2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)

Effective Free Field of View Scene Exploration in VR and AR

Lili Wang, Antong Cao, Zhichao Li, Xuefeng Yang/ Beihang University
State Key Laboratory of Virtual Reality Technology and Systems
Beijing, China
Email: wanglily@buaa.edu.cn

Voicu Popescu/ Purdue University
Computer Science
West Lafayette, US
Email: popescu@purdue.edu

***Abstract*—We propose to improve virtual reality (VR) and optical see-through augmented reality (AR) head-mounted display scene exploration efficiency by allowing the user to adapt the field of view interactively. This way the user can zoom in to examine parts of the scene in more detail without having to translate the viewpoint forward, as would be required in conventional fixed field of view scene exploration. The user can also zoom out, to gain a more comprehensive view of the scene and to examine distant parts of the scene in parallel, without the need to translate the viewpoint backward. Zooming in is supported with a focus+context visualization approach that integrates a distortion-free magnified focus region seamlessly into context. For AR, the higher resolution focus region is resampled from the video feed acquired by a head-mounted high-resolution camera. We demonstrate the benefits of our free field of view scene exploration in the context of VR and AR tasks, where it brings a substantial reduction of viewpoint translation, view direction rotation, and task completion time.**

***Keywords*-Augmented reality; Virtual reality; Navigation; Field of view;**

I. Introduction

Tracked head-mounted displays (HMDs) are intuitive interfaces in virtual reality (VR) and augmented reality (AR) applications. The user can change the view naturally by walking and by rotating their head. However, relying exclusively on such an interface can be inefficient. When the user explores a region of interest (ROI) in a 3D scene far away, especially small targets, he/she has to examine potential ROIs one at a time by walking closer to and then retreating from each ROI sequentially. Furthermore, it is often the case that the physical space that hosts the VR application is smaller than the virtual world, so physical barriers might prevent the user from walking to distant ROIs. When scene understanding requires examining several ROIs simultaneously, the user might not be able to assume a view that shows all of them at once. A solution to this problem is to resort to flying mechanisms, but these can disorient the user and even induce nausea.

In this paper we propose a method to enhance VR and AR scene exploration by allowing the user to adapt the field of view interactively. The user can zoom in, to see a potential ROI in greater detail, bypassing the need for inefficient and sometimes impossible forward translation (Figure 1, top). Zooming in is supported with a focus+context visualization, which provides a higher resolution over a focus region, while integrating the focus region seamlessly into context. The context region does not change when the user decides to engage the focus+context visualization, the stable context region is in agreement with the user's lack of motion perception, which prevents disorientation and nausea. The user can also zoom out, to benefit from a more comprehensive visualization afforded by the larger field of view (Figure 2). We have measured the benefits of our free field of view VR and AR visualization in a user study.

Figure 1. Top: conventional (left) and our focus+context (right) visualization in VR. The focus region (inner circle) has a distortion-free 5x magnification factor, and it is connected with C^1 continuity to the surrounding context. Bottom: conventional user view (left) and our AR focus+context visualization (right). The AR HMD is enhanced with a high resolution camera that provides the magnified view of the real world, which is integrated continuously into the user's view.

II. Interactive Field of View for VR and AR

In this section we describe our technique for focus+context visualization in VR (Section II-A) and in AR (Section II-B), and our technique for zooming out in VR and AR (Section II-C).

A. Focus+context visualization in VR

We have developed a focus+context pinhole camera (FPC) that allows increasing the sampling rate over a region of focus. The FPC is based on the general pinhole camera [2] , with the change that the focus region is now circular,

978-1-5386-7592-2/18/$31.00 ©2018 IEEE
DOI 10.1109/ISMAR-Adjunct.2018.00043

Figure 2. A conventional VR visualization that uses the HMD's natural field of view cannot show both shelves at the same time (left). Both shelves are visible when the user zooms out, and the red-highlighted cradles with identical motion pattern can be matched (right).

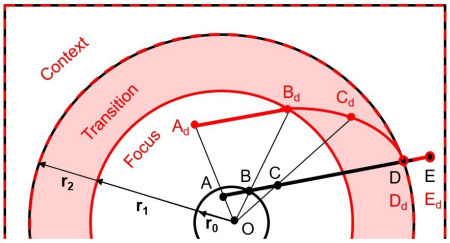

Figure 3. The projection displacement of the FPC model.

and not rectangular, based on polar, and not conventional, image plane coordinates. A circular focus region provides the magnification factor at the center without the significant distortions at the corners of the rectangular region.

