Tablet Fish Tank Virtual Reality: a Usability Study

Sirisilp Kongsilp [*] Mintra Ruensuk [†]

Matthew N. Dailey [‡] Takashi Komuro [§]

October 19, 2017

[*] Sirisilp Kongsilp (corresponding author) is with Faculty of Engineering, Thammasat University, Pathumthani, Thailand (sirisilp@engr.tu.ac.th) and with Department of Computer Science and Information Management, Asian Institute of Technology, Pathumthani, Thailand.

[†] Mintra Ruensuk is with Graduate School of Creative Design Engineering, Ulsan National Institute of Science and Technology, Korea.

[‡] Matthew N. Dailey is with Department of Computer Science and Information Management, Asian Institute of Technology, Pathumthani, Thailand.

[§] Takashi Komuro is with Graduate School of Science and Engineering, Saitama University, Japan.

Abstract

In an effort to further expand the impact of VR technology, we developed a new implementation of an existing technique that allows widely accessible consumer-level tablets to display perspective-corrected 3D (Fish Tank Virtual Reality or FTVR) images. To assess the usability of the technique, we conducted a human study using a visual search task previously developed for desktop FTVR systems. We recorded participants' task performance, subjective level of presence, visual fatigue, and informal feedback. In this paper, we identify challenges and opportunities for the adoption of tablet-based FTVR and point toward appropriate directions for future research.

Keywords: *Virtual Reality Fish Tank Virtual Reality Head coupled display Human factors User behavior*

1 Introduction

Fish Tank Virtual Reality (FTVR) was first introduced by Ware, Arthur, and Booth (?, ?), who defined it as "a stereo image of a three dimensional (3D) scene viewed on a monitor using a perspective projection coupled to the head position of the observer." FTVR relies on head position tracking and stereoscopic display techniques. Tablets have become a part of everyday life. Many people use tablets for productivity activities, entertainment, and more. Although tablet applications can present either 2D or 3D environments, most, if not all, applications simply render 2D projections. This perhaps explains why many researchers have introduced new hand-held VR systems built from scratch based on FTVR techniques. Bringing new 3D display capabilities to hand-held devices may bring new opportunities for user experience, a new range of applications, and improved productivity. However, while developing a new device from scratch may overcome current devices' limitations, it also precludes exploitation of the wide accessibility of existing devices to put the technology in peoples' hands quickly and easily. An alternative approach, then, is to develop 3D user interfaces based on FTVR techniques that do not require additional enhancements to a basic tablet. This approach allows us to start immediately, exploiting the ease of use and wide accessibility of existing devices as much as possible. For example, Francone (?, ?) developed a perspective-corrected interface for smartphones and tablets, but it only uses motion parallax for depth cues. Unuma and Komuro (?, ?) propose a technique for natural 3D interaction using a see-through mobile AR system. Lastly, Cuaresma and MacKenzie (?, ?) conducted a study that compares between the tilt-input and facial tracking as input for mobile games on a Google Nexus 7 HD tablet.

In this paper, we describe the development a tablet FTVR prototype that incorporates both motion parallax and stereo cues. We take some inspiration from the work of Li et al. (?, ?) and Rekimoto (?, ?), who advocate the use of easy-to-find hardware to enable head tracking and stereoscopic display for desktop computers. As a simple way of adding stereoscopic cues to existing

