DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-111

Comparing object recognition from binary and bipolar edge features

Jae-Hyun Jung[1], Tian Pu[1, 2], Eli Peli[1]*
1. Schepens Eye Research Institute, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA
2. School of Optoelectronic Information, University of Electronic Science and Technology of China, Chengdu, China

Abstract

Edges derived from abrupt luminance changes in images carry essential information for object recognition. Typical binary edge images (black edges on white background or white edges on black background) have been used to represent features (edges and cusps) in scenes. However, the polarity of cusps and edges may contain important depth information (depth from shading) which is lost in the binary edge representation. This depth information may be restored, to some degree, using bipolar edges. We compared recognition rates of 16 binary edge images, or bipolar features, by 26 subjects. Object recognition rates were higher with bipolar edges and the improvement was significant in scenes with complex backgrounds.

Object recognition in edge images

Human object recognition is 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, which assumes 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 either purely based upon a view-invariant structural description (object-centered models) or upon view-specific features (view-based models) is arguable, however, edges and cusps are presumed to be visual system primitives in object recognition [1-5]. Therefore, a scene filtered to edge representation may be an effective visual descriptor for object recognition [4, 5].

Numerous edge detection methods have been developed [6-8] and said to effectively convey the essential feature of a scene to observers [5, 6, 9, 10]. However, Sanocki et al. [5] argued that edge representations in human vision were fundamentally different from those used in computational algorithms. Generally, edge detection algorithms merely locate edge pixels defined by luminance or color differences within a small region of the image [11] Differences exceeding a threshold are represented in black or white pixels on a contrasting background (i.e., binary edges). In comparison, edge or contour extraction in human vision is thought to be an abstraction of the scene using global information to combine edges that form regions, volumes, or some other intermediate-level processing (e.g., grouping, segmentation, etc.) [1-5].

Using cluttered scenes as stimuli may illustrate the difference between the efficacy of unipolar (i.e., binary edge) and bipolar features. The human vision converts the scene to edge information components by segregating the key objects and suppressing background clutter using global information and intermediate level processing. However, if an observer sees the edge image rather than the original image, segregating and the target object is more challenging in cluttered scenes because all edges from the scene are represented without the benefit of intermediate-level processing and global information. Edges from background clutter do not contribute to object recognition because they frequently interfere with the object edges [5, 10].

Sanocki et al. [5] compared object recognition between full-color images and binary edge images, with and without manual removal of background clutter. The average recognition rate of binary edge images was only 45.7% (69% without background), significantly lower than full-color images (90.4% and 90.8% with and without background, respectively). Removal of background clutter significantly improved object recognition of edge representations but not of color images, presumably in part because humans are able to effectively suppress background clutter in color images. Processing that facilitates background clutter suppression in edge images [10] may improve object recognition. Additional information may indicate whether edges belong to background clutter or to the target object.

Bipolar feature images

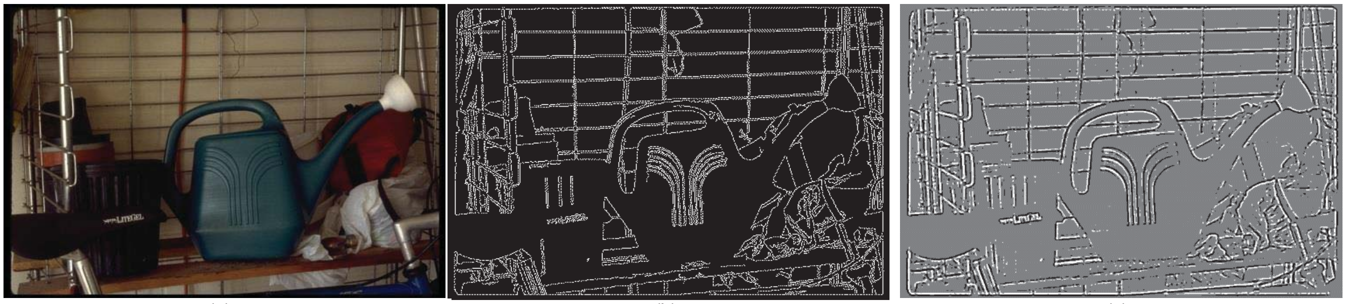
Peli proposed a bipolar edge/feature detection algorithm motivated by a model of the human visual system [8]. Whereas unipolar edges only where luminance changes occur, bipolar edges represent the location and the polarity of the luminance transition (the bright and dark side), a cusp, or very narrow bars relative to their background. Based on phase congruence across scales in a multi-scale structure, the bipolar feature representation marks the darker and brighter sides of the edge as black and white lines, respectively, over a gray background.

