第 27 回日本バーチャルリアリティ学会大会論文集（2022 年 9 月）

Predicting finger position during midair visual-haptic interaction

Yue Zhang[1)], Masahiro Fujiwara[1)], Yasutoshi Makino[1)], and Hiroyuki Shinoda[1)]

1) 東京大学 新領域創成科学研究科 (〒 277-8561 千葉県 柏の葉 5-1-5, zhang@hapis.k.u-tokyo.ac.jp)

Abstract: This paper proposes a method to utilize the prediction of finger movement in haptic interaction through airborne ultrasound. Using the Autoregressive model confirmed that the predicted position of the fingertip at 200ms ahead can be estimated with an accuracy of 5mm or less with finger velocity provided less than 10cm/s.

Keywords： Aerial interface, movement prediction

1. Introduction

When considering a visual-haptic interactive system, synchronization between visual and haptic information is important to enrich the quality of the experience. This study focuses on reducing the delay for ultrasonic tactile display when presenting tactile feedback while moving the finger.

Ultrasonic tactile display uses numbers of ultrasonic transducers to creates an ultrasonic focus at an arbitrary position in the air. The size of the focus can be reduced to about a wavelength. The current 40kHz transducer can provide focus approximately 1cm. Assuming that the width of the fingertip is also 1 cm, an 1cm shift from the center of the fingertip will result in loss of the tactile sensation. When the speed of the finger is assumed to be 5cm/s, the position of the finger 0.2s ahead shifts 1cm. This will cause the problem that if there is a delay of more than 0.2s from the measurement of the fingertip position, the tactile sensation will not be presented in ultrasound midair tactile display. To overcome this time latency, we propose a method to estimate the fingertip position and present the tactile sensation at intended location.

There are three types of time delays in the visual-haptic interactive system we considered. One is sensing delay. The latest hand tracking system has minimum latency of 13ms [1]. However, hand reconstruction error due to various conditions like blurring and occlusion could reduce the tracking accuracy [2] and cause asynchronous in between visual-haptic interaction.

The second type is the delay in the calculation of the ultrasound phase. To form complex sound field, GPU-based high-speed algorithm such as GS-PAT [3] have been proposed. This paper assumed a single focus and the computational load in this case is negligible compared to other delays. Also, when predicting the fingertip position, the computation time will be added as a delay to the processing. In this study, a simple Autoregressive model was introduced to reduce this time delay.

The third type of delay is the time from the system sends a actuation signal to the ultrasonic wave is actually generated. Hardware limitations also cause temporal latency ranges from 33.27ms with a standard deviation of 17.8ms [4] to 200ms [5] for different haptic devices.

Several prediction methods have been proposed to address such delay problems. An existed method proposed to resolve the issue of time latency between visual-haptic interactions by accurately determining interaction timing between the hand and the contact object. A model has been proposed [6] for predicting interaction timing and sending the actuation signal to haptic devices in advance to achieve optimal realism. Moreover, Hopping-Pong [7] was proposed, the system has succeeded in applying a computationally controlled ultrasound force to a high-speed moving ping-pong ball to change the trajectory by predicting the trajectory of the ball based on the assumption that it follows a parabolic trajectory.

Our proposal augments the existing visual-haptic system by predicting fingertip position in advance. The visual-haptic system we consider takes Airborne Ultrasound Tactile Display (AUTD) [8] as haptic hardware. Consider a series of processes such as measurement of the depth camera, phase calculation, and actuation of the hardware, assume the whole process is 0.2 seconds, the position of the fingertip needs to be predicted and the output of AUTD needs to be directed to the fingertip to aim the ultrasonic wave at the intended fingertip. Therefore, we use 0.2 seconds as the criterion and evaluate the shift of the fingertip based on the judgment that the predicted fingertip position does not deviate by 1 cm from the intended fingertip position. This prediction can be used to reduce visual-haptic interaction latency to achieve the result of increasing realism in mixed reality.

2. Method

This paper proposed a method to use the prediction of finger movement to overcome the time difference produced during the visual-haptic interaction. Finger movement is a time series problem, especially the future position of a finger that can be calculated based on the previous finger position. Our proposal provides the possibility of calculating future finger position by using the AR (Autoregressive) model which is a model using a linear combination of past values of finger position to forecast the future finger position. The model follows

$$y_{t+\delta t} = \beta_0 + \beta_1 \cdot y_t + \beta_2 \cdot y_{t-1} + \beta_3 \cdot y_{t-2} + \epsilon_t, \qquad (1)$$

where the y_t represents the current finger position. Therefore, $y_{t+\delta t}$ is the predicted position in the time interval δt in advance. y_{t-2} represents finger position at the 3 frame before. This model's output is the predicted finger position value in the next time interval. The Autoregressive model was performed by using the data from previous 3 frames, which is AR(3). Bayesian information criterion (BIC) was used for calculate the optimal values of β_i. When the model with order over 2, the converges of the results stay stable. Since three frames are used, information corresponding to velocity and acceleration is incorporated into the model. In general, the higher the order, the better the prediction result, however, higher order could also cause over-fitting. Therefore, we used order of 3 to increase prediction accuracy but not to over-fit the results.

