Dynamic Eye Convergence for Head-mounted Displays

Andrei Sherstyuk [‡]
Avatar Reality Inc.

Andrei State [§]
InnerOptic Technology Inc.
and University of North Carolina at Chapel Hill

Figure 1: *Close encounters in VR are a challenge for parallel eyes.*

Abstract

A virtual hand metaphor remains by far the most popular technique for direct object manipulation in immersive Virtual Reality (VR). The utility of the virtual hand depends on a user's ability to see it correctly in stereoscopic 3D, especially in tasks that require continuous, precise hand-eye coordination.

We present a mechanism that dynamically converges left and right cameras on target objects in VR, mimicking the effect that naturally happens in real life. As a result, the system maintains optimal conditions for stereoscopic viewing at varying depths, in real-time. We describe the algorithm, implementation details and preliminary results from pilot tests.

CR Categories: I.3.7 [Computer Graphics]: Three-dimensional Graphics and Realism — Virtual Reality I.3.6 [Computer Graphics]: Methodology and Techniques — Interaction techniques

Keywords: hand-eye coordination, stereoscopic vision, virtual hand.

1 Introduction

Studies in experimental neurophysiology show that human eyes always automatically focus and converge on task-relevant objects: a hand, hand-held tools or locations of tool application [Biguer et al. 1982]. Examples are: a hammer hitting a nail, a hand pointing at or touching objects. Fixating the object of interest brings that object into the center of the field of view. Projected into the eye, the object's image falls onto a special area of the retina, called fovea, where most photoreceptors are located and spatial resolution is highest. As the eyes converge on objects of interest, they line up the target object, the entrance pupil and the fovea, for each eye. As a result, both left and right eyes provide images of the highest possible quality, making it easier for the brain to fuse them into one.

As in real life, stereoscopic vision in VR provides important clues for object position and orientation, needed for most tasks that require direct object access. In immersive VR systems, stereo viewing is often implemented with head-mounted displays (HMDs), whose left and right channels represent virtual cameras co-located with the viewers eyes. It is typical, however, for both cameras to be controlled by a single motion sensor mounted on the user's head. The convergence distance usually remains fixed at some predefined value, frequently at infinity but sometimes only a few feet away from the head (whereas the focus distance is fixed at a few feet in contemporary HMDs). As a result, stereo image pairs rendered for the fixed (relative to each other) virtual cameras will significantly differ from images that real eyes would produce if the object of fixation were located closer or farther than the predefined convergence distance. Figure 1 illustrates the problem: fixed parallel cameras make the closely located butterfly appear in opposite corners of the image pair, making it very hard to fuse into a single stereo-view. Distance to hand 70 cm, distance to butterfly 15 cm.

We propose to improve the visual response of HMD-based VR systems by adding dynamic convergence to the virtual cameras when required by the current task. We build on the expectation that human eyes will mostly maintain convergence on the hand-object contact point while performing direct object access and manipulations. Automatic convergence of VR cameras will ensure that rendered images will be closer to the real-life imagery they simulate, which in turn should improve users' ability to perceive both depth and orientation of objects in the area of current activity.

2 Previous work

In the work that introduced virtual convergence for see-through HMDs [State et al. 2001], the authors explored the manipulation of eye convergence through rotation (yaw) of the virtual cameras in the augmented-reality HMD of their guidance system for medical procedures. The location of current user activity was predicted through heuristics that seemed to work quite well within the scope of the application; no formal user study was conducted.

More recently, [Sherstyuk et al. 2008] suggested to use the dominant hand to slide the viewport along the viewing plane in im-

[‡]e-mail: andrei@avatar-reality.com
[§]e-mail:andrei@cs.unc.edu

VRST 2010, Hong Kong, November 22 – 24, 2010.
© 2010 ACM 978-1-4503-0441-2/10/0010 $10.00

mersive VR settings, for conventional, non-see-through (opaque) HMDs. In a user study, participants with a hand-driven camera showed significant improvement in their control over a virtual hand and virtual tools, in the context of a medical VR simulator.