In natural scenes, the luminance difference across an edge can be caused by reflection changes within objects (caused by pigments) or differences in brightness between the background and an occluding object above, or in front of, the background. The polarity of such edges provides additional information about the relative reflections of objects and their components. In addition, edges can indicate the interaction between illumination and the shape of objects (shading). Occluding objects in front or above background surfaces will frequently cast shadows on the background resulting in the bright side of the occluding edge inside the object. This information may be revealed by bipolar edges as useful depth information that may aid in object and ground segmentation and thus help suppress clutter and aid in object recognition.

As the binary edge (Fig. 1b) presents only the location of luminance transitions it is harder for the viewer to segregate the object (watering pot) from background clutter (wire mesh and bags). In the bipolar feature image (Fig. 1c), the watering pot's outer edges are white on the outside and black inside because of the contrast between the dark object and its bright background.

Polarity at occluding edges may be the same as the luminance and shading, or it may conflict (darker object in front of a brighter

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-111



(a) (b) (c)

***Figure 1**. Image of 'Watering pot' (a) Color image of the watering pot in front of wire mesh and other clutter (b) Canny [7] binary edge result, 5 of 13 (40%) subjects identified the watering pot. (c) Bipolar edge [8] result. The bipolar edges and cusps act as a depth cue and may help segregate the object from the background clutter. Recognition rate was noticeably improved to 92.3% (12 of 13). Note that the images were scanned from prints resulting in artifactual outside edges interfering with the presentation.*

ground surface), for example, the dark watering pot in front of the lighter background in Fig. 1. Here (Fig. 1c) the dark pigmentation overcomes the weaker depth information with bipolar edges, causing the pot to appear more distant than the background. However, the white pot nozzle has an edge polarity consistent with its position in front of the relatively darker background, making it appear closer than the background.

Due to the assumption that illumination typically comes from above, objects are expected to cast shadows below, therefore, the bottom side of shading edges are expected to be dark while the upper side of edges are bright. This shading effect lost in binary edges is preserved in bipolar edges and therefore provides a depth cue that could aid in object recognition [12]. A uniformly pigmented object like the body of the watering pot in Fig. 1 may have specular points or lines on glossy surfaces, represented as cusps in the bipolar edge image (e.g. the decorative flourish on the watering pot) or it may illustrate internal corners, which in a bipolar edge representation will preserve the object's 3D structure. The various polarity effects may help observers recognize the scene and objects correctly, as they combine to provide additional depth cues.

Methods

We compared object recognition rates of bipolar edge images and binary edge images using the paradigm and image data set used by Sanocki et al. [5] This dataset contains 16 different office and household items at the center of a scene with varying levels of background clutter. The dataset is shared on our webpage (http://serinet.meei.harvard.edu/faculty/peli/), including original and filtered images. Twenty six normally sighted subjects (nine men) aged 21–67 participated. The study was approved by the Human Studies Committee of Massachusetts Eye and Ear and written informed consent was obtained from all participants.

Object recognition performance is highly dependent on the difficulty of the image dataset. Since we were looking for relative improvement in object recognition with bipolar feature images, we were concerned about a possible ceiling effect. Therefore, we used the Sanocki dataset, in which subjects had 45.7% average recognition rate with binary edge images [5].

The 16 images were split into two groups (A and B) based on 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 group A and the other to group B. The average recognition rate in Sanocki et al. was 46.5% and 44.9% for groups A and B, respectively, as shown in Fig. 4. Each subject viewed binary edge images from one of groups and bipolar feature images from the other group. The group presentations were counter balanced between subjects.

