AR Interfaces for Disocclusion–A Comparative Study

Shuqi Liao*
Purdue University

Yuqi Zhou†
Purdue University

Voicu Popescu‡
Purdue University

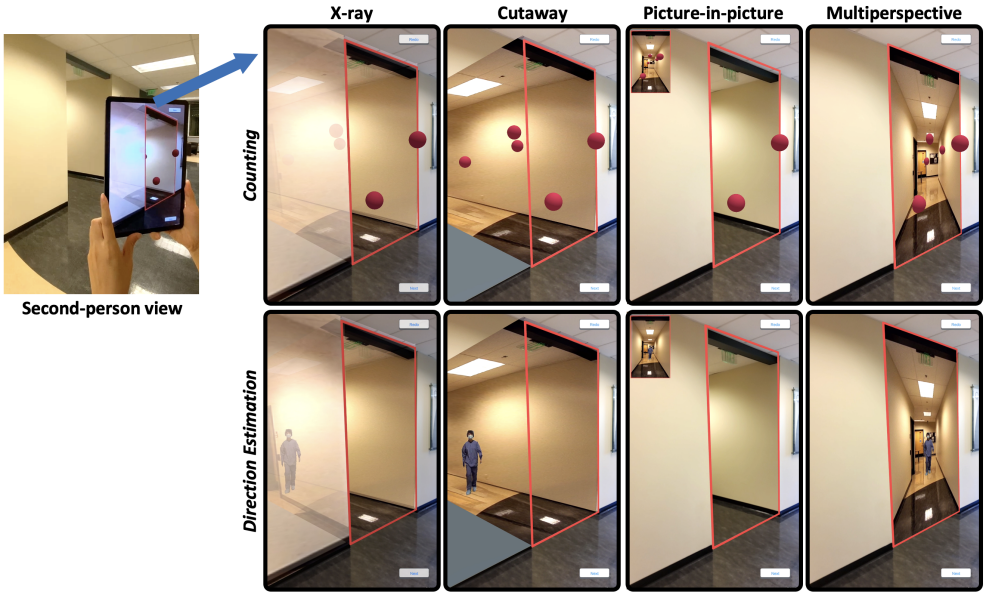

Figure 1: A user holds a tablet to see around the corner (second-person view). The user is asked to count moving spheres (top row), and to estimate the direction to the person in the side corridor (bottom row), using each one of four AR interfaces (columns).

Abstract

An important application of augmented reality (AR) is the design of interfaces that reveal parts of the real world to which the user does not have line of sight. The design space for such interfaces is vast, with many options for integrating the visualization of the occluded parts of the scene into the user's main view. This paper compares four AR interfaces for disocclusion : X-ray, Cutaway, Picture-in-picture, and Multiperspective. The interfaces are compared in a within-subjects study ($N = 33$) over four tasks: counting dynamic spheres, pointing to the direction of an occluded person, finding the closest object to a given object, and finding pairs of matching numbers. The results show that Cutaway leads to poor performance in tasks where the user needs to see both the occluder and the occludee; that Picture-in-picture and Multiperspective have a visualization comprehensiveness advantage over Cutaway and X-ray, but a disadvantage in terms of directional guidance; that X-ray has a task completion time disadvantage due to the visualization complexity; and that participants gave Cutaway and Picture-in-picture high, and Multiperspective and X-ray low usability scores.

Index Terms: Computing methodologies—Computer graphics—Graphics systems and interfaces—Mixed / augmented reality. Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality.

*e-mail: liao201@purdue.edu
†e-mail: zhou1168@purdue.edu
‡e-mail: popescu@purdue.edu

1 Introduction

The visualization of real world scenes is hindered by occluders that block the user's line of sight to potential regions of interest. The problem is particularly severe in the case of large and complex scenes which the user explores from within. Occlusions can make scene exploration inefficient, ineffectual, and even unsafe. Consider an application where a user walks through a building in search of a target object. In order to check a side corridor, the user has to walk up to the corridor entrance. If the side corridor proves to be empty, the user has to retrace their steps and the navigation is wasted. If the target is dynamic, the user might never be in the right place at the right time to find it. If the task is to find two similar objects that the user cannot see simultaneously, the user has to compare the object they currently see to the memory of an object seen earlier. Finally, the requirement of establishing line of sight to a dangerous target could reveal the user's presence, placing the user at risk.

Using an Augmented Reality (AR) visual interface that lets the user inspect the hidden part of the scene from the current location has the potential to overcome these challenges. If the hidden part of the scene is of no interest, the user can go on to explore other parts without wasting time and energy. A dynamic target is more likely to be found if the user also sees parts of the scene to which they do not line of sight. Furthermore, using a visual disocclusion method allows the user to compare visible and hidden objects directly, in the same image, which is more reliable than comparing from memory. Finally, the ability to see a target before the target has line of sight to the user improves user safety in the exploration of dangerous scenes.

In this paper we examine the design of discocclusion AR interfaces in support of scene exploration efficiency and effectiveness. We investigate the integration of the hidden parts of the scene into the user's view of the real world. We compare four AR interfaces

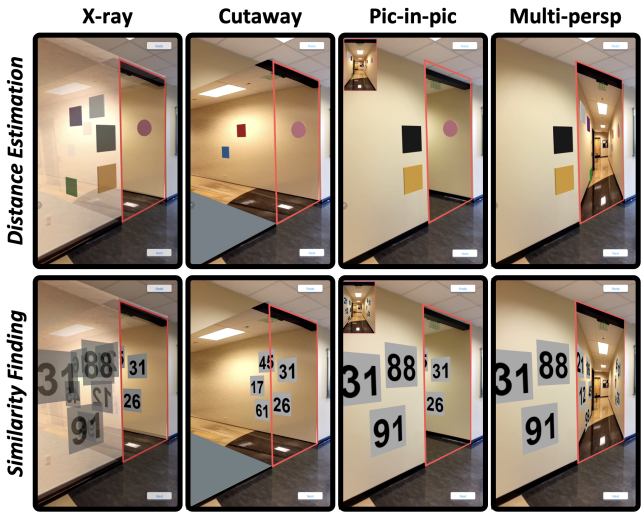

Figure 2: Disocclusion with four AR interfaces (columns) in two additional tasks (rows). For one task, the user is asked to find the rectangular color patch closest to the pink disk (top row). For the second task, the user is asked to find the number that appears twice.

implemented by handheld tablet (Figure 1). Two interfaces let the user see through the occluder (X-ray and Cutaway). The other two resort to a multi-view visualization, i.e., Picture-in-picture and Multiperspective. For Multiperspective the pixels of the secondary video feed are integrated seamlessly into the main video feed, e.g., over the entrance of the side corridor in Figure 1. The disocclusion capabilities of the AR interfaces are compared on four tasks: counting dynamic objects (top row in Figure 1), estimating the direction to an occluded object (bottom row), estimating the distance between objects (top row in Figure 2), and finding similarities among parts of the scene (bottom row). The four tasks cover the main design considerations when building a visual interface for disocclusion : (1) visualization redundancy, (2) discontinuity, and (3) comprehensiveness, (4) directional guidance accuracy, (5) scalability with occlusion depth complexity, and (6) familiarity to the user.

