Mono-glass for Providing Distance Information for People Losing Sight in One Eye

Masahiro Toyoura* Kenji Kashiwagi Atsushi Sugiura Xiaoyang Mao†
University of Yamanashi

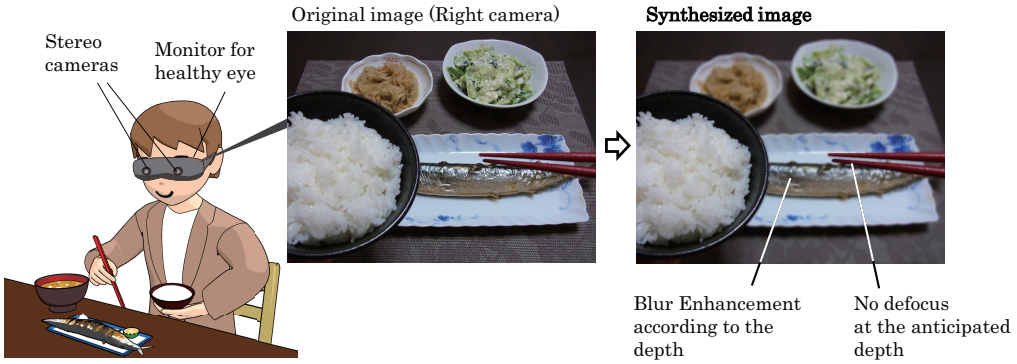

Figure 1: *System overview of mono-glass with an example application to dining assistance.*

Abstract

We propose *mono-glass* for providing distance information for people losing sight in one eye. We implemented a pilot system of mono-glass under careful consideration of precision, real-time processing and intuitive presentation. The loss of sight in one eye disables binocular disparity processing and makes short-range activities difficult. The proposed mono-glass is a wearable device with two cameras and one display. The two cameras capture the images on behalf of user's eyes. Depth information is then reconstructed from the captured images and visualized with defocusing for the healthy eye. Experimental results supported that our system could represent depth information in single-channel images.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image Generation—Display algorithms; I.4.3 [Image Processing and Computer Vision]: Enhancement—Filtering

Keywords: Augmented reality, medical image, information visualization, sense of distance

Links: DL PDF

1 Introduction

The people losing sight or having weak sight in one eye cannot do binocular disparity processing. It means they cannot perceive the sense of distance from the binocular disparity, which makes grasping an object or sewing clothes difficult. It can significantly reduce the quality of life.

The number of visually impaired patients is 285 million, which includes 246 million people of poor vision and 39 million people of blindness [Pascolini and Mariotti 2010]. The number of people losing sight in one eye is not published, however it is considered to be the same order to that of the visually impaired patients.

We implemented *mono-glass* to provide the function of binocular disparity on behalf of user's eyes. Figure 1 shows the overview of our system. Two cameras in front of the user eyes capture images. The images contain depth information what they could have seen. The system displays single-channel defocused images synthesized according to the images from the both cameras and estimated depth from them. The defocused images enable the user to perceive a sense of distance from the degree of defocus.

In experiments, we use WrapAR920 of Vuzix Corporation, a video-see-through head mount display with two cameras and two displays. The cameras capture the images from the viewpoint of a user. GPGPU and CUDA support real-time processing of stereo depth estimation and image synthesis.

The main contributions of this work are the followings.

1. Perceptual and intuitive representation of depth information.
2. Validation of mono-glass through a subject study and its analysis.

To enable user to perceive the depth information with mono-glass, the displayed images should be intuitive and perceptually natural. In other words, the accuracy of depth estimation is required for mono-glass implementation. In the benchmark images for conventional stereo matching works, many textured objects are included, which makes it easy for stereo matching algorithms to find corresponding points between two images. However, we cannot expect such rich textures in captured images when using mono-glass. Special consideration should be given on the accuracy of depth estimation for the scenes captured in daily life.

The result of a subject study indicates the validity of mono-glass,

*e-mail: mtoyoura@yamanashi.ac.jp
†e-mail: mao@yamanashi.ac.jp

although there was difference among individuals on how much they perceive the provided distance. The result also provides a good perspective for the following works of *computational ophthalmology*.

2 Related Works

To the best our knowledge, mono-glass is the first device which can assist patients losing sight in one eye to perceive distance information. For visually impaired patients, the distance information has been provided by virtual white sticks [Okayasu 2009; Okayasu 2010]. The sticks provide the depth of obstructions with bumps or vibration on their handles. Our mono-glass, meanwhile, provides the sense of distance to the users in a way as if it was perceived by both of their own eyes. In other words, the proposed the system can complement the lost sense of distance, which is the best way to enhance the quality of life.

