IEICE TRANS. INF. & SYST., VOL.E96–D, NO.12 DECEMBER 2013

PAPER

Depth Perception Control during Car Vibration by Hidden Images on Monocular Head-Up Display

Tsuyoshi TASAKI[†a)], **Akihisa MORIYA**[†], **Aira HOTTA**[†], **Takashi SASAKI**[†], ***Nonmembers***, ***and*** **Haruhiko OKUMURA**[†], ***Member***

SUMMARY A novel depth perception control method for a monocular head-up display (HUD) in a car has been developed, which is called the dynamic perspective method. The method changes a size and a position of the HUD image such as arrow for depth perception and achieves a depth perception position of 120 [m] within an error of 30% in a simulation. However, it is difficult to achieve an accurate depth perception in the real world because of car vibration. To solve this problem, we focus on a property, namely, that people complement hidden images by previous continuously observed images. We hide the image on the HUD when the car is vibrated very much. We aim to point at the accurate depth position by using see-through HUD images while having users complement the hidden image positions based on the continuous images before car vibration. We developed a car that detects big vibration by an acceleration sensor and is equipped with our monocular HUD. Our new method pointed at the depth position more accurately than the previous method, which was confirmed by t-test.

key words: *HUD, augmented reality, depth control*

1. Introduction

Many augmented reality systems are expected to be applied in cars [1], [2]. Specially, many works use head-up displays (HUDs) for car application [3], [4]. The HUD superimposes images that are reflected on a combiner, such as a windshield, onto real world. A driver can observe both the outside information and the HUD information visible at the same time since the HUD image is see-through. Therefore, the driver can get information such as a speed of the car or navigation information with minimum eye movements while driving [5], [6]. In previous works, the HUD that has the drawback of a binocular parallax was used, which means the driver cannot watch far outside clearly as shown in Fig. 1. Therefore, we have developed a monocular HUD and accurate depth perception control method for navigation and pointing at obstacles. We called the method the dynamic perspective method [7], [8]. The monocular HUD does not have the drawback of the binocular parallax because optical systems of the monocular HUD enable a user to observe images with only one eye, as shown in Fig. 2 [9]. The dynamic perspective method uses a size and a position of an object image as depth cues that are important factors in psychology [10]. An example of an object image is an arrow for navigation and pointing at obstacles. When we want users to perceive near position, the object image on the monocular HUD is displayed bigger and lower. When we want users to perceive far position, the object image is displayed smaller and higher as the object image moves to the far position. Examples of object images based on the dynamic perspective method are shown in Fig. 3. The dynamic perspective method achieved a depth perception position of 120 [m] within an error of 30% in the simulation.

In the case of applying the dynamic perspective method to a moving car, the object image position sometimes changes randomly because of car vibration. Differences of the object image position make users perceive wrong depth positions. In this paper, we deal with the problem of the depth control in the case of car vibration. There are some methods that decide the image positions based on the result of measuring the HUD position [11], [12]. However, the

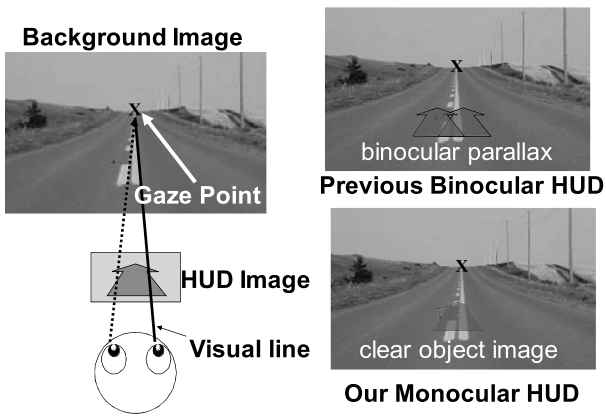

Fig. 1 The drawback of the previous binocular HUD.