We have conducted a within-subject study ($N = 33$) with IRB approval. Participants performed the four tasks with each of the four AR interfaces (see Figure 1, Figure 2, and accompanying video). The study confirms that: (1) Cutaway has the highest usability score, but that it can discard important parts of the scene, leading to poor task performance; (2) compared to X-ray and Cutaway, Multiperspective and Picture-in-picture have the advantage of visualization comprehensiveness, and the disadvantage of inaccurate directional guidance; (3) X-ray visualization clarity suffers from blending together occluding and occluded layers, which leads to lower accuracy and longer completion times for tasks that require examining the scene in detail, compared to Multiperspective and Picture-in-picture.

2 Prior Work

Our paper compares several approaches for occlusion removal in AR. The occlusion removal approaches investigated were first developed in the context of visualization (Sec. 2.1). We then discuss occlusion removal using AR interfaces (Sec. 2.2), we contrast AR occlusion removal based on the type of AR interface (Sec. 2.3), we discuss the acquisition of the occluded part of the scene (Sec. 2.4), we end the discussion of prior work with a brief overview of diminished reality techniques that are relevant to disocclusion (Sec. 2.5).

2.1 Occlusion Removal in Visualization

Occlusion management has long been studied in visualization. A classical approach of alleviating occlusions is to give the user X-ray vision, i.e., to render occluding layers semi-transparently to reveal the occluded layers. Researchers have shown that X-ray vision is superior to overcoming occlusions conventionally, through interactive viewpoint changes, in terms of both efficiency and correctness in the context of perceptual tasks [13]. In addition to the see-through methods of X-ray and Cutaway, visualization research has also developed multi-view methods that combine samples from multiple acquisition viewpoints. Picture-in-picture and matrix visualizations simply juxtapose the multiple views, without attempting to remove the redundancy and discontinuity between individual views. Multiperspective visualization integrates the views for a comprehensive visualization that is also non-redundant and continuous [35].

In our work we compare a set of four occlusion removal techniques that were first developed in the context of visualization: X-ray, Cutaway, Picture-in-picture, and Multiperspective. This set provides good coverage of the space of disocclusion methods, with two see-through and two multi-view methods.

2.2 Occlusion Removal in AR

Researchers have investigated the potential of AR interfaces to remove occlusions in the visualization real world scenes. Owing to the many possibilities for designing AR interfaces that let the user see objects to which they do not have line of sight, there is an abundance of prior studies that compare multiple possible designs. A study [16] compared four variants of the Cutaway technique on the task of estimating the absolute distance to a sphere hidden by a wall. One variant conveys to the user the distance to the sphere directly, using a numerical label; a second variant draws a ground-plane regular grid, and a vertical regular grid through the sphere, so the user can estimate the distance by counting the number of horizontal grid cells to the sphere (Grid); a third variant draws a hole in the wall that reveals the sphere (Hole); the fourth variant deletes the wall and renders it in wireframe, through which the sphere can be seen (Wireframe). As expected, the results showed better performance for Tape and Grid than for Hole and Wireframe, and no differences between Tape and Grid, or between Hole and Wireframe.

A second study compares Cutaway variants in their ability to scaffold depth judgment in underground exploration [14]. The study finds that an *excavation box* approach that simulates digging out the underground pipes to be serviced outperforms overlaying the pipes on top of the occluding layer. A third study compares in-view to out-of-view visualizations of the occluded object and finds the former more useful in object selection and the latter in object searching tasks [17]. A fourth study examines how best support a user wanting to manipulate–and not just see–an occluded object, where a see-through method performed best [24].

In X-ray vision for AR, a concern is conveying to the user the correct depth ordering of layers [27, 43], which was addressed by enhancing the visualization with edges or with a frame that evokes a hole (a tunnel) in the front layer [4, 5]. Another approach shows that depth sorting can be conveyed using a perceptual, salience, and edge analysis of the real world image [64]. Real-time X-ray vision has been demonstrated outdoors with the help of security cameras that sample the occluded parts of the scene [22].

Surgeons are a category of users who benefit from preexisting familiarity with X-rays. Guiding biopsies is an application of AR X-ray vision that has been receiving attention from early on: the surgeon cannot directly see the target tissue and conventional visualization on a nearby monitor hinders the all-important hand-eye coordination [41, 42, 47]. As laparoscopic surgery has evolved allowing for evermore complex surgeries to take place without large incision wounds, AR has been used to improve the surgeon's indirect view of the operating field [45]. One challenge is the hands-free adjustment of the AR visualization, which has been attempted with the help of a brain-computer interface [6].

Researchers have investigated AR disocclusion methods that go

beyond X-ray and cutaway. One study compared a variable perspective to a multi-view AR technique on object finding in an outdoor scene, and found that although task completion times were larger, the variable perspective technique provided more visualization flexibility in terms of adjusting the overview amount and of unveiling occluded objects [46]. Another study showed that AR multiperspective visualization improves scene exploration efficiency and dynamic target tracking effectiveness [58].

Like prior work that compared multiple AR disocclusion methods [16], we also rely on simple, abstract tasks, which are building blocks of more complex tasks used in actual applications. Unlike prior work, our tasks are designed to evaluate multiple aspects of the disocclusion effect, including non-redundancy, continuity, comprehensiveness, spatial awareness, scalability with occlusion complexity, and familiarity to the user.

2.3 Head-Mounted vs. Handheld AR Displays

An essential concern when designing an AR interface with occlusion removal capabilities is whether to resort to a handheld or to a head-mounted display. Prior work has investigated disocclusion with both head-mounted [4, 17, 24] and handheld [14] displays. Optical see-through AR head-mounted displays (HMDs) have the advantage of placing the annotations directly into the user's view of the real world, which provides accurate directional guidance, whereas the video see-through effect of handheld displays creates the dual-view problem [9]. However, optical see-through interfaces cannot completely erase a bright occluder to reveal a dark occludee, as they can only add, and cannot block light from the scene. Depth perception is an important advantage of the stereoscopic visualization of the scene provided by HMDs [1]. Depth perception allows users to place the graphical annotations at their correct 3D location within the real world scene, even when cues such as shading or shadows are not as strong or consistent as they are in a fully synthetic scene [2].

In this paper we investigate disocclusion AR interfaces deployed on handheld displays. The lack of depth cues make tasks such as direction and depth estimation more challenging. Deployment on HMD AR interfaces is beyond the scope of this paper.

2.4 Acquisition of Occluded Parts of the Scene

A straightforward way of acquiring the occluded scene is with additional cameras, but this comes at the cost of having to deploy additional hardware. Camera research is rapidly developing devices that can see occluded parts of the scene, opening the door to self-contained AR devices that acquire the scene not only from the user viewpoint, but also from the translated viewpoints needed to capture hidden objects of interest. One camera for imaging behind occluders uses two-bounce light: the camera emits a light pulse which bounces on a directly visible object A to cast a shadow of the occluded object on a directly visible object B, shadow from which the occluded object is reconstructed [19]. Time-of-flight cameras record the light that leaves the occluded object and bounces on a visible surface to reach the sensor [48]. Non-line-of-sight imaging has also been demonstrated with thermal cameras, leveraging the fact that many surfaces become specular under long-wave IR illumination, acting like mirrors that give the user a glimpse of occluded objects [28].

We acquire the occluded part of the scene straightforwardly with an additional RGB camera. Our work focuses on the visual integration of the occluded part of the scene into main user view, and not on the acquisition of the occluded part of the scene.

