Effect of Coupling Haptics and Stereopsis on Depth Perception in Virtual Environment

Laroussi Bouguila, Masahiro Ishii and Makoto Sato
Precision and Intelligence Laboratory, Tokyo Institute of Technology
4259 Nagatsuta cho Midori ku Yokohama shi 226-8503 – Japan
Tel: +81 45 924 5050 Fax: +81 45 924 5016
Email : laroussi@pi.titech.ac.jp

ABSTRACT
The present work aims to overcome human's instability of stereopsis depth perception in virtual environment provided with stereo graphics. We investigated the possibility of integrating both haptic sensation and stereopsis cues to improve the task of localizing and manipulating objects in virtual environment.

Keywords
Stereopsis, Haptic, Depth perception, Stereogram.

INTRODUCTION
Our understanding of the three dimensional environment around us is inferred from a variety of depth cues. The information received from our diverse perceptual modalities is consistently and reliably integrated into a unitary perception of the world. To reach such capability of complete and accurate perception in multi-modal virtual environment, see figure 1, artificial display systems and feedback cues have to work into concert with each other so as to create the illusion of natural sense of interaction.

While the needs and the roles of multi-modal virtual environment are well documented by a large body of theoretical development, relatively little is known about the perceptual issues resulting from the combination of different artificial display systems. In real world, human's senses such as vision, audition, haptic, etc. are almost always in agreement with each other, so an accurate depth perception is possible. In virtual reality, however, technological limitations are usually such that only a small subset of available cues can be fully implemented. Other cues are either missing or not well displayed. As well, the use of different technology with different bandwidths and mechanisms make the integration between modalities not consistent and in correspondence with human sensitivity. Uncontrolled feedback cues in a virtual scene can end up providing false depth information and may create a sensory conflict, which can lead to distorted perceptions and unskillful interaction.

RESEARCH BACKGROUND AND MOTIVATION
It has been proven in many researches that stereoscopic images are very effective in improving interactions in virtual environment. As well, haptic sensations are known to impart users with realistic feeling about physical interactions, which improve the control over virtual objects. However, in most virtual reality systems, when these two modalities are coupled together, the real world (users) and the virtual world are separated, and do not interact directly with each other. For example, when the task of the user is to grasp a virtual object, usually we provide an imaginary graphic hand to interact directly with the virtual object, whereas the real hand is used to control the interaction, see figure 2-a. Such indirect control makes the user think in terms of manipulating objects in remote site, and do not have the conviction that he is within the virtual environment itself. Although, representing body's parts involved in the virtual environment with similar graphics has sufficient accuracy for interactions (Butts & McAllister [1], Spain [2], Beaton [3] and Reinhardt [4]), yet, it keeps the real and the virtual worlds non fused, and felt as being two different environments.