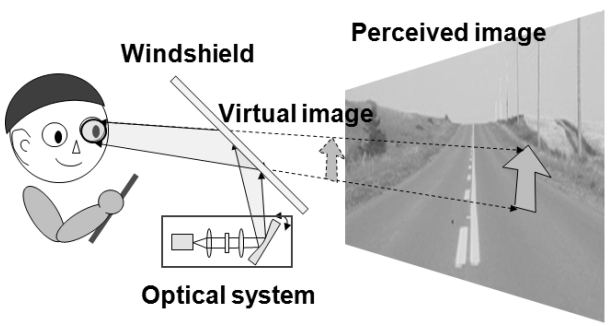

Fig. 2 Schematic illustration of the monocular HUD.

Manuscript received March 27, 2013.
Manuscript revised June 3, 2013.

[†]The authors are with Toshiba Corporation, Kawasaki-shi 212–8582 Japan.
a) E-mail: tsuyoshi.tasaki@toshiba.co.jp

DOI: 10.1587/transinf.E96.D.2850

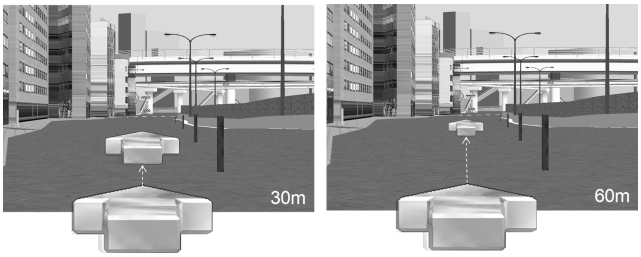

Fig. 3 Examples of the object images based on the dynamic perspective method.

methods cannot work well in the case of a moving car because the position of HUD in the moving car is not measured fast and accurately. While the car vibrates, the positions of not only the HUD but also the observer change. We cannot measure accurately the both positions by using a cheap sensor and cannot fix the image accurately. Our contribution is more accurate depth perception control by using only cheap sensor and a method which is easy to implement.

To solve the problem, we focus on a property of human perception, namely, that continuous observed images complement hidden images [13]. We detect big vibration by using an acceleration sensor mounted in the car. When we detect big vibration, the object image is temporarily hidden and we expect users to perceive the complemented depth position. We aim to point at more accurate depth positions by images on the monocular HUD based on the position complemented by using the continuous images before car vibration. We will apply our monocular HUD to point at a road to turn at or to point at obstacles on the road when the car is close to them.

Section 2 describes the new dynamic perspective method based on the car vibration. In Sect. 3, we confirm the ability of our new method by using simulator. In Sect. 4, we evaluate our new method by using a real moving car. Section 5 concludes this paper.

2. The Dynamic Perspective Method Based on Car Vibration

In this paper, we achieve the accurate depth control by hiding the object image during big car vibration. In order to detect big vibration, we use an acceleration sensor because it is small, inexpensive, and sensitive to vibration.

We can regard car vibration as the height displacement of the car. The height displacement is calculated by means of acceleration data along a vertical direction as shown in (1). Here, a denotes the acceleration data along a vertical direction and is obtained by the acceleration sensor mounted in the car. x denotes the height displacement of the car. f denotes a frequency of the vibration.

As shown in (1), x is directly proportional to a. Therefore, we can regard a big difference of a as big vibration. Figure 4 shows continuous acceleration data while the car is moving. The circle in Fig. 4 denotes the biggest difference of a per unit of time. Certainly, the car passed a small step

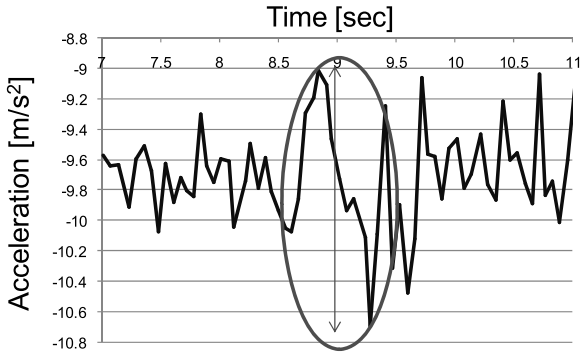

Fig. 4 The examples of the acceleration data.

