Predator-Prey Vision Metaphor for Multi-Tasking Virtual Environments

Andrei Sherstyuk* Anton Treskunov† Marina Gavrilova ‡
University of Hawaii Samsung University of Calgary

ABSTRACT

Narrow field of view of common Head Mount Displays, coupled with lack of adaptive camera accommodation and vergence make it impossible to view virtual scenes using familiar eye-head-body coordination patterns and reflexes. This impediment of natural habits is most noticeable in applications where users are facing multiple tasks, which require frequent switching between viewing modes, from wide range visual search to object examination at close distances. We propose a new technique for proactive control of the virtual camera by utilizing a predator-prey vision metaphor. We describe the technique, the implementation, and preliminary results.

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

1 INTRODUCTION

Human eyes is a remarkably flexible instrument capable of performing a wide range of tasks, from watching "bigger-than-life" 3D IMAX movies to reading fine print. The human vision is tightly coupled with the propriocentric system; acting together, they dynamically integrate the visual input with body, head and eye movements into one continuous image of reality.

In Virtual Reality (VR) systems that employ Head Mounted Displays (HMD), the range of eye movements is restricted by the field of view of the display device. With the exception of ultra-panoramic models, (e.g., piSight HMD series from Sensics) horizontal field of view of most HMDs ranges from 30° to 50°, compared to 180° of natural eye span. When HMD users attempt wide eye movements their gaze stops short at the black frame of the display, resulting in diminished sense of presence. Consequently, during tasks that require wide-range viewing, HMD users tend to compensate the limited utility of eye movements by increased amount of head rotation. This effect received much attention in research community [1, 4, 16]. On the other hand, during near range interactions with virtual objects, users were observed to position themselves so that objects of interest appear at the center of the display. Such arrangement provides optimal conditions for stereoscopic viewing and higher accuracy of eye-hand coordination during handling of objects [13]. However, this practice often fails at close distances when left and right images of the object differ so much that stereo fusion becomes impossible. When that happened, HMD wearers were reported to resort to monoscopic viewing, using their dominant eye only and ignoring the input from the other eye [15].

Evidently, different visual tasks and conditions in virtual environments call for different types of optimization of scene rendering. For search tasks, stereoscopic viewing is of a lesser importance compared to the size of the visible area. For direct manipulation of objects at close range such as in grabbing tasks, the object of interest is already selected and fixated upon. In this case, correct depth perception becomes a priority, which requires proper convergence of cameras on the object, at the expense of reduced field of view and increased diplopic appearance of objects that are located closer or farther than the current convergence distance.

Such polarization of viewing conditions closely resembles the two types of visual systems, known as predator and prey vision. We suggest using this metaphor in VR, by adaptively changing virtual camera settings according to the current visual task.

2 PREDATOR/PREY VISION IN REAL AND VIRTUAL WORLDS

Due to its anatomical configuration, the human visual system is better suited to operate in predator rather than prey mode. Our eyes are oriented in forward direction, with a large stereoscopic overlap of 60° and approximately 180° horizontal field of view (FOV), combined for both eyes. Using fast eye movements, we can locate and fixate objects in the visible area very quickly. Looking outside the visible field requires head and torso rotations that are relatively slow and less frequent. As a result, wide area viewing requires more efforts. However, during the course of normal daily activities, we routinely use both viewing modes, changing them as needed.

One example of such automatic shifting of visual gears comes from everyday driving experience. While moving on a busy road, drivers are mainly concerned with potentially hazardous objects located directly ahead, such as surface obstacles, other cars and similar. For safety, it is important to be able to estimate the distance to these objects accurately and react promptly. Thus, a good stereo overlap is needed for better depth cues, which prompts drivers to keep their eyes on the road. When coming to a stop at an intersection, the driver's attention is redirected towards objects in peripheral areas, such as pedestrians approaching the intersection. The prey vision mode is engaged that covers a wider area, at the expense of reduced stereo perception. In this mode, detection of presence of dangerous objects takes priority over the quality of their image.

In virtual environments, the use of non-panoramic HMDs defacto enforces the predator viewing mode, because the peripheral vision is no longer available. This brings up the question: how can one support wide-area viewing while using HMDs with hardware-limited FOV? More generally, is it possible for the rendering system to recognize the current task and adjust itself accordingly to provide optimal viewing conditions? In the next section, we review the previous work on optimizing rendering for HMD-based VR systems.