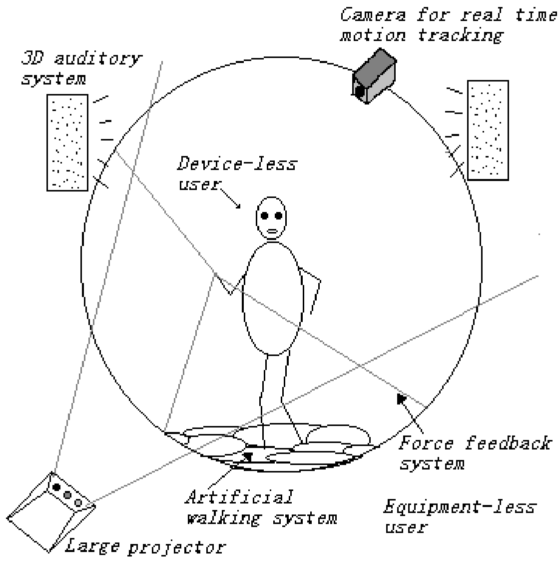


Figure 1: Multi-modal virtual environment

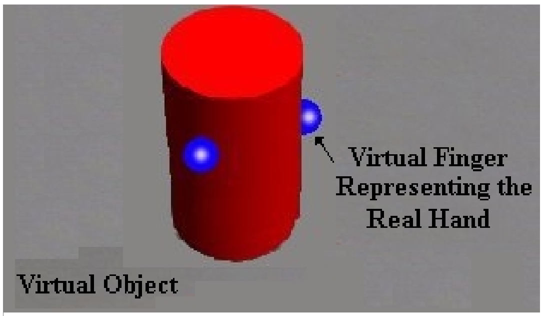


Figure 2-a: Indirect interaction between the real hand and the virtual object

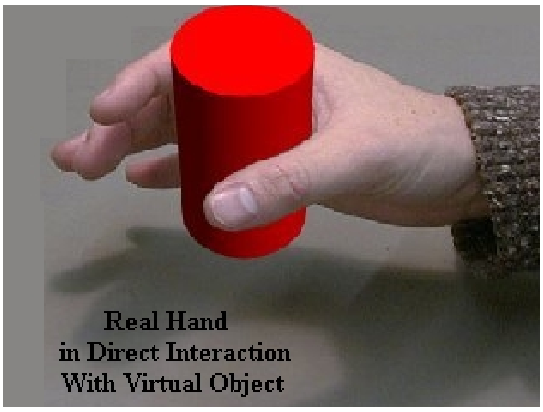


Figure 2-b: Direct interaction between the real hand and the virtual object

ISSUES OF DIRECT COUPLING OF HAPTICS AND STEREOPSIS

In order to provide an accurate integration between stereopsis and haptics cues, three facts have to be considered.

a- Misperception of the binocular stereopsis cue. This phenomena is common where the location of the same object can be perceived at different depth each time.

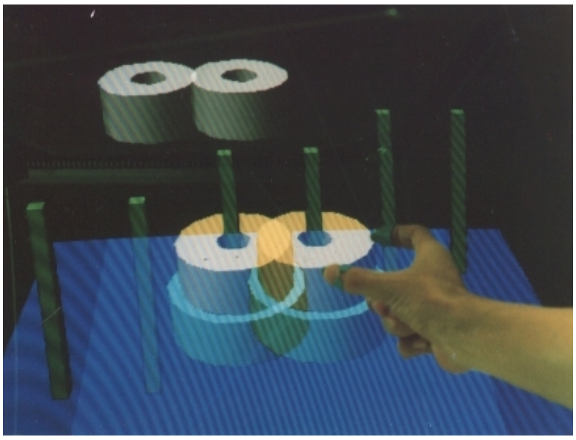


Figure 3: differences between stereoscopic and real display

The depth instability of stereopsis affects directly the coordination between the visual and haptic modalities. A simple reaching to grasp task cannot be done appropriately. Drascic & Milgram [5]

b- Perception of real hand and stereoscopic images differ in a variety of ways. The main difference is that we use accommodation to perceive depth for real objects, whereas we use convergence to perceive depth for virtual objects. Miyashita & Uchida [6]. See figure 3.

c- Occlusion and focus issues between real and virtual objects. Real objects have much stronger visual cues than stereographic virtual objects. For example the hand may occlude the virtual object, but the inverse situation is not possible, which violate interposition cue.

The first issue is of particular interest in the development of mixed and augmented reality systems. The problem of misperceiving the location of objects at different depths is especially important if one of the principal tasks of the user is to reach and grasp an object, or to align objects. The force feedback sensation has to be always displayed in accordance with the virtual environment; any mismatching between the two modalities may cause the failure of the tasks. The present paper presents the results of our investigation about this issue.

PURPOSE OF THE RESEARCH

In the current research, we are interested in giving the user the ability to reach and touch virtual objects by his hand as in real world, see figure 2. We aim to display haptic sensation at the same position where objects are perceived. This entails establishing whether or not the haptic display can overcome the stereopsis distance-scaling problem, and be consistently and accurately matched with the stereoscopic display.

At our knowledge there has been no work reported in the literature, which is directly relevant to this particular coupling in virtual environment systems. The results of this study can lead to improve the accuracy of perception and coordination between haptic sensation and visual cues in augmented and mixed reality applications.

In the next section, we present the overall condition of the experiments. In the next sections, we describe and discuss about our experiments. There are four experiments. The first one was dedicated to state the variance of stereopsis depth perception without any force feedback. The second experiment was to the frame the haptic depth consistency without any visual display. The third experiment was the combination of the previous experiments and to study the effect of coupling both haptic and stereopsis on depth perception. The last experiment was carried out to study the depth threshold of force feedback within which the two modalities can be fused. In the last section the remaining problems are discussed.

METHOD

To get the positions of perceived depth we used the Big SPIDAR interface, B. Laroussi & M. Sato.[7], see figure 4. The device can track the real position of subject's hand as well as provides both stereoscopic images and force feedback sensation. The system keeps the transparency of the working space and do not hide any part of the screen, see appendix A for more details. The stereoscopic image was displayed on a 120-inch large screen and observers viewed it by wearing liquid-crystal-shuttered glasses. In order to limit the perceptual capabilities of the observer's eyes to only a single depth cue, all experiments were performed in a completely dark room, and the stereoscopic image was reduced to a basic random dot stereogram, which display a simple positional depth distance. We adopted such approach to isolate the role of other visual cues from the acquisition of depth information.

