edge location between the binary edge and bipolar edge images.

Peli's bipolar edge detector has no free parameters that were fitted to the images or adjusted in any way. The only adjustable parameter is the assumed angular size of the image. Due to the contrast sensitivity function and bandpass filtering, more features are detected by Peli's bipolar edge detector in the nearer image (wider angular size) than in the farther image (narrower angular size) as by the human visual system, if we ignore the changes in contrast sensitivities with retinal eccentricity.³⁵

Bipolar edge representation has been used for image enhancement for visually impaired people by superimposing the high contrast edges on the original image.^{7,24,26,34} Although the binary edge enhancement may have highest contrast on the contrasting background, the background in the superimposed image is usually complex, which results in different contrast in the enhanced image. Bipolar edges work better than binary edges as they would enhance feature visibility over both dark and bright backgrounds.^{6,24,26,34}

2.2 Shape from Shading in Bipolar Edge Image

We hypothesize that the bipolar edge image can support better object recognition performance than the binary edge image with normal subjects even with background clutter. In an image of natural scenes, the luminance difference across features can be caused by reflection changes within objects (caused by varying pigmentation) or by shading in the shaded background or shaded side (or casted shadow) of an occluding object. Due to the assumption that the illumination source is typically above, objects are expected to cast shadows below;³⁶ therefore, the bottom side or farther side of edges is expected to be dark while the upper side or closer side of edges is bright. This shading effect is lost in binary edges but is preserved in bipolar edges. Therefore, bipolar edges provide shape and depth cues that could aid in object recognition.³⁷

Figure 2 shows bipolar edge results of uniformly pigmented objects as an example. In uniformly pigmented objects, the polarity of the bipolar edge is only affected by the illumination condition. Using the shading effect, the bipolarity of edge lines then spontaneously presents the edge location and relative depth and/or shape of the object, whereas the binary edge results only show the location of



Fig. 1 Bipolar edge representation: (a) original color image of the target object (champagne flute), (b) binary edge image filtered by Canny detector,²³ and (c) bipolar edge image filtered by Peli's bipolar edge detector,⁶ which represents both the location of a feature and the polarity of luminance on both sides of the feature. There are also unipolar edge lines of either polarity that represent thin features of one polarity over a contrasting background (e.g., letters on books, specular highlight on the flute, and the boundary between books). (d) Binary representation derived from the bipolar edge image (c) by eliminating the polarity information shows the similarity of edge location with the binary edge image in (b). Due to the double edges converted from bipolarity, the edge lines in (d) look thicker and brighter.



Fig. 2 Bipolar edge representation of uniformly pigmented objects: (a) original color image of multiple sculptures, (b) binary edge image filtered by Canny detector, and (c) bipolar edge image filtered by Peli's bipolar edge detector. Due to the uniform pigmentation of the sculptures, only the shape from shading affects the polarity of edge lines. Bipolar edge lines represent the boundaries between sculptures and background and may help to segregate the objects from the background. Unipolar lines due to protruding cusps (bright) and recessed (dark) details on the sculptures also provide depth information. For example, the squirrel sculpture in the bipolar edge image (c) is much easier to segregate and recognize than the binary edge image (b).

edges and cannot distinguish the protruding and recessed features.

In addition, bipolar edge images can represent more aspects of features than binary edge images which represent only edge line location. As seen in Fig. 2(c), bipolar double lines represent boundaries of two surfaces (or two pigments) and unipolar lines (white or black single lines) represent positive or negative cusps that may result from protruding or recessed features. Although the white unipolar lines may also represent specular highlight, there is no geometrical feature and shape information.

The contrast polarities of features may be helpful as they provide more information about the scene. That bipolarity may be interpreted and is frequently and spontaneously perceived as shape from shading (black lines for shadows and white lines for lighting surfaces).³⁶ When such perception is correct, it may be helpful in segregating an object from the background.