It is well known that defocus provides a cue of distance for human [Vishwanath and Blaser 2010]. A technique has been proposed to synthesize a defocused image based on estimated depth from a single image in order to enhance the sense of distance [Okatani and Deguchi 2007; Held et al. 2010]. However, the technique, which is used for making entertainment movies and so on, has a limitation that the structure of captured scene should be quite simple. It cannot deal with small objects in hand which is required for our application.

Hardware implementation for synthesizing defocus has also been proposed. Consumer cameras, FinePix F300EXR [FinePix] and Panasonic LUMIX GF2 [LUMIX], have a function of synthesizing defocused image by capturing multiple images with different focal distances. LYTRO light field camera [LYTRO] realizes it by using camera lens array. Such cameras cannot provide images with an arbitrary degree of defocusing, because they do not reconstruct depth information, but synthsize pixel-level composite images from real captured images. We adopted software implementation for mono-glass for this reason.

The depth map is calculated as the disparity of stereo cameras in our method. The method proposed by Wang *et al*. [Wang et al. 2006] provides a good result by referring corresponding points and their neighborhoods in stereo paired images. GPU implementation can accelerate the speed of the method. Since calculating depth maps is a fundamental problem of computer vision, new even better methods have been proposed in recent years [Gong et al. 2007]. Whereas Wang *et al*. adopted DP matching for calculating the depth map, DP with 4 states [Criminisi et al. 2007] and belief propagation [Yang et al. 2010] could improve the results. Other than stereo cameras, pattern code projection [Kinect], combined with stereo cameras [Yang et al. 2007] and coded aperture [Zhou et al. 2009] may provide better solution to our application in near future. For the meantime, however, we adopt the recent fundamental method by Wang *et al*., by which depth map estimation with two small cameras only can balance the precision, speed, and physical size well.

An implementation on IC tips, such as in Kinect, can reduce the computation cost. But we choose software implementation for assuring the flexibility of personal adaptation of defocusing parameters. More suitable devices rather than stereo cameras might appear in near future but at this moment, a personal 3D range scanner like Kinect is still too large to wear on the head. We employed stereo matching for estimating the depth, and GPGPU for acceleration of the computation.

3 Artificial Defocused Image Synthesis

Using estimated depth map, we synthesize single-channel images representing depth information. By defocusing images according to the estimated depth, a user can perceive the distance from the single-channel images.

Figure 2 shows the concept of how to change the degree of defocusing. The defocusing is controlled by blur radius σ. The solid line represents the original defocus of a real camera. σ is defined by the focal length z_0 and the estimated distance from lens z with constant coefficient A as

$$\sigma = A\frac{|z - z_0|}{z_0}. \tag{1}$$

Two dotted lines represent the defocus controlled by parameters α, β, γ and δ as the following:

$$\sigma' = \alpha\left(\gamma\frac{|z - z_0|}{z_0} + \delta\right)^{\beta}. \tag{2}$$

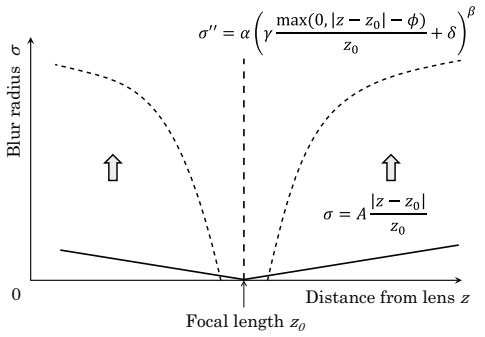

Figure 2: *Blur radius with defocus controlling parameters.*

Note that δ is a small positive number for assuring σ'' to be positive. In addition, we introduce a parameter ϕ for representing non-defocusing range.

$$\sigma'' = \alpha\left(\gamma\frac{\max(0, |z - z_0| - \phi)}{z_0} + \delta\right)^{\beta}. \tag{3}$$

The pixel value $I_{\sigma''}(x, y)$ in the defocused image is calculated with the pixel value $I(x, y)$ in the original image and the blur radius σ'' is given in the following way:

$$I_{\sigma''}(x,y) = \frac{\sum_{(x',y')\in\Omega(x,y)} I(x,y)exp(-g(x,y,x',y')^2/\sigma''^2)}{\sum_{(x',y')\in\Omega(x,y)} exp(-g(x,y,x',y')^2/\sigma''^2)}, \tag{4}$$

where $\Omega(x, y)$ is a set of neighbors of the pixel (x, y) and $g(x, y, x', y')$ the Euclid distance between (x, y) and (x', y').