3 PREVIOUS WORK ON RENDERING FOR HMD

The loss of peripheral vision experienced by users was widely recognized as the most objectionable drawback of HMD-based immersion. Multiple studies demonstrated that restricted FOV has negative impact on perception in VR, including size and location of objects, motions and user spatial orientation [1, 10, 12]. Therefore, it is generally accepted that narrow FOV values of display devices limit users ability to interact with the environment. Several solutions were proposed to compensate for this deficiency by means of using non-linear mappings between user head motion and virtual camera controls, aiming to increase the perceived size of the visible field. One such approach employed amplified head rotation [5], where the virtual cameras co-located with the eyes eyes of the immersed user were turning faster than the user head, in horizontal direction. The authors conducted an experimental study and observed a 21% increase in performance for "amplified" users. Sim-

---

*e-mail:andreis@hawaii.edu
†e-mail: anton.t@sisa.samsung.com
‡e-mail:mgavrilo@ucalgary.ca

IEEE Symposium on 3D User Interfaces 2012
4-5 March, Orange County, CA, USA
978-1-4673-1205-9/12/$31.00 ©2012 IEEE

ilarly, amplification of pitch and roll head rotations was suggested and tested experimentally, resulting in estimated 25% reduction of user efforts during visual search [2]. Effects of extreme amplification of horizontal head rotation on scanning and counting tasks were examined, with positive results for scanning and negative for counting task [6]. Variations on amplified camera rotation were also applied in other types of VR architectures, with the CAVE [9] and surround-screen [8] visualization devices.

Unfortunately, despite the promising results reported in multiple experimental studies, amplification of camera rotation has two serious drawbacks. Firstly, amplifying an arbitrary 3-D rotation does not satisfy two important user-interface guidelines, known as nulling and directional compliance [11]. The former demands that returning an object into its initial position must also cancel the amplified response. The latter tells that both amplified and source rotations must happen around the same axis. Violation of these requirements breaks the generally accepted principles of user interface design [3], which limits amplification of head rotation to 1-D cases only, where both compliances still hold. Note, that in all examples described above the head rotation was amplified in one dimension: yaw [5, 6], pitch and roll, separately [2]. Another serious limitation comes from users' inability to operate their own virtual hands while head rotation amplification is in effect. The head and the hands move independently from each other, and their free-style rotations can not be synchronized using the same 1-D amplification scheme. As the result, the hands will be out of sight most of the time, causing inevitable user confusion.

In order to reconcile the eye and hand coordination for narrow FOV HMDs, a view sliding technique was developed and evaluated experimentally [13]. During view sliding, both left and right camera viewports are dynamically shifted on the screen space, following the current point of fixation, keeping it in continuous view. As the results, the efforts required for maintaining visual contact with the current object is reduced and more attentional resources become available for object manipulation tasks. The experiments demonstrated more than 50% increase in performance during grabbing and placements tasks for objects in close range.

Another optimization for near-range viewing improves the quality of stereoscopic rendering by using dynamic camera convergence [14, 15]. This technique converges both cameras at the distance where the current object of interest is located. As the result, the left and right images may be fused into a single stereo view with minimal effort required from the user wearing an HMD. With the inter-pupillary distance properly adjusted for the current viewer, stereo fusion can be achieved with parallel eyes now, because the rendered images are already pre-converged in software.

The view sliding and dynamic camera converging techniques [14, 13] facilitate close range viewing and interactions with virtual objects. By design, they do not lend themselves for an easy extension to wide-area visual tasks, which makes them complementary to optimizations that employ amplified head rotation [2, 5, 6].

As the studies demonstrated, the amplified head rotation methods [2, 5, 6] and close-range rendering optimizations [14, 13, 15] proved to be effective in their own application domains. The former methods support the prey mode of vision, when users are mostly concerned with wide area viewing. The latter realize the predator mode, helping users fixate on and keep continuous eye contact with a single object.