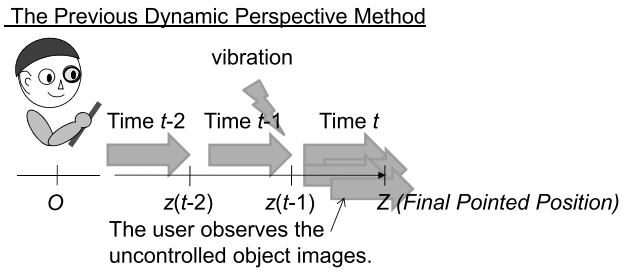

Fig. 5 The previous method during car vibration.

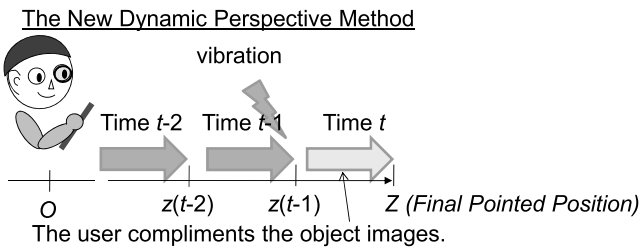

Fig. 6 The new method during car vibration.

at the time indicated by the circle.

$$x = a/(2\pi f)^2 \tag{1}$$

In particular, we use big influential vibration that occurs at key positions. The key positions are the car's positions when the object image approaches the position where the object image should point. If users observe the object image during influential vibration, they do not perceive the correct distance pointed by the uncontrolled object image, as shown in Fig. 5. We consider the hidden complemented object image achieves more accurate depth control than the uncontrolled object image does, as shown in Fig. 6. We hide the object image from the time t that satisfies both (2) and (3).

$$Z - z(t) < C_z \tag{2}$$

$$|a(t) - A| > C_a \tag{3}$$

Here, $z(t)$ denotes the object image position where we want to point in the world at time t, and Z denotes the position where we want to point finally. We define Z and $z(t)$ on

the coordinates whose origin located at a given observation position O fixed on the world coordinates (Figs. 5 and 6.). We call Z a final pointed position. When the Eq. (2) is satisfied, the position where the object image points now is close to Z. Moreover, the observer will soon perceive the position where we want to point finally. Such positions affect the depth perception in the dynamic perspective method. $a(t)$ denotes the acceleration value at time t, and A denotes the acceleration data measured while the car stops before experiments. C_z and C_a are constant thresholds.

3. Depth Perception Experiments in the Simulator

3.1 Experimental Settings

In order to confirm the ability of our new method, we performed experiments by using a monocular HUD simulator as shown in Fig. 7. In the simulator, a subject uses polarized glasses and sees HUD images (objects) by monocular eye through the glasses. A video is projected to the screen as a background image. The video was previously recorded from an assistant driver's seat in the moving car which was driven on the test course as shown in Fig. 8. Figure 9 shows some frames in the video. The size of the video is 800×600 pixels. The test course is very flat and there are no big car vibrations. Therefore, we virtually add a vibration at the position of 3.3 [m] before the end position as shown in Fig. 8. The end position denotes the position where the object image on the combiner arrives at the final pointed position. When the virtual vibration occurs, the background image moves up by 12 pixels during 150 [ms]. After moving up, the image moves down by 12 pixels during 150 [ms]. During the virtual vibration, background images move as shown in Fig. 10. The setting of virtual vibration is determined as the vibration becomes sufficiently big.