3 Methods

3.1 Dataset

To evaluate the impact of bipolarity on object recognition, we compared object recognition rates of bipolar edge and binary edge images by normally sighted subjects. Comparing the object recognition rate between two different representation methods required an image dataset that contains one target object at the center of each image with natural background clutters. Object recognition performance is highly dependent on the difficulty of the image dataset. We were concerned about a possible ceiling or basement effect limiting our ability to find an effect even if it exists. For that reason, the object recognition rate in one condition should be intermediate to enable either an increase or decrease recognition rate to be found. We selected the Sanocki et al.'s⁵ dataset, in which 68 subjects had 41.2% average object recognition rate across 16 objects using binary edge images.

The Sanocki et al.'s dataset contains 16 different office and household items at the center of a scene with varying levels of background clutter (Fig. 7). The dataset is shared on data sharing section on our web page (Ref. 38), including original and edge images. We attempted to follow Sanocki et al.'s⁵ method in an effort to achieve similar recognition rates for binary edge images. The resolution of the color images in this dataset is 768×512 and the images were converted to grayscale images before edge detection processing.

Binary edge images were calculated with the Canny edge detector²³ in Image Processing Toolbox of MATLAB[®] R2013b (MathWorks, Natick, Massachusetts). Sanocki et al.⁵ manually adjusted three parameters (sigma, upper threshold, and lower threshold) and the experimenter selected those to achieve a representation of the target object, which included edges important to the target object while having minimal noise edges. However, the exact parameters used by Sanocki et al. were not reported and not available to us. We adjusted the high threshold and sigma parameters for each image while the low threshold was set to 0.4 of the high threshold. We used the few images presented in their paper as a guide and similarly selected the best representation of each image. The dataset of our binary edge images is in the left column in Fig. 8.

The dataset of bipolar edge images (the right column in Fig. 8) was generated using Peli's bipolar edge detector.⁶ This algorithm has no free parameters, other than the presumed angular image span which we adjusted for each image in an effort to visually match the level of extracted features to the binary edge images. The presumed angular image span affects the level of details represented by the detected features and the threshold being used. Note that we did not change the resolution of the image and only

applied a different angular size span, which adjusts the threshold applied at different scales. Although the bipolar edge images have the additional information of contrast polarity, we tried to have the same contents and a similar level of details in both binary and bipolar edge images with adjusting the presumed image span in the bipolar algorithm. We visually compared the location and detail of features between the binary edge and bipolar edge images. For example, the binary image converted from bipolar edge image without polarity information [Fig. 1(d)] represents only the location and detail of features, which closely correspond to the information in the binary edge image except for double lines and noises. As verification, the second column in Fig. 8 shows the binary dataset converted from the bipolar edge dataset for comparing the location of features.

To present only one type of processing per object for each subject, the 16 images were split into two groups of similar difficulty based on the recognition rates reported by Sanocki et al.⁵ Images were sorted by the reported recognition rate, and image pairs were formed from those with consecutive recognition rates. One image from each pair was assigned randomly to one group and the other image to the other group. The average recognition rate by Sanocki et al. 8% and 40.6% for each group, respectively, as shown in Fig. 3. Paired sample *t*-test (p = 0.35) showed that the two groups were not significantly different. Each subject viewed binary edge images from one of the two groups and bipolar edge images from the other group. The group presentations were counterbalanced between subjects.

3.2 Object Recognition Test

Twenty-six normally sighted subjects (nine men) aged 21 to 67 participated. The study was approved by the Human Studies Committee of Massachusetts Eye and Ear, and written informed consent was obtained from all participants. Subjects were seated 33 in. from an LCD monitor and the image width was 8 in. to approximately match the 14-deg angular image span used by Sanocki et al. We explained the task to subjects during a training session where we presented two images ("umbrella" and "camera") in both binary edge and bipolar edge versions. We also explained the binary edges and bipolar edges (the meaning of edge polarity) but did not suggest the polarity was a cue to depth. The likely position (i.e., image center) and size (i.e., the biggest object in the image) of the target objects were indicated to subjects during training.

The test was performed in a dimly lit room. The 8 binary edge images were presented first followed by 8 bipolar edge images. At the beginning of the test in each condition, the training umbrella image in that condition was displayed as an example. Each test image was preceded by an audible alerting beep and disappeared 1 s later, concurrent with a second beep. The subjects were then asked to name the object at the center or describe the use of the object if they could not name it. The operator wrote down the subjects' responses and provided no feedback. In determining the response veracity when the name of the object was not given, describing object usage was valued more than a general description of the object's shape. The next image trial started after the subject pressed any button on the keyboard.



