An Interactive 4D Vision Augmentation of Rapid Motion

Tao Tao
Graduate School of Information Science and Technology, Osaka University
Osaka, Japan
tao.tao@lab.ime.cmc.osaka-u.ac.jp

Photchara Ratsamee
Graduate School of Information Science and Technology, Osaka University
Osaka, Japan
photcyber@gmail.com

Yuki Uranishi
Graduate School of Information Science and Technology, Osaka University
Osaka, Japan
uranishi@ieee.org

Kiyoshi Kiyokawa
Nara Institute of Science and Technology
Nara, Japan
kiyo@is.naist.jp

Tomohiro Mashita
Graduate School of Information Science and Technology, Osaka University
Osaka, Japan
mashita@ime.cmc.osaka-u.ac.jp

Haruo Takemura
Graduate School of Information Science and Technology, Osaka University
Osaka, Japan
mashita@ime.cmc.osaka-u.ac.jp

ABSTRACT

We propose an interactive 4D visualization (3D space with an additional time dimension) for the purpose of understanding rapid motion in dynamic 3D information. With an interactive 4D visualization, the user can not only observe dynamic 3D motion from different viewpoints, but also freely adapt the speed of visualization on each viewpoint. We developed a system by reconstructing 3D data from an RGBD camera into a VR environment so that the user can visualize the 3D information via an HMD. The experiment results show that our proposed system outperforms conventional 2D and 3D visualizations in terms of both the user's recognition accuracy and view counts when observing rapid motion.

CCS CONCEPTS

• **Human-centered computing** → **Visualization design and evaluation methods**;

KEYWORDS

Augmented Reality, 3D Motion Data, Interactive 4D Visualization, Slow Motion

ACM Reference Format:
Tao Tao, Photchara Ratsamee, Yuki Uranishi, Kiyoshi Kiyokawa, Tomohiro Mashita, and Haruo Takemura. 2018. An Interactive 4D Vision Augmentation of Rapid Motion. In *AH2018: The 9th Augmented Human International Conference, February 7–9, 2018, Seoul, Republic of Korea.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3174910.3174934

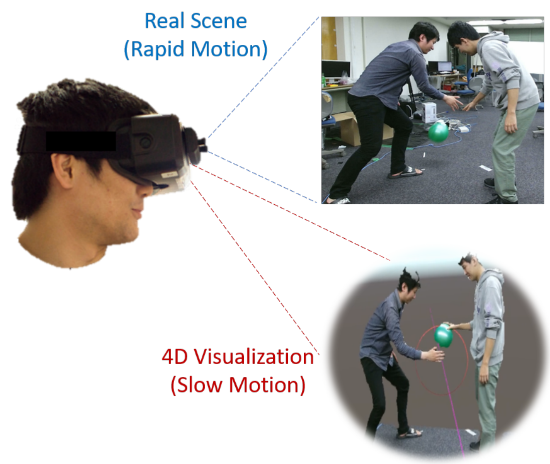

Figure 1: 4D visualization for Augmented Reality (the red circle shows the slow motion area)

1 INTRODUCTION

Human eyes are used to perceive information and recognize motion in the real world. However, due to the limited capabilities of vision, humans have difficulty understanding rapid motion happening in real time. For example, during a game of basketball, a player tries to steal the ball from their opponent (as presented in Fig. 1). Because this action is extremely fast, a referee would have a difficult time determining whether the player touched the ball or their opponent's hand. Augmented Reality (AR) is one of the key technologies which enables a user to improve their vision capabilities through wearable AR displays [3, 8, 13]. A number of new AR visualizations using wearable displays have been proposed. For instance, Orlosky et al. [8] utilized fisheye lenses to expand a user's vision in AR systems. Yano et al. [13] utilized wide lenses and evaluated four different augmentative methods. Ito et al. [3] proposed a system which can predict future trajectory information to assist the user in understanding an object's motion. Mori et al. [5] proposed a furniture arrangement system which grants the user the ability to freely pick their viewpoint. Finally, Orlosky et al. [7] proposed a visualization of dynamic text in an AR system. However, none of these researchers dealt with motion data.