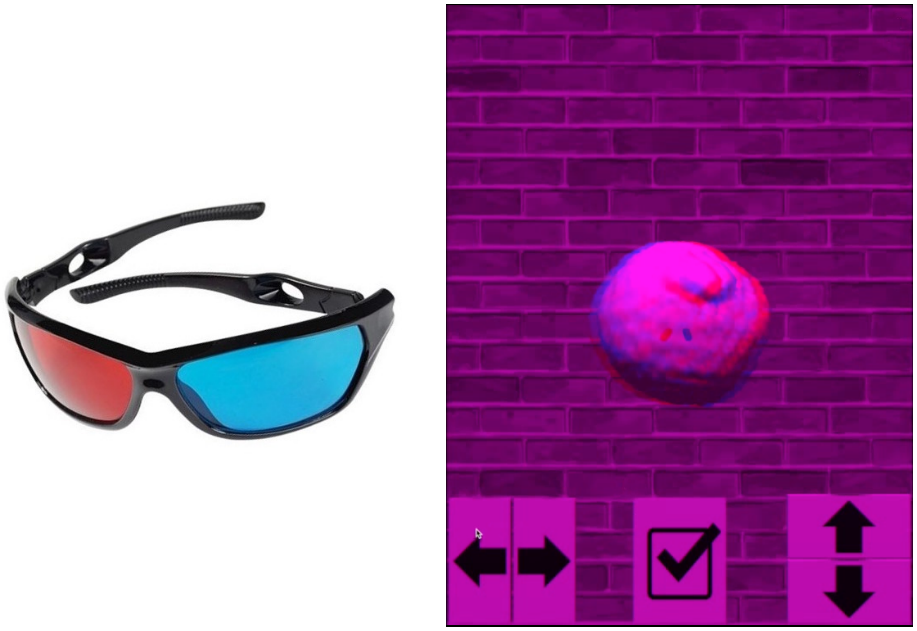

Figure 1: Anaglyph glasses, tablet FTVR screenshot.

tablet technology, we use Anaglyph 3D glasses. We conducted a usability study based on our prototype. We measured visual fatigue and subjective level of presence with standard questionnaires (?, ?, ?) and followed up with our own qualitative questionnaire on user experience. Our goals are to identify challenges and opportunities for the adoption of tablet-based FTVR and to determine directions for future research.

2 Implementation of Tablet FTVR

To achieve tablet FTVR without any enhancement to the hardware itself, we combine Anaglyph 3D for stereopsis with head position tracking from the tablet's front camera. For stereo, we use Anaglyph 3D images multiplex two color-filtered images (red and cyan). The user views Anaglyph 3D images as shown in Figure 1. For motion parallax, following previous studies (?, ?, ?), we use face tracking. The face tracking system tracks the user's face in real time using images from the tablet's front camera. We use the well-known Haar face detection cascade technique to find the face and track the detected face region using the Camshift tracking algorithm. See Algorithm 1 for more detail. We used the Unity game engine to develop the application, and ran it on an iPad Air (model number A1474). Based on the face tracking system and Anaglyph 3D, the application operates in four view modes. 1) *Normal 2D* view mode (2D): the application displays a static scene in 2D. 2) *Head-coupled display* view mode (HCD): the application shows perspective-corrected image according to the user's head position in 2D. 3) *Anaglyph 3D* view mode (Anaglyph): the application displays a static anaglyph 3D scene. 4) *Combined* view mode (Combined): the application shows an Anaglyph 3D scene according to the user's head position. We found that Algorithm 1 was efficient enough to not introduce any lagging issues. The system runs at 60 fps in all view modes.

3 Experiment

We conducted an experiment on the usability of tablet FTVR using the visual search task from the comparative study between CAVE and FTVR (?, ?). We recruited 40 participants (30 male and 10 female). All participants were

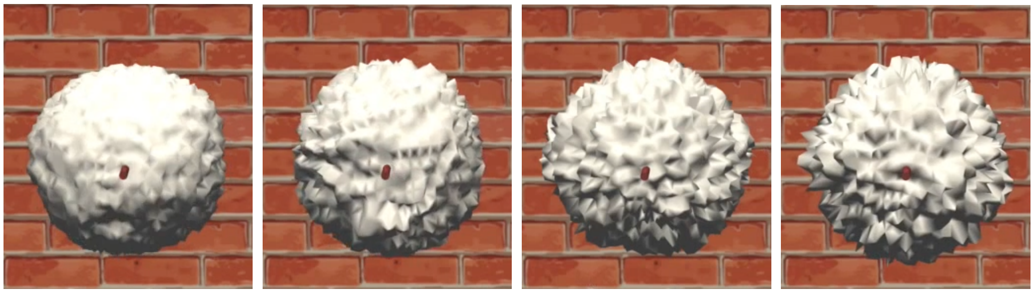

Figure 2: Four levels of noise used in the experiment.