Fig. 3 The average recognition rate for each object in the test conditions and in Sanocki et al.⁵ For counterbalancing, we divided Sanocki's dataset into two groups ("Lamp" to "Compact Disc" as one group, and "Wastebasket" to "Sprinkler" as the other group) based on their recognition results. Our subjects' recognition rates with the binary edge images were much higher than theirs, leaving little room for improvement with the bipolar processing (ceiling effect).

4 Results

Average overall object recognition rates for binary and bipolar edge images are given in Table 1. Recognition rates for bipolar and binary images were 79.3% and 71.6%, respectively. The modest average improvement of 7.7% approached significance (p = 0.069). Note, however, that the recognition rate of our subjects, in the binary edge condition, for 4 of the 16 images was 100% (ceiling effect) preventing any possible improvement (Fig. 3). For three more images (thus 7 of 16), the recognition rate reached 100% with the bipolar edge images, which may also represent a ceiling effect and limit the ability to achieve higher improvement. The ceiling effect was due to our subjects' recognition rate with the binary edge images (71.6%) being much higher than Sanocki et al.'s subjects (41.2%) although we used the same dataset, edge filtering method, and presentation paradigm. Our subjects' results and Sanocki et al.'s results for recognizing various objects were correlated moderately (Pearson's correlation coefficient $\rho = 0.53$ with p = 0.035). Note that the recognition rate of a binary edge in the result of Sanocki et al. was derived from Figure B1 in their paper.⁵

The results for images in which a ceiling effect was not suspect, "Watering Pot" (92.3% versus 46.2%, p < 0.01; Fig. 4) and "Sprinkler" (53.8% versus 0%, p < 0.001; Fig. 5), showed significantly improved recognition rates in bipolar edge images.

We further analyzed the results using a binary logistic regression model in SPSS 11.5. The model correctly classified 94.9% of the correct recognitions and 78.4% of incorrect

Table 1 Average recognition rate, %, standard error, SE, and significant level, *p*, between conditions.

	Bipolar edge	Binary edge	Binary edge (Sanocki et al.)
Average (%)	79.3	71.6	41.2
SE (%)	2.8	3.1	3.3
p	0.069		NA

recognitions. Odds ratio (O_R) is defined as the ratio of the odds with a variable to the odds of a reference, where Ris the probability of the binary event (recognized or not), the complementary probability is 1 - R, and the odds of an event are defined as R/(1-R). Therefore, where R is the recognition rate with a variable and R' is the recognition rate of the reference for this variable in this experiment, the odds of reference are defined as R'/(1-R'). Note that sometimes the odds ratio is regarded as the multiplication factor of the event, but it is only the case if the probability of the binary event is very small (then the complementary probability is ~1). The odds ratio $(O_R)^{21}$ was 1.52 and approached significant level (p = 0.069), indicating that the odds of recognition in the bipolar edge image are 1.52 times more than the odds of recognition in the binary edge image when holding all other variables constant This means the recognition rate in the bipolar edge image (R_{Bipolar}) could be predicted by the odds ratio (O_R) and the recognition rate in the binary edge image (R_{Binary}) using Eq. (1). This ratio is also limited by the ceiling effect

$$R_{\rm Bipolar} = \frac{O_R\left(\frac{R_{\rm Binary}}{1-R_{\rm Binary}}\right)}{1+O_R\left(\frac{R_{\rm Binary}}{1-R_{\rm Binary}}\right)}.$$
(1)

For example, if the recognition rate in the binary edge image was 41.2% as found by Sanocki et al., the recognition rate for the bipolar edge images would be expected to be 51.6%.