We attempted to follow Sanocki et al.'s [5] method closely. Binary edges were calculated with the Canny edge detector [7] in Matlab R2013b (MathWorks, Natick, MA) and the three parameters (sigma, upper and lower thresholds) were manually selected as described by Sanocki et al. [5] (the exact parameters used by Sanocki et al. were not available to us). Bipolar feature images were generated by Peli's method [8]. Peli's algorithm has no free parameters, other than presumed angular image span, which we adjusted to attempt to equate the level of extracted features, from the binary and bipolar algorithms. The presumed image span affects the level of details represented by the detected features and the de-noising threshold. Although the bipolar feature images have more information of contrast polarity, we tried to have same contents and similar level of details in both edge images with adjusting the presumed image span in the bipolar algorithm. As shown in Figs. 1b and 1c, the details and level of features are similar in the binary edge and bipolar edge images.

Subjects were seated 33 inches from a LCD monitor and the image width was 8 inches to approximately match the 14° angular size 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 feature versions. We also explained the difference between the binary edges and bipolar features (the meaning of edge polarity in bipolar feature: bar, line, and cusp) 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 dark room. The 8 binary edge images were presented first and followed by 8 bipolar feature images. At the beginning of the test in each condition, the trained 'umbrella' image in that condition was displayed as an example. Each test image was preceded by an audible beep and disappeared 1 second 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 gave neither feedback nor correction. In determining the response veracity, describing object usage was valued than more than a general description of the object's shape. The next image was displayed after the subject pressed any button on the keyboard.

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-111

(a) (b) (c)

***Figure 2**. Image of 'Briefcase' that was better recognized form 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) Bipolar feature was detected by Peli's method [8]. Due to the black color of outer briefcase, the depth from shading is inconsistently perceived in the bipolar feature. This might have reduced the recognition rate to 38.5% (5 of 13 subjects) although contrast polarity is consistent with the depth for numerous bright objects in the same scene (e.g. papers on the shelves).*

Results

Average overall object recognition rates for binary and bipolar images are given in Table 1 for both our subjects and Sanocki et al.'s results. Recognition rates for bipolar and binary images were 79.3% and 71.6% respectively. The modest improvement was 7.7% and approached significance (p = 0.069). Note, however, that the recognition rate of our subjects, in the binary edge condition, was much higher than Sanocki et al.'s subjects although we used same dataset and edge filtering method (Canny detector). Our results and Sanocki et al.'s were correlated moderately (ρ = 0.53, p = 0.035).

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

	Bipolar Feature	Binary Edge	Binary Edge (Sanocki et al.)
Average (%)	79.3	71.6	45.7
SE (%)	2.8	3.1	4.4
p	0.069		.

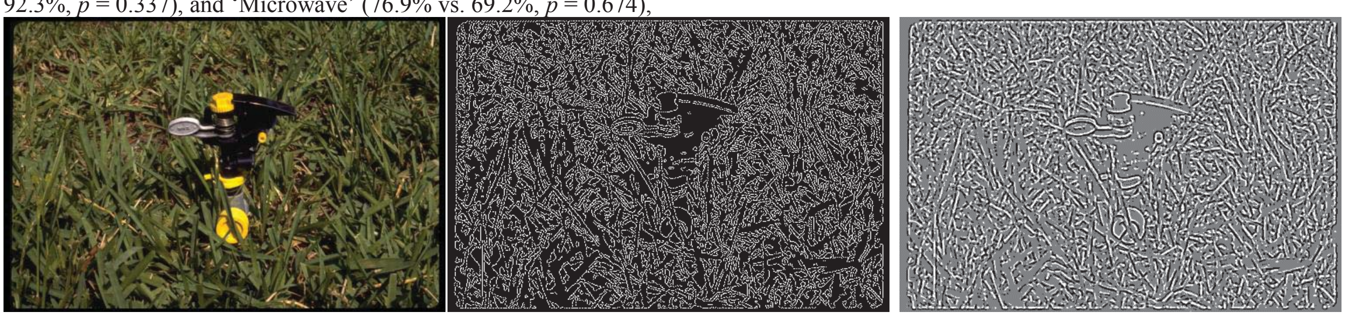
Figure 4 shows recognitions rate for all images in each condition. Except for a few images that were recognized better in the binary edge condition, such as 'Briefcase' (binary edge = 53.9% vs. bipolar feature = 38.5%, p = 0.452; Fig. 2), 'Lamp' (100% vs. 92.3%, p = 0.337), and 'Microwave' (76.9% vs. 69.2%, p = 0.674), the recognition rates of bipolar images were better or similar. Nine of the 16 images were recognized at above 80% from the binary edges, implying a ceiling effect that may limit the possibility of the getting an improvement with the bipolar edges. In the results for images in which a ceiling effect is not suspect, 'Watering Pot' (92.3% vs. 46.2%, p < 0.01) and 'Sprinkler' (53.8% vs. 0%, p < 0.001) had significantly improved recognition rates in bipolar feature images.