In this work, we combine the ideas of [Sherstyuk et al. 2008] using the location of the virtual hand to predict the current point of user attention, with automatic virtual convergence [State et al. 2001], aiming to produce more effective stereoscopic imagery for activities performed in immersive HMD environments.

It should be mentioned that there exists a separate body of research aiming to improve certain aspects of stereoscopic rendering through hardware solutions. One of the recent results in that field [Liu et al. 2008] employed active optical elements capable of presenting imagery at various focus depths. Even with a display that allows natural accommodation of very close objects, the stereoscopic convergence problem remains, since it is caused by the physical near-impossibility of constructing an HMD with generous nasal-side display pixel areas. In this paper, we offer a software solution for the latter.

3 Dynamic eye convergence implementation

In real life, each eye's viewing direction is the result of the combination of two rotations, the head's and each eye's. To a certain extent, the same happens in VR, especially if the field of view (FOV) of the display device is sufficiently large to allow wide eye movements. To determine a user's gaze direction and point of convergence, one would have to track eye gaze; doing that reliably requires special hardware such as near-eye cameras or electromagnetic sensors that are attached to a user's temples and capture the orientations of the eyeballs.

We suggest a simpler approach to estimating user gaze direction. Most mid-range HMD models suffer from what is called a "tunnel vision" effect, which is due to the limited FOVs of the displays. Tunnel vision is regarded as one of the most objectionable features of HMD-based VR systems. However, in our case, this deficiency turns into an advantage, because it allows us to approximate the gaze direction by orientation of the user head. For non-panoramic HMDs, the horizontal field of view ranges from 20 to 50 degrees. In contrast, the human visual range spans 180 degrees horizontally, for both eyes combined, with 60 degrees of stereo overlap. This allows enough room for active eye movements while the head direction remains fixed. When viewing the scene through a narrow HMD window, users are forced to rotate their head instead of and in addition to moving their eyes. Basing on that observation, we propose to approximate the user's gaze direction by orientation of his or her head.

Therefore, we elected to exclude free eye rotations from our implementation of simulated eye convergence. Instead, we will only modify the convergence angle between the eyes' virtual cameras, keeping the target object (in our case, the virtual hand) at the intersection point. In this approach, the target object will only move in the screen space horizontally, and will not snap into the centers of the view frames from an arbitrary position in screen space. This helps to avoid possible perceptional conflicts due to head rotations. Also, users will not feel that they lost control over their virtual hand. Finally, this approach allows us to keep the simulated convergence active most of the time (except as explained later), without heuristic rules for engaging and disengaging the convergence mechanism. Such rules might have been based on current hand position, velocity, direction of movements, head orientation, etc. These rules may require individual calibration for each user, which would add unnecessary complexity to using the VR system.

Algorithm. The algorithm for dynamic eye convergence is straightforward and consists of the following steps, executed in each cycle of the main graphics loop:

1. Test visibility of the target object and its distance z in a single cyclopic head camera. If the test fails, reset the rotation angles for the virtual eye cameras to default values (zero rotation for both eyes) and return.

2. Find the position of the target object in cyclopic camera space (x, y, z), and its azimuthal and elevation angles from the center of the screen: $a_x = \arctan(x/z)$, $a_y = \arctan(y/z)$.

3. Compute angular attenuation, $f = (1 + s^2d^2)^2$, where d^2 is the squared distance between the target and the screen center: $d^2 = a_x^2 + a_y^2$. Parameter s^2 defines the width of the bell-shaped function, shown in Figure 2. There, $s^2 = 0.28$, which makes the convergence negligible at distances larger than 10 degrees from the screen center in all directions.