Building on best practices from the previous research, we present a new system for dynamic camera control for HMD-based virtual environments. The proposed system covers the entire range of visual tasks, from predator to prey mode, by tracking and interpreting user activities in VR and altering camera settings accordingly.

4 IMPLEMENTATION OF PREDATOR-PREY VISION

The proposed system consists of two components: a task-dependent camera control mechanism and a user interface module that maps user actions and movements onto appropriate viewing task.

4.1 Dynamic camera control mechanism

We implemented the predator-prey vision metaphor by simulating vergence of virtual cameras, co-located with the viewers' eyes. For close-range viewing, we used the automatic camera convergence technique [14, 15]. For wide-range viewing, we extended the same technique by adding simulated divergence, the outwards movement of eyes. While the converging mechanism of dynamic camera control is well understood and has a direct analogue in real life, the diverging part has not been explored yet in VR interface design.

Due to anatomical constraints, human eyes can diverge until their lines of sight become parallel, which happens when looking at distant objects. Unlike human eyes, virtual cameras are not bound by this restriction and can be rotated outwards beyond the natural limits. Moreover, vergence of human eyes for healthy individuals is always symmetric, while the virtual cameras can, in principle, rotate independently from each other. In order to minimize disorientation of users, we propose to keep the symmetric property of eye rotation and alter the divergence angle only, when rotating the cameras farther outwards. The maximum angle of such forced divergence that immersed users will be able to accept is still to be measured experimentally. This angle will certainly be device specific and probably user-specific, too. The upper bound for simulated divergence angle D may be estimated as $D = HS/2$, where H is the horizontal field of view and S is the stereo-overlap fraction of the HMD. For example, for V8 HMD by Virtual Research with 48° horizontal FOV and 100% stereo overlap, left and right cameras can be rotated up to 24° to each side, without creating a blind spot in frontal visible area. However, after experimenting with stereo pairs produced at various levels of simulated divergence, we found that for V8-like displays the acceptable values of D lie somewhere between 10° and 15°. These values should be further verified and corrected, if necessary, in a proper user study. The proposed system can be conveniently described as a task-based state machine, illustrated in Figure 1. In each state, the cameras are controlled as described below.

(A) Conventional viewing mode. This is the default state, initialized at system start-up. The left and right cameras are set to default orientation, for example, parallel to each other. In this state, the system provides general-purpose viewing, without any task-related optimizations.

(B) Near field-viewing mode. The cameras rotate inwards symmetrically converging on the current point of fixation, which is located at close range, typically, within arm's reach from the user.

(C) Wide-area viewing mode. The cameras temporarily diverge outwards horizontally, giving the user a chance to "peek outside the frame", as set by the field of view of their HMD.

(D) "Cyclopic camera". This is a special mode for monoscopic viewing. Both cameras assume the default orientation, and move to the centrally located "third eye" position. This mode is engaged when users are working with special-purpose objects, such as UI elements, i.e., on-screen menus. Also, monoscopic mode is used in aiming, pointing or steering tasks.

To ensure smooth change of viewing conditions, all transitions between states go via the default mode (A). The transitions are handled by the user interface module, described next.

4.2 User interface for mode switching

The interface module provides mapping between user actions and movements and the states of the virtual camera control system, described above. This module reads all available user input channels,

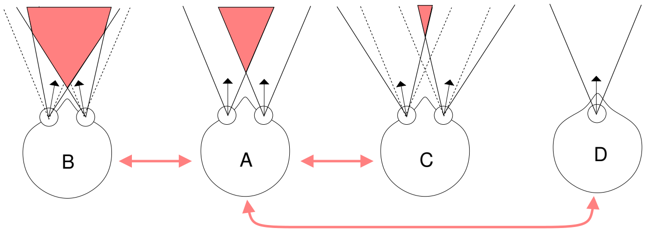

Figure 1: States and transitions of the predator-prey vision control system: (A) default mode with parallel cameras; (B) predator mode, cameras turning inwards; (C) prey mode with diverging cameras; (D) cyclopic camera for monoscopic viewing. Shaded areas show zones of stereo vision. Dotted lines in (B) and (C) show the original FOV.

such as data from the trackers, and triggers transitions between the states, according to rules, that must be set beforehand according to the configuration of the VR system and the application tasks.