2.5 Diminished Reality

Diminished reality is a subarea of AR related to occlusion removal. In diminished reality the goal is improving the user's view of the real world not by adding to it, but rather by subtracting from it. We refer the reader to a survey of diminished reality techniques [32]. Diminished reality approaches can be categorized based on how they

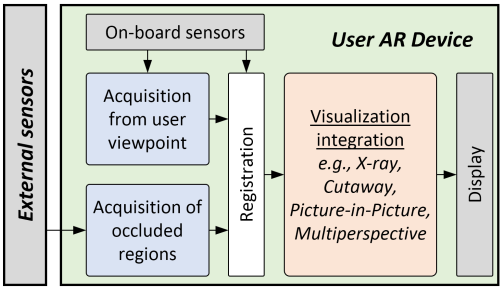

Figure 3: Architecture of AR interfaces for occlusion removal .

acquire the occluded scene that needs to be restored after the occluder is erased. One approach is to rely on 2D [33] or 3D priors [37]. The inpainting approach uses samples from regions neighboring the occluder [21, 26, 63]. Another approach is to search for the missing scene samples in earlier frames [18, 61]. The texture synthesis approach generates the patch that should cover the occluder algorithmically [10, 50]. A dynamic occluded scene has to be acquired in parallel, from a secondary viewpoint [3, 23, 31].

In diminished reality the focus is on completely eliminating the occluder to the benefit of the occludee, whereas our work operates under the assumption that there is no a priori knowledge of which objects are useless occluders and which objects are of interest. Some of tasks we use in our study specifically require the visualization of both occluding and occluded objects.

3 Overview

AR interfaces allow overlaying graphical annotations directly into the user's view of a real world scene. The typical use of an AR interface is to convey additional information about elements of the real world and to do so with annotations that are anchored to the real world elements that they describe. For example, text labels remain anchored to each of several persons moving trough a scene, conveying their name to the user. Another use of AR is *not* to add, but rather to *remove* from the user's view of the scene. Indeed, the user's understanding of the scene can benefit not only from adding information, but also from removing clutter or noise from the user's view of the scene. One such diminished reality interface could, for example, help a technician learn a jet engine repair procedure by removing from the technician's view engine components that are not involved in this particular repair procedure, increasing the visual salience of the relevant components.

In this paper we are concerned with letting the user see more than what is visible from their current position. Such disocclusion-capable AR interfaces are in-between diminished and augmented reality since they both *diminish* the visibility of occluders while *augmenting* the visibility of occludees. Figure 3 gives a general pipeline for for disocclusion using AR. The handheld device acquires the scene from the user viewpoint and registers each frame to the scene using visual features, possibly with input from auxiliary on-board sensors (e.g., depth cameras, gyros, and accelerometers.

The occluded parts of the scene are acquired with the use of external sensors. Once non-line-of-sight acquisition technology becomes available in a portable format, more compact architectures of disocclusion-capable AR interfaces will become practical, with all sensing performed by the user device. The data for the occluded parts of the scene is registered to the scene using off-line calibration when the external sensors are fixed with respect to the scene, or using real-time tracking, like for the user device, when the external sensors are mobile. In this paper we assume that the occluded part of the scene is acquired with a conventional RGB video camera mounted at a fixed position in the 3D scene. The additional camera is fully calibrated, with known intrinsic and extrinsic parameters,

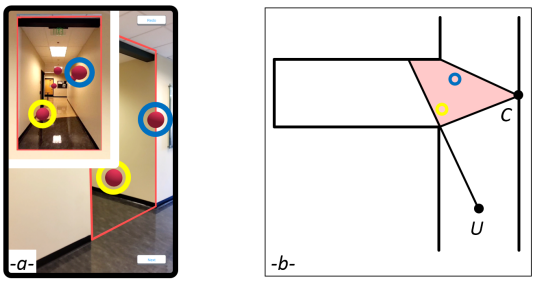

Figure 4: Visualization redundancy. Picture-in-picture shows two spheres twice (yellow and blue highlights in (*a*)), since the frusta of the primary and secondary views overlap (red region in *b*).

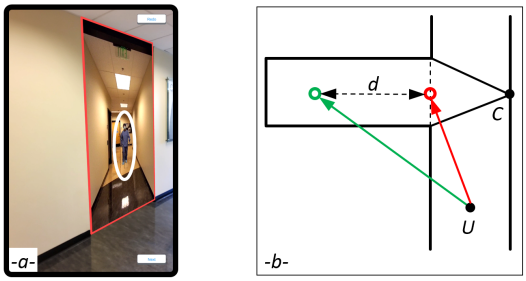

Figure 5: Approximate directional guidance. The Multiperspective visualization shows the person in the side corridor (white ellipse in *a*). The person appears projected onto the side corridor entrance (red dot in *b*), and not at their true location (green dot in *b*), so the visualization does not provide the true direction to the person.

which allows to effectively turn the camera into a projector that can augment the user's primary view with information about the occluded part of the scene.

The data for the occluded parts is integrated into the user's view of the scene to provide the user with an AR interface for disocclusion. The AR interface is constructed in one of several ways.

4 Visualization of Occluded Scene Parts

We first provide a set of design considerations for building disocclusion-capable AR interfaces (Sec. 4.1). We then give an overview of general occlusion management approaches (Sec. 4.2), followed by a description of the four disocclusion AR interfaces selected for comparison in our study (Sec. 4.3).

4.1 Design Considerations

Based on (1) extensive, empirically validated prior work in occlusion management in visualization [11, 30, 34, 36, 38–40, 57], in virtual reality [49, 51–53, 55, 60], and specifically in augmented reality [25, 59, 62], and based on (2) an analysis of basic tasks that benefit from disocclusion, including search, counting, similarity finding, distance estimation, and direction estimation, we have distilled a set of six considerations to follow when designing disocclusion AR interfaces.

Visualization non-redundancy. It is challenging to cover a 3D scene with RGB cameras such that no 3D scene point is captured by more than one camera. In other words, the raw feeds of the multiple cameras are partially redundant. Redundancy is an important problem in visual interface design, because the user cannot easily distinguish between the case when the interface shows two similar but distinct scene regions, and the case when the interface shows the same region twice [39]. One option is to alleviate the problem by choosing camera poses that minimize acquisition overlap and therefore minimize visualization redundancy. Another option is to process the raw camera feeds in order to eliminate redundancy. It is relatively straight forward to adjust the view frustum of a camera by cropping the image left-right and top-bottom. In Figure 1, the Picture-in-picture visualizations were built by cropping the raw secondary video feed left-right and top-bottom to match the entrance of the corridor. However, this does not eliminate *all* redundancy (Figure 4). In order to eliminate all redundancy, one also has to crop the view frustum with general clipping planes that do not pass through the user viewpoint. Implementing general clipping planes for real world cameras is challenging. One approach is to estimate scene geometry, which allows defining any desired clipping plane. In Figure 1, the Multiperspective visualization avoids all redundancy by clipping the secondary view frustum to retain only what is on the far side of the side corridor entrance plane, which is done leveraging the known depth of the virtual spheres.

Visualization continuity. We call a visualization of a 3D scene continuous if any pair of nearby 3D scene points have nearby projections in the visualization. If the distance between the two 3D points decreases to 0, so should the distance between their projections. This way, the user can estimate relative distances in the scene accurately. Furthermore, an object moving with a continuous trajectory will have a continuous trajectory in the visualization, allowing the user to track the object as it moves. If, on the other hand, the visualization is discontinuous, objects close to one another might appear at different locations in the visualization giving the user the incorrect sense that the objects are far apart. The Picture-in-picture visualization in Figure 2 is discontinuous across the frame of the secondary view, and the user might not be able to estimate the distance between an object seen in the primary view and one seen in the secondary view. Furthermore, a moving sphere jumps from the secondary to the primary view as it leaves the side corridor. The Multiperspective visualization does maintain visualization continuity between the secondary and primary views (see accompanying video).

