Touch-Consistent Perspective for Direct Interaction under Motion Parallax

Yusuke Sugano
Institute of Industrial Science
The University of Tokyo
4-6-1 Komaba, Meguro-ku
Tokyo, 153-8505 JAPAN
sugano@iis.u-tokyo.ac.jp

Kazuma Harada
Institute of Industrial Science
The University of Tokyo
4-6-1 Komaba, Meguro-ku
Tokyo, 153-8505 JAPAN
dani@iis.u-tokyo.ac.jp

Yoichi Sato
Institute of Industrial Science
The University of Tokyo
4-6-1 Komaba, Meguro-ku
Tokyo, 153-8505 JAPAN
ysato@iis.u-tokyo.ac.jp

ITS'12, November 11–14, 2012, Cambridge, Massachusetts, USA.

Abstract
A 3D display is a key component to present virtual space in an intuitive way to users. A motion parallax-based 3D display can be easily combined with multi-touch surfaces, and it is expected to bring a natural experience of viewing and controlling 3D space. However, since virtual objects are rendered in accordance with the head position of the user, their projected positions are not fixed on the display surface. We propose a novel formulation of head-coupled perspective that adaptively changes the position of the projection image plane to maintain touch consistency of direct interaction.

Author Keywords
Head-coupled Perspective, Direct Manipulation

ACM Classification Keywords
H.5.2 [**Information interfaces and presentation**]: User Interfaces. - Graphical user interfaces..

General Terms
Theory, Human Factors

Introduction
Techniques for intuitive interaction with 3D virtual environments have long been investigated. Direct 3D interaction is expected to provide the most natural user

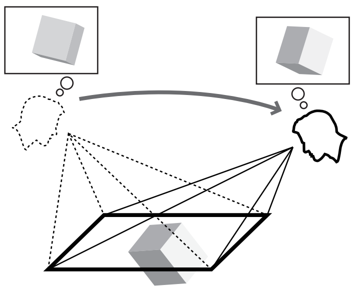

Figure 1: Illustration of head-coupled perspective

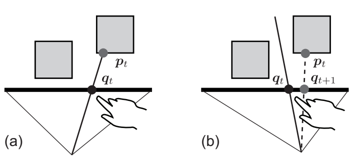

Figure 2: Touch inconsistency during viewpoint change with head movement

experience in many scenarios, such as 3D desktop environments, and to appeal to not only novices but also experts.

Technical elements for 3D interaction are mainly divided into two components: display and input [1]. For the display technique, a 3D display is obviously the most intuitive way to present the 3D space to users. Many existing techniques utilize the mechanism of binocular stereopsis for depth perception; however, monocular cues like motion parallax are another important factor for humans to perceive the 3D space.

To present motion parallax cues, the eye or head position of the viewer is used to render 3D objects under a perspective consistent with the viewpoint, which we refer to as *head-coupled perspective* (Figure 1). Conventionally, head-coupled perspective has been studied in the context of VR under the name of Fish Tank VR [8]. Owing to the recent advance in head tracking technologies, it is becoming much easier to achieve head-coupled perspective without forcing users to wear additional devices.

For the input technique, a multi-touch surface provides a great advantage by providing an experience of direct manipulation on virtual objects. Several research efforts have attempted to utilize 2D touch to directly manipulate 6-DoF position and orientation of 3D objects [5, 3, 4]. As discussed in [5], one of the key factors in designing touch-based 3D interactions is directness. It is easy to define pointing and manipulation gestures separately, but such indirect interaction significantly spoils the benefit of touch-based interaction.

However, as illustrated in Figure 2, the projected position of virtual objects are not fixed on the display surface under head-coupled perspective and direct touch interaction technique cannot be applied in a straightforward manner. In other words, if the user moves her head while touching a 3D object, projected position of the object moves away from the touched position. Since most existing touch-based 3D manipulation techniques implicitly assume a static projection mapping, they cannot handle the touch inconsistency problem and the interaction cannot be direct under head-coupled perspective. Some researches have studied 3D manipulation on stereoscopic displays [2, 7, 6]. However, the case of head-coupled perspective has not been discussed well in previous researches. The view of a virtual camera is closely related to users' depth perception, and it is not a trivial task to resolve the touch inconsistency problem without affecting rendered scene.