Below, we describe the implementation of the user interface module, based on a conventional VR setup. We assume that the user is immersed with a stereo HMD and their head and hands are tracked. The presence of at least one virtual hand is essential because it allows close-range operations on virtual objects. Along with wide-range visual search tasks, that are common for most virtual environments, the use of virtual hand enables near-distance visual tasks, that require precise eye-hand coordination, such as object grabbing and placement. In these tasks, the hand location gives a sufficient approximation of the point of fixation, which allows to obtain the current camera convergence distance and angle [14, 13].

For our purposes, we require that the user neck rotation is also available. That can be achieved by placing an additional motion sensor on user torso and comparing its readings with the head rotation. We propose to use the neck rotation as a condition for engaging the wide-area viewing. It is known that in real life eye movements larger than a few degrees are typically accompanied by neck rotation, which contributes up to one third of the total gaze rotation [7]. Thus, large values of neck rotation can be used as a signal that wide horizontal eye movement is in progress. In certain cases, the neck rotation can be approximated by head rotation if users are restricted in their torso movements, for example, when seated in a non-rotating chair.

Using the information on neck rotation and hand position, switching between viewing modes can be implemented as follows:

Transition	Conditions
A → B	1. Virtual hand is fully visible in both cameras
	2. Neck rotation $\leq T$ threshold
A → C	3. Virtual hand is not visible
	4. Neck rotation $> T$ threshold
A → D	5. User operates his/her virtual hand as a pointer in steering or object selection tasks

For any transition from the default state (A), all listed conditions must be satisfied. If any condition becomes false, the system returns to the default state. The parameter T, that appears in conditions 2 and 4, must be determined experimentally. The value of 20° seems to be a reasonable initial guess. Also, an additional condition on rotation speed may be useful for filtering out slow head movements.

The last rule that turns on the cyclopic camera mode, may be implemented in various ways, depending on other UI metaphors already employed. For example, in our system the virtual hand is actively used for travel. During travel, the hand may be used for steering, providing continuous relocation. Also, the hand can be used to point and mark distant objects as destinations for teleportation. In both cases, the system treats the hand as a pointing device, when it is extended for more than 80% of the total arm's length and is currently visible. These two conditions prompt the transition to monoscopic rendering, in order to eliminate unwanted diplopia.

The rules for dynamic camera controls, as described, are designed for background execution, without direct user intervention. However, they can be augmented by adding explicit commands for setting required viewing modes. When extending and modifying these rules, a few general guidelines should be observed, such as use of damping for convergence and divergence angles, avoiding oscillation between the states and never allowing states to compete over camera controls. The rules presented above satisfy these guidelines, as they are based on user hand and head motions, which are naturally smooth and continuous.

5 PROOF OF CONCEPT PROTOTYPE

We prototyped the proposed system in *VR-Triage* trainer for teaching first responders basic triage skills [17]. The preliminary tests were conducted in desktop mode, using an LCD screen as a display device and mouse and keyboard for camera control. Sample images are shown in Figures 2 and 3, illustrating prey and predator vision modes, respectively. All images are left-right pairs.

The virtual environment used in these tests depicts a simplified case of a training scenario. In VR-Triage, students are required to locate, approach and medically examine virtual victims, scattered on the scene. The scenario requires students to perform a variety of tasks, including a wide area visual search and close-range interaction with the victims, including visual examination and treatment, using hand-held virtual tools. In advanced cases, hostilities may be also simulated by placing virtual enemies on the scene. In the simplified example shown in Figures 2 and 3, the sitting and lying characters are the victims. The standing character visible in Figure 2 plays a role of an enemy, that must be timely noticed and handled. In this situation the prey vision appears to be helpful.