4. Compute the convergence angle, $A = \arctan(D/2z)$, where D is the camera separation distance, which must be initialized to the viewer's inter-pupillary distance (IPD), and z is obtained in step 1. Rotate left and right cameras inwards by fA amount. The camera nodes are parented to the head node, facing in $-Z$-direction, separated by D.

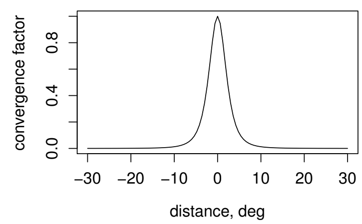

Figure 2: *Attenuation function. Convergence is strongest at the center of user view and rapidly falls off towards the screen edges.*

The results of the hand-driven convergence mechanism are demonstrated in Figure 3. Note that with convergence, the left and right images of the target object appear closer to the center of the current frame, so viewers can fuse these images with less effort. Conversely, images of distant objects (in this case, the umbrella), are moved further apart. In stereo, these objects will appear in double-vision, as expected.

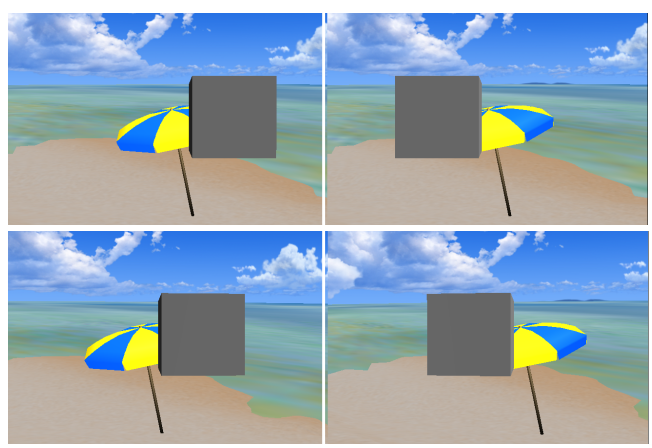

Figure 3: *Test cube as a target object, rendered with parallel eyes (top) and with eye convergence (bottom). Distance to target object 20 cm, convergence angle 9 degrees. The cube is attached to the virtual hand, which is not rendered in this example. IPD 7 cm, cube size 5 cm. (Fusible left-right image pairs)*

Special cases. Since the attenuation function has infinite support (Figure 2), dynamic convergence can remain active at all times while users are performing near-field tasks. However, there are application-specific situations and tasks when auto-convergence on the hand should be either switched off, or re-targeted to another object. Some examples are listed below:

- The user is performing a task that requires bimanual operations. To continue using dynamic convergence, a point halfway between the two hands must be computed and used as a new target location.
- The hand, or a or hand-held tool, is operated as a pointer, aiming at distant objects, for example, selecting a destination location for virtual travel or shooting a hand-gun. In these cases, stereoscopic rendering should be disabled by switching to a single cyclopic camera.
- The user is operating a tool which affects objects that are located far from the hand. Since the objects of interest are already known (unlike to the previous case), the system may converge on the average location of distant objects the hand is operating on.

4 Pilot tests

VR system. The dynamic convergence mechanism was implemented in the open source VR engine Flatland, developed at the Albuquerque High Performance Computing Center [1] and tested in non-immersive mode. Both user head and virtual hand transformations were operated by a mouse; the left and right viewports were displayed as a stereo left-right pair on a laptop screen. After some practice, one of the authors learned to run the system, using mouse and keyboard, while maintaining the stereo view by freely fusing the stereo pairs on the screen.

Virtual hand. In the Flatland-based system, users interact with the environment by pointing at and "touching" objects with the index finger on the virtual hand. A 1 cm hidden cubic shape, attached to the last joint of the index finger, was used for processing hand-object collisions and served as Flatland's target object for eye convergence. Figure 4 shows details of the virtual hand implementation.