The FPC model is shown in Figure 3. The outer rectangle is the image frame, which is the same before and after the focus+context visualization effect is applied, hence the alternating black (original) and red (with effect) dotted line. The image is not modified in the context region, i.e. beyond the large dotted circle. The inner most circle (solid black) corresponds to the focus region in the original image, i.e. the part of the original field of view that is magnified. In the focus+context visualization, the focus region is magnified to the solid red circle. The magnification factor is the ratio of the radii of these two circles, i.e. r_1/r_0. There are no distortions in the focus or in the context regions. The focus region is connected to context with C^1 continuity, i.e. with C^0 sampling rate continuity. Consider a scene straight line segment with projection AE in the original image (black line from A to E). Point A is displaced at A_d in the focus+context visualization image. Point B, which marks the segment's exit from the focus region, is displaced to B_d at the boundary between the focus and transition regions. A point C in the transition region, is displaced to point C_d. Points D and E are not displaced since they are on the boundary of, and inside the context region. The projection of the line in the focus+context visualization consists of two line segments A_dB_d and D_dE_d, connected by a curve.

We render using the FPC on the GPU with a custom vertex projection, which first conventionally projects a vertex V' to V, and then displaces V to V_d according to Algorithm 1. The focus+context visualization is obtained by displacing radially away from the origin of the polar coordinate system, which is the center of the focus region, by maintaining the angular coordinate θ constant (line 1). The displacement of a vertex depends on the location of its projection. If the vertex projects in the focus region, the displacement scales the distance r from the center by the magnification factor (line 3). If the vertex projects in the context region, then the vertex is not displaced (line 7).

Algorithm 1 Computation of vertex displacement.

Input: image projection $V(r,\theta)$ of vertex V', focus region radius r_0, magnified focus region radius r_1, and radius r_2 of outer boundary of transition region
Output: displaced image location $V_d(r,\theta)$
1: $V_d.\theta = V.\theta$ // the displacement acts radially
2: **if** $V.r < r_0$ **then** // focus region
3: $\quad V_d.r = V.r \cdot (r_1/r_0)$
4: **else if** $r_0 <= V.r <= r_2$ **then** //transition region
5: $\quad V_d.r = (1-t)^2 r_1 + 2t(1-t)((r_2-r_1)f+r_1)+t^2 r_2$
6: **else if** $V.r > r_2$ **then** // context region
7: $\quad V_d.r = V.r$
8: **end if**

A vertex in the transition region is displaced according to a Bezier curve expression in the (r, r_d) domain (line 5). The three control points of the Bezier curve are: $P_0(r_0, r_1)$, $P_1((r_0 r_2/r_1 - r_0)f + r_0, (r_2 - r_1)f + r_1)$, and $P_2(r2, r2)$, where f is a parameter that modulates the location of the second control point. f influences the shape of the Bezier curve, which translates to different distortion patterns at the transition region. f can be any fractional number, and we set it to 0.5, which distributes the distortion at the transition region as uniformly as possible. The Bezier curve ensures C_1 continuity between focus and context regions.

The user invokes the focus+context visualization mode hands free, by rolling their head left, and returns to conventional visualization by rolling their head right. A predefined minimum rolling angle is used to exclude unintentional head movements. The location of the focus region is fixed to the center of the image, and the maximum value of magnification factor is fixed (i.e. 5x in our experiments). The size of the focus region (i.e. r_0 in the original image) is set by the user based on the amplitude of the roll left. The size of the transition region is fixed (i.e. r_2 is double of r_1 in our experiments).

B. Focus+context visualization in AR

The requirements for focus+context visualization in AR are the same as those for VR, but meeting them is complicated by the fact that an AR visualization has to show the real world scene. We achieve focus+context visualization in AR based on the general pinhole camera [2], which has a

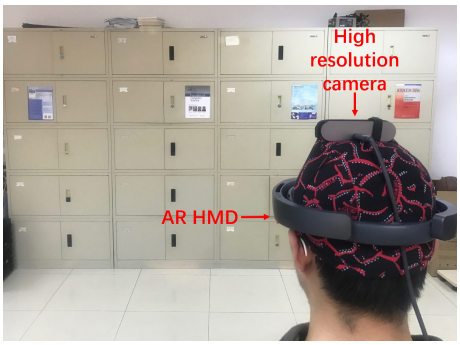

Figure 4. User wearing the AR HMD enhanced with a high resolution camera.