AH2018, February 7–9, 2018, Seoul, Republic of Korea

ACM ISBN 978-1-4503-5415-8/18/02…$15.00
https://doi.org/10.1145/3174910.3174934

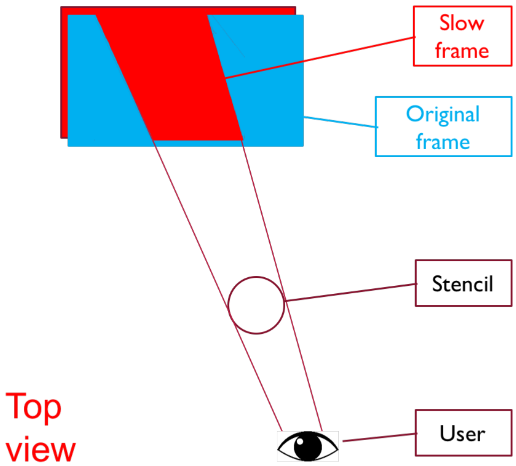

Figure 2: The layout of 4D visualization

Recently, various methods for visualizing motion in AR have been proposed. Potthast [9] developed a helmet which allows the user to perceive the world in slow motion. Langlotz et al. [4] proposed a new visualization method utilizing flash-trail effects applied in real time to visualize the path and motion within 2D videos. However, those methods converted the 3D motion data into 2D data, resulting in the loss of 3D information. 3D visualization allows users to freely observe the 3D motion data from any viewpoint. To visualize rapid motion data, this method is not sufficient, as the user can easily lose track of rapid motion data. As a result, object recognition accuracy will be decreased, requiring more views than even using 2D visualization. We need to allow the user to visualize 3D motion data in an immersible 3D environment using wearable displays. Therefore we propose a novel 4D visualization method (3D with an additional time dimension) where the user can observe 3D motion data from a preferred viewpoint and freely adapt the visualization speed on each viewpoint as presented in Fig. 1. We also compare 2D, 3D, and 4D visualization, and analyze the user experience among each visualization to verify the advantage of the proposed visualization.

2 METHODOLOGY

2.1 Recording and 3D Reconstruction

The complete 3D reconstruction of static and dynamic objects is required for perfect visualization. Recently, Newcombe et al. [6] and Dou et al. [2] made an achievement in visualizing the motion by reconstructing 3D data in real time from multiple cameras installed around the environment. However, in our study, we reconstruct the 3D motion data in terms of a 2.5D pointcloud using an RGBD sensor (int our case, a Kinect V2) with existing software [1]. We import the pointcloud and display it frame by frame in a virtual reality environment (in our case, Unity) from a fixed position to record the 3D motion data.

2.2 4D Visualization

For 4D visualization, we used two 3D video frame streams. Firstly, we decelerated one frame stream to a play speed that the user can define in advance. Secondly, we rendered this frame stream in the Unity. Finally, we rendered the other frame stream in the same coordinates as the original play speed. Due to the rendering order, the slow motion frame was occluded by the original play speed frame. We used the shader's stencil buffer [12] in an area chosen by the user in real time. In this case, we manually created the stencil with a sphere shape in advance which can ensure the user has the same circle shaped slow motion area from any different viewpoint. In this stencil area, the user can see the slow motion frames through the original speed frame. This layout is presented in Fig. 2.

2.3 Interaction

An HTC Vive [10] is used for supporting the interaction between the user and system. The user can use the HTC Vive controller to replay the 3D motion video. In addition, we supported a function where the user chooses a point of interest using raycast collision [11]. Finally, the user can control the play speed of the 3D motion video.

3 EXPERIMENT

To evaluate our interactive 4D visualization system and compare the performance of the 4D visualization and conventional visualization, we conducted an experiment to test a rapid motion video. Our system utilized an HTC Vive on a desktop with an Intel Core i7-7700 CPU, NVIDIA GeForce GTX 1070 graphic card, and 16 GB RAM. The running performance was approximately 40 fps.

3.1 Experiment Setup

We recruited a total of 8 participants with an age range of 23∼27. The task was to watch a quick typing scene presented in Fig. 3 and then recall the typing order as quickly and correctly as possible. Every video lasted for approximately 5 seconds. Participants were divided into two groups of 4 participants. One group was tested using 2D modes followed by 3D modes, and the other group was tested on 3D modes first followed by 2D modes. Participants were asked to recall each typing order orally, and the result, completion time, and views were recorded.