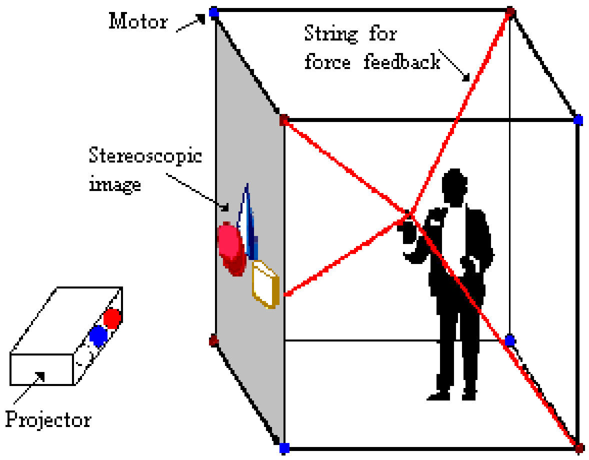


Figure 4:The haptic device Big SPIDAR

EXPERIMENT I

Subjects

Four males served as observers. One was experienced observer knowledgeable about stereopsis. The others were naïve. All observers have normal or corrected to normal vision. None of them reported any haptic deficiencies. Although it was not necessary for the experiment, all subjects were familiar with haptic devices and virtual environment.

Apparatus

In all experiments we used the human-scale haptic device Big SPIDAR (Figure 4) to get hand positions as well as to display force feedback sensation. The device is coupled with a large screen where computer generated random-dot stereogram is displayed. The observer was seated on a chair inside the device facing the screen. To provide constant viewing distance and avoid cues known to affect depth perception such as perspective and motion parallax, the observer's head was stabilized with a chinrest at predetermined distance of 70 cm from the screen. All experiments were carried out in a completely dark room to discard any aiming point or background information that the observer may use as depth reference. The random-dot stereograms were made up of a square matrix consisting of 230x230 square dots, each of which had an equal probability of being displayed or not. The square has no background, and all dots were displayed in only red color. The square displayed to the left eye had a range of disparities added to it by shifting its horizontal position. Eight crossed disparities were employed in this experiment. The disparity ranges from 20 dots to 160 dots, and each disparity was in integer multiple of 20 dots. When the observer fuse left and right image he always perceive the square being in front of the screen and at hands reach. Figure 5 gives more detail about the apparatus of the system.

Procedure

Each observer was tested individually. In each session the observer sat on the chair, wear the fingering provided by the haptic device Big SPIDAR, the positioned his head on the chinrest and looked straight ahead to the square. The task was to move his right hand forward until it become aligned with the right side of the square. Observers were not able to see their hands, because the room was dark and also instructed previously to not occlude the image by heir hands, figure 6. When the observer subjectively believes that his hand is at the same depth as the square he report this judgment vocally by saying "HERE". As the hand position is tracked in real time via the Big SPIDAR device, the experimenter on clicking a mouse button recorded it instantly. The same task was repeated randomly at least 4 times for each disparity (8 disparities in all). After each trial the observer was asked to close his eyes while the

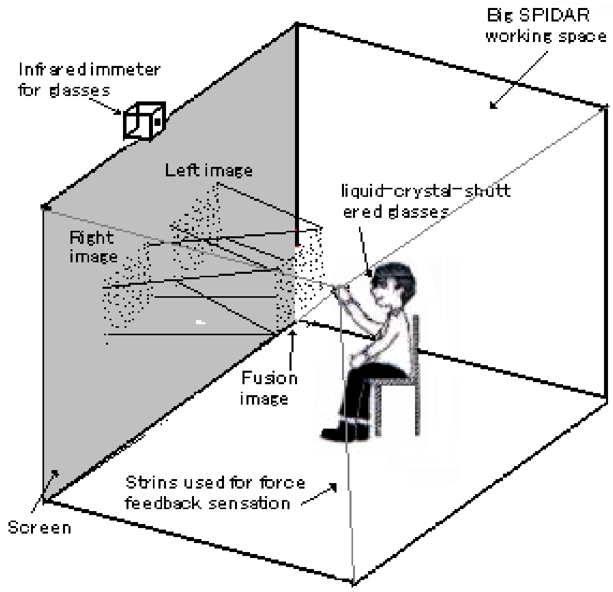


Figure 5: Apparatus of the experiments installation

experimenter change the disparity by clicking another mouse button, this pause is about 2 to 3 seconds.