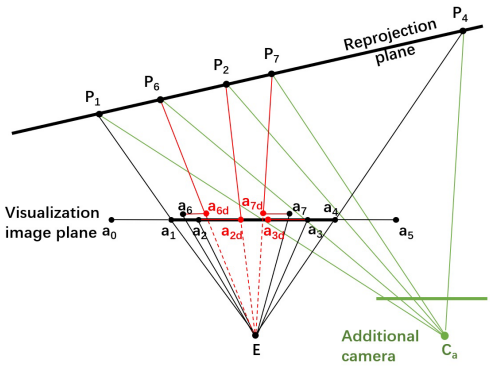

Figure 5. The sampling of the general pinhole camera model for zooming in for AR scenes

rectangular focus region connected to context continuously through a transition region. Both the focus and the context regions are undistorted. Figure 1 (bottom) shows the focus+context visualization we achieve in the AR context.

We describe our focus+context visualization for an optical see-through AR HMD system, which allows the user to see the real world directly, and which can overlay graphics over a small portion of the user's field of view. The main challenge is to be able to show the real world to the user at a higher resolution than what they can see directly. The AR HMD does have an integrated camera, but the resolution and the focal length of this camera are insufficient. Therefore, we attach to the AR HMD a high-resolution additional camera (Figure 4). This additional camera captures the real world in greater detail than what the user can see, and it provides the needed input for magnifying the visualization over the focus region.

The focus+context visualization is implemented by mapping the focus and transition regions to the active part of the optical see-through AR HMD. Since the additional camera has a different viewpoint and a different view direction compared to the user eyes, the additional camera image has to be reprojected to the view of each of the user's eyes. This reprojection requires (1) the pose of the additional camera with respect to the user's eyes, and (2) an approximation of the scene geometry.

(1) The additional camera is fixed with respect to the AR HMD, and its pose with respect to the HMD are recovered in a standard preliminary calibration step that shows a grid to both the additional camera and the HMD integrated camera. At run time, this calibrated transformation is concatenated with the API provided transformations from the integrated camera to each of the user's eyes, which allows mapping the additional camera to the two output images.

(2) We reproject the additional camera frame to the user's viewpoint using a reprojection plane that is aligned with scene geometry in a preliminary step. For the example in Figure 1 bottom, the reprojection plane matches the plane of the metal cabinets.

Figure 5 describes the AR focus+context visualization for one of the user's eyes, E. Unlike for Figure 3 where effect is shown in terms of the displacement of 3D scene point projections, this figure shows the effect using camera rays. The AR HMD is transparent throughout, i.e. from a_0 to a_5. The active part of the AR HMD, i.e. where graphics can be overlaid, is from a_1 to a_4. The focus and transition regions map to the active part of the HMD.

The rays at the focus region a_2a_3 are displaced inward to $a_{2d}a_{3d}$, which increases their scene sampling spatial resolution, achieving the desired magnification effect. The magnification factor is the ratio $a_2a_3/a_{2d}a_{3d}$. For example, a ray that reaches the focus region before displacement at a_7 is displaced closer to the view direction to a_{7d}. After displacement the rays continue to originate from E.

The rays at the transition region before displacement correspond to a_1a_2 and a_3a_4, and after displacement to a_1a_{2d} and $a_{3d}a_4$. Within the transition region, the spatial sampling rate of the rays is decreased to compensate for the denser rays at the focus region. The amount rays are displaced decreases towards the periphery of the transition region. For example the ray through a_6 is displaced less than the ray through a_2. The final ray of the transition region, i.e. the ray through a_1, is not displaced at all, which achieves continuity with the context region a_0a_1. Unlike for VR where the sampling rate decreases linearly from the focus to the context region, for AR we use a constant sampling rate over the transition region, i.e. a_1a_2/a_1a_{2d}. This provides C^0 continuity at both the focus to transition and the transition to context borders.

So far we have described the ray pattern needed to achieve the focus+context visualization. The color samples along these rays are retrieved from the additional camera image (green in Figure 5) by projective texture mapping on the reprojection plane. For example the color at ray a_7 is found by computing the intersection P_7 between the ray and the reprojection plane, and by projecting P_7 onto the image plane of the additional camera. The user controls the field of view

interactively with a hand-held controller.

C. Zooming out in VR and AR

Whereas focus+context allows the user to see part of the scene in greater detail, the user might also benefit from a larger field of view that provides a more comprehensive visualization of the scene. In the context of VR, zooming-out is trivially done by rendering the scene with a wider field of view.