Every participant was tested using 4 different modes of visualization, consisting of a normal 2D mode, normal 3D mode, slow motion 2D mode, and slow motion 3D mode (4D visualization). The participants were asked to watch 6 videos for every mode. In every video, 6 letters were typed in random order. Participants were asked to answer the typing order after finishing watching one video. To control the experiment time, we limited the number of views to 5 per video. The participant could answer the result in less than 5 views if they had confidence in recognizing the correct typing order. When the participants finished watching all the videos, they were asked to fill the user experience questionnaire. In the normal 2D mode, participants were asked to finish the task on the monitor using the keyboard. They can press the key "R" for starting or replaying the video. In the slow motion 2D mode, participants can also move the mouse to choose the slow motion area. If the mouse remained still for more than 1 second, slow motion would start

Figure 3: Experiment task

automatically; if the mouse moved, the slow motion would end. In the normal 3D mode, participants were asked to use the HTC Vive. They could use a Vive controller to interact with the system. Pressing the trackpad button was used for starting or replaying the video, similar to the key "R". In the slow motion 3D mode, participants could also press the menu button once to select a slow motion area through raycasting. When they had chosen an area, they could keep pressing the trigger button to begin slow motion. Both the slow motion 2D and 3D modes had a slow motion speed that was 5 times slower the recorded video. After all the participants finished testing, we summed up and analyzed the results.

3.2 Experiment Result

In this section, we analyzed the accuracy of recognized characters, number of views, complete time, and the participants' questionnaire results.

3.2.1 Accuracy. As every participant watched 6 videos in every mode, and 6 characters were typed in every video, the sum of observed characters per mode is 36. We sum all the correctly recognized characters within every mode, and then calculate the average of the 8 participants' results.

As shown in Fig. 4, 4D visualization had the highest accuracy with 29.125 correctly recognized characters, which is 81% of the total. We used analysis of variance (ANOVA) in the comparison of 4 modes, and used the Holm method for p-value adjustment to evaluate the participants' average accuracy. The three asterisks, one asterisk, and plus symbols between bars in the Fig. 4 indicates a significance level of $p < 0.001$, $p < 0.01$ and $p < 0.05$ respectively.

From this analysis, we found significant differences between the slow motion 3D mode and normal 3D mode with $p < 0.001$. Also, we found significant differences between slow motion 3D mode and the slow motion 2D mode with $p < 0.01$. We calculated the Cohen's d as effect size by using the means and standard deviations of two groups. The Cohen's d was 1.28 between the slow motion 2D mode and the slow motion 3D mode. Moreover, between the normal 2D mode and the slow motion 3D mode, the Cohen's d was 4.21, and between the normal 3D mode and the slow motion 3D mode, the Cohen's d was 5.65.

3.2.2 Views. To control the experiment time, we limit the maximum views as 5. We recorded the participants' views for each video. We then calculated all the participants' average views of every mode. Because of the maximum limitation, we don't use the

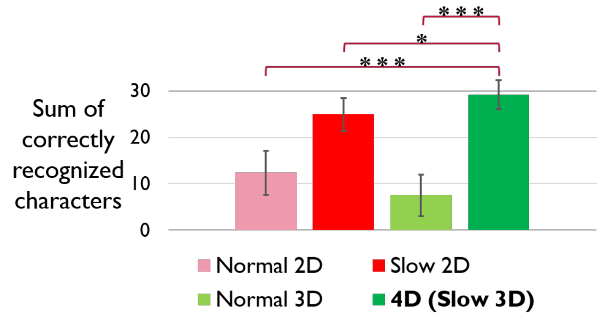

Figure 4: The accuracy of correctly recognized characters

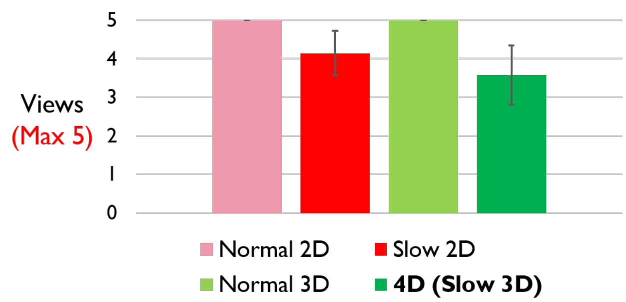

Figure 5: Views of 4 modes

Table 1: Average completion time of 4 modes

Mode	Average Complete Time (Min)
Normal 2D	00:37.98
Slow 2D	01:15.89
Normal 3D	00:38.48
4D (Slow 3D)	01:07.70