When the car arrives at a start position where the object image starts to be displayed on the combiner, the object image is displayed through the combiner as users perceive it 20 [m] from the position of the car in the video. After

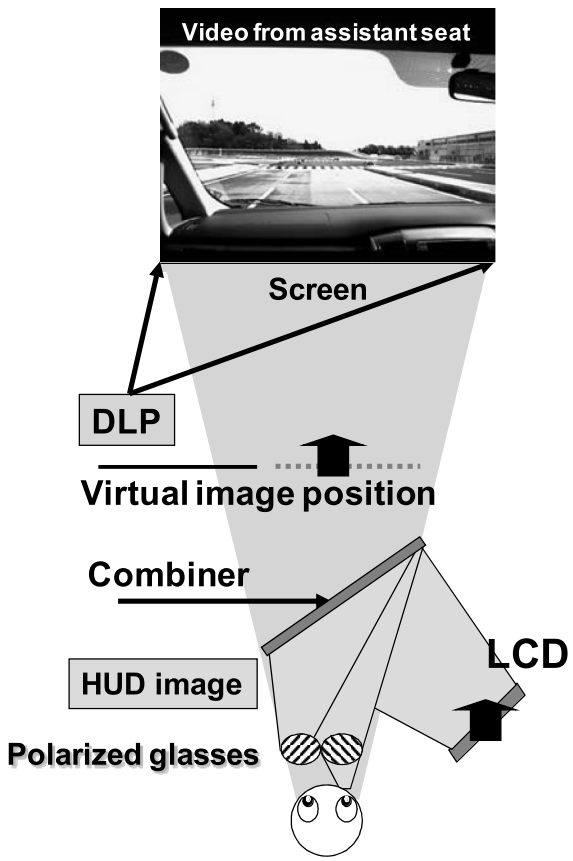

Fig. 7 The monocular HUD simulator.

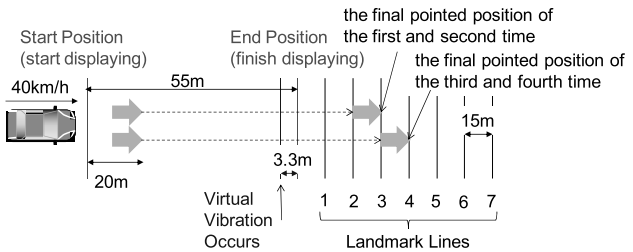

Fig. 8 The test course for capturing video.

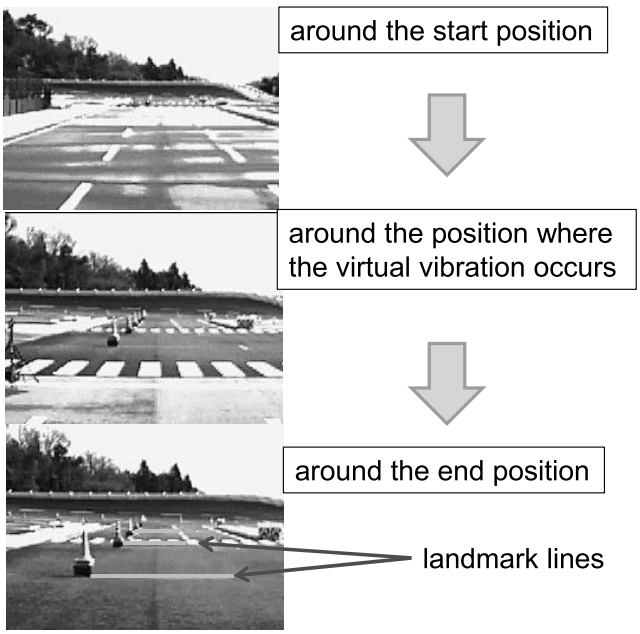

Fig. 9 Some frames in the video.