In the context of AR, with a see-through HMD display, the horizontal field of view approaches that of human vision. However, zooming out can still be beneficial to the user because of two important reasons. First, the active part of the AR HMD display has a rather limited field of view, so any visual augmentation of the real world scene over a large field of view can only be done when the larger field of view is mapped to the active part of the display. For example if the application wants to circle elements of the scene over a large field of view, this can only be done when the large field of view is mapped to the active part of the display. Second, mapping a large visualization field of view to a small user field of view has the benefit that the user can see a greater part of the scene in front of them, where the human visual system has highest resolution.

III. User Study

We have measured the VR and AR HMD scene exploration benefits of our free field of view visualization in a controlled user study where participants (n = 32) were asked to perform two VR and one AR task. We recruited participants from the graduate student population of our institution. Their age is between 22 and 26 years, and six participants were female. Four of the participants had experience with HMD VR applications. Only one participant had experience with HMD AR applications. Each participant performed all three tasks.

A. VR methods

System implementation. We used an HTC Vive system that has a tracked HMD, an external tracker, and a wireless hand-held controller. The HMD is tethered to a desktop PC (Intel i7 processor, 16GB RAM, and NVIDIA 1070-ti graphics card). The virtual environments were rendered at 60fps for each eye.

Conditions. Each of the 32 participants was randomly assigned to one of four groups of eight: two control groups and two experiment group. Both control groups used conventional fixed field of view VR visualization. All four groups selected the desired view direction through the tracked HMD. However, the first control group (VRCG1) and the first experiment group (VREG1) selected the desired viewpoint by actually walking in the physical space and by teleportation. The second control group (VRCG2) and the second experiment group (VREG2) selected the desired viewpoint by flying along the view direction using the controller. We tested our system against these two control conditions since both viewpoint selection approaches are commonly used in VR applications. The flying approach is used in the case when the physical mobility of the user is limited by physical world constraints such as size, presence of obstacles, limited tracked area, and ergonomy.

First VR task (VR1). Participants were asked to find a cell phone inside a home (Figure 1 top). The home VR environment has a living room, a reading room, a dining room, and two bedrooms. The cellphone was placed in a location randomly selected from a set of 10 locations where people commonly place their cellphone. The floor space is 17m x 9.5m, with a ceiling height of 2.8m. The tracked physical space hosting the VR applications is 4m x4m, which is large enough for any of the rooms, but not large enough to accommodate the entire floor plan. VRCG1 and VREG1 participants were allowed to teleport from one room to another room using the controller. VRCG2 and VREG2 participants could not teleport, as they had the ability to travel any distance in VR without moving in the real world by flying along the view direction. For this task, participants in VREG1 and VREG2 could use our VR focus+context visualization.

Second VR task (VR2). Participants were asked to find pairs of Newton cradles with matching motion pattern. Nine pairs of matching cradles were placed on two perpendicular bookshelves in a room (Figure 2). A cradle is selected with a virtual hand-held pointer aimed with the controller. When a pair is found, the pair is removed. The task is complete when all pairs are found. The room is 8m x 8m, with the central 4m x4m being tracked, which was sufficient for the participant to move around in search of the matching cradles, so no teleportation ability was needed, nor provided.

B. AR method

System implementation. For the AR study, we used a Microsoft HoloLens, enhanced with a Logitech BRIO additional camera, with a resolution of 4,096 x 2,160 pixels. The additional camera is rigidly attached to the AR HMD, and it is connected through a wire to the same desktop computer described above. The additional camera video is transferred to the desktop at 14fps. The user controls the field of view with a hand-held controller (Logitech Gamepad F710) connected to the desktop. The desktop communicates with the AR HMD wirelessly (Wi-Fi interface). The AR HMD sends the current user view to the desktop. The desktop computes the focus+context or the zoom-out visualization based on the high resolution frame, the current user view, and the desired field of view. Finally, the desktop sends the visualization to the AR HMD where it is displayed over its active region.

Conditions. Each of the 32 participants was randomly assigned to one of two groups: one control group and one

Figure 6. AR task scene, 10 labels on the left and right walls

Table I
TOTAL VIEWPOINT TRANSLATION, IN METERS

Task	Group	Avg± std.dev.	Red.	p	Cohen's d	Effect size
VR1	VRCG1	134 ± 27				
	VREG1	83 ± 9.4	38%	< 0.01	2.5	Huge
	VRCG2	122 ± 14				
	VREG2	59 ± 2.8	52%	< 0.01	3.1	Huge
VR2	VRCG1	4.4 ± 1.6				
	VREG1	2.3 ± 1.1	48%	< 0.01	1.5	VLarge
	VRCG2	8.7 ± 2.8				
	VREG2	3.0 ± 0.4	66%	< 0.01	2.9	Huge
AR	ARCG	61 ± 2.6				
	AREG	5.2 ± 1.2	91%	< 0.01	28	Huge

Table II
TOTAL VIEW DIRECTION ROTATION, AS A MULTIPLE OF FULL 360° ROTATIONS.