We further analyzed the results using binary logistic regression model in SPSS 11.5. The model correctly classified 94.9% of the correct recognitions and 78.4% of incorrect recognitions. The odds ratio [10] was 1.52 and approached significant level (p = 0.069), indicating that the odds of recognition in bipolar feature image is 1.52 times more than the odds of recognition in binary edge image when holding all other variables constant. For example, if the recognition rate in binary edge image was 45.7% as found by Sanocki et al., the recognition rate for bipolar feature images is expected to be 54.5%.

Discussion

Bipolar features contain more information than binary edges because they represent the contrast polarity of contrast at edges. The bipolar feature images also distinguish edges from cusps and thin bars. Importantly, polarity can provide shading information and thus be a depth cue. The improvement we found in object recognition with bipolar features was modest and only approached significance. This might have been caused by a ceiling effect, (>80% correct recognition) that was present for 9/16 images. Two of the 3 images that were recognized at less than 50% from the edge images were much better recognized from the bipolar features.

The improvement in recognition might be caused by better



(a) (b) (c)

***Figure 3**. Image of 'Sprinkler' that was much better recognized from bipolar images (a) Original color image (b) Binary edges image that no subject correctly recognized. (c) Bipolar edge image was recognized correctly in 53.9% (7 of 13 subjects) of presentations. The depth cues based on contrast polarity changes in the sprinkler might have helped to segregate the background and object.*