ANOVA for analyzing the views. The number of views result is shown in Figure 5. In normal 2D and 3D modes, every participant used the maximum available views to observe the quick typing in all videos. In the slow motion modes, the slow motion 3D mode has a mean of 3.583 views which is lower than the slow motion 2D mode views of 4.146.

3.2.3 Completion Time. We recorded the completion time of every video, then calculated the average value in each mode. The result is shown in Table 1. In this table, we found that the slow motion mode's completion time is approximately twice as much as the normal modes because the play speed is 5 times slower. In the slow motion modes, we found that the completion time of the 3D mode is shorter than that of the 2D mode. In addition, we also noticed that the completion of the normal 3D mode was even longer than that of the normal 2D mode.

3.2.4 Questionnaire Result. All the participants were asked to fill a questionnaire after finishing the experiment. This questionnaire consisted of 16 questions including each participant's basic

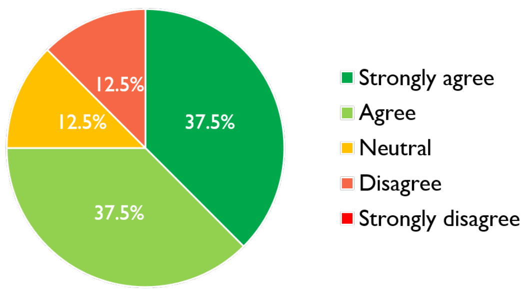

Figure 6: The result of the question 'Would you want to use the 3D slow motion mode for this task'

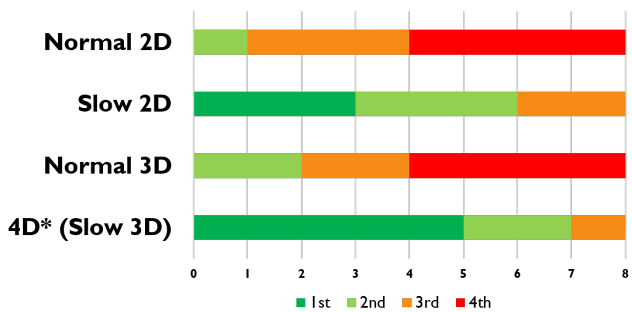

Figure 7: The result of question rank the 4 modes

information, HMD experience, and subjective evaluation of our system. Most participants had experience using HMDs. The result of each user's experience in each mode shows that 75% of participants want to use the proposed slow motion 3D mode, with 37.5% strongly agreeing and 37.5% agreeing (Fig. 6). When ranking the 4 modes, as shown in Fig 7, more than half of participants (5 participants) considered the slow motion 3D mode to be the best visualization method of the 4 modes. Furthermore, 2 participants ranked the slow motion 3D mode in second place, and no participant ranked it in last place.

3.2.5 Discussion. According to the experiment's result, we found that in normal speed modes, the accuracy of the 3D mode is lower than the 2D mode. A reason for this can be that for a participant observing a rapid scene it is difficult to find the best viewpoint to observe a target, requiring more time in a 3D environment.

Furthermore, we found that in slow motion modes, the accuracy of the 3D mode is higher than the 2D mode. It can be interpreted that in the 4D mode, the participant has enough time to find a better viewpoint to observe the target. For this experiment task (as presented in Fig 3), the participant can use the camera's viewpoint to observe the position of pressed key clearly, and use the reverse direction viewpoint to understand the key label easily.

We found that 2 participants didn't agree with the statement "Would you want to use the 3D slow motion mode for this task" from the questionnaire. The feedbacks from these participants noted difficulty in finding a viewpoint in the 3D environment and unclear detail of motion due to low frame rate of RGBD sensor.