Task	Group	Avg± std.dev.	Red.	p	Cohen's d	Effect size
VR1	VRCG1	70 ± 16				
	VREG1	48 ± 10	31%	< 0.01	1.6	VLarge
	VRCG2	22 ± 1.2				
	VREG2	15 ± 7.2	32%	0.02	1.2	VLarge
VR2	VRCG1	7.5 ± 3.6				
	VREG1	3.3 ± 1.5	56%	< 0.01	1.5	VLarge
	VRCG2	6.1 ± 1.2				
	VREG2	3.9 ± 0.56	36%	< 0.01	2.2	Huge
AR	ARCG	35 ± 13				
	AREG	8.9 ± 3.9	75%	< 0.01	2.8	Huge

Table III
AVERAGE LEFT-RIGHT TURN TIMES PER SUBJECT FOR VR2

Task	Group	Avg± std.dev.	Red.	p	Cohen's d	Effect size
VR2	VRCG1	57 ± 22				
	VREG1	26 ± 9.2	54%	< 0.01	1.8	VLarge
	VRCG2	43 ± 1.9				
	VREG2	34 ± 13	21%	0.07	0.95	Large

experiment group. The experiment group (AREG) used the controller to invoke and tune the free field of view visualization. The control group (ARCG) wore the AR HMD display, which was turned off. In other words, the participants saw the real world through the optical see-through part of the AR HMD, without any variation of the field of view.

AR task. In the AR task, the participants were asked to read words on paper labels that were placed in an office scene, and to arrange the words in a sentence. There were ten labels posted on the opposite walls (Figure 6). Each label had a number such that the participant would know right away whether the word is the next word they need to find. In other words, the task didn't test the participant's ability to make sentences from disparate words, which would have been a confounding factor for our study.

C. Results and discussion

We measured task performance using the following metrics. *Viewpoint translation* is defined as the total distance covered by the participant's viewpoint. For VRCG1 and VREG1, viewpoint translation was measured as the sum of the physical world distance walked by the participant, as recorded by the tracked HMD. For VRCG2 and VREG2, viewpoint translation is measured as the total flying distance. For ARCG and AREG, viewpoint translation is the total distance walked by the participant. *View direction rotation* is defined as the total angular change in view direction. For all participant groups, view direction rotation was measured using the tracked HMDs. *Task completion time* is defined as the total time a participant took to complete a task. All participants finished all tasks.

We detect and measure any benefit of our method compared to the conventional visualization using the p value of t-test and Cohens effect size d [1]. Effect sizes are qualified as 'very small', 'small', 'medium', 'large', 'very large', and 'huge' based on the following d thresholds: 0.2, 0.5, 0.8, 1.2,

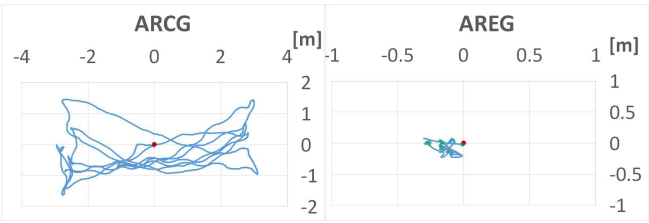

Figure 7. Control (left) and experiment (right) participant viewpoint 2D trajectory visualization for the AR task.

Table IV
TASK COMPLETION TIME, IN SECONDS

Task	Group	Avg± std.dev.	Red.	p	Cohen's d	Effect size
VR1	VRCG1	348 ± 72				
	VREG1	289 ± 46	17%	0.07	0.97	Large
	VRCG2	148 ± 6.4				
	VREG2	115 ± 48	22%	0.13	0.95	Large
VR2	VRCG1	119 ± 58				
	VREG1	75 ± 24	37%	0.07	1.01	Large
	VRCG2	77 ± 4.3				
	VREG2	65 ± 8.8	16%	< 0.01	1.71	Vlarge
AR	ARCG	106 ± 9.2				
	AREG	95 ± 19	10%	0.05	0.76	Medium

and 2.0 [3].