5 Discussion

Bipolar edge images contain more information than binary edge images because they represent the contrast polarity of features. The bipolar edge images also distinguish edges from cusps and mark the polarity of cusp features (protruded or recessed). Further, the bipolarity can provide three-level shading information, thus serving as a shape-from-shading cue. Therefore, we were expecting improved recognition with the bipolar edge images especially in the limited display condition of the visual prostheses. The improvement we found in object recognition with bipolar edges was modest

Journal of Electronic Imaging

061619-5



Fig. 4 Image of "Watering Pot" that resulted in significant improvement in the bipolar edge image. Displayed angular size was 14 deg horizontally; that is the size watching this printed figure at 21.7 cm away. (a) Color image of the watering pot in front of wire meshes and other clutter, (b) binary edge image results that 5 of 13 (40%) subjects identified the watering pot, and (c) the bipolar edges and cusps act as a depth cue from shading and may have helped segregate the object from the background clutter. Recognition rate was substantially improved to 92.3% (12 of 13). Note that the images were scanned from prints resulting in artifactual outside edges that slightly interfere with the presentation.



Fig. 5 Image of "Sprinkler" that was much better recognized from bipolar images: (a) original color image, (b) binary edge image that no subject was recognized, and (c) bipolar edge image was recognized correctly in 53.9% (7 of 13 subjects) of presentations. While the location of the object of interest is similarly apparent in both cases, apparently the structure of the sprinkler is easier to note in the bipolar edge image. Presumably, this is affected by the pronounced three-dimensional structure notable in the bipolar image.

and only approached significance. This might be caused by a ceiling effect. Indeed the recognition rate of 2 of the 3 images that were recognized by less than 50% of the subjects from the binary edge images was much better recognized from the bipolar edges (Figs. 4 and 5). Our interest is in using the bipolar representation in an imaging system for visual prostheses.²¹ In this application, the very limited resolution will severely restrict recognition, thus eliminating the ceiling effect, and may show the higher level of improvement.

Improvement in recognition rate might be supported by better segregation of the object from the background and by shape from shading cues for uniformly pigmented objects. As shown in Fig. 4(c), the watering pot's edges are white on the outside and black inside because of the contrast between the dark pigment object and its bright pigment background. Because black lines can be perceived as shadows and white lines can be perceived as brighter side, if the object is surrounded by white inner edge lines with black outer edge lines the object may be perceived as closer than the background clutter and vice versa. This effect may help to segregate the object from background clutter in scenes with a complex background (Figs. 4 and 5).

A uniformly pigmented object, such as the watering pot and the basket in Fig. 4, may have specular points or lines on glossy surfaces represented as unipolar edges in the bipolar edge image (e.g., the decorative pattern on the watering pot and the curves on the basket). Understanding global illumination direction may provide a reference to shape from shading.

Luminance differences caused by a pigment (darker than background) of the local object could confound the depth cues (Fig. 6). Due to the black color of an outer briefcase, the shape from shading is perceived inconsistently in the bipolar edge. The border between the inside and outside of a uniformly pigmented briefcase would cause a black line of a bipolar edge inside the case (due to shading) and a white line edge in the outer area. However, the polarity of edge lines pointed by the red arrows in Fig. 6(c) is reversed due to the black color of the briefcase, which may have caused a misperception (the interior of the briefcase may appear higher than it should be) that might affect its recognition rate. If we correct the effect of the pigment to match the shape from shading in the scene [Fig. 6(d)], the interior of the briefcase may appear lower than the edge and this may help to recognize the object.

We did not investigate directly the impact of background clutter or depth cues from illumination in this pilot study. In future studies, to reveal the impact of bipolar edges in segregating the object from background clutter, controlling the complexity of background or the illumination (e.g., direction





Fig. 6 Image of "Briefcase" that was better recognized from binary edges than from bipolar edges. (a) Original color image. (b) Binary edge image calculated using the Canny detector. Average recognition rate was 53.9% (7 of 13 subjects). (c) The bipolar edge detected by Peli's method.⁶ The recognition rate was reduced to 38.5% (5 of 13 subjects). Contrast polarity in the bipolar edge image is consistent with the shape from shading for numerous bright objects in the same scene (e.g., papers on the shelves). However, on the top of the front and right side of the briefcase (pointed by red arrows), the polarity is inconsistent with shape from shading. (d) The bipolar edge image was manually edited to reverse the polarity of edge lines pointed by the red arrows. The edited polarity restores the correct shape of the briefcase and may help to recognize it better (this prediction was not tested). Displayed angular size was 14 deg horizontally; that is the size watching this printed figure at 31 cm away.

and surface material: glossy or matte) may be necessary. To reduce the ceiling effect, shorter display times or the use of lower resolution images may be implemented. In future studies, we plan to match the binary edge images to the bipolar edges by extracting the binary edges from the bipolar edge image [Fig. 1(d) and the middle column in Fig. 8].