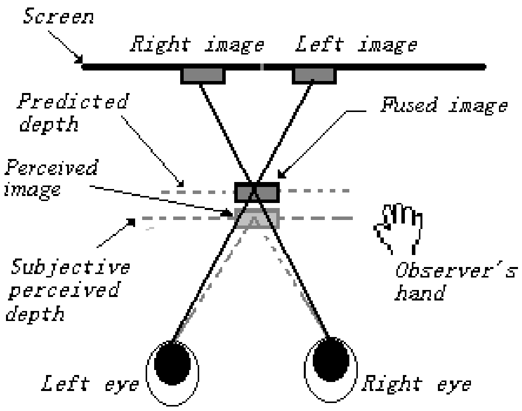
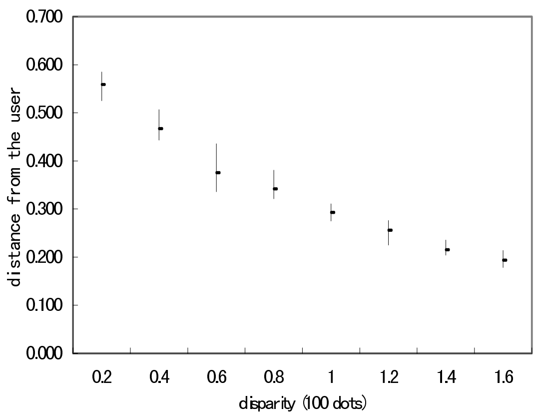


Figure 6: The observer subjectively position his at the same depth as the perceived square

The procedures were explained beforehand to each observer and given a short time of practice. Each session was preceded by at least 1 minute of dark-adaptation. Observers were allowed free eyes movement and as much time as required to estimate the depth of the square. Observers responded to the all disparities, repeated each 4 times at least (46 trials in all)

Results

The results are summarized in figure 7. The graph shows the means and deviations of perceived depth as a function of disparity. Depth is expressed in terms of the distance interval between the observer and the perceived square. The small rectangles in the graph represent the mean of depth for each disparity. The vertical line represents the



Graph 7: Perceived depth for each disparities

depth variation related to each disparity. The standard errors across disparities ranged from 1.5 cm to 5 cm of the mean depth and averaged 2.3 cm. Nearly all subjects responded quickly and confidently to all trial, usually viewing them for about 2 to 4 seconds before making a response. The results from the naïve observers did not differ in any systematic way from those from the experienced observer.

Discussion

We notice the clear relationship between the perceived depth and the physical disparity for all trials. This clear trend is evident for all subjects, and merely confirms that the magnitude of perceived depth increase with disparity as estimated by the following equation deduced from figure 8.

$$\frac{I}{2d} = \frac{R}{2(D-d)} \tag{1}$$

Where ***I*** is the interocular distance. ***D*** is the viewing distance. ***R*** is the disparity distance and ***d*** is the depth distance from the observer. Equation (2) gives the depth ***d*** expressed as function of variable **R**.

$$d = \frac{ID}{R+I} \tag{2}$$

As ***D*** and ***I*** are constant values, the depth is affected only by the change of disparity. That is, when disparity becomes smaller the square tends to be farther and inversely. Assuming that the interocular distance is 6 cm, equation (2) was represented by figure 9. Based on this theoretical prediction, the results presented in the graph 1 can be

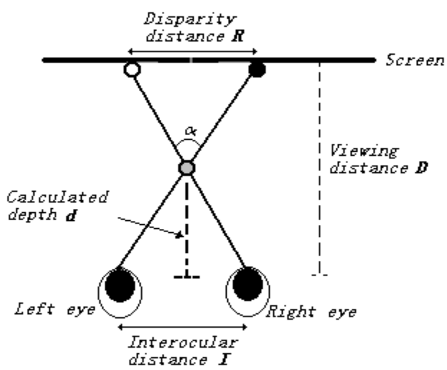


Figure 8: Simple positional disparity

considered as stable and reliable.

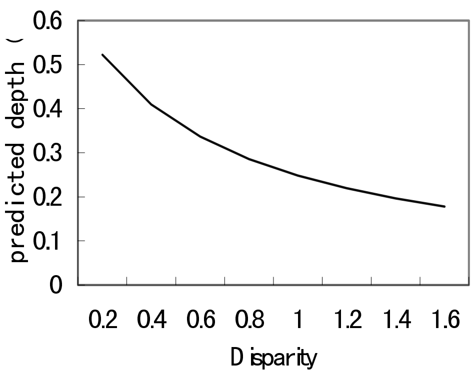


figure 9: Predicted depth

If we look to the accuracy of observers' depth perception in regard to disparities, we find that most of the user showed strong sensitivity to depth variation when the disparity was small. This ability decrease when the disparity become bigger, i.e. the change of disparity from 20 dots to 40 dots generate about 9 cm of depth variation, whereas the same change of 20 dots from 140 to 160 dots gives only 3 cm of depth variation. Ogle [8], Sperling [9] and others researchers classified this phenomena into two kinds of stereopsis, "patent" and "qualitative". The first represent accurate stereopsis and occurs when disparity is small. The later represent imprecise stereopsis and occurs when disparity is big. This observation is validated by equation (3), which is the derivative function of equation (2). When the disparity is small the slop is big and gradually flattened while disparity is increasing, see figure 10.