Table 2: The result of evaluation experiment

	View Point	Accuracy	Average Views
Normal 2D	Fixed	34%	5.000
Slow 2D	Fixed	69%	4.146
Normal 3D	Freedom	21%	5.000
4D (Slow 3D)	**Freedom**	**81%**	**3.583**

To solve these issues, we would like to develop an automatic viewpoint searching algorithm to support the user in finding the optimal viewpoint, and to utilize a high-speed camera to increase the detail of motion.

4 CONCLUSIONS

In this paper, we proposed the concept of a 4D visualization for rapid 3D motion where a user can control the speed of 3D motion data. We planned an evaluation experiment and conducted the evaluation experiment with 8 participants. The result of this experiment (Table 2) shows that our proposed 4D visualization improves the rapid motion's recognition accuracy and reduces the required number of views. All the results show that our visualization has better performance than conventional visualization.

REFERENCES

[1] Jasper Brekelmans. 2014. Brekel Pro PointCloud v2. (2014). http://brekel.com/brekel-pro-pointcloud-v2/
[2] Mingsong Dou, Sameh Khamis, Yury Degtyarev, Philip Davidson, Sean Ryan Fanello, Adarsh Kowdle, Sergio Orts Escolano, Christoph Rhemann, David Kim, Jonathan Taylor, et al. 2016. Fusion4d: Real-time performance capture of challenging scenes. *ACM Transactions on Graphics (TOG)* 35, 4 (2016), 114.
[3] Yuta Itoh, Jason Orlosky, Kiyoshi Kiyokawa, and Gudrun Klinker. 2016. Laplacian vision: Augmenting motion prediction via optical see-through head-mounted displays. In *Proceedings of the 7th Augmented Human International Conference 2016*. ACM, 16.
[4] Tobias Langlotz, Thanh Nguyen, Dieter Schmalstieg, and Raphael Grasset. 2014. Next-generation augmented reality browsers: rich, seamless, and adaptive. *Proc. IEEE* 102, 2 (2014), 155–169.
[5] Mami Mori, Jason Orlosky, Kiyoshi Kiyokawa, and Haruo Takemura. 2016. A Transitional AR Furniture Arrangement System with Automatic View Recommendation. In *Mixed and Augmented Reality (ISMAR-Adjunct), 2016 IEEE International Symposium on*. IEEE, 158–159.
[6] Richard A Newcombe, Dieter Fox, and Steven M Seitz. 2015. Dynamicfusion: Reconstruction and tracking of non-rigid scenes in real-time. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 343–352.
[7] Jason Orlosky, Kiyoshi Kiyokawa, and Haruo Takemura. 2013. Dynamic text management for see-through wearable and heads-up display systems. In *Proceedings of the 2013 international conference on Intelligent user interfaces*. ACM, 363–370.
[8] Jason Orlosky, Qifan Wu, Kiyoshi Kiyokawa, Haruo Takemura, and Christian Nitschke. 2014. Fisheye vision: peripheral spatial compression for improved field of view in head mounted displays. In *Proceedings of the 2nd ACM symposium on Spatial user interaction*. ACM, 54–61.
[9] Lorenz Potthast. 2012. The Decelerator Helmet - A slow motion for Real Life. (2012). https://www.designboom.com/technology/the-decelerator-helmet/
[10] Lily Prasuethsut. 2016. HTC Vive: Everything you need to know about the SteamVR headset. *Retrieved January* 3 (2016), 2017.
[11] Unity Technologies SF. 2010. Physics.Raycast. (2010). https://docs.unity3d.com/jp/540/ScriptReference/Physics.Raycast.html
[12] Unity Technologies SF. 2010. ShaderLab: Stencil. (2010). https://docs.unity3d.com/Manual/SL-Stencil.html
[13] Yuki Yano, Jason Orlosky, Kiyoshi Kiyokawa, and Haruo Takemura. 2016. Dynamic view expansion for improving visual search in video see-through AR. In *Proceedings of the 26th International Conference on Artificial Reality and Telexistence and the 21st Eurographics Symposium on Virtual Environments*. Eurographics Association, 57–60.