According to Eq. (3), even when an object is positioned at only a very small distance from the focal length z_0, the object can be observed as largely defocused. The degree of defocus is controlled with α, β and γ non-linearly. This can only be realized by software implementation using depth estimation.

4 Experiments

In experiments, we used WrapAR920 of Vuzix Corporation. It has two independent display monitors and stereo cameras at the position of user's eyes. Captured images from the stereo cameras are transferred to a PC. Depth maps are generated and defocus is enhanced according to the depth maps. Defocused images are displayed on the monitor for the healthy eye. Our system was implemented with Visual Studio 2010, CUDA4.2 and OpenCV2.4.1, and executed on a desktop PC (OS: Windows7 64bit, CPU: Intel Core2Quad 2.83GHz, GPU: NVidia GeForce GTX 460 1GB, MM: 2GB).

The length of the baseline of stereo cameras was 60mm. The physical focal length of each camera was set to 400mm. An initial z_0 is also set to 400mm, and we allow the subjects to change z_0 as they like. Parameters in Eq. (3) were empirically set as $\alpha = 3.0$, $\beta = 0.5$, $\gamma = 5.0$, $\phi = 20.0$ and $\delta = 0.001$.

4.1 Defocused Image Synthesis

Figure 4 shows examples of synthesized images. The original images as shown in Figure 4(a) do not have noticeable defocus. The images as shown in Figure 4(b) are synthesized with the original images and their corresponding depth maps. The images except for the rightmost image include not defocused regions, which indicate that the regions are positioned at the anticipated distance. In the rightmost image, almost all regions with rich texture are defocused. It means there are no regions with rich texture at the anticipated distance.

It took 46.75ms on average for capturing and synthesizing a single image and frame rate was 21.93fps. We can regard the processing time as real-time. We used the resolution of 80×60 when generating depth maps for achieving real-time processing and for suppressing artifacts.

4.2 Subject Study

The synthesized images were provided to right eyes of 8 subjects denoted by A to H (5 male and 3 female in their 20s or 30s with both eyes being healthy) with mono-glass. The dominant eye of C is left, and those of the others are right. Corrected visions were between 0.6 and 1.2. No images were displayed to the left eyes.

The experimental environment is shown in Figure 3. The subjects are asked to perform the task of inserting a cylinder of 20mm in diameter and 85mm in length into two holes on a box placed on a table. The diameters of the holes are 25mm.

The subjects repeat the following steps 30 times.

1. Put the cylinder on the front of the box.
2. Insert the cylinder into the right hole.
3. Look ahead.
4. Put the cylinder on the front of the box.
5. Insert the cylinder into the left hole.
6. Look ahead.

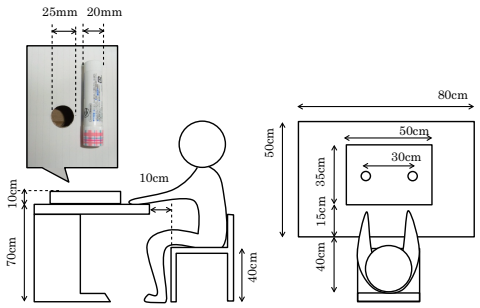

Figure 3: *Environment for subject study.*

The task requires the perception of distance. The information of distance is reset by step 3 and 6. We instructed the subjects that the task can be easily done by setting both of the cylinder and the holes at a matching distance which is indicated by without being defocused. We had the subjects grasp the top side of the cylinder. The cylinder was wrapped with rich texture and printed papers with parallel rules were also put on the box so that the subjects can perceive the defocus effect. We believe that it is reasonable in real application to assume that handling tools have rich-texture. In the example of Figure 1, the user is forced to use chopsticks with rich-texture.

Subjects A to D did the task in an order of synthesized, original, synthesized, and original images. Subjects E to H did the task in an order of original, synthesized, original, and synthesized images.

The results are shown in Figure 5. We adopt t-test for examining the difference of time for completing the second task. The second time of subjects C, D, E, G and H, and the first time of A, C, D, G, and H became significantly shorter in the case that the synthesized images were displayed. Only the first time of subject F significantly became longer in the case that the synthesized images were displayed. There are also differences among individuals even in the second study.

To summarize the subject study, our synthesized images could provide distance information for some subjects, while there were differences among individuals.

5 Discussion and Future Work

We proposed mono-glass to provide distance information for losing sight in one eye. The distance information is represented by defocusing. The result of the subject study supported the validity of mono-glass even though there was difference among individuals.

As an important future work, we will continue to improve the real-time depth estimation. The time required for synthesizing defocused images of resolution 80×60 in GPGPU with CUDA is 21.39fps. Finer resolution with larger window size could give us more accurate result. Since the system requires a laptop computer at least, currently it is not easy to wear the system. However, we expect GPU computing will be realized in mobile environments in near future.