$$|d_R| = \left| \frac{-ID}{(R+I)^2} \right| \tag{3}$$

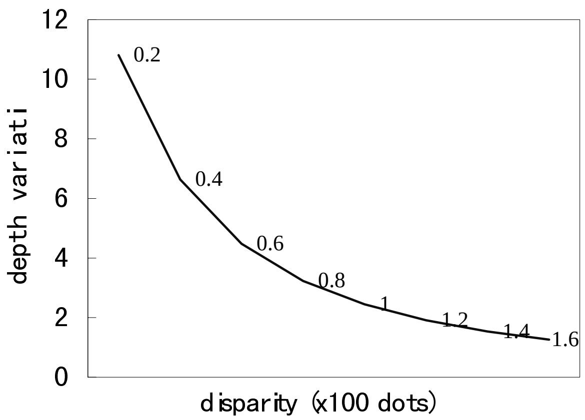


Figure 10: Sensitivity to depth variation

An important question arises immediately in regard to all these data and observations. At which position the force feedback should be displayed in such a way it can be perceived at the same position as the stereoscopic image?. The following experiment was conducted to address this issue.

Experiment II
Coupling Stereopsis and Haptics

As discussed above, for each disparity, the square can be perceived randomly at different depth positions. It is impossible to predict at which distance the observer will perceive the stereoscopic square. In this experiment we investigate whether coupling force feedback with stereopsis can lead to improve the depth perception. We aim to display both haptic and stereopsis cues approximately at the same depth, in such a way they can be perceived by the subject as the same thing.

Apparatus and procedure

The same experimental apparatus and procedure as in experiment (I) were used. With the extent of displaying haptic sensation. The observer was asked to move forward his hand until it comes into contact with the virtual wall; the square and the haptic wall were supposed to be at the same depth. The observer then has to judge whether he could perceive either one single fused haptic and stereopsis depth cue or definitely two different depth cues. For each trial the observer was asked to state the haptic depth position in regards to the stereopsis. Subjective feeling about the situation was reported orally by saying "Front", "Same" or "behind". "Front" situation occurs when the haptic wall stops the hand before it reaches the same depth as the visual square. "Same", means both the haptic wall and the viewed square are perceived at the same distance. Observer responds "Behind" when his hand moves forward until it passes behind the square to reach the haptic wall. The position of haptic wall was within the range of depth deviation determined by the previous experiment. For each disparity we displayed force feedback at eight different depth positions each of which was tried at least twelve times. Both disparity and haptic depth was displayed randomly.

Results and discussion

The result of this experiment showed that some haptic depth position are more representative and coincide often with stereoscopic depth than others. These positions are usually located somewhere in between the depth range of each disparity. At the extremities of these ranges simultaneously "Front" or "Behind" situation were dominant, see figure 11. The figure represents the case of disparity equal to 40 dots. As you can see from the figure, the range of positions where both modalities are perceived as one is smaller than the depth variation caused by binocular cue only. We find also that, displaying force

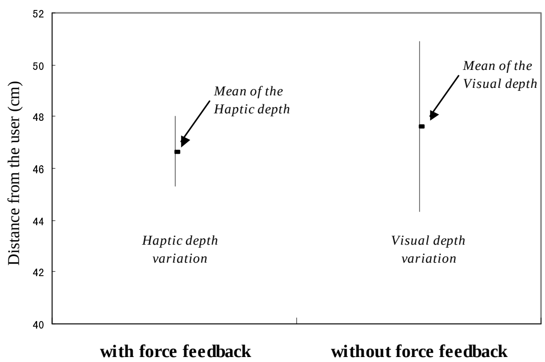


Figure 11: Threshold of force feedback position

feedback at the same position at the mean of visually perceived depths is the most representative position. Usually, there was 70% of chance to reach a complete integration between both modalities. Whereas, displaying the force feedback far from the mean created a conflict situation between the two modalities and usually the user have to ignore one of them depending on the dominance of the cue. As conclusion, we consider that coupling force feedback with stereopsis increase the stability of depth perception.

EXPERIMENT III

While the previous experiments most of the subjects reported that they perceived depths faster when they were provided by force feedback sensation. The current experiment was carried to state this fact.