Edge enhancements by adding high contrast edge information for visually impaired patients have been proposed.^{7,26,34,38,39} Both binary edges and bipolar edges have been employed in such studies.^{6,7,24,26,34} The approach has also been implemented in augmented reality where high contrast edges are added virtually to objects. Only binary (bright edges) can be added in optical see-through systems;^{8,40,41} however, in video see-through systems, bipolar edges may be used.^{7,42} A number of studies have demonstrated a preference for enhanced images, but performance improvements have not been clearly demonstrated.³⁹ This paper is a first attempt to compare performance with bipolar and binary edge images.

All edge images were using the same dynamic range on the same monitor. The bipolar edges are reproduced with slightly higher contrast. This is more apparent for the object of interest in most of the images. The higher contrast which is apparent in the bipolar images and easier to gauge in the binarized bipolar images (middle column in Fig. 8), may also contribute to the better segregation of the object and to the recognition improvement. This may be an artifact of the intentional photographic aiming for the center object in the images of the dataset. In a future study, we will also plan to test the impact of the bipolar filtering to the object recognition in low-resolution images close to the current visual prostheses (up to 1500 electrodes).

Appendix: Image Dataset

A1 Image Dataset

Sixteen image dataset captured by Sanocki et al.⁵ Sanocki approved the public release of this dataset in this appendix. The original dataset is divided into two groups (left and right columns, Fig. 7) for counterbalancing. Sanocki et al.'s recognition rates in binary edge image⁵ are presented in the bracket.

A2 Different Edge Filtering of Image Dataset

In Fig. 8, the left column shows 16 binary edge image dataset filtered by Canny edge detector.²³ The middle column shows



'Lamp' (100%)



'Wastebasket' (83%)



'Coffee pot' (74%)



'Watering pot' (64%)

'Motorcycle' (56%)



'Cinder block' (52%)



'Phone' (61%)



'Pan' (49%)

Fig. 7 The Sanocki et al.'s dataset contains 16 different office and household items at the center of a scene with varying levels of background clutter. Left and right columns indicates two different groups in the study. The name of object and recognition rates in Sanocki et al.'s binary edge image are noted.



'Fire extinguisher' (34%)



'Computer' (34%)



'Briefcase' (18%)



'Picnic table' (14%)



'Chair' (10%)



'Microwave' (10%)



'Compact disc' (0%)



'Sprinkler' (0%)

Fig. 7 (Continued)

Journal of Electronic Imaging

061619-9



'Phone' (92.3%)

(100%)

Fig. 8 Edge filtering of image dataset. Binary edge image filterd by Canny edge detector (left column), bipolar edge image filtered by Peli's bipolar edge detector (right column), and binary edge image converted from bipolar edge image (middle column) are presented. The name of object and recognition rates in our study are noted.

16 binary images converted from bipolar edge image dataset filtered by Peli's bipolar edge detector⁶ in the right column. The recognition rates are represented under the image, and the images are sorted by the recognition rate in the bipolar edge image. Due to the 2-in. width of the image dataset, 8.1in. watching distance from the image will make the same angular size (14 deg) of the image with the experiment. In the binary image (middle column) converted from the bipolar edge image, the edge lines may seem thicker and to have higher contrast than the binary edges in the left column due to the double edges from bipolar lines, but the contrasts of the edge lines are the same.



Fig. 8 (Continued)

Jung, Pu, and Peli: Comparing object recognition from binary and bipolar edge images for visual prostheses



Fig. 8 (Continued)

Acknowledgments

This work was supported in part by the Office of the Assistant Secretary of Defense for Health Affairs, through the Clinical and Rehabilitative Medicine Research Program. Joint Program Committee 8, Neurosensory and Rehabilitative Research Program under Award No. W81XWH-16-1-0033. Opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the Department of Defense. We thank Merve Tuccar for help with data collection and processing and Dr. Thomas Sanocki for sharing the image dataset.