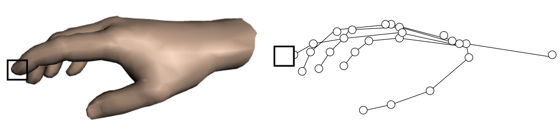

Figure 4: *Virtual hand: deformable visible mesh (left) and skeleton (right). The skeleton can be animated using prerecorded motion data, making the hand shape assume different poses: open-hand (shown), closed-hand, pointing-with-finger and others. The target object, a small invisible cube, is attached to the index finger.*

Task. To evaluate the effects of dynamic convergence, a nearly empty beach scene was used, with few static objects and a flock of butterflies flying around the user and making occasional stops for rest. The user's task was to reach and capture butterflies by touching them with the tip of the index finger. The system automatically converged the cameras on the index finger at all times while the hand was in view.

[1] http://www.hpc.unm.edu/homunculus

Findings. During multiple trials, we observed that

1. The dynamic convergence algorithm does not produce any disturbing or unpleasant sensations;
2. Dynamic convergence helps reduce diplopia (double vision) significantly for close objects.

The second result was confirmed by numerous tests, performed as follows: after approaching a resting butterfly at a very close distance (25 cm or closer), all animations were paused, so the butterfly remained seated as long as necessary. Then, the hand was extended to 1 m away from the camera, far beyond the location of the butterfly; the hand's images were fused into one. As the results, the butterfly would appear highly diplopic. Then, without losing focus, the hand was slowly retracted back towards the camera. Without convergence, the left and right images of the butterfly remained separated by the same distance. With convergence, the images were moving towards each other, until fused. Three stereo pairs from the pilot tests are shown in Figure 5.

5 Discussion and future work

We consider these results very promising, although more tests are needed in a full-fledged immersive environment. To obtain quantifiable results, we are planning a formal between-subjects user study, with a Virtual Research Systems V8 head mounted display and an Ascension Flock of Birds magnetic system for head and hand tracking. The task will be the same as above, or similar: to catch as many butterflies as possible in a given time period. We hope to obtain statistically significant results for the subject groups, counting successful and failed captures for each participant. Also, personal reports collected from test subjects will help determine whether the proposed technique feels comfortable and natural.

In our version of Flatland, all logged events are time-stamped at the frequency of the graphics loop. This makes it possible to observe trends in user skills as the users progress through their mission. Time-based analysis of log files will help determine whether dynamic eye convergence has any effect on the learning curve. This information, in turn, can help us design and implement better VR simulators for motion-intensive near-field tasks such as suturing, biopsy, and other fine-motor skill training applications.

References

BIGUER, B., JEANNEROD, M., AND PRABLANC, P. 1982. The coordination of eye, head and arm movements during reaching at a single visual target. *Experimental Brain Research 46*, 301–304.

LIU, S., CHENG, D., AND HUA, H. 2008. An optical see-through head mounted display with addressable focal planes. In *ISMAR '08: Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, IEEE Computer Society, Washington, DC, USA, 33–42.

SHERSTYUK, A., VINCENT, D., AND JAY, C. 2008. Sliding viewport for interactive virtual environments. In *Proceedings of ICAT 2008: 18th International Conference on Artificial Reality and Telexistence*.

STATE, A., ACKERMAN, J., HIROTA, G., LEE, J., AND FUCHS, H. 2001. Dynamic virtual convergence for video see-through head-mounted displays: Maintaining maximum stereo overlap throughout a close-range work space. In *Proceedings of the International Symposium on Augmented Reality (ISAR) 2001*, 137–146.