Apparatus and procedure

The same experimental apparatus and procedure as in experiment (II) were used. The experiment had two sessions, in the first session we don't display force feedback. The observer was asked to move forward his hand until it comes into the same depth position as the stereoscopic square. For each trial the user was asked to close his eyes at first until a beep sound is displayed. Then the subject open his eyes try to fuse both right and left images and position his hand as fast and accurate as possible. The time required for each trial, disparity magnitude and hand position were recorded. Once the subject finish estimating the depth, he was asked to close again his eyes. The experimenter changes meanwhile the disparity magnitude. The second session was identical with the first one except we displayed this time force feedback sensation approximately at the same level of the stereoscopic square. So the subject moves his hand forward until it comes into contact with the haptic wall.

Results and discussion

The results of this experiment are presented in figure 12. We can observe that the average time required to perceive the depth positions is shorter when the force feedback is provided. When the haptic wall stops the hand, the eyes converge directly at the hand's position, which supposed to be at the same depth as the stereoscopic square. I think that information of depth preceded the stereopsis; this information is inputted by posture of the subject's arm.

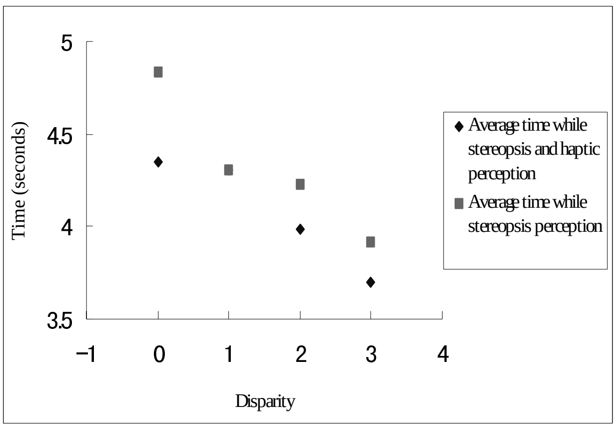


Figure 12: the small square represents the average time needed to perceive depth while using only stereopsis. The dark dots represents the perceived depth when both visual and haptic cues are provided.

When there is no force feedback the subject has to scan the image more times so as to succeed the fusion of both left and right images. Also, we can see that the time needed to decide the position of the stereoscopic square is proportional to the depth. This fact may be caused by the distance that the hand has to move to reach the same depth as the visual or haptic wall. Figure 13 gives more detail about the stability of depth perception in term of time. If we look to the case of disparity magnitude equal to 80 dots, the variation of perception time is about 3 seconds when no force feedback is displayed. This time delay is dropped to less than one minute when haptic sensations are provided. The same gap can be seen for disparity magnitude equal to 20 dots. This gap is smaller when the disparity is about 40 or 60 dots. These two disparities represents respectively the depth of about 30 to 40 cm, which is considered as the preferred depth of many subjects, may be this can be the raison. We can say that adding force feedback sensation will speed up the depth perception especially when the depth to perceive is too close or too far.

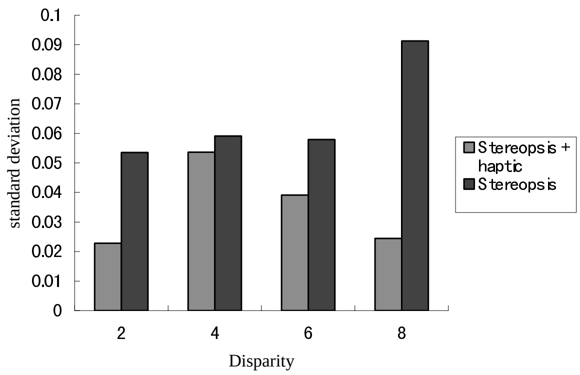


Figure 13: Standard deviation of perception time

CONCLUSION

There are a wide variety of factors that affect the achievement of a complete integration between haptics and stereopsis in virtual environment. However, we proposed a preliminary approach that can fairly improve the perception of virtual objects location by adjusting the position of haptic display so as to match the stereoscopic image. Also It was clear that adding force feedback sensation within appropriate threshold distances improve the integration of both haptic and stereopsis modalities. This supports our assumption that haptics may overcome stereopsis scaling distance problem. As well we showed that adding force feedback improved the time needed to perceive depths, this, is of special interest in augmenting the reality of virtual environment.

However, there are still many issues to investigate, especially to find a model that can couple haptics and stereopsis with high level of integration when we use real image and not random dot stereograms. Farther studies are necessary about the effect of viewing distance and the effect of occlusion on the depth perception.