References

 M. J. Tarr and H. H. Bülthoff, "Image-based object recognition in man, monkey and machine," *Cognition* 67(1–2), 1–20 (1998).

- 2. M. Riesenhuber and T. Poggio, "Models of object recognition," Nat. Neurosci. 3, 1199–1204 (2000).
- I. Biederman, "Recognition-by-components: a theory of human image understanding," *Psychol. Rev.* 94, 115–147 (1987).
 I. Biederman and P. Kalocsai, "Neurocomputational bases of object and
- face recognition," Philos. Trans. R. Soc. B: Biol. Sci. 352(1358), 1203-1219 (1997).
- 5. T. Sanocki et al., "Are edges sufficient for object recognition?" J. Exp. ychol. Human Percept. Perform. 24(1), 340–349 (1998).
- 6. E. Peli, "Feature detection algorithm based on a visual system model," *Proc. IEEE* **90**, 78–93 (2002).
- E. Peli et al., "Wide-band enhancement of television images for people with visual impairment," *J. Opt. Soc. Am. A* 21(6), 937–950 (2004).
 A. D. Hwang and E. Peli, "An augmented-reality edge enhancement application for Google Glass," *Optom. Vis. Sci.* 91(8), 1021–1030 (2014). (2014).
- È. Margalit et al., "Retinal prosthesis for the blind," Surv. Ophthalmol. 47(4), 335–356 (202).
 10. J. M. Ong and L. da Cruz, "The bionic eye: a review," *Clin. Exp.*
- Ophthalmol. 40(1), 6-17 (2012).
- A. K. Ahuja and M. R. Behrend, "The Argus II retinal prosthesis: factors affecting patient selection for implantation," *Prog. Retinal Eye Res.* **36**(4), 1–23 (2013).
- 12. A. Nau, M. Bach, and C. Fisher, "Clinical tests of ultra-low vision used to evaluate rudimentary visual perceptions enabled by the BrainPort vision device," *Transl. Vision Sci. Technol.* **2**(3), 1 (2013).
- 13. M. A. Singer et al., "Improving quality of life in patients with end-stage clin. Ophthalmol. **6**, 33–39 (2012).
- 14. K. Stingl et al., "Functional outcome in subretinal electronic implants depends on foveal eccentricity," Invest. Ophthalmol. Vis. Sci. 54(12), 7658-7665 (2013).
- V. G. Chouvardas, A. N. Miliou, and M. K. Hatalis, "Tactile displays: overview and recent advances," *Displays* 29(3), 185–194 (2008).
- M. S. Humayun et al., "Interim results from the international trial of second sight's visual prosthesis," *Ophthalmology* 119(4), 779–788 (2012).
- 17. E. Zrenner et al., "Subretinal electronic chips allow blind patients to read letters and combine them to words," Proc. R. Soc. B: Biol. Sci. 278(1711), 1489-1497 (2011).
- W. Al-Atabany et al., "A processing platform for optoelectronic/opto-genetic retinal prosthesis," *IEEE Trans. Biomed. Eng.* 60(3), 781–791 (2013).
- 19. N. Parikh, L. Itti, and J. Weiland, "Saliency-based image processing for
- N. Fairki, E. Hi, and J. Weirald, "Satisfy-based image processing for retinal prostheses," J. Neural Eng. 7(1), 16006 (2010).
 J. D. Weiland et al., "Smart image processing system for retinal pros-thesis," in Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC'12), pp. 300–303 (2012).
 J. H. Jung et al., "Active confocal imaging for visual prosthesis," Vision p. 11107: p. 162 (2015).
- Res. 111(Pt. B), 182–196 (2015).
 M. D. Heath et al., "A robust visual method for assessing the relative
- performance of edge-detection algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(12), 1338–1359 (1997).
- 23. J. Canny, "A computational approach to edge detection," IEEE Trans. Pattern Anal. Mach. Intell. PAMI-8(6), 679-698 (1986).
- L. Itan and Y. Yitzhaky, "Performance of visual search tasks from various types of contour information," *J. Opt. Soc. Am. A* **30**(3), 392–402 24. (2013).
- 25. H. L. Apfelbaum et al., "Inattentional blindness and augmented-vision displays: effects of cartoon-like filtering and attended scene, *Ophthalmic Physiol. Opt.* **28**(3), 204–217 (2008).
- 26. E. Peli, "Wide-band image enhancement," US Patent No. 6,611,618 B1 (2003).
- J. A. Dowling, A. Maeder, and W. Boles, "Mobility enhancement and assessment for a visual prosthesis," *Proc. SPIE* **5369**, 780–791 (2004). C. Zhou and B. W. Mel, "Combining multiple cues for contour detec-27.
- 28. tion: lessons from (and to) the visual cortex," J. Vision 3(9), 252-252 (2003).