Visualization comprehensiveness. Given a visualization canvas resolution, the pixel budget has to be distributed among the multiple views. One approach is to throw away the occluding layers, which is only acceptable when the occluding layers have no contribution to the visualization. For example, in Figure 2, the Cutaway visualizations discard two walls to let the user see inside the side corridor. However, these walls could include the rectangular patch closest to the pink disk or the second instance of the repeated numbers, so the drastic approach of making room for the occluded parts by removing occluders might hinder visualization effectiveness. Similarly, for the Picture-in-picture visualizations in Figure 1, the frame of the secondary view hides a part of the primary view, which is not always acceptable. Furthermore, selecting the size of the secondary view frame faces the competing requirements of achieving secondary view legibility while not deleting too much of the primary view. The multi-view visualizations are comprehensive, as they show the sum of the primary and secondary views.

Directional guidance accuracy. Disocclusion requires deviating from the rules of conventional single-perspective visualization in order to show parts of the scene to which there is no line of sight. This deviation could lower the accuracy of the directional guidance provided by the visualization. For example, in Figure 1, the Picture-in-picture visualization shows the person in the side corridor at the top left of the primary view. Clearly, the person is not located along a ray that starts at the user viewpoint and passes through the top left corner of the tablet. For the Multiperspective visualization, the person is shown within the side corridor entrance, providing better but still approximate directional guidance (Figure 5). The see-through visualizations do provide accurate directional guidance, i.e., users sees the occluded parts of the scene where they would see them in the absence of occluders.

Scalability with occlusion depth complexity. Disocclusion becomes more difficult when reaching the occluded part of the scene requires traversing multiple occluding layers. In the examples in Figures 1 and 2, the occluded part of the scene is reached with a single

additional camera that captures the side corridor. The side corridor ends with a T intersection with another corridor, and disoccluding this additional corridor exacerbates the challenges described above. Even if the disocclusion effect is limited to first-order occlusions that can be eliminated with a single additional view, integrating the additional view can create an occlusion depth complexity greater than one. For example, the X-ray visualizations in Figures 1 and 2, the side corridor is revealed by blending two occluding layers, which reduces the readability of the visualization. The multi-view visualizations show both the occluded and occluding layers clearly. Picture-in-picture does so at the cost of not showing a part of the main view, which is presumed of low interest, whereas Multiperspective shows the union of the two views.

Familiarity to the user. Disocclusion interfaces confer to users superpowers that let them examine parts of the scene to which they have no line of sight. The deviation from conventional human vision can make the visualization cryptic. Whereas X-ray visualizations might be somewhat familiar to the user from medical X-rays or from physical transparent surfaces such as glass walls, the user is likely to be less familiar with a visualization like Multiperspective for which there are no immediate real world counterparts. The user has to learn to interpret the Multiperspective visualization, for example through an analogy where the secondary view is projected onto a curtain placed at the side corridor entrance.

4.2 General Approaches for Occlusion Management

Visualization research has developed two main approaches for occlusion management: let the user *see through* the occluder, and let the user see in parallel *multiple views* of the scene.

Multi-view approaches display the primary view along with the secondary views. For one or a small number of secondary views, one option is to present the secondary views in small rectangles overlaid on top of the primary view, to obtain a Picture-in-picture visualization. Another option is to avoid overlap between the views and to display each view in its own rectangle, similar to the matrix visualization used by a security guard to monitor multiple camera feeds simultaneously. One advantage of these options is implementation simplicity, without the requirement of any geometric acquisition or camera calibration. Another advantage is familiarity to the user, since the visualization is the equivalent of a set of conventional images. The main disadvantage is a poor integration of the individual views. The user is left with examining each one of the views in turn, without the benefit of a true parallel visualization. The relationship between the poses of the multiple views is cryptic. Given a moving object currently visualized in one of the views, it is hard for the user to develop an intuition as to in what view the moving object will appear next. Scene understanding is also complicated by the redundancy between the views.

Another option is to improve the integration of individual views to obtain a Multiperspective visualization. When building a Multiperspective visualization the goals are to achieve visualization continuity and to avoid redundancy. These goals are achieved by constructing the Multiperspective visualization either by warping the 3D scene geometry and then rendering the warped scene with a conventional camera, or, conversely, by replacing the conventional linear camera rays with piecewise linear or even curved rays that are routed around occluders and then rendering the 3D scene with the generalized camera. If the warp of the 3D scene is continuous, so is the resulting visualization. Furthermore, the warp does not replicate scene geometry, and therefore the resulting visualization is non-redundant. When the Multiperspective visualization is constructed using generalized cameras, visualization continuity is achieved by making sure that nearby rays have similar trajectories; this way two 3D points that are close to one another will be projected by rays of visualization pixels that are close to one another. Redundancy is avoided by avoiding that the curved rays intersect. Compared to warping, the generalized camera approach has the advantage of a more direct construction of the desired disocclusion effect, allowing to pick a specific bundle of rays and guiding it to a specific occluded part of the scene. With the warping approach, the disocclusion effect is built by trial and error, modifying the scene geometry and then rendering the warped scene to inspect the results. Overall, an important advantage of both Multiperspective approaches is the better integration of the multiple views. This comes at the cost of construction complexity, which requires some knowledge of scene geometry, and of a visualization that is complex and not familiar to the user.

The see-through approach to occlusion management is to gain line of sight not *around* but *through* the occluder. The X-ray approach renders the occluding layers semi-transparently, whereas Cutaway removes them altogether. Advantages include showing an occluded part of the scene where the user would see it in the absence of the occluder, which can help the user build an accurate mental map of the scene. Another advantage is the simplicity of the visualization, as it builds upon constructs familiar to the user. The disadvantage of the X-ray approach is that the visualization is less clear, using the same output image real estate to show multiple layers on top of each other; the blending results in blurriness, i.e., in a loss of visual salience. The Cutaway method restores the sharpness of the visualization, but this comes at the cost of not showing the occluding layers at all, which is only acceptable when the occluding layers are merely a barrier preventing access to the important parts of the scene, and they do not contain important information of their own. Both methods require knowing the 3D scene geometry to know what to render semi-transparently or to cut away.

4.3 Four Disocclusion Capable AR Interfaces

Based on prior work we built four AR interfaces that provide a good coverage of the space of occlusion management techniques.

X-ray. Given the occluded region of interest and a geometric proxy of the scene, the AR interface renders semi-transparently all geometry layers between the user and the region of interest. For the examples in Figures 1 and 2, the scene proxy is limited to two tubes with a rectangular cross section, one for the main and one for the side corridor. The person in the side corridor is modeled with a video sprite computed by background subtraction and placed at the appropriate depth in the side corridor using the feet contact point.

Cutaway. The Cutaway AR interface is implemented like the X-ray interface, except that the occluding layers are *not* rendered semi-transparently but are instead completely removed.

Picture-in-picture. We use the simplest form of Picture-in-picture visualization: the secondary view is cropped to the side corridor entrance and displayed in the top-left corner of the primary view. The visualization does have some residual redundancy at the entrance of the side corridor, as discussed above (Figure 4).