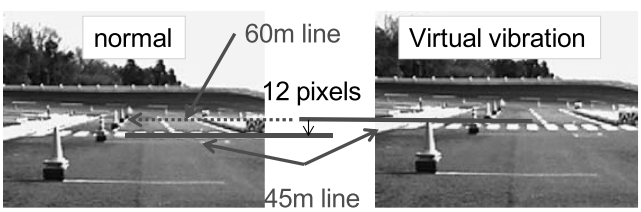

Fig. 10 The virtual vibration.

starting to display the object, the object image is displayed as users perceive it moving to the final pointed position while the car in the video moves from the start position to the end position. A subject sees the HUD image at 4 times. The first time and the third time, the subject observes the object image that moves to the final pointed positions of 45 [m] and 60 [m], respectively, based on the previous dynamic perspective method. Therefore, the first time and the third time, the object image is not hidden even if virtual vibration occurs. The second time and fourth time, the subject observes the object image based on the new method. That is, the object image is hidden at the virtual vibration position. The second time, the object image moves the same as the object image the first time, before virtual vibration. Once the vibration is occurred, the object image is hidden. The fourth time, the object image moves the same as the object image the third time, before the virtual vibration. Once the vibration is occurred, the object image is hidden.

For the evaluation, all subjects tell us the perceived depth position where the subject thinks the object image points based on the landmark lines printed on the test course, as shown in Figs. 8 and 9, each time. We compare the averages of the distances between the perceived depth positions of all subjects and the final pointed positions, in order to evaluate our method. We regard the distance as an error.

Here, subjects are 8 men from 20 to 40 years of age.

3.2 Results of Depth Perception and Discussions

The averages (avg.) and the standard deviations (sd.) of the distances between the perceived depth positions and the final pointed positions (fpp.) are shown in Table 1.

The results listed in Table 1 show that the error of our new method based on vibration was less than that of the previous method, which is confirmed by t-test ($p < 0.01$). Our new method can point at the final pointed position of 60 [m] 3.4 times more accurately than the previous method does. The values (3.4 times) are not meaningless statistically. We use the values as just reference data for discussion from now. We think that hiding the object image during vibration can control the depth position more accurately than displaying the uncontrolled object image. These results show hiding the object image during simple vibration is effective. However, in this experiment, the subjects do not move and the vibration is occurred only in the background image. In the next section, we will evaluate the ability in the real moving car.

Our new method can point at the final pointed position of 45 [m] 3.7 times more accurately than the previous method does. 3.7 times are a little larger than 3.4 times which is calculated in the case of the final pointed position of 60 [m]. This is not a big difference. However, we want to discuss the difference after real car evaluation in the next section because the property is important in order to consider when to use our method. We also discuss this point in the next section.

Table 1 Comparison of our new dynamic perspective method and the previous dynamic perspective method.

Method	Previous		New	
fpp.	45 [m]	60 [m]	45 [m]	60 [m]
error avg. [m]	7.1	10.2	1.9	3.0
error sd. [m]	5.7	4.5	3.2	3.0

4. Depth Perception Experiments in the Simulator

4.1 Experimental Settings

In order to confirm the ability of our new method in the real car, we performed experiments by using a car equipped with an acceleration sensor. The position of the car is measured accurately by an infrared (IR) sensor mounted externally, as shown in Fig. 11. The acceleration sensor is mounted on the ceiling of the car. Our monocular HUD is mounted in the glove compartment in front of the front passenger seat.

The implemented system is shown in Fig. 12. Our system calculates the position of the car continuously by IR sensors and a speed meter of the car. The previous dynamic perspective method decides the size and the position of the object image. Before making the graphics of the object image, we decide whether the image is displayed or not based on the method described in Sect. 2 that uses the acceleration sensor. When the system decides to display the image, the system makes graphics and displays them by the monocular HUD. An example of a displayed object image is shown in

Fig. 11 The car for experiments and the infrared sensor for measuring positions of the car.