- 29. Y. Yitzhaky and E. Peli, "A method for objective edge detection evaluation and detector parameter selection," *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(8), 1027–1033 (2003).
- 30. J.-H. Jung et al., "Reconstruction of three-dimensional occluded object using optical flow and triangular mesh reconstruction in integral imag-Opt. Express 18(25), 26373–26387 (2010). ing,'
- R. A. El-laithy, H. Jidong, and M. Yeh, "Study on the use of Microsoft Kinect for robotics applications," in *IEEE/ION Conf. Proc. of the IEEE* Position Location and Navigation Symp. (PLANS'12), Myrtle Beach,
- South Carolina, pp. 1280–1288 (2012).
 D. Aloni and Y. Yitzhaky, "Automatic 3D object localization and isolation using computational integral imaging," *Appl. Opt.* 54(22), 6717–

- lation using computational integral integral, *Appendix Representation of the second state of the* 1762-1774 (1991).
- 36. S. K. Khuu, V. J. Honson, and K. L. Challinor, "The influence of cast shadows on the detection of three-dimensional curved contour struc-ture," *Perception* **45**(3), 425–442 (2016).
- 37. R. P. O'Shea, S. G. Blackburn, and H. Ono, "Contrast as a depth cue," Vision Res. 34(12), 1595–1604 (1994).
- 38. J. S. Wolffsohn, D. Mukhopadhyay, and M. Rubinstein, "Image enhancement of real-time television to benefit the visually impaired," m. J. Ophthalmol. 144(3), 436-440 (2007).
- 39. M. Kwon et al., "Contour enhancement benefits older adults with simulated central field loss," *Optom. Vision Sci.* **89**, 1374–1384 (2012). 40. J.-H. Jung et al., "Dataset ZIP," 2016, *Comparing Object Recognition*
- from Binary and Bipolar Edge Images, http://serinet.meei.harvard.edu/ faculty/peli/shared/Bipolar_filtering_dataset.zip (16 December 2016). 41. E. Peli et al., "Applications of augmented vision head-mounted systems
- J. P. Rolland and H. Fuchs, "Optical vision read-monitor systems in vision rehabilitation," J. Soc. Inf. Display 15(12), 1037–1045 (2007).
 J. P. Rolland and H. Fuchs, "Optical versus video see-through head-mounted displays in medical visualization," *Presence Teleoperators Virtual Environ.* 9(3), 287–309 (2000).

Jae-Hyun Jung is an investigator at the Schepens Eve Research Institute, Massachusetts Eye and Ear, and an instructor in ophthalmology at Harvard Medical School. He received his BS degree in electrical engineering from Pusan National University, Republic of Korea, in 2007 and his PhD in electrical engineering from Seoul National University in 2012. Since then he has worked at the Schepens Eye Research Institute, Massachusetts Eye and Ear, Harvard Medical School. His work was focused on optical engineering for low vision rehabilitation and imaging systems for visual prostheses.

Tian Pu received his PhD in optical engineering from Beijing Institute of Technology in 2002. Afterward, he worked as senior research engineer in Dayang Imaging Company, Beijing. After several years of experience in industry, he has worked as an assistant professor at the School of Optoelectronic Information, the University of Electronic Science and Technology of China since 2008. His research interests include image processing, computer vision, and signal processing.

Eli Peli received his BSc and MSc degrees in electrical engineering from Technion-Israel Institute of Technology, Haifa, Israel, in 1978, and his OD from the New England College of Optometry, Boston, Massachusetts, in 1983. Since then he has been at the Schepens Eye Research Institute, where he is the Moakley Scholar in Aging Eve Research and a professor of ophthalmology at Harvard Medical School.