university students and frequent users of computers in their daily life. Their age ranged from 17 to 31 years old. The participants had normal eyesight or wore eyeglasses to correct their eyesight to normal. None had previous experience using FTVR displays before they participated in the experiment. We used a 2 × 2 experimental design in which each participant was assigned to the *Normal 2D* group, the *Head-coupled* group, the *Anaglyph 3D* group, or the *Combined* group. To perform the task, participants had to identify the location of a rectangular bump on the surface of a noisy potato-shaped object then move it under a pole by rotating the potato using the arrow keys at the bottom of the display, as shown in Figure 1. Once the participant believes that the rectangular bump is correctly placed, he or she must tap on a checkmark button on the display. Then the application informs the participant whether the bump is correctly placed or not. The user must keep performing the task until he/she places the bump correctly. In the meantime, the application records the participant's performance time and number of false identifications. To avoid a ceiling effect, we made the task harder by applying four levels of noise to the potato-shaped object, as shown in Figure 2.

At the beginning of the experiment, a researcher briefly described how the system works to the participant. The researcher then calibrated the system for the participant by measuring his/her pupillary distance and entering it into the system. Next, the participant was allowed to practice the task at the easiest difficulty level until satisfied. Once ready, the participant performed the visual search task in a controlled sequence. There were 20 trials for each participant (1 view mode × 4 difficulty levels × 5 repetitions giving 20 trials). The participant had to complete all trials in a random order. When the participant completed the task, the researcher immediately asked the participant to evaluate his or her level of visual fatigue with the Simulation Sickness Questionnaire (SSQ) (?, ?), followed by the Presence Questionnaire (PQ) (?, ?) to evaluate the level of presence he or she experienced. After the participant completed both questionnaires, the researcher asked the participant to use the system in the normal 2D and in the combined view modes regardless of which group he or she was originally in. As the participant freely used the system, the researcher asked the participant to compare the two view modes and give his or her preference for each view mode along the seven dimensions (Overall experience, Visual comfort, Shape perception, Depth perception, Natural interaction, Feeling that the object is there, and Preference), because we are interested in investigating the contrast between the two view modes. At the end of the session, the researcher interviewed

Table 1: Results summary. Mean and standard deviation of task performance time, error rates, SSQ scores, and PQ scores.

	Average task performance time (second)	*Average number of error*	*SSQ score*	*PQ score*
2D	$11.57\ ^{*}$ (9.09)	0.14^{*} (0.46)	29.50^{*} (37.75)	76.11 (16.34)
HCD	$16.43\ ^{*}$ (15.84)	$0.66^{*\star\Diamond}$ (1.79)	$31.42^{\star}$ (42.90)	70.25 (15.34)
Anaglyph	13.70 (30.85)	$0.16^{\star}$ (0.66)	48.62 (28.21)	83.70 (16.26)
Combined	12.64 (8.64)	$0.16^{\Diamond}$ (0.62)	$74.43^{*\star}$ (40.83)	74.80 (11.67)

$*, \star$ and $\Diamond$ indicate statistically significant differences between two means in the same table column.

the participant about his or her experience with and opinion of the system. We hypothesized that the combined view mode will convey more depth information and provide a better level of presence more than the normal 2D view mode, because the combined view mode uses both stereopsis and motion parallax cues. However, we also hypothesized that the combined view mode will cause more visual fatigue than the normal 2D view mode, due to color distortion caused by the Anaglyph images.

4 Results

Here we present the results of the experiment. We begin with results on task performance (time and misidentification), then move to results on visual fatigue, level of presence, the comparison between *Normal 2D* and *Combined* view modes, and participants' feedback. The objective data are summarised in Table 1. We dropped the data for one participant from all analyses because the time the individual took to complete the task was many standard deviations beyond the mean.