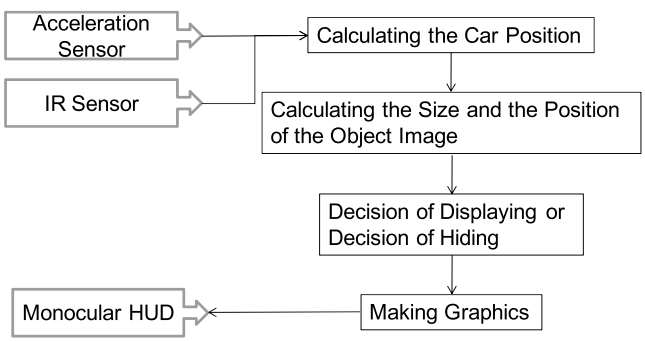

Fig. 12 The implemented system.

Fig. 13.

The car moves along a test course as shown in Fig. 14 at 40 [km/h]. There is a step as shown in Fig. 15 at the key position 3.3 [m] before the end position where the object image finishes being displayed in spite of vibration. The position of step is decided experimentally as vibration affects the depth perception drastically. The height of the step is 13 [mm].

When the car arrives at a start position where the object image starts to be displayed, the object image is displayed as users perceive it 20 [m] from the car. After starting to display the object, the object image is displayed as users perceive it moving to the final pointed position while the car moves from the start position to the end position. An operator drives the car along the course 4 times with a subject who sits on the front passenger seat. The first time and the third time, the subject observes the object image that moves to the final pointed positions of 45 [m] and 60 [m], respectively, based on the previous dynamic perspective method. Therefore, the first time and the third time, the object image is not hidden even if influential vibration occurs. The second time and fourth time, the subject observes the object image based on the new method. That is, the object image is hidden at the step position. The second time, the object image moves the same as the object image the first time, before detecting influential vibration caused by the step. Once the vibration is detected, the object image is hidden. The fourth time, the object image moves the same as the object image the third time, before detecting the influential vibration. Once the vibration is detected, the object image is hidden.

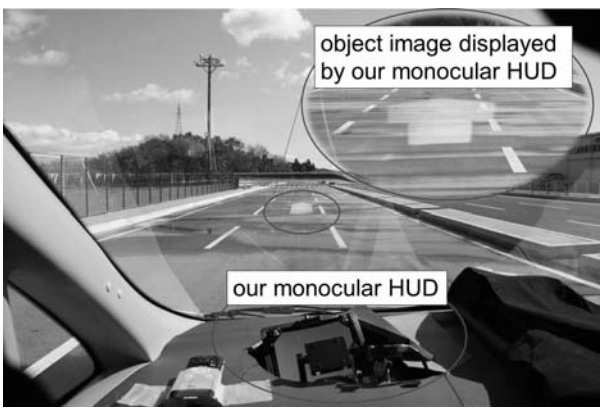

Fig. 13 Our monocular HUD and an example of the object image.

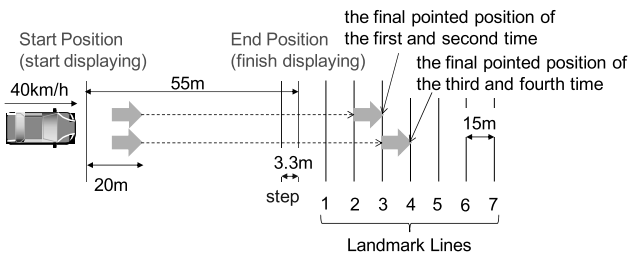

Fig. 14 The test course for experiments.

Fig. 15 The step on the course.

For the evaluation, all subjects tell us the perceived depth position where the subject thinks the object image points based on the landmark lines printed on the test course, as shown in Fig. 14, each time. We compare the averages of the distances between the perceived depth positions of all subjects and the final pointed positions, in order to evaluate our method. We regard the distance as an error.

Here, threshold C_z and C_a are 22 [m] and 1.3 [m/s^2], respectively. Constant value A is −9.5 [m/s^2]. These parameters are decided experimentally. Subjects are 8 men from 20 to 40 years of age. Specially, three of them participated in a similar experiment 6 times before the evaluations. We confirmed the standard deviation of the error for each person was small (less than 0.2 [m]). Other 5 subjects did not participate in the experiment before the evaluations because we wanted to reduce a load of them.