Multiperspective. We chose a Multiperspective method based on a one-bend graph camera [35]. The primary rays that hit the side corridor entrance bend to the left to assume the secondary viewpoint of the camera capturing the side corridor. The implementation requires the intrinsic and extrinsic calibration of the side camera, as well as the implementation of a near clipping plane for dynamic scenes. We implement the near clipping plane leveraging the known geometry of the moving spheres or the position of the rectangular video sprite modeling the person in the side corridor.

Table 1 shows our hypotheses regarding potential strengths and weaknesses of the four disocclusion AR interfaces , with respect to the design considerations enumerated in Sec. 4.1. Picture-in-picture can suffer from redundancy, which can be a problem in tasks that require disambiguation between similar and repeated parts of the scene. Picture-in-picture can also suffer from discontinuity between the main and secondary views, which can be a problem in tasks that require judging 3D proximity across the two views, or tasks that require tracking dynamic objects that cross from one

view to another. Cutaway reveals occluded parts of the scene by removing occluders, which results in an incomplete visualization that might not be suitable for a task that relies on both the occluder and the occludee. The multi-view methods, i.e., Picture-in-Picture and Multiperspective, disoccluded by perturbing the global spatial relationships in the scene, which is likely to be a problem in tasks that require conveying the real-world (absolute) direction to the disoccluded parts of the scene. The X-ray visualization is less clear when two or more surfaces blend together, which can become problematic in tasks that require identifying parts of the scene based on subtle characteristics that are obscured by the transparency effect. Finally, the Multiperspective approach is less familiar to the user, with a steeper learning curve.

5 User Study

We have conducted a user study to compare the X-ray, Cutaway, Picture-in-picture, and Multiperspective AR interfaces.

5.1 Methods

Participants. We have recruited 34 participants from the undergraduate and graduate student population of our University. The minimum participant age range was 20 years, and 5 participants were over 30 years old. 8 participants marked the "female" gender box, 25 the "male" box, and one the "no answer" box. 8 participants indicated that they had never used AR before, 12 had used AR once, 13 occasionally, and one frequently. We excluded the data of one participant because of a technical malfunction.

Study design. We have opted for a within-subjects design where all participants performed all tasks in all conditions, which, compared to a between-subjects design, has the advantage of greater statistical power for the same number of participants.

Conditions. There were four conditions corresponding to the four AR interfaces: X-ray (XR), Cutaway (CA), Picture-in-picture (PP), and Multiperspective (MP). Our comparison study hypothesizes strengths and weakness for all conditions (see Sec. 4.3), hence we do not label the conditions as control or experimental.

Tasks. We have chosen four tasks to shed light on the potential AR interface advantages and disadvantages discussed in Sec. 4.3. Each participant repeated each task three times, i.e., in three trials, for each of the four conditions. The participant stood in a main hallway 2.39m wide, 3.50m away from the entrance of a side corridor, the width of the side corridor entrance was 1.60m, and the length of the side corridor was 7.31m (Figure 1).

The *Counting* task asked a participant to count the number of moving identical spheres (Figure 1, top row). The spheres changed color from one trial to the next. The spheres were virtual, lit in a way consistent with the lighting of the real world scene. Two trials had a low level of difficulty, with the number of spheres randomly selected in the [4, 6] interval, and one trial had a high level of difficulty, with the number of spheres randomly selected in the [9, 11] interval. The possible answers were 3, 4, 5, 6, 7 for the low difficulty trials and 8, 9, 10, 11, 12 for the high difficulty trials.

Design consideration \ Interface	X-ray	Cutaway	Picture-in-picture	Multi-perspective
Non-redundancy	++	++	--	++
Continuity	++	++	--	++
Comprehensiveness	++	--	+	++
Directional guidance	++	++	--	--
Scalability w/ layers	--	-	+	+
Familiarity to user	+	+	++	--

Table 1: Potential strengths and weaknesses of the four AR interfaces selected for comparison in this paper.

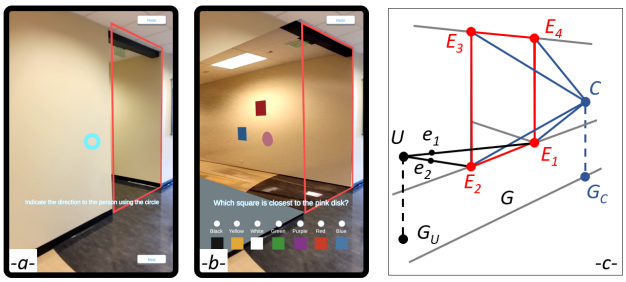

Figure 6: **Panel a.** In the *DirectionEstimation* task the user is asked to point in the direction of the hidden person using the blue circle at the center of the frame. **Panel b.** In the *DistanceEstimation* task the user is asked to find the square closest to the pink disk. **Panel c.** The additional camera capturing the side corridor from C is registered to the world coordinate system using the precalibrated side corridor entrance rectangle $E_1E_2E_3E_4$.

The *DirectionEstimation* task first showed the participant a person in the side corridor using the disocclusion capabilities of the AR interface (Figure 1, bottom row); an on-screen message asked the participant to "memorize the position of the person within the side corridor". Then the disocclusion effect was removed, reverting to the regular tablet video camera feed in which the person is occluded, and the participant was asked to "indicate the direction to the person using the circle" (Figure 6*a*).

The *DistanceEstimation* task affixed 8 virtual stickers to the walls of the corridors (Figure 2, top row). The stickers were 7 squares of different colors and one pink disk. The participant was asked "which square is closest to the pink disk". The distance differences were sufficiently large for the answer to not depend on the precise distance definition, e.g., distance between centers, or smallest distance between any two points. The pink disk changed position from one trial to the next. The pink sticker was always visible in the Cutaway condition, but the true closest square was cut away in two of the three trials. The participant was provided the possible answers using squares of the same colors as the stickers (Figure 6*b*).

The *SimilarityFinding* task showed the participant 13 virtual white stickers with black double-digit numbers (Figure 2, bottom row). 11 stickers showed unique numbers and two stickers showed the same number. The participant was asked "which number appears on two different stickers". The possible answers provided included all 12 double digit numbers, as well as "None". For Cutaway, the repeated number was culled away in two of the three trials, so the appropriate–but overall incorrect–answer was "None".

The position of the participant and the field of view of the camera of the tablet were such that the participant had to change the view direction of the tablet to scan the entire side corridor in the XR and CA conditions. In other words, the side corridor was deep enough to not fit in a single "default" view (see X-Ray and Cutaway frames in Figure 1. Aiming the tablet implies a significant translation of the tablet, which in turns results in motion parallax in the visualization. The motion parallax is a depth cue that can compensate somewhat for the lack of interpupillary disparity in the tablet's monoscopic visualization. The comprehensiveness of the PP and MP visualizations preclude the need to translate the tablet and hence devoid the user from the motion parallax clue.

Implementation. We have implemented the interfaces and the tasks in an Android application using Unity [44]. We used a Samsung Galaxy S6 tablet, tracked using AR core. The additional camera capturing the side corridor from viewpoint C in Figure 6*c* was preregistered to the side corridor entrance $E_1E_2E_3E_4$. Then, at the

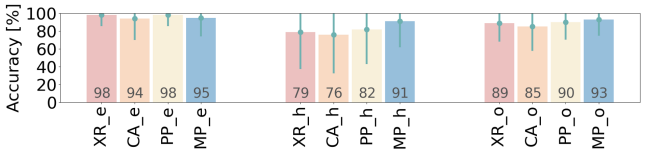

Figure 7: *Counting* accuracy over low and high difficulty trials.