Our basic idea is adaptively changing the depth of the image plane so that the projected position of the target virtual object is always fixed on the display surface. Similar techniques are sometimes used in the context of VR and 3D user interfaces. However, since the position of the virtual camera always have to be consistent with the user's view position, it is not clear how to incorporate the depth adjustment into head-coupled perspective. In our formulation, the rendered image does not change depending on the touch action, and can be naturally combined with touch-based interaction techniques.

Touch-consistent Perspective

Touch Inconsistency

As briefly discussed earlier, the problem arises when the head-coupled perspective is combined with touch-based interaction. At the time of the initial touch (Figure 2 (a)), touched 2D position $\boldsymbol{q}_t$ can be mapped to a 3D position $\boldsymbol{p}_t$ through back projection. When the viewpoint moves to a new position (Figure 2 (b)), $\boldsymbol{p}_t$ is also projected to a new position $\boldsymbol{q}_{t+1}$, and the initial position $\boldsymbol{q}_t$ does not correspond to $\boldsymbol{p}_t$. If the one-to-one mapping is forced, i.e., the target object is moved in accordance with the viewpoint change so that $\boldsymbol{p}_t$ is projected to $\boldsymbol{q}_t$, the user's head movement can produce unintended object movement. In both cases, interaction techniques that rely on the one-to-one mapping between $\boldsymbol{p}$ and $\boldsymbol{q}$ cannot be direct.

Touch-consistent Relocation of the Image Plane

Let us denote the 3D head position of the viewer in the virtual world coordinates as $\boldsymbol{h} = {}^t(x_h, y_h, z_h)$. The virtual camera is assumed to be directly placed at $\boldsymbol{h}$ and facing the negative Z direction in a right-hand coordinate system. The target plane, which is perceived as fixed on the screen, is defined as a rectangle parallel to the X-Y plane whose four corners in the world coordinates are $\boldsymbol{a} = {}^t(x_l, y_t, z_d)$, $\boldsymbol{b} = {}^t(x_r, y_t, z_d)$, $\boldsymbol{c} = {}^t(x_l, y_b, z_d)$, and $\boldsymbol{d} = {}^t(x_r, y_b, z_d)$. The distance from $\boldsymbol{h}$ to the target plane becomes $\delta_z = z_h - z_d$.

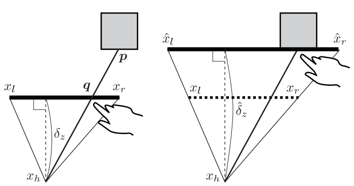

Figure 3: Touch-consistent perspective with adaptive relocation of the target plane

Our key observation is that the inconsistency does not occur when and only when the target position $\boldsymbol{p}$ is on the image plane. To resolve the touch inconsistency problem, we propose changing the position of the target plane adaptively during touch-based interaction. As illustrated in Figure 3, the depth z_d of the target plane is set to be the same as the depth $\hat{z}_d$ of the back-projected 3D position $\boldsymbol{p}$ of the touched position $\boldsymbol{q}$.

To maintain the current visual field of the camera and to avoid the jump of the rendered image depending on the touch action, the size of the target plane is also changed as $\hat{x}_l = x_h + \frac{\hat{\delta}_z}{\delta_z}(x_l - x_h)$ and $\hat{x}_r = x_h + \frac{\hat{\delta}_z}{\delta_z}(x_r - x_h)$, where $\hat{x}_l$ and $\hat{x}_r$ are the new values of x_l and x_r, and $\hat{\delta}_z = \hat{z}_d - z_h$. y_t and y_b are also changed in the same way in accordance with y_h.

Experiments

To compare user experience, users were asked to execute a 3D positioning task with and without the proposed head-coupled perspective. The task was to align the pose of the target dice whose initial position and orientation were randomly defined to the goal. A translucent dice indicates the goal object, and users were asked to align both position and orientation of the dice. If both position and orientation errors became lower than a predefined threshold, the color of the goal dice changed to red. Then users could finish the task in their own time at the best position. There were also randomly placed obstacles, and users were instructed to avoid them as much as possible, so that depth perception of the 3D space had more importance during the experiments.

Reisman et al.'s screen space formulation [5] was used as the baseline method for touch-based 3D manipulation. In their method, 6-DoF position $\boldsymbol{t}$ and orientation $\boldsymbol{r}$ of the target object can be directly controlled by 2D touch through projection error minimization. By changing the projection function at each time step in accordance with the current head position and the target plane, this method can be combined with our proposed touch-consistent perspective without any restriction.