4.2 Results of Depth Perception and Discussions

The averages (avg.) and the standard deviations (sd.) of the distances between the perceived depth positions and the final pointed positions (fpp.) are shown in Table 2.

The results listed in Table 2 show that the error of our new method based on vibration was less than that of the previous method, which is confirmed by t-test ($p < 0.01$) as same as Table 1 shows. Our new method can point at the final pointed position of 60 [m] 2.1 times more accurately than the previous method does. We think that hiding the object image during vibration can control the depth position more accurately than displaying the uncontrolled object image in the real car, too. In the case of pointing at the road to turn at, if the error average is more than 10 [m], the driver

Table 2 Comparison of our new dynamic perspective method and the previous dynamic perspective method.

Method	Previous		New	
fpp.	45 [m]	60 [m]	45 [m]	60 [m]
error avg. [m]	12.2	7.1	4.7	3.4
error sd. [m]	4.6	8.2	4.8	7.9

will not determine where to turn because there are a lot of roads whose widths are less than 10 [m]. The experimental result shows that our new method can work well in the case of pointing at the road to turn at. However, we cannot estimate that our method works well in the case of pointing at the obstacles by using only this result. The obstacles are much smaller than the width of the road. In future work, we want to verify that our new method works well in the case of pointing at obstacles.

The results listed in Table 2 also show that the error in the case of pointing at 60 [m] position is less than the error in the case of pointing at 45 [m] during vibration. Moreover, our new method can point at the final pointed position of 45 [m] 2.6 times more accurately than the previous method does. 2.6 times are a little larger than 2.1 times which is calculated in the case of the final pointed position of 60 [m], which is similar result in the case of simulation. This is because people can observe the image object clearly at the near position. The position of the object image whose size is big at the near position affects the depth perception greatly. We think using the position of the complemented object image at the near final pointed position is more effective than using it at the far final pointed position. Therefore, changing how to hide the object image based on the final pointed positions could be effective.

These results show hiding the object image during simple vibration (only pitching vibration) is effective. However, these experiments do not deal with yawing vibration, rolling vibration and random complex vibration. In future work, we will confirm the ability of our new dynamic perspective method during various vibrations.

5. Conclusion

This paper has dealt with the problem of the depth control by using the object image displayed by monocular HUD while a car vibrates. Our aim is to point at the positions in the real world accurately by the see-through object image on the monocular HUD in a moving car. To solve the problem, we focus on the fact that people complement the hidden image by the continuous images observed previously. We usually point at the positions by using previous dynamic perspective method. The method points at the positions by changing the size of the object image from big to small and the height of it from low to high. Additionally, our new method hides the object image when big vibration is detected by the acceleration sensor. People can perceive the pointed positions accurately by complementing the position of the hidden image.

We performed experiments in order to confirm our new dynamic perspective method could point at the positions more accurately than the previous method did while the car vibrates. We implemented our monocular HUD and the acceleration sensor in a car. 8 subjects observed the object image displayed by two methods while the car moved at 40 [km/h] along the test course where the simple vibration occurred. The two methods were the previous dynamic perspective method and our new method that hides the object image during the vibration. The results showed that our method pointed at the depth position of 60 [m] within a 3.4 [m] error. Our new method pointed at the depth position more accurately than the previous method, which was confirmed by t-test. In future work, we intend to confirm the effect of our method when the car moves along a street where more complex vibration occurs. Moreover, we want to verify that our new method works well in the case of pointing at obstacles on the road.