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-111

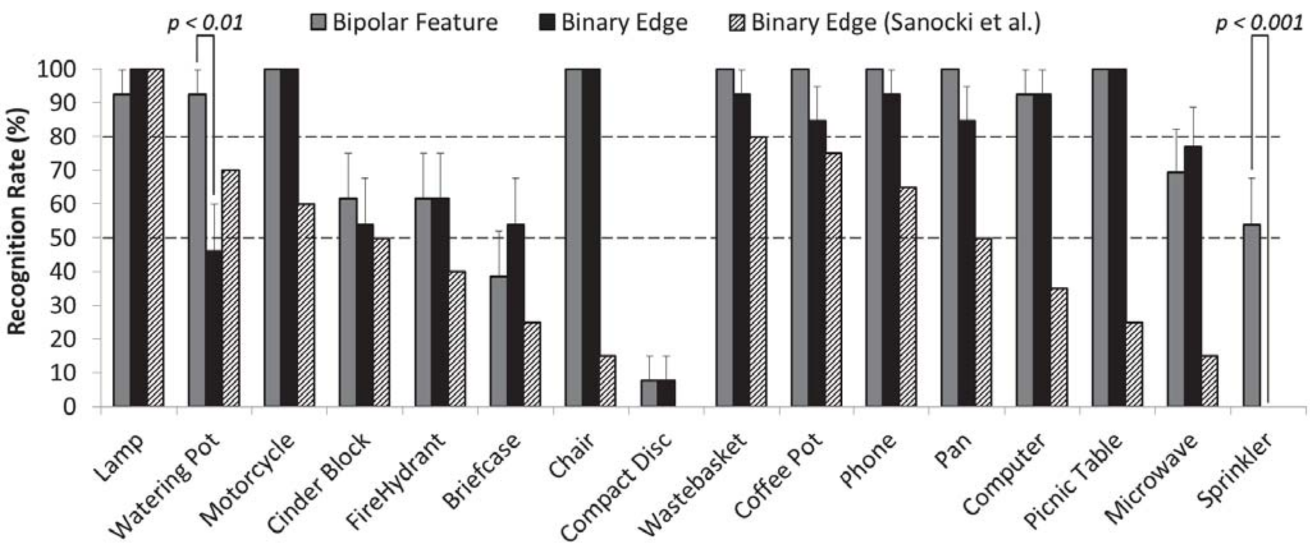


***Figure 4**. Recognition rate for each object in all the test conditions. We divided Sanocki's dataset into two groups based on their recognition results for counter balancing. As seen our subjects' recognition rates were higher with the binary edge leaving little room for improvement with the bipolar processing (ceiling effect).*

segregation of object and background based on depth cues. Using contrast polarity, the subjects may have perceived depth from shading and more easily distinguish the object contrast polarity, the subjects may have perceived depth from shading and more easily distinguish the object of interest. In images with complex background clutter (e.g. Figs. 1 and 3), depth cues caused by illumination may have helped segregate object from background. However, luminance differences caused by pigment (darker than background) of local object could confound the depth cue.

In this study, we only used luminance differences, and no color differences, in binary and bipolar feature detection algorithms. The bipolar features within the briefcase (Fig. 2c) caused incorrect perceived depth due to the reversed contrast polarity of the black color of the briefcase's exterior. In typical illumination, the border between the inside and outside of the briefcase would have a black edge in the inner area (due to shading) and a white edge in the outer area. However, the polarity in Fig. 2c was reversed due to the black color of the briefcase, which may have caused a misrepresentation that affected its recognition rate.

We did not investigate directly the impact of background clutter, nor depth cues from illumination, in this pilot study. In future studies, to reveal the impact of bipolar features in segregating the object from background clutter, controlling the complexity of background, or the illumination (e.g., direction 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.

We used Sanocki et al.'s images in effort to prevent ceiling effect, as in their study low recognition rate (45.7%) was found. We used processing parameters similar to theirs and expected similar recognition rate. However, the recognition rate of our subject with binary filtered edge images was higher than theirs and the correlation between our and their results was just moderate. This might have caused lower impact of the bipolar filtering on object recognition. In future studies, we plan to use the binary edge images extracted from the bipolar feature image with removing negative or positive polarity edges.

Since edge images are believed to provide useful representations for object recognition, adding high contrast edge information has been proposed as a way of enhancing image visibility for the visually impaired [13, 14]. Such enhancements have been implemented for video displays [15-17]. Both binary edges and bipolar edges have been employed in such studies. 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 [18-20], however in video see-through systems bipolar edges may be used. A number of studies have demonstrated preference for enhanced images but performance improvements have not been frequently demonstrated. For the same reasoning, and due to the limited dynamic range of visual prosthetics (such as retinal implants being developed for the blind), many have proposed using binary edge representations for these systems. We have suggested that for systems that can provide more than two levels of stimulations (e.g., visual prostheses [10]) the bipolar edge representation may provide an advantage. This paper is a first attempt to test this hypothesis in high resolution images, and we predict a larger effect for lower resolution images typical in such systems.