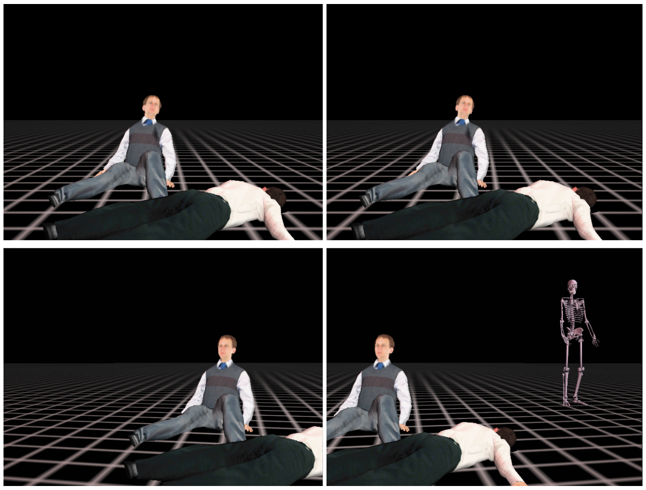

Figure 2: Diverging cameras simulate the prey vision mode. As the cameras rotate outwards (bottom pair) the visible area expands and the standing character moves into view. Divergence angle 12°.

For close-range viewing, the predator mode is engaged, converging cameras on the object of interest. In this case, it is the victim's head (Figure 3). Because the distance to the object is very small, parallel cameras produce images that are very hard to fuse in one stereo view (Figure 3, top). With converging cameras, the head is rendered centrally. Note that in this particular case, the victim has a severe neurological condition, indicated by asymmetric dilation of his pupils. Without camera convergence, the person in training could be tempted to look at the object with one dominant eye (as described in Introduction) and miss this important evidence.

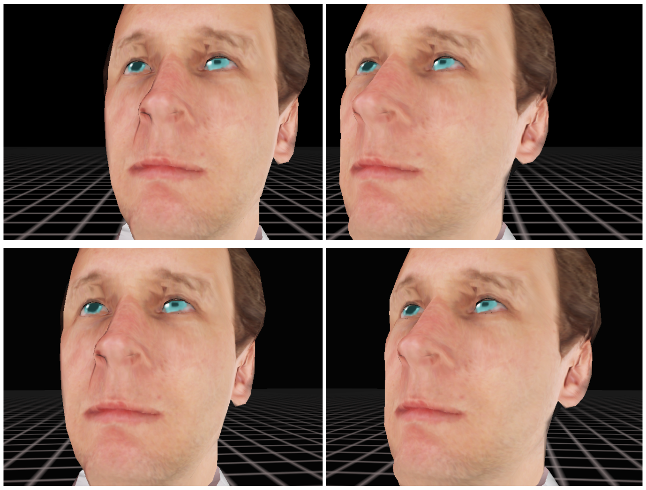

Figure 3: Converging cameras optimize viewing for extreme close-ups. Parallel cameras (top) render the object close to the inner edge of the display. Images produced with converging cameras (bottom) are located centrally and can be easily fused. Convergence 8°.

6 EXPERIMENTAL EVALUATION

In order to evaluate usability and effectiveness of the proposed technique, an immersive VR system must be used, as described in 4.2. Presently, we are considering extending the existing medical trainer *VR-Triage* [17], by adding virtual enemy characters that would approach the immersed trainee at random times and directions. In such extended scenario, all vision modes depicted in Figure 1 will be engaged. Another possibility is to implement a virtual version of Elimination game, played on Paint-ball fields. In this case, autonomous enemy agents will be moving across large virtual terrain with multiple occluders. The objective is to locate and shoot the agents, using wide area and cyclopic vision modes, respectively.

In both scenarios, the user tasks must be carefully designed and balanced, to emphasize the visual component of user performance in VR. For example, navigation and travel should be made as easy as possible, to allow users direct all their attention to visual tasks. All user actions and vision-related events must be time-stamped and recorded for post-session analysis and comparisons between the effect and control groups of subjects.

7 DISCUSSION AND FUTURE WORK

We have presented a novel system that implements predator-prey vision metaphor for immersed virtual environments, aiming to provide optimal viewing conditions for multiple tasks. The system is based on a novel approach to wide area viewing by artificially diverging left and right cameras, enlarging the size of users' peripheral vision. The rules for switching viewing modes were discussed for conventional VR platforms that employ stereo HMDs and body tracking. The proposed predator-prey vision metaphor is simple to implement and optimize for a wide range of VR applications and hardware components.