The system consisted of a 3M Multi-Touch Display M2256PW, and a Logitech QuickCam Pro 9000. faceAPI library from Seeing Machines [1] was used to obtain users' 3D head positions $\boldsymbol{h}$. The system runs at more than 60 [fps] with a desktop PC with Intel Core i7 3.33GHz processor, 8GB RAM, and nVidia GeForce GT230 graphics.

Ten people took part in the experiments, and each person

[1] http://www.seeingmachines.com/product/faceapi/

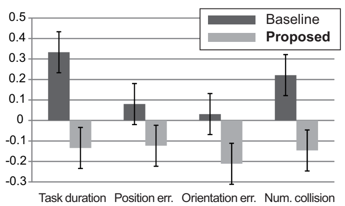

Figure 4: Mean scores of five criteria (duration, position and orientation error, and number of collisions to the obstacles). Since the criteria of the task completion was left to each user's decision and relative difference between modes is important. Absolute values of each criterion were first normalized per user so that they have zero mean and one variance. Lower scores indicate better performance. Error bars indicate standard errors.

executed ten trials for two modes alternately. In the first mode, Reisman et al.'s original method was used as it is and users' head poses were not used to change viewpoint. Instead of using head poses, users could change the camera position using arrow keys. For fair comparison, however, the camera was not allowed to be rotated more than 45 degrees from the initial orientation. The second mode corresponded to the proposed method.

As shown in Figure 4, proposed method achieved better scores in task duration (t-test: $t(99) = 3.21$, $p < 0.01$) and number of collisions (t-test: $t(99) = 2.36$, $p = 0.02$) while pose errors remained the same. After experiments, five-grade subjective scores from 5 (best) to 1 (worst) were collected from each user. The score of the proposed method ($M = 3.7$) was higher than the score of the baseline method ($M = 2.5$) (Wilcoxon signed-rank test: $p = 0.02$).

Conclusion

This paper presented a novel projection method suited to touch-based interaction under head-coupled perspective. In the proposed formulation, The position of the image plane is adaptively changed without causing change of the rendered virtual scene. The consistency between touched positions and target 3D positions is maintained during interaction, and any kind of touch-based interaction techniques can be naturally combined with the motion parallax-based 3D display. Through a user study, interaction using the proposed formulation was shown to enable more smooth and natural user experience for 3D manipulation.

References

[1] Grossman, T., and Wigdor, D. Going deeper: a taxonomy of 3D on the tabletop. In *Proc. Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems* (2007), 137–144.

[2] Hachet, M., Bossavit, B., Cohé, A., and de la Rivière, J. B. Toucheo: multitouch and stereo combined in a seamless workspace. In *Proc. 24th annual ACM symposium on User interface software and technology*, UIST '11 (2011), 587–592.

[3] Hancock, M., ten Cate, T., and Carpendale, S. Sticky tools: full 6DOF force-based interaction for multi-touch tables. In *Proc. ACM Intl. Conf. on Interactive Tabletops and Surfaces*, ITS '09 (2009), 133–140.

[4] Martinet, A., Casiez, G., and Grisoni, L. The effect of DOF separation in 3D manipulation tasks with multi-touch displays. In *Proc. 17th ACM Symposium on Virtual Reality Software and Technology*, VRST '10 (2010), 111–118.

[5] Reisman, J. L., Davidson, P. L., and Han, J. Y. A screen-space formulation for 2D and 3D direct manipulation. In *Proc. 22nd annual ACM symposium on User interface software and technology*, UIST '09 (2009), 69–78.

[6] Valkov, D., Steinicke, F., Bruder, G., and Hinrichs, K. 2D touching of 3D stereoscopic objects. In *Proc. 2011 annual conf. on Human factors in computing systems*, CHI '11 (2011), 1353–1362.

[7] Valkov, D., Steinicke, F., Bruder, G., Hinrichs, K., Schöning, J., Daiber, F., and Krüger, A. Touching floating objects in projection-based virtual reality environments. In *Proc. Joint Virtual Reality Conference (JVRC 2010)* (2010), 17–24.

[8] Ware, C., Arthur, K., and Booth, K. S. Fish tank virtual reality. In *Proc. INTERACT '93 and CHI '93 conf. on Human factors in computing systems*, CHI '93 (1993), 37–42.