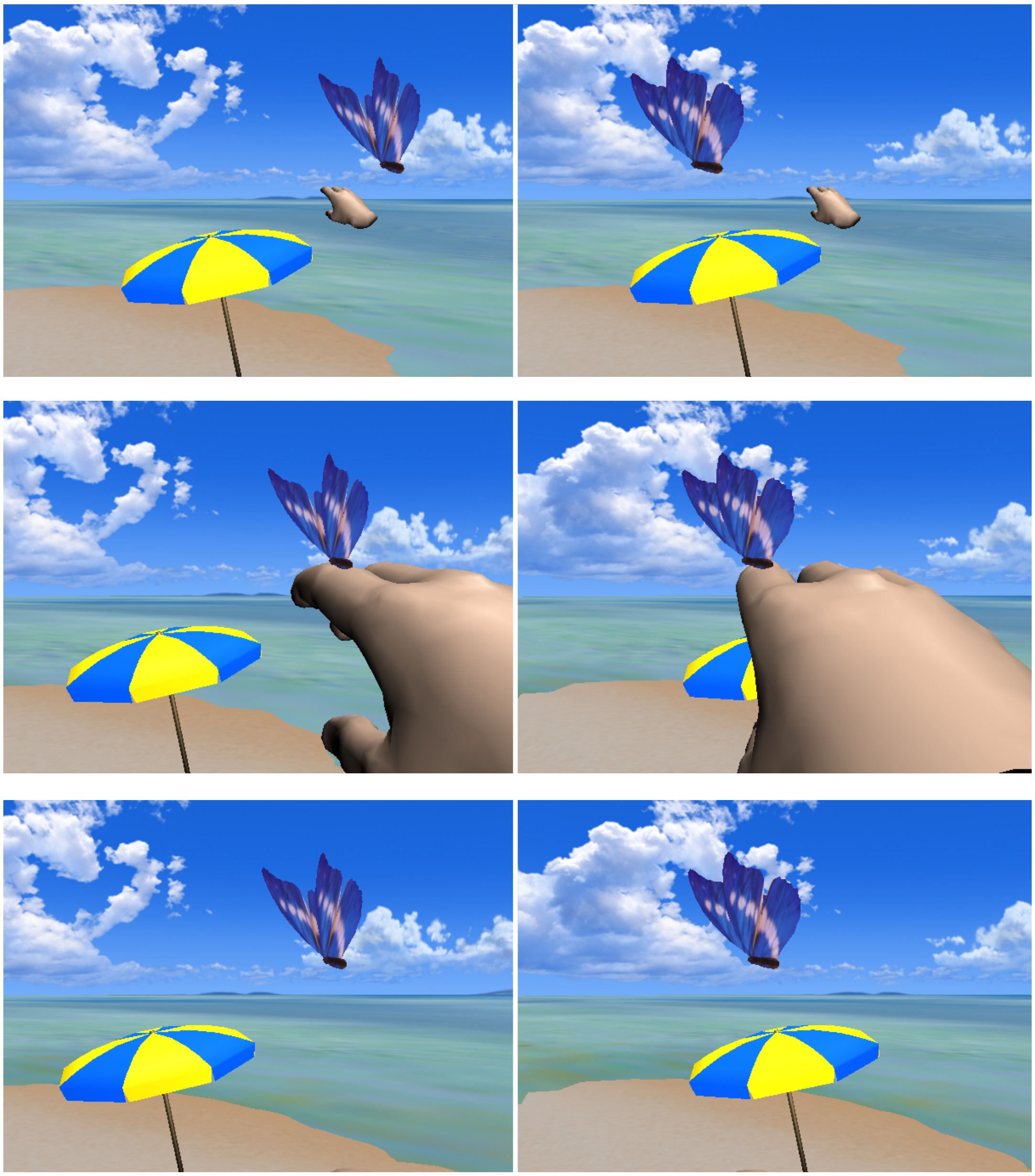

Figure 5: *Pilot tests: catching butterflies on a tropical beach, distance to butterfly 23 cm. Top image: the hand is placed at 100 cm from the camera, resulting in nearly parallel views. It is very difficult, although possible, to fuse the images of the butterfly onto one. Middle and bottom images: the hand is retracted back to the camera and is co-located with the target. The eyes converge by 8.5 degrees, making it relatively easy to fuse the butterfly stereoscopically. For better view, the hand is not rendered in the bottom image. (IPD 7 cm)*