We implemented and tested the proposed system under non-immersive conditions, and produced usable stereo pairs for wide and near viewing tasks. We have yet to explore the effects of prolonged use of artificial camera divergence in immersive mode. However, the preliminary tests indicate that continuity of visual perception can be maintained between task switching, if the divergence angle is set to a moderate value, between 10 and 15 degrees. In addition, the camera divergence is only engaged temporarily during wide neck rotations, which naturally prevents immersed viewers from overusing this mode. Because the camera controls are directly coupled with user motions, it seems reasonable to expect a self-regulating effect will take place, and will help users to find the optimal combination of head and torso rotations naturally. We are planning to fully implement the proposed system and evaluate it in a user study under immersive conditions.

We believe that with proper calibration and optimization the newly introduced predator-prey vision metaphor for dynamic camera control will enhance user performance in VR applications with multiple viewing tasks.

REFERENCES

[1] P. L. Alfano and G. F. Michel. Restricting the field of view: perceptual and performance effects. *Perceptual and Motor Skills*, 70:35–45, 1990.

[2] B. Bolte, G. Bruder, F. Steinicke, K. H. Hinrichs, and M. Lappe. Augmentation techniques for efficient exploration in head-mounted display environments. In *ACM Symposium on Virtual Reality Software and Technology (VRST)*, pages 11–18. ACM Press, 2010.

[3] E. G. Britton, J. S. Lipscomb, and M. E. Pique. Making nested rotations convenient for the user. *SIGGRAPH Comput. Graph.*, 12:222–227, August 1978.

[4] H. Dolezal. *Living in a world transformed: Perceptual and performatory adaptation to visual distortion.* New York, NY: Academic Press, Inc., 1982.

[5] C. Jay and R. Hubbold. Amplifying head movements with head-mounted displays. *Presence: Teleoperators and Virtual Environments*, 12:268–276, June 2003.

[6] R. Kopper, C. Stinson, and D. Bowman. Towards an understanding of the effects of amplified head rotations. In *The 3rd IEEE VR Workshop on Perceptual Illusions in Virtual Environments*, 2011.

[7] M. F. Land and B. W. Tatler. *Looking and acting: vision and eye movements in natural behaviour*. Oxford University Press, 2009.

[8] J. LaViola and M. Katzourin. An exploration of non-isomorphic 3d rotation in surround screen virtual environments. *3D User Interfaces*, 0, 2007.

[9] J. J. LaViola, Jr., D. A. Feliz, D. F. Keefe, and R. C. Zeleznik. Hands-free multi-scale navigation in virtual environments. In *Proceedings of the 2001 symposium on Interactive 3D graphics*, I3D '01, pages 9–15, New York, NY, USA, 2001. ACM.

[10] D. C. Neale. Head-mounted displays: Product reviews and related design considerations. In *Considerations, Hypermedia Technical Report HCIL-98-02, HumanComputer Interaction Laboratory, Department of Industrial and Systems Engineering, Virginia Tech*, 1998.

[11] I. Poupyrev, S. Weghorst, and S. Fels. Non-isomorphic 3d rotational techniques. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '00, pages 540–547, New York, NY, USA, 2000. ACM.

[12] M. B. Russo, T. R. Letowski, E. Schmeisser, and C. E. Rash. *Helmet-mounted Displays: Sensation, Perception, and Cognition Issues*. U.S. Army Aeromedical Research Laboratory, 2009. ch. Guidelines for HMD Design.

[13] A. Sherstyuk, C. Jay, and A. Treskunov. Impact of hand-assisted viewing on user performance and learning patterns in virtual environments. *The Visual Computer*, 27:173–185, 2011.

[14] A. Sherstyuk and A. State. Dynamic eye convergence for head-mounted displays. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology*, VRST '10, pages 43–46, New York, NY, USA, 2010. ACM.

[15] A. State, J. Ackerman, G. Hirota, J. Lee, and H. Fuchs. 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*, pages 137–146, 2001.

[16] M. Venturino and M. J. Wells. Head movements as a function of field-of-view size on a helmet-mounted display. In *Human Factors and Ergonomics Society Annual Meeting*, pages 1572–1576, 1990.

[17] D. Vincent, A. Sherstyuk, L. Burgess, and K. Connolly. Teaching mass casualty triage skills using immersive three-dimensional Virtual Reality. *Academic Emergency Medicine.*, 15(11):1160–5, 2008.