Acknowledgement

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

[1] M. J. Tarr and H. H. Bülthoff, "Image-based object recognition in

man, monkey and machine," Cognition, vol. 67, no. 1-2, pp. 1-20, 1998.
[2] M. Riesenhuber and T. Poggio, "Models of object recognition," Nature Neuroscience, vol. 3, pp. 1199-1204, 2000.
[3] I. Biederman, "Recognition-by-components: A theory of human image understanding," Psychological Review, vol. 94, pp. 115-147, 1987.
[4] I. Biederman and P. Kalocsai, "Neurocomputational bases of object and face recognition," Philos. Trans. R. Soc. Lond. B, Bio. Sci., vol. 352, no. 1358, pp. 1203–1219, 1997.
[5] T. Sanocki, K. W. Bowyer, M. D. Heath, and S. Sarkar, "Are edges sufficient for object recognition?" Journal of Experimental Psychology: Human Perception and Performance, vol. 24, no. 1, pp. 340-349, 1998.
[6] M. D. Heath, S. Sarkar, T. Sanocki, and K. W. Bowyer, "A robust visual method for assessing the relative performance of edge-detection algorithms," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.19, no.12, pp.1338-1359, 1997.
[7] J. Canny, "A computational approach to edge detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, no. 6, pp. 679-698, 1986.
[8] E. Peli, "Feature detection algorithm based on a visual system model," Proceedings of the IEEE, vol. 90, pp. 78-93, 2002.
[9] H. L. Apfelbaum, D. H. Apfelbaum, R. L. Woods, and E. Peli, "Inattentional blindness and augmented-vision displays: Effects of cartoon-like filtering and attended scene," Ophthalmic and Physiological Optics, vol. 28, no. 3, pp. 204-217, 2008.
[10] J.-H. Jung, D. Aloni, Y. Yitzhaky, E. Peli, "Active confocal imaging for visual prostheses," Vision Research, vol. 2015; no. 111, pp. 182–196, 2015.
[11] C. Zhou and B. W. Mel, "Combining Multiple Cues for Contour Detection: Lessons from (and to) the Visual Cortex [Abstract]," Journal of Vision, vol. 3, no. 9, pp. 252, 2003.
[12] R. P. O'Shea, S. G. Blackburn, and H. Ono, "Contrast as a depth cue," Vision research, vol. 34, no. 12, pp. 1595-1604, 1994.
[13] E. Peli, J. Kim, Y. Yitzhaky, R. B. Goldstein, R. L. Woods, "Wide-band enhancement of television images for people with visual-impairments," Journal of Optical Society of America A, vol. 21, no. 6, pp. 937-950, 2004.
[14] E. Peli, "Wide-band image enhancement," US Patent No 6,611,618 B1, Washington, DC: U.S. Patent and Trademark Office, 2003.
[15] P. Satgunam, R. L. Woods, G. Luo, P. M. Bronstad, Z. Reynolds, C. Ramachandra, B. W. Mel, and E. Peli, "Effects of contour enhancement on low-vision preference and visual search," Optometry and Vision Science, vol. 89, pp. E1364-1373, 2012.
[16] J. S. Wolffsohn, D. Mukhopadhyay, and M. Rubinstein, "Image enhancement of real-time television to benefit the visually impaired," American Journal of Ophthalmology, vol. 144, pp. 436-440, 2007.
[17] M. Kwon, C. Ramachandra, P. Satgunam, B. W. Mel, E. Peli, and B. S. Tjan, "Contour enhancement benefits older adults with simulated central field loss," Optometry and Vision Science, vol. 89, pp. 1374-1384, 2012.
[18] H. L. Apfelbaum, D. H. Apfelbaum, R. L. Woods, and E. Peli, "Inattentional blindness and augmented-vision displays: Effects of cartoon-like filtering and attended scene," Ophthalmic and Physiological Optics, vol. 28, no. 3, pp. 204-217, 2008.
[19] E. Peli, G. Luo, A. Bowers, N. Rensing, "Applications of augmented vision head-mounted systems in vision rehabilitation," Journal of Society Information Display, vol. 15, no. 12, pp. 1037-1045, 2007.
[20] A. D. Hwang and E. Peli, "An augmented-reality edge enhancement application for Google Glass," Optometry and Vision Science, vol. 91, no. 8, pp. 1021-1030, 2014.

Authors Biography

Jae-Hyun Jung *received his BS in electrical engineering from the Pusan National University in Korea (2007) and his PhD in electrical engineering from Seoul National University (2012). He is an investigator at the Schepens Eye Research Institute and an instructor of ophthalmology at Harvard Medical School. His work has focused on optical engineering for low vision rehabilitation and imaging systems for visual prostheses.*

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

Eli Peli *received his BSc and MSc in electrical engineering from Technician-Israel Institute of Technology (IIT), Haifa, Israel (1978), and the OD from the New England College of Optometry, Boston, MA (1983). Since then he has been at the Schepens Eye Research Institute where he is the Moakley Scholar in Aging Eye Research and a professor of Ophthalmology at Harvard Medical School.*