References

[1] K.H. Kim and K.Y. Wohn, "Effects on productivity and safety of map and augmented reality navigation paradigms," IEICE Trans. Inf. & Syst., vol.E94-D, no.5, pp.1051–1061, May 2011.
[2] S. Kim, H. Kim, S. Eom, N.P. Mahalik, and B. Ahn, "A reliable new 2-stage distributed interactive TGS system and augmented reality navigation paradigms," IEICE Trans. Inf. & Syst., vol.E89-D, no.1, pp.98–105, Jan. 2006.
[3] M. Tonnis, C. Lange, and G. Klinker, "Visual longitudinal and lateral driving assistance in the head-up display of cars," Proc. ISMAR '07, pp.91–94, 2007.
[4] T. Nakatsuru, Y. Yokokohji, D. Eto, and T. Yoshikawa, "Image overlay on optical see-through displays for vehicle navigation," Proc. ISMAR '02, pp.109–112, 2002.
[5] M. Tonnis, L. Klein, and G. Klinker, "Perception thresholds for augmented reality navigation schemes in large distances," Proc. ISMAR '08, pp.189–190, 2008.
[6] G. Weinberg, B. Harsham, and Z. Medenica, "Evaluating the usability of a head-up display for selection from choice lists in cars," Proc. AUI 2011, pp.39–46, 2011.
[7] A. Hotta, T. Sasaki, and H. Okumura, "Depth perception effect of dynamic perspective method for monocular head-up display realizing augmented reality," Proc. IDW '11, pp.521–524, 2011.
[8] T. Sasaki, A. Hotta, A. Moriya, T. Murata, H. Okumura, K. Horiuchi, N. Okada, K. Takagi, Y. Nozawa, and O. Nagahara, "Nobel depth perception controllable method of WARP under real space condition," Proc. SID2011, pp.244–247, 2011.
[9] T. Sasaki, A. Hotta, A. Moriya, T. Murata, H. Okumura, K. Horiuchi, N. Okada, M. Ogawa, and O. Nagahara, "Hyperrealistic display for automotive apllication," Proc. SID2010, pp.953–956, 2010.
[10] H.E. Burton, "The optics of Euclid," J. Opt. Soc. Am., vol.35, no.5, pp.357–372, 1945.
[11] K.H. Park, H.N. Lee, H.S. Kim, H. Lee, and M.W. Pyeon, "Perception thresholds for augmented reality navigation schemes in large distances," International Symposium on Virtual Reality Innovation, pp.261–266, 2011.
[12] A. Moriya, T. Sasaki, and H. Okumura, "Development of an eye tracking system for the monocular head-up display," Proc. IDW '11, pp.525–528, 2011.
[13] M. Wertherimer, "Experimentelle studien uber das sehen von bewegung," Zeitschrift fur Psychologie, pp.161–265, 1912.

Tsuyoshi Tasaki received the B.S. and M.S. degrees in Informatics from Kyoto University in 2004 and 2006, respectively. He works Toshiba Corporation from 2006. He is currently pursuing his PhD degree at the Graduate School of Informatics, Kyoto University, too. His research interests include the robot sensing based on the robot vision and augmented reality. He is a Member of RSJ.

Akihisa Moriya received the M.S. degrees in biophysical engineering from Osaka University in 2003. He works Toshiba Corporation from 2003. His research interests include the human sensing and augmented reality display. He is a Member of ITE and SCIE.

Aira Hotta received the M.S. degrees in Inorganic Materials from Tokyo Institute of Technology in 1994. She works Toshiba Corporation from 1994. Her research interests include the augmented reality display. She is a Member of ITE.

Takashi Sasaki received the M.S. degrees in biosphere science from Hiroshima University in 1990. He works Toshiba Corporation from 1990. His research interests include the augmented reality display. He is a Member of ITE.

Haruhiko Okumura received the B.S., M.S. and Ph.D. degrees in Electrical Engineering from Waseda University in 1981, 1983 and 1995, respectively. Joined Toshiba R&D Center in 1983. He has been engaged in developing image-pickup equipment, TV telephone and convention equipment. He is now working on research and development of driving technologies for flat panel displays. From 2008 to 2009, he has been an Electronic Information Display (EID) Committee Chair, Institute of Electronics, Information and Communication Engineers of Japan (IEICE). Since 2007, he also has been Display Electronic Systems (DES) Workshop chair of the International Display Workshop (IDW).