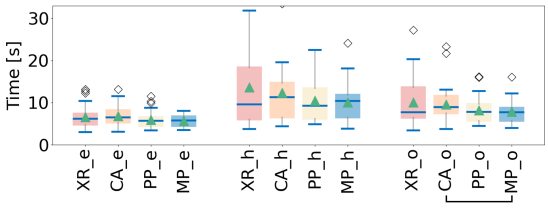

Figure 8: Box plots of trial completion times for the *Counting* task. Circles stand for outliers, green triangles for average values, horizontal segments for median values, rectangles for the middle quartiles, and vertical capped segments for value ranges.

beginning of each application session, the entrance $E_1E_2E_3E_4$ is registered to the world coordinate system of the AR Core tablet tracking module with the following setup procedure: the experimenter taps the tablet screen at e_1 where E_1 appears, the tablet camera ray Ue_1 is intersected with the ground plane G tracked by AR core to obtain the 3D coordinates of E_1, the same procedure is repeated to find the 3D point E_2, and then E_3 and E_4 are constructed above E_2 and E_1, in the vertical plane through E_1E_2, using the known height of the side corridor entrance. For *DirectionEstimation*, the video stream of the additional camera is background subtracted with a fragment shader to build the video sprite of the person (Sec. 4.3).

Experimental procedure. A participant first fills out the biographical data questionnaire which asks for the participant's age, gender, and level of AR experience. The experimenter performs the setup procedure to register the additional camera to the AR Core world coordinate system as described above. The experimenter hands the tablet to the participant. The participant is located in the hallway of an office building (Figure 1). The participant starts with a practice run of all tasks under all conditions, under the guidance of the experimenter, which takes approximately 10min. During this practice run a participant had the option of redoing a task if they felt like the task or the visualization condition is not clear. The participant performs the experimental run, which includes all tasks under all conditions, with different number of spheres, different positions of the person in the side corridor, different sticker placement, and different sticker numbers. A participant goes over the four tasks in the same order from the practice to the experimental run. A participant performs a task in all four conditions before moving on to the next task. A participant has to provide an answer to move on to the next trial. The experimental run takes 10min, after which the participant fills out a system usability scale (SUS) questionnaire [7] for each of the four AR interfaces. The total participant time is 40min.

Data collection. For *Counting*, *DistanceEstimation*, and *SimilarityFinding*, the application records the one bit correctness of the answer provided by the participant. For *DirectionEstimation* the application records the pointing direction error as the angle between the direction indicated by the participant and the actual direction to the person. The angle is measured in the horizontal plane to not distinguish between pointing to the head or to the center of the participant. For each task, the application also records the time the participant needs to provide the answer. In total, there were five continuous dependent variables, i.e., time for each of the four tasks and pointing direction error, and three dichotomous dependent variables, i.e., counting, distance estimation, and similarity finding accuracy. The answers to the 10 SUS questions were provided on a five-point Likert scale, and were treated as continuous variables.

Data analysis. We analyzed the dichotomous dependent variables by computing the percentage of correct answers, as well as standard deviations. Furthermore, we performed Cochran's Q test [8] for an initial comparison between multiple methods. If differences proved to be significant, we performed a posthoc analysis to investigate pairwise differences using McNemar's test [29], with a Bonferroni corrected significance level. The distribution of the continuous dependent variables was not normal so we analyzed differences between the four conditions using Friedman's non-parametric test [15]. When a statistically significant difference was found between the four conditions, we performed a posthoc analysis of pairwise differences using Wilcoxon's signed-rank test [54], with a significance level Bonferroni corrected [12] to $\alpha = 0.05/6 = 0.0083$, where 6 is the number of unsorted pairs of conditions. We used SPSS [20].

5.2 Results and Discussion

We first present and discuss the results for each task, and then we discuss the results over all tasks, relating back to the potential strengths and weaknesses of the four AR interfaces (Table 1).

Counting. The accuracy of the counting task is given in Figure 7. Each bar shows the average, including numerically, as well as the standard deviation (green line), capped at 100%. The four interfaces perform similarly on the low difficulty trials (subscript e for "easy"), with high accuracy values between 94% and 98%. For the high difficulty trials (subscript h for "hard"), MP has a substantial advantage, i.e., 91% vs the runner up method PP which is at 82%. The "overall" (subscript o) accuracy is the average of the "easy" and "hard" trials, and it ranks the methods in the order MP, PP, XR, and CA. None of the differences in counting accuracy are statistically significant.

Figure 8 shows that, as expected, the easy trials take substantially less time than the hard ones. The average and median values are similar across methods. For the difficult trials, the 75th percentile times, indicated by the top edges of the vertical rectangles, decrease substantially from XR to CA to PP and to MP. For example, for MP, 75% of the times are faster than 12.6s, whereas for XR, the upper limit of the 75% shortest times is 19.2s. The Friedman test reveals a significant difference across the four conditions ($\chi^2(3) = 8.745, p = 0.033$). The posthoc analysis (Table 2) shows that MP has significantly shorter overall *Counting* task completion times than CA. The other three multi-view to see-through pairwise differences, i.e., MP-XR, PP-CA, and PP-XR, are not significant using the Bonferroni $\times 6$ correction. The differences within the same category, i.e., MP-PP and CA-XR, are far from significant.

We conclude that the multi-view methods have an advantage over the see-through methods in terms of accuracy for difficult trials and in terms of completion times. This is probably due to the fact that with see-through methods the user has to pan the tablet left-right to examine the entire scene, which multi-view methods show in a single frame. MP has a slight advantage over PP in terms of accuracy for difficult trials, as PP suffers from redundancy when the sphere is in the wedge of the side corridor visible in the main frame.

DirectionEstimation. The see-through methods CA and XR have smaller direction estimation errors than the multi-view methods MP and PP (Figure 9, left). Within the same method category, PP has smaller errors than MP and CA has smaller errors than XR.

		MP-CA	MP-XR	PP-CA	PP-XR	MP-PP	CA-XR
Wilcox.	*Z*	-2.653	-2.260	-2.475	-1.742	-0.384	-0.956
	px6	0.047*	0.144	0.078	0.486	4.206	2.034

Table 2: Comparisons of *Counting* times between pairs of conditions using Wilcoxon's signed-rank test. All *Z* values are based on positive ranks, i.e., the first condition has a shorter average completion time.

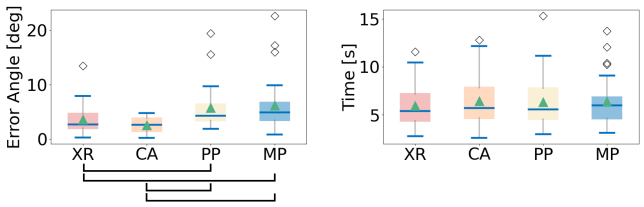

Figure 9: *DirectionEstimation* errors (left) and times (right).

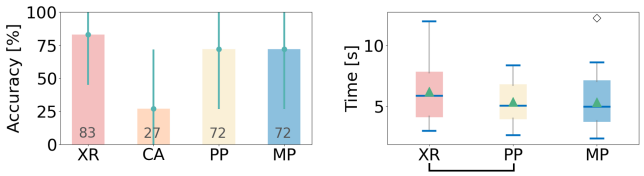

Figure 10: *DistanceEstimation* accuracy (left) and time (right). CA times are omitted because CA discards the closest sticker, resulting in inconsistent time measurements.