In this paper, our proposed finger position prediction method is based on a haptic-optic display (see Fig. 1). An aerial 3D image (blue square displayed in Figure 1) presents in midair by the display, while the finger moves toward the image, the depth camera located above the display measures all the contact positions between the virtual surface and the finger and calculate the average position from them. Our goal is to identify prediction errors to verify the prediction accuracy.

3. Experiment

3.1 Experimental setup and procedure

To verify the prediction errors of finger movement, we use a haptic-optic display (Fig. 1) to collect data for finger movement. While the speed of moving a finger varies from the person, we include 2 males and 3 females, ages 23-35 to make a contact with the aerial 3D image. They were asked to contact the image with their right index finger starting from the left and continue to move their finger to the right with two kinds of speed, "fast" and "slow", all based on their understanding of "fast" and "slow".

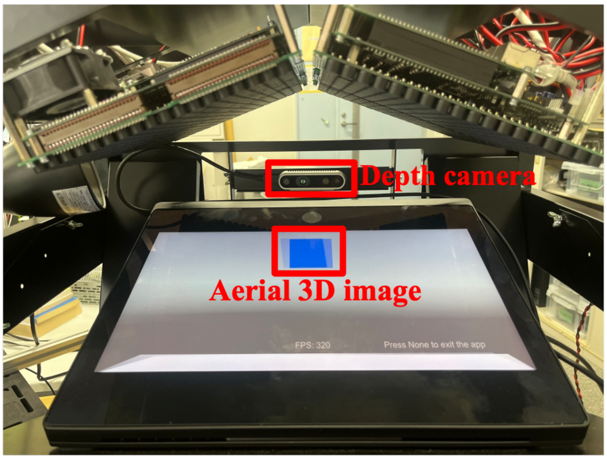


Fig.1: Haptic-optic display

The haptic-optic display (Fig. 1) was designed to display both optical image and tactile feeling. A light field display to present an aerial 3D image (Holo Player One, Looking Glass Factory), with a depth camera (Intel Real Sense D435) to detect the position of a finger. In this study, we only used the function of optical image display and a finger position measurement. When a finger makes a contact with the presented midair image, the depth camera tracks the contact position of the fingertip. And record the finger position during the moving.

The experiment image (Fig. 2) shows each subject contacting the image starting from the left side, moving their finger horizontally to the right. While the finger moves, the contact position is recorded in a constant time interval. One movement represents one trial. Each subject needs to complete at least 3 valid trials for each kind of speed, "fast" and "slow".

3.2 Results

While the contact between a finger and an aerial 3D image was made, the position of the finger was recorded. The moving speed of the finger can be calculated based on the recorded finger position and the recorded time. Since people tend to slow down their moving speed during the experiment's starting and finishing periods. We consider the maximum instantaneous speed in a trial as the moving speed of the finger. Based on the results of recorded finger movement speed, we distinguish "slow" movement (≤10cm/s, each trial normally includes more than 80 data points) and "fast" movement (>10cm/s,

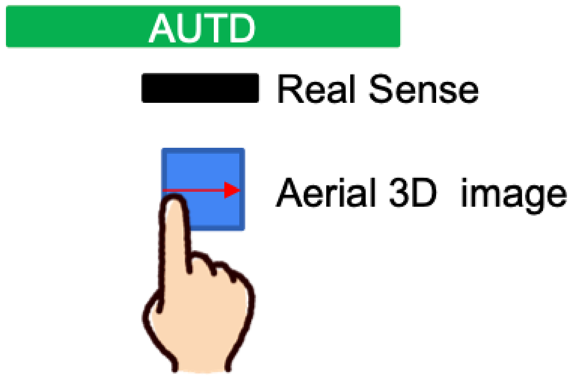


Fig.2: Experiment image

each trial normally includes less than 50 data points).

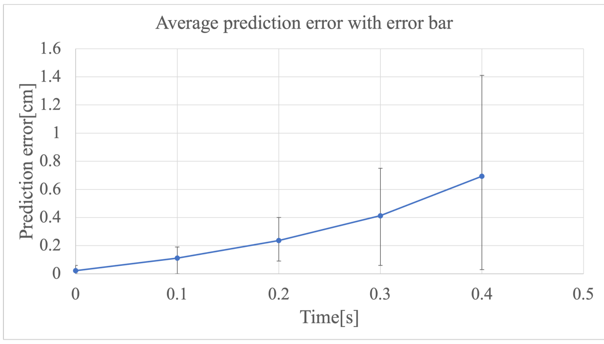
We take the previous 70% of recorded data as input data to train the model and use the trained model to forecast the future position of the finger. Calculate the difference between forecast value and test data (the last 30% of recorded data) as prediction error. Results are separated into two sets, one for "slow" movement and one for "fast" movement.

The average prediction error was calculated, and the results for "slow" movement shows in Fig. 3 (a). From the result, the average prediction error is 0.24cm while predicting the finger position is 0.2s ahead. The maximum value of prediction error at 0.2s is 0.4cm, which is less than 0.5cm. By contrast, the average prediction error for "fast" movement shows in Fig. 3 (b). From the result, the average prediction error is 0.54cm at 0.1s, and the maximum prediction error is over 1.2cm.