For the task performance time, a two-way ANOVA revealed view condition to be a main effect ($F(3, 764) = 2.628$; $p = 0.049$). A Tukey analysis for the view modes showed that participants complete the task significantly slower in the *Head-coupled display* view mode than in the *Normal 2D* view mode. As expected, the ANOVA also revealed difficulty level to be a main effect ($F(3, 764) = 13.761$; $p < 0.001$). For the misidentification rate during the task, a two-way ANOVA revealed view condition to be a main effect ($F(3, 764) = 12.391$; $p < 0.001$). A Tukey analysis for the view modes showed that participants incorrectly identified the rectangular bump more in the *Head-coupled display* view mode than in all other view modes. Again, as expected, the ANOVA also revealed difficulty level to be a main effect ($F(3, 764) = 8.597$; $p < 0.001$). For the SSQ score, a one-way ANOVA revealed view condition to be a main effect ($F(3, 35) = 2.96$; $p = 0.045$). A Tukey analysis for the view modes showed that participants from the *Combined* group experienced more visual fatigue than those from the *Normal 2D* group and the *Head-coupled display* group. For the PQ score, besides the previously mentioned outlier participant, we also dropped the data for two additional participants because they failed to complete the questionnaire.

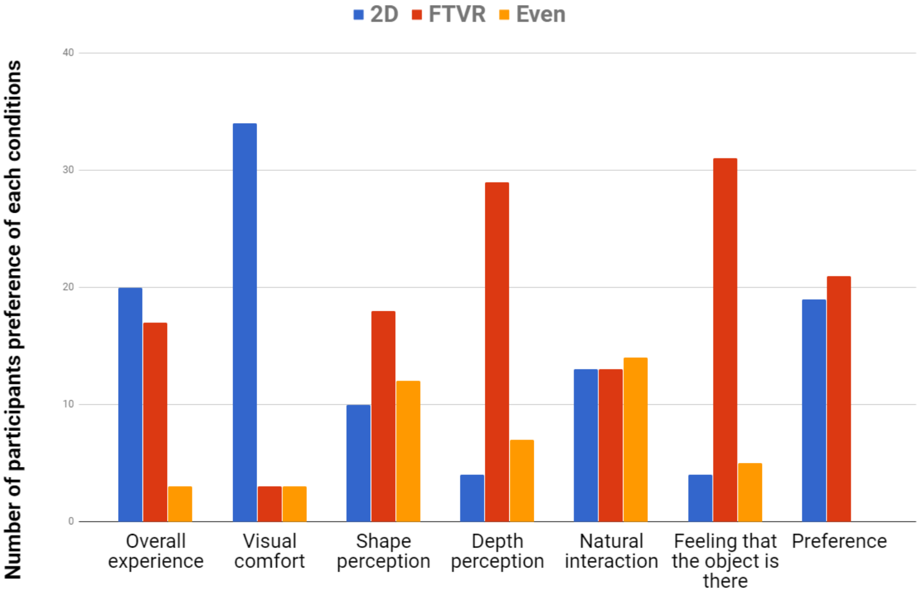

Figure 3: Users' preference between the *Normal 2D* and the *Combined* view modes along the seven dimensions.

The ANOVA did not reveal view condition to be a main effect.

The results of the comparison between the *Normal 2D* view mode and the combined view mode are summarized in Figure 3. A majority of participants rated the *Normal 2D* view mode more highly than the combined view mode in terms of *visual comfort*. In the interview session, many participants told us that they felt uneasy using the red-cyan glasses. A majority of participants also rated the *Combined* view mode over the *Normal 2D* view mode in terms of *depth perception* and *feeling that the object is there*. The participants told us that they found the *Combined* view mode to be more realistic, interactive, challenging, and fun. Nonetheless, at the end of the questionnaire, the participants rated their *preference* between the two view modes evenly. In the interview session, we found that the participants determined their preference by relative weighting of visual comfort from the *Normal 2D* view mode and the more realistic and fun features of the *Combined* view mode.

5 Discussion and Conclusion

The results of the usability study and our experience in developing the prototype suggest that in order to make tablet FTVR suitable for everyday usage, we must overcome two significant challenges. First, the front camera of a tablet has a relatively narrow field of view. Users tend to use tablets differently from desktop systems. They may use a device while lying down or sitting up, with the device very close to their face, and they may move or rotate the device often. Tablet FTVR thus requires a wide field of view for the tracking area, but front-facing cameras of commodity tablets are not built for that purpose and provide very limited fields of view. There are two potential solutions to this problem. The first solution is to add a wide-angle lens to the front-facing camera, as suggested by Lpez et al (?, ?). This should increase the tracking area, but a new tracking algorithm may be needed to process distorted images. While this approach preserves some of the benefits of using an existing device, it is to be tested whether the new field of view is large enough and the new tracking technique is robust enough to accommodate tablet users' behavior. The second solution is to develop a new tracking technique from scratch. We propose an infrared-based tracking system composed of an infrared light array and infrared cameras. This system can be built into a custom tablet case and communicate with a tablet via a USB port, so that it can be easily equip to existing devices. Infrared-based tracking systems are well developed and should be accurate and robust enough for tablet FTVR usage. We recommend tracking the user's head position in the landscape orientation setting, because the portrait setting provides a very limited horizontal viewing angle, which may not be suitable for tablet FTVR usage.