The Friedman test indicates a significant difference in direction estimation error between the four methods ($\chi^2(3) = 29.54, p < 0.001$). The posthoc analysis (Table 3) shows that any see-through method (CA, XR) has a significant advantage over any multi-view method (MP, PP). There are no significant differences between the two methods in the same category, i.e., between MP and PP, and CA and XR. The task completion times are given in Figure 9, right. None of the time differences are significant.

This confirms that direction estimation is problematic for multi-view methods which perturb spatial relationships, and that direction estimation is easier for see-through methods that show the target where it would be seen in the absence of the occluder. The lack of differences between times indicates that participants do not find that they could refine the direction estimation based on the multi-view visualization and opt to provide a rough guess right away.

DistanceEstimation. The distance estimation accuracy is given in Figure 10, left. As expected, CA performs poorly since in some trials the closest sticker is missing, so the participant has no chance of answering correctly. Among the remaining three methods, accuracy is better for XR and the same for MP and PP, with no significant differences. Completion times are given in Figure 10, right. CA is omitted since Cutaway greatly simplifies the scene, at the cost of removing the closest sticker, so CA times are inconsistent with those of the other methods. XR has the longest times, and MP and PP times are similar. The Friedman test reveals significant differences between the MP, PP, and XR times ($\chi^2(2) = 6.606, p = 0.037$). The posthoc analysis (Table 4) reveals that the PP-XR difference is significant after the $\times 3$ Bonferroni correction , i.e., the PP times are significantly shorter than the XR times.

We conclude that the CA method has the challenge of potentially removing parts of the scene that are essential for the task at hand. Furthermore, the spatial perturbation of multi-view methods makes distance estimation more challenging. The XR method offers the best opportunity to judge distance accurately, but this comes at the cost of longer task completion times, which the user needs to disambiguate the complex multi-layered visualization.

		CA-MP	XR-MP	CA-PP	XR-PP	PP-MP	CA-XR
Wilcox.	*Z*	-4.244	-3.082	-4.279	-3.082	-0.018	-2.269
	px6	<0.001*	0.012*	<0.001*	0.012*	5.916	0.138

Table 3: Comparisons of *DirectionEstimation* errors between pairs of conditions. All Z values are based on positive ranks, i.e., the first condition has a smaller average direction estimation error.

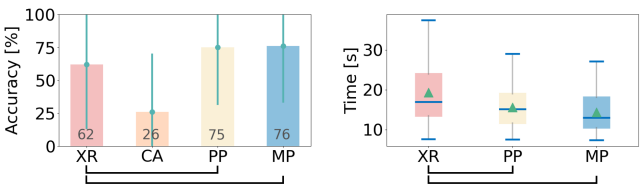

Figure 11: *SimilarityFinding* accuracy and times. CA times are omitted because CA discards the sticker with the matching double-digit number, resulting in inconsistent time measurements.

SimilarityFinding. As expected, CA performs poorly since in some trials the sticker with the repeated number is removed, so the participant has no chance of answering correctly (Figure 11, left). Among the remaining three methods, accuracy is worse for XR and similar for MP and PP. The differences between MP, PP, and XR are significant (Cochran's Q test $\chi^2(2) = 11.091, p = 0.004$). A posthoc analysis using McNemar's test, with an $\times 3$ Bonferroni correction, reveals that accuracy is significantly higher for MP versus XR ($p \times 3 = 0.012$), and for PP versus XR ($p \times 3 = 0.045$).

Figure 11, right, gives the *SimilarityFinding* completion times, omitting CA which can remove the repeated sticker. XR has the longest times, and MP has a slightly shorter time than PP. The Friedman test reveals significant differences between the MP, PP, and XR times ($\chi^2(2) = 10.424, p = 0.005$). The posthoc analysis with Bonferroni $\times 3$ correction (Table 4) reveals that the PP-XR and MP-XR differences are significant, i.e., the PP and MP times are shorter than the XR times.

We conclude that, like for *DistanceEstimation*, CA is also unsuitable for the *SimilarityFinding* task, because CA could discard a part of the scene that is essential for the task. Furthermore, the visual complexity of the XR method due to the multiple blended layers makes finding similarity with the XR method more challenging than with the multi-view methods, resulting in significantly lower accuracy and higher task completion times. The multi-view methods are well suited for the similarity finding task since they provide a comprehensive visualization of the scene that facilitates comparing distant parts of the scene. Unlike for *DistanceEstimation*, *SimilarityFinding* is not affected by the perturbation of global spatial relationships introduced by the multi-view methods.

The overall SUS scores for XR, CA, PP, and MP are 64, 82, 76, 58, corresponding to "Poor", "Excellent", "Good", and "Poor", respectively [7]. The differences are significant (Friedman $\chi^2(3) = 24.708, p < 0.001$). There are significant pairwise differences in favor of CA over XR ($p \times 6 < 0.001$) and MP ($p \times 6 < 0.001$) and in favor of PP over MP ($p \times 6 = 0.0032$). Users prefer CA for its clarity, but this comes at the cost of poor task performance when Cutaway removes important parts of the scene.

Considering the hypothesized strengths and weaknesses of the four methods outlined in Table 1, our results provide strong evidence for the comprehensiveness shortcoming of CA. We have also found strong evidence that the multi-view methods investigated (PP and

		Distance Estimation			Similarity Finding		
		PP-XR	MP-XR	PP-MP	PP-XR	MP-XR	MP-PP
Wilcox.	*Z*	-2.385	-1.787	-0.018	-2.385	-3.297	-1.438
	px3	0.051*	0.222	2.958	0.051*	0.003*	0.450

Table 4: Comparisons of *DistanceEstimation* (left) and *SimilarityFinding* (right) times between pairs of conditions. All Z values are based on positive ranks, i.e., the first condition has a shorter average time.

MP) do not provide accurate directional guidance, as the user cannot easily reverse engineer the true direction to a target that does not appear where the user would see it in the absence of the occluding layer. Our results indicate that comprehensiveness is also a challenge for see-through methods such as CA. Since CA cannot show the entire side corridor in a single frame, the user has to pan the tablet, which leads to longer task completion times than for multi-view methods (Tables 2 and 4). Our results support the hypothesis that the complexity of the XR visualization stemming from blending multiple layers together affects performance in tasks where the user has to closely examine distant parts of the scene, such as *SimilarityFinding*, i.e., XR has significantly lower accuracy and longer times compared to MP and PP (Figure 11 and Table 4). The frustrating complexity of the XR visualization is also echoed in the answers to the SUS questionnaire (SUS2, SUS3, and overall). The lack of user familiarity with MP is confirmed by SUS (SUS7, SUS9, SUS10, and overall). Our results did not find any significant difference between PP and MP to confirm the non-redundancy and continuity challenges hypothesized for PP. We attribute this to the fact that the PP visualization used in the study is relatively simple, with one additional view that is well aligned and mostly non-redundant with the main view. Counting accuracy for the high difficulty trials is higher for MP (91%) vs PP (82%), but only one in three trials had a high difficulty. The small level of redundancy and discontinuity of the PP used in the study was not sufficient to become a problem for the low difficulty *Counting* trials (accuracy better than 94%).