REFERENCES

1. Butts, D.R.W. and D.F. McAllister, “Implementation of true 3D cursors in computer graphics”. SPIE Volume 902: 1988. P74-84
2. E H Spain, “A psychophysical investigation of the perception of depth with stereoscopic television displays”. Phd of dissertation, University of Hawaii, may 1984.
3. R J Beaton, DeHoff, N Weiman, W Hildebrandt “ An evaluation of input devices for 3D computer display workstations”. SPIE vol. 761, p 94-101, 1987.
4. F W Reinhardt, “Effect of depth cues on depth juddgements using a field-sequential stereoscopic CRT Display”, PhD dissertation, Industrial engineering & operations research dept, Virginia polytechnic institute 1990
5. D. Drascic & P. Milgram “Positioning Accuracy of virtual stereographic pointer in a real stereoscopic video world” SPIE vol. 1457, p 58-69, San Jose Sep 1991
6. T. Miyashita and T. Uchida “Fatigueless stereoscopic display” 3D 映像 Vol. 4 No. 4 April 1990
7. B. Laroussi Y. Cai & M. Sato “New haptic device for human-scale virtual environment: Scaleable SPIDAR” ICMR’97 volume IXB, p 67-72, Tamperre Finland 1997
8. Ogle, K. N. “The optical space sense”. In The Eye vol. 4 ed. Davson, H. New York: Academic
9. Sperling, G “Binocular vision: A physical and neural theory” Am. f. Psychol. P 462-534, 1983
10. B. Laroussi & M. Sato “Object recognition potentiality of haptic virtual environment”. Proceeding of ICPA’9, Toronto , July 1997
11. Y. Hirata, M.Sato and H.Kawarada, ``A Measuring Method of Finger Position in Virtual Work Space" Forma, Vol.6, No.2, pp.171-179(1991)

Appendix A

CONCEPT OF SCALEABLE-SPIDAR

The device is derived from the original desktop SPIDAR device, which was introduced late in 1990 by professor Makoto sato *et al* [11]. As shown in figure 14-a, Scaleable-SPIDAR is delimited by a cubic frame that enclose a cave-like space, where the operator can move around to perform large scale movements. The experimental prototype is 27m^3 size (3m x 3m x 3m). Within this space, different aspect of force feedback sensations associated mainly with weight, contact and inertia can be displayed to the operator's hands by means of tensioned strings. The front side of the device holds a large screen, where a computer-generated virtual world is projected. Providing such a combination of haptic and visual feedback cues is indispensable to lets the operator's eyes and hands work in concert to explore and manipulate objects populating the virtual environment.

The device uses tensioned string techniques to track hands position as well as to provide haptic feedback sensations. The approach consists mainly on applying appropriate tensions to the four strings supporting each fingering worn by the operator. The force feedback felt on the operator's hand is the same as the resultant force of tension from strings at the center of the fingering; next subsection gives more detail about forces and position computation. In order to control the tension and length of each string, one extremity is connected to the fingering and the other end is wounded around a pulley, which is driven by a DC motor. By controlling the power applied to the motor, the system can create appropriate tension all the time. A rotary encoder is attached to the DC motor to detect the string's length variation, Figure 14-b. The set of DC motor, pulley and encoder controlling each string is fixed on the frame.

Force Control

Scaleable-SPIDAR uses the resultant force of tension from strings to provide force display. As the fingering is suspended by four strings, giving certain tensions to each of them by the means of motors, the resultant force occurs at the position of the fingering, where transmitted to and felt by the operator's hand.

Let the resultant force be $\vec{f}$ and unit vector of the tension be $\vec{u}_i$ (*i=0,1,2,3*), figure 14-a, the resultant force is :

$$\vec{f} = \sum_{i=0}^{3} a_i \vec{u}_i \qquad (a_i > 0)$$

Where a_i represents the tension value of each string. By controlling all of the a_i the resultant force of any magnitude in any direction can be composed [5].

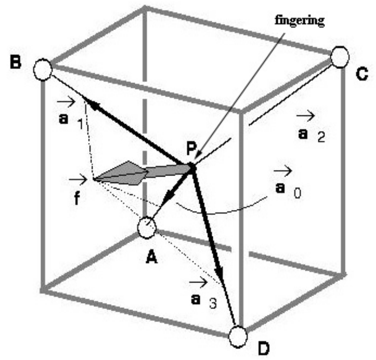


Figure 14-a: Resultant force of tension

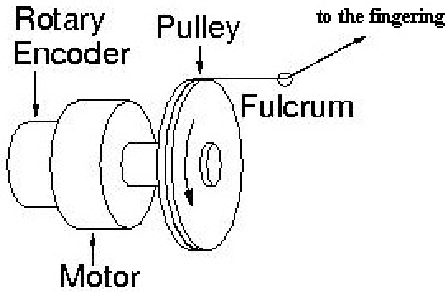


Figure 14-b: Motor and rotary encoder

Position Measurement

Let the coordinates of the fingering position be *P (x,y,z)*, which represent in the same time the hand position, and the length of the i^{th} string be l_i *(i=0, ..., 3)*. To simplify the problem, let the four actuators (motor, pulley, encoder) A_i be on four vertexes of the frame, which are not adjacent to each other, as shown by figure 15. Then *P (x,y,z)* must satisfy the following equations (Eqs).