Experiment group participants translated the viewpoint considerably less than control group participants (Table I). For example, for VR1, VREG1 participants translated the viewpoint an average of 83m, as opposed to the VRCG1 participants who translated the viewpoint an average of 134m, which corresponds to a translation reduction of 38%, a Cohen's d of 2.5, indicating a huge effect. The smallest effect size, but still very large, is for VR2/VREG1, where the participants translate frequently but smaller amounts. For the VR tasks, the improvement brought by our free field of view exploration is statistically significant with p-values below 0.01.

Participants also benefited from our method for the AR task (i.e. last row in Table I), where they could read the labels from the center of the room, without having to magnify the labels by walking from one wall to the other, which eliminated the need to walk (i.e. a reduction of 91% in total translation, $p < 0.01$). Figure 7 shows that this particular participant completed the task almost without any viewpoint translation (i.e. walking).

Our method reduces the amount of view direction rotation (Table II). All p values are less 0.01 except for $p = 0.02$ for VREG2 of VR1. The rotation is reduced the most for the VR2 and the AR tasks. For VR2, control participants have to pan their head left/right in search of a match, which experiment participants don't have to do as a benefit of the larger field of view. For AR, control participants have to turn around 180° as they walk from one wall to the other, which experiment participants don't have to do.

For VR2 we have measured the number of times the participant changed their head rotation direction, in search for matching cradles. Table III shows that the experiment groups participants panned their view direction fewer times than those in the control groups. A rotation change is counted if the rotation amplitude is above a threshold of 5°.

Table IV shows the average task completion times over all participants. The experimental group finished the tasks considerably faster. The task completion time reduction was smaller for the AR task. We attribute this to the participants unfamiliarity with the zoom in ability in the AR context. Being able to zoom in with the sunglass-like see through display was surprising to the participants, who took some time to get used to it.

After the experimental group participants completed their tasks, we collected their opinion on the free field of view navigation technique through a questionnaire. The questionnaire consisted of 8 questions covering motion sickness, usability and suggestion. No participant reported motion sickness or nausea for our free field of view exploration. About 80% participants reported that the field of view changes were easy to invoke, and that the resulting visualization was easy to understand. Four VREG2 participant reported that the sudden increase in the amount of information in the image after a zoom out operation requires a pause to readjust cognitively to the image. Six VREG1 participants reported that teleportation is disorienting. The implementation of teleportation are issues orthogonal to our free field of view exploration technique. Furthermore, our technique reduces the need for teleportation, therefore alleviating the issue over conventional navigation. Future work will examine user perception and spatial orientation in further detail. Eleven AREG participates reported the latency of the focus region visualization, and three wished for a higher resolution still on the focus region. The best solution for addressing these limitations is to integrate a high resolution small field of view video camera into the AR HMD.

IV. Conclusion

We have presented a method for increasing VR and AR HMD scene exploration efficiency by allowing the user to modify the field of view interactively. Our focus+context visualization allows zooming in without user disorientation, by keeping the context region unchanged. The focus region is integrated into context seamlessly. We have implemented focus+context visualization for AR HMD by resampling in real time an on-board, high-resolution video feed. We have conducted a user study that confirms important benefits of our free field of view scene exploration in VR and AR tasks.

The AR system we introduced has some limitations. One limitation of our AR focus+context visualization is that the resampling of the high-resolution video feed relies on a known and simple scene geometry proxy, i.e. rectangles placed in the 3D scene to match large planar surfaces. One possible line of future work is to remove this limitation by leveraging the approximate scene geometry acquired by the AR HMD (i.e. HoloLens), which will also reduce the latency caused by data transfer from PC to HMD. The second limitation is that zooming out with our AR HMD is inherently discontinuous, since the active part is just a small part of the total see-through display. Furthermore, our HMD can only hide the user's direct view of the real world if the active part of the display is brighter than the scene. Future work could integrate diminished reality research results that modify the overlaid image to best hide the background.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China through Project 61772051.

References

[1] J. Cohen. *Statistical power analysis for the behavioral sciences: Jacob Cohens*. Psychology Press, 2009.

[2] V. Popescu, P. Rosen, L. L. Arns, X. Tricoche, C. Wyman, and C. M. Hoffmann. The general pinhole camera: Effective and efficient nonuniform sampling for visualization. *IEEE Trans. Vis. Comput. Graph.*, 16(5):777–790, 2010.

[3] S. S. Sawilowsky. New effect size rules of thumb. In *Journal of Modern Applied Statistical Methods*, p. 597C599, 2009.