Optimization of parameters for defocusing is an important future work. We will try to explorer the parameters in addition to making intuitive interfaces for controlling them.

References

CRIMINISI, A., BLAKE, A., ROTHER, C., SHOTTON, J., AND TORR, P. 2007. Efficient dense stereo with occlusions for new

(a) Original images captured by the right camera.

(b) Synthesized defocused images according to estimated distance maps.

Figure 4: *Results of defocused image synthesis according to estimated distance.*

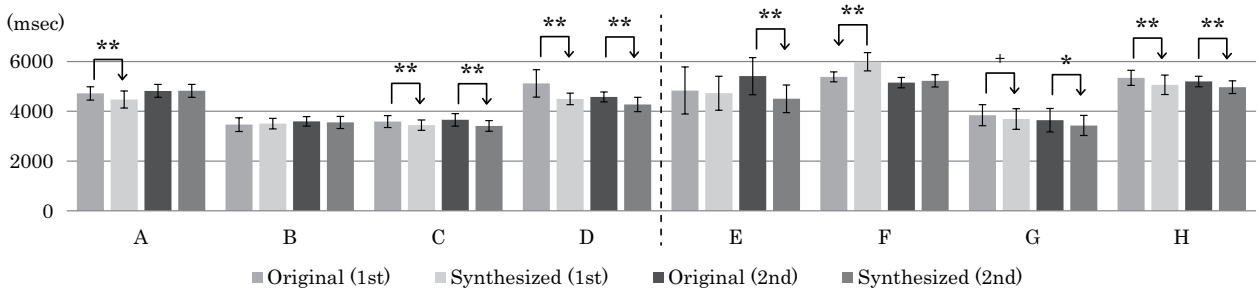

Figure 5: *Difference of execution time between original and synthesized images. Subjects A, B, C and D executed the task with original images first, and E, F, G and H did with synthesized images first. Average execution time of the task with a textured stick. The leftmost bars indicate the execution time in the case of original images, the second bars from the left are synthesized images, the third bars from the left are for original images of the second time, and the rightmost bars are for synthesized images of the second time. The results were analyzed with t-test.* $^{+}$*:* $p < 0.10$*,* **:* $p < 0.05$*, and* ***:* $p < 0.01$*. The directions of arrows indicate the smaller elements.*

view-synthesis by four-state dynamic programming. *International Journal of Computer Vision 71*, 1, 89–110.

FujiFilm, FinePix F300EXR. http://www.fujifilm.co.jp/corporate/news/articleffnr_0414.html.

GONG, M., YANG, R., WANG, L., AND GONG, M. 2007. A performance study on different cost aggregation approaches used in real-time stereo matching. *International Journal of Computer Vision (IJCV) 75*, 2, 283–296.

HELD, R. T., COOPER, E. A., O'BRIEN, J. F., AND BANKS, M. S. 2010. Using blur to affect perceived distance and size. *ACM Transactions on Graphics 29*, 2, 1–16.

Microsoft Corporation, Kinect for Xbox 360. http://www.xbox.com/kinect/.

Panasonic, LUMIX GF2. http://panasonic.jp/dc/gf2/.

LYTRO, Light Field camera. https://www.lytro.com/camera.

OKATANI, T., AND DEGUCHI, K. 2007. Estimating scale of a scene from a single image based on defocus blur and scene geometry. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

OKAYASU, M. 2009. The development of a visual system for the detection of obstructions for visually impaired people. *Journal of Mechanical Science and Technology 23*, 10, 2776–2779.

OKAYASU, M. 2010. Newly developed walking apparatus for identification of obstructions by visually impaired people. *Journal of Mechanical Science and Technology 24*, 6, 1261–1264.

PASCOLINI, D., AND MARIOTTI, S. P. 2010. Global estimates of visual impairment. Tech. rep., BJO Online First.

VISHWANATH, D., AND BLASER, E. 2010. Retinal blur and the perception of egocentric distance. *Journal of Vision 10*, 10, 1–16.

WANG, L., LIAO, M., GONG, M., YANG, R., AND NISTER, D. 2006. High-quality real-time stereo using adaptive cost aggregation. In *International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)*.

YANG, Q., YANG, R., DAVIS, J., AND NISTER, D. 2007. Spatial-depth super resolution for range images. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1–8.

YANG, Q., WANG, L., AND AHUJA, N. 2010. A constant-space belief propagation algorithm for stereo matching. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1458–1465.

ZHOU, C., LIN, S., AND NAYAR, S. 2009. Coded aperture pairs for depth from defocus. In *Computer Vision, 2009 IEEE 12th International Conference on*, 325–332.