$$(x+a)^2 + (y+a)^2 + (z+a)^2 = l_0^2 \qquad (1)$$

$$(x-a)^2 + (y-a)^2 + (z+a)^2 = l_1^2 \qquad (2)$$

$$(x-a)^2 + (y+a)^2 + (z-a)^2 = l_2^2 \qquad (3)$$

$$(x+a)^2 + (y-a)^2 + (z-a)^2 = l_3^2 \qquad (4)$$

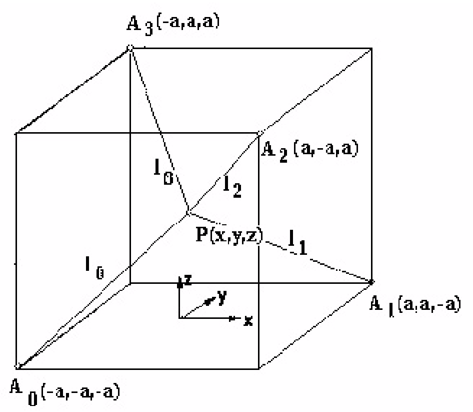


Figure 15: Position measurement

After differences between the respective adjacent two equations among equation (1)-(4) and solve the simultaneous equations, we can obtain the position of a fingering (hand) as the following equation (5):

$$\begin{cases} x = \dfrac{(l_0^2 - l_1^2 - l_2^2 + l_3^2)}{8a} \\ y = \dfrac{(l_0^2 - l_1^2 + l_2^2 - l_3^2)}{8a} \\ z = \dfrac{(l_0^2 + l_1^2 - l_2^2 - l_3^2)}{8a} \end{cases} \qquad (5)$$

EXPERIMENTAL PROTOTYPE

The experimental prototype provides two fingerings to be worn by the operator on both hands, Figure 16-b. The fingerings are made of light plastic material and the size can fit to any operator. As well, this small device leaves the hand free and easy to put on and off. Although the operator can wear the fingering on any finger, middle finger is most recommended. The bottom of this finger is close to the center of hand, and the force feedback applied on this position is felt as being applied to the whole palm.

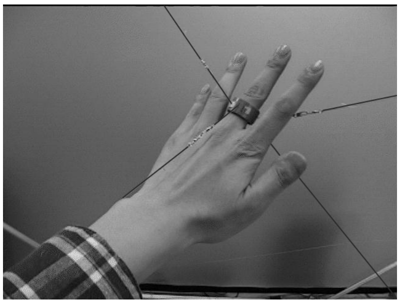


Figure 16-b: The fingering

To provide the appropriate tensions and lengths of the strings, a personal computer (PC) is used to control an 8-bits D/A, A/D converter and a VME bus, which control respectively the currents entering the motors and detect the changes occurred on each rotary encoder. The PC is connected to a graphics workstation that provides a real-time video image of the virtual world. The apparatus of the prototype is shown by Figure 16a.

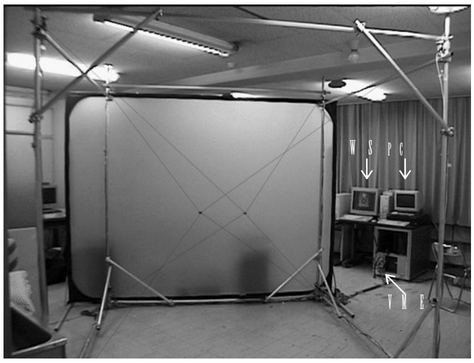


Figure 16-a: Apparatus of the Scaleable-SPIDAR

Performance of Scaleable-SPIDAR

Position Measurement Range: the coordinates origin is set to the center of the framework. The position measurement ranges of all x, y and z in[-1.50m, +1.50m].

Static Position Measurement Error: the absolute static position measurement errors are less than 1.5cm inside the position measurement range.

Force Feedback Range: within the force displayable sphere, force sensation range is from 0.005*N* (minimum) to 30*N* (maximum) for all directions.

System Bandwidths:

- ✧ Video: 10 ~ 15 Hz
- ✧ Audio: 22 kHz (stereo)
- ✧ Position measurement and force display: > 1200 Hz (depends also on hardware installation)