6 Conclusions, Limitations, and Future Work

We have presented a study that compares four disocclusion AR methods : two see-through methods, i.e., X-ray and Cutaway, and two multi-view methods, i.e., Picture-in-picture and Multiperspective. The AR interfaces allow the user to perform tasks that require seeing parts of the scene to which the user has no line of sight. The results indicate that Cutaway scores highest in terms of usability, with the important caveat that the method might discard elements of the scene that are essential to the task, and should therefore only be used when application domain knowledge can be used to avoid an incomplete visualization. The results also indicate that see-through methods cannot avoid eliminating the occluding layers by blending them together, as the resulting X-ray visualization is hard to parse. Compared to see-through methods, multi-view methods have the advantage of comprehensiveness, leading to faster task completion times, but this comes at the cost of perturbing spatial relationships needed, for example, for accurate direction estimation. Within the multi-view method category, our results show that Multiperspective does not have a time disadvantage compared to Picture-in-picture. This is against the findings of earlier studies [46] that found that users require more time to adjust the many parameters of a Multiperspective visualization. We explain this discrepancy by the fact that in our case the Multiperspective effect is deployed automatically, without the need for time consuming user interaction.

Our study has several limitations that should be addressed by future work. We have opted to compare four different disocclusion approaches to sample the design space as well as possible. This comes at the cost of comparing only one variant of each approach. Future studies could lead to additional insights. For example, one could also include in the comparison a Multiperspective method that disoccludes by moving the occluder away and keeping the occludee in place [56], which is likely to provide more accurate direction guidance to the occludee. Furthermore, one could also include in the comparison a Picture-in-picture method that shows more than two views, e.g., a 4x4 security-guard like visualization, where the redundancy and discontinuity of the visualization is likely to lead to a significant reduction in task performance.

Our study relied on synthetic objects and on simple modeling of real world objects using a video sprite. Future work could employ more sophisticated depth acquisition, e.g., using an RGBD camera, or depth inference, e.g., structure from motion, approaches could be used to improve the visualization of real world, dynamic objects.

The tasks were chosen to illuminate on all potential strengths and weaknesses of the four interfaces investigated. Specifically, the similarity task cannot be completed accurately in the CA condition if one of the two numbers was culled. The similarity task is easier in the PP and MP conditions due to the comprehensiveness of the visualization, but it is not trivial, since the visualization comprehensiveness comes at the cost of lowering the resolution on objects that are distant and or seen at an angle. Furthermore, the direction estimation task is challenging for the multi-view PP and MP conditions which perturb global spatial relationships. On the other hand, participants could "bookmark" the direction to the person in the side corridor in the XR and CA conditions by centering the person on the tablet while the person was visible; this would "pre-aim" the tablet in the correct direction, an important advantage over MP and PP. Future studies could calibrate tasks to avoid under- and over-flow errors in performance measurement for various conditions.

In this work we relied on a tablet implementation of the AR interface. Compared to an optical see-through AR head-mounted display (HMD), the tablet has the advantage of lower hardware cost, wider field of view, robustness to scene lighting conditions, and social acceptability. Future work could investigate HMD AR interfaces, as these bring the important advantage of depth perception, which might help users parse complex visualizations. One challenge is to provide the left-right eye disparity for the "pixels" corresponding to the disoccluded parts of the scene, i.e., for the side corridor in our example. One solution is to acquire the side corridor with an RGBD camera, which, for X-ray and Cutaway allows reprojection to where the user would see the side corridor in the absence of the occluding wall. For Picture-in-picture, the left and right eye positions can be offset half the pupillary distance from the position of the acquisition camera. For Multiperspective, the RGBD frame can be warped and then rendered conventionally for the left and right eyes [49, 52]. The depth cue provided by stereoscopic visualization is likely to strengthen the separation between objects at different depths, and users are likely to count, track, find, compare, and estimate distances and directions to objects more easily.

In this paper we compared AR disocclusion methods on basic tasks such as searching, counting, finding similarities, and direction and distance estimation. These tasks are building blocks of more complex tasks encountered in real world applications. Future studies should gauge the pros and cons of the various disocclusion methods in the context of actual applications, such as providing driver assistance in dangerous intersections, situational awareness during emergency management and law enforcement in indoor and urban environments, and occlusion-free visualization in multi-camera laparoscopic surgery.

Finally, there is also a need for longitudinal studies that give participants the exposure length needed to become proficient with the powerful AR interfaces needed for solving complex tasks.

Acknowledgments

The authors would like to thank the anonymous reviewers for their help with improving this manuscript, and the members of the Computer Graphics and Visualization Laboratory of the Computer Science Department of Purdue University for their feedback. This work was supported in part by the United States National Science Foundation through awards 2212200 and 2219842, and by the Purdue University Computer Science Department.

References

[1] H. Adams, J. Stefanucci, S. Creem-Regehr, and B. Bodenheimer. Depth perception in augmented reality: The effects of display, shadow, and

position. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 792–801. IEEE, 2022.

[2] H. Adams, J. Stefanucci, S. Creem-Regehr, G. Pointon, W. Thompson, and B. Bodenheimer. Shedding light on cast shadows: An investigation of perceived ground contact in ar and vr. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):4624–4639, 2021.

[3] B. Avery, W. Piekarski, and B. H. Thomas. Visualizing occluded physical objects in unfamiliar outdoor augmented reality environments. In *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 285–286. IEEE, 2007.

[4] B. Avery, C. Sandor, and B. H. Thomas. Improving spatial perception for augmented reality x-ray vision. In *2009 IEEE Virtual Reality Conference*, pp. 79–82. IEEE, 2009.

[5] R. Bane and T. Hollerer. Interactive tools for virtual x-ray vision in mobile augmented reality. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 231–239, 2004. doi: 10.1109/ISMAR.2004.36

[6] T. Blum, R. Stauder, E. Euler, and N. Navab. Superman-like x-ray vision: Towards brain-computer interfaces for medical augmented reality. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 271–272, 2012. doi: 10.1109/ISMAR.2012.6402569

[7] J. Brooke et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.

[8] W. G. Cochran. The comparison of percentages in matched samples. *Biometrika*, 37(3/4):256–266, 1950.

[9] K. Čopič Pucihar, P. Coulton, and J. Alexander. The use of surrounding visual context in handheld ar: device vs. user perspective rendering. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 197–206, 2014.

[10] A. Criminisi, P. Perez, and K. Toyama. Object removal by exemplar-based inpainting. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 2, pp. II–II. IEEE, 2003.

[11] J. Cui, P. Rosen, V. Popescu, and C. Hoffmann. A curved ray camera for handling occlusions through continuous multiperspective visualization. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1235–1242, 2010.

[12] O. J. Dunn. Multiple comparisons among means. *Journal of the American statistical association*, 56(293):52–64, 1961.

[13] N. Elmqvist, U. Assarsson, and P. Tsigas. Employing dynamic transparency for 3d occlusion management: Design issues and evaluation. In *IFIP Conference on Human-Computer Interaction*, pp. 532–545. Springer, 2007.

[14] M. T. Eren and S. Balcisoy. Evaluation of x-ray visualization techniques for vertical depth judgments in underground exploration. *The Visual Computer*, 34(3):405–416, 2018.

[15] M. Friedman. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the american statistical association*, 32(200):675–701, 1937.

[16] U. Gruenefeld, Y. Brück, and S. Boll. Behind the scenes: Comparing x-ray visualization techniques in head-mounted optical see-through augmented reality. In *19th International Conference