The second challenge is that Anaglyph 3D causes visual discomfort and is not suitable for everyday usage. This observation coincides with previous research on desktop FTVR systems (?, ?). Besides Anaglyph 3D, there are other two technologies that can be used to enable stereoscopy in a tablet: polarized 3D and active shutter 3D systems. Unfortunately, neither of these systems can be easily added to existing devices: ordinary LCD displays contain polarizers, which would disrupt the polarized 3D technique, and active shutter 3D systems require the display to be manufactured specifically for the technique. Stereoscopy is thus the biggest barrier to developing tablet FTVR based on existing devices.

Finally, we answer the underlying questions motivating this research. The first question is *How effective is tablet FTVR?* To answer the this question, we look to the results of the experiment. The comparison results suggest that participants perceived depth and felt that a virtual object existed in front of them more in the *Combined* view mode, when compared to the *Normal*

2D view mode. This coincides with the hypothesis based on our experience developing the prototype that tablet FTVR should convey depth information and presence more than a normal display. Although there were no statistically significant differences between the view mode for PQ scores, we suspect that this was more because of the visual discomfort from Anaglyph 3D and the front-facing camera-based tracking technique's limitations than anything else. Our findings coincide with those of Li et al. (?, ?), who used consumer-level hardware to simulate 3D displays on desktop systems. They found that the *Combined* view mode achieves the best depth perception. On the other hand, our experiment did not find the *Combined* view mode to be statistically different from the *Anaglyph* view mode in any of our measures. As previously mentioned, we suspect that participants were unable to perform the task better in the *Combined* view mode because of the front-facing camera-based tracking technique's limitations. This coincides with a study by Kongsilp and Dailey (?, ?), who found that in desktop FTVR settings, the combination of motion parallax and stereopsis cues produces lower visual discomfort and higher subjective level of presence when compared to the stereopsis cue only. Overall, we believe that there are opportunities for tablet FTVR development. If done correctly, we believe that tablet FTVR can convey depth information and immerse users in a scene, displaying virtual objects as if they really existed in front of them. This capability would enable a new range of applications, interactions, and user experiences.

The last question is *If it is useful, should we develop a new system or* *enhance existing devices?* To answer this question, we look at the challenges mentioned previously. When dealing with head-tracking capability only, we believe that it is better to enhance existing devices with the infrared-based approach. This should be robust enough while best preserving the benefits of using the existing device. However, we believe that it would be best to develop a new system from scratch if we absolutely require stereoscopic displays. Both polarized 3D and active shutter 3D technologies would require a fair amount of hardware changes to today's commodity tablets.

Algorithm 1 Face tracking algorithm

```
procedure FaceTracking
    if a face was not detected in the previous frame then
        use *Haar Feature-based Cascade Classifier* to search for a face in the frame.
        if face found then
            $camShiftCounter = 0$.
            return new face's size and position
        end if
    else
        $searchPosition \leftarrow lastKnownFacePosition$.
        $searchArea \leftarrow lastKnownFaceSize \times 2$.
        use *Haar Feature-based Cascade Classifier* to search for a face in the *searchArea* at the *searchPosition* of the frame.
        if face found then
            $camShiftCounter = 0$.
            return new face's size and position
        else
            use *Camshift* to approximate the face region in the *searchArea* at the *searchPosition* of the frame.
            $camShiftCounter = camShiftCounter + 1$.
            if $camShiftCounter <= 5$ then
                return new face's size and position
            else
                $camShiftCounter = 0$.
                return no face detected
            end if
        end if
    end if
end procedure
```