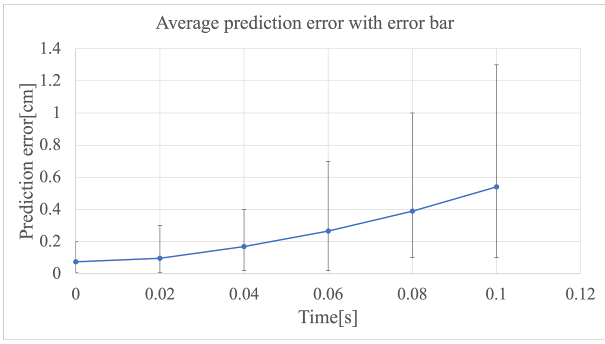
Measuring finger position by using the haptic-optic device, for "slow" movement with the maximum instantaneous speed of 10.6cm/s across 5 subjects was obtained. And measured prediction error results for "slow" movement fit within 0.5cm at 0.2s ahead. Considering the width of the human finger is around 1cm, a 0.5cm error makes it possible for providing haptic feedback to the finger even under the 0.2s delay condition. For "fast" movements, we also observed that the average prediction error exceeded 0.5 cm within 0.1 second.

4. Discussion

The proposed method conducted the experiment by using the haptic-optic display. We record and calculate the position of the finger movement; the result shows the position was recorded at around 58fps. As the speed of the finger increases, while the sampling rate and record rate keep the same, it leads to fewer data points collected in each experiment trial. Also, while the finger moves at a fast pace, the variation in the distance that the finger tries to move in one frame can be large. Addressed reasons are possible explanations for the results for the finger moving faster than 10cm/s.



(a) "slow" movement



(b) "fast" movement

Fig.3: Results of taking average of prediction error at different timestamps across different subject with upper and lower bounds. (a)Average for "slow" movement, (b)Average for "fast" movement

5. Conclusion

In this paper, we show that it is possible to forecast linear finger movement in a visual-haptic system, a haptic-optic display to track the movement of the finger and provide haptic feedback correspondingly. The result we obtained shows that while the finger movement is within 10cm/s, the prediction error for finger position at 0.2s ahead can be fit into 0.5cm, which is half of the finger's width.

Moreover, we observed prediction hardness while the finger moved at a fast pace. The overall prediction error for finger moves faster than 10cm/s is much greater than the results we obtained for "slow" movement. In order to control the prediction error within 0.5cm while using haptic-optic display, it needs to control the speed of finger movement within 10cm/s to improve the realism of the system.

In future work, we will evaluate the finger movement in 2D, and 3D space separately. And looking forward to evaluate the actual performance of the proposed prediction method by combining a midair tactile feedback system.

References

[1] F. Mueller *et al.*, “Ganerated hands for real-time 3D hand tracking from monocular RGB,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 49–59.

[2] A. Armagan *et al.*, “Measuring generalization to unseen viewpoints, articulations, shapes and objects for 3D hand pose estimation under hand-object interaction,” in *Proc. Eur. Conf. Comput. Vis.*, Cham: Springer, 2020, pp. 85–101.

[3] Diego Martinez Plasencia, Ryuji Hirayama, Roberto Montano-Murillo, and Sriram Subramanian. 2020. "GS-PAT: High-Speed Multi-Point Sound-Fields for Phased Arrays of Transducers," *ACM Transactions on Graphics*, 39, 4, Article 138 (August 2020), 12 pages.

[4] Tatsuya TAKEI, Shun SUZUKI, Masahiro FUJIWARA, Yasutoshi MAKINO and Hiroyuki SHINODA, “Reduction of delay in midair haptic interaction by hand position prediction,” *24th Annual Conference of the Virtual Reality Society of Japan, 5C-05, Hongō campus, Tokyo*, Sep. 11-13, 2019.

[5] E.J.Gonzalez *et al.*, “Reach:Extending the reachability of encountered-type haptics devices through dynamic redirection in VR,” *Proc. 33rd Annu. ACM Symp. User Interface Softw. Technol.*, 2020, pp. 236–248.

[6] M. Salvato, N. Heravi, A. M. Okamura and J. Bohg, "Predicting Hand-Object Interaction for Improved Haptic Feedback in Mixed Reality," in *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 3851-3857, April 2022

[7] Tao Morisaki, Ryoma Mori, Ryosuke Mori, Yasutoshi Makino, Yuta Itoh, Yuji Yamakawa, and Hiroyuki Shinoda. 2019. "Hopping-Pong: Changing Trajectory of Moving Object Using Computational Ultrasound Force." In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces (ISS '19)*, Association for Computing Machinery, New York, NY, USA, 123–133.

[8] S. Suzuki, S. Inoue, M. Fujiwara, Y. Makino and H. Shinoda, "AUTD3: Scalable Airborne Ultrasound Tactile Display," in *IEEE Transactions on Haptics*, doi: 10.1109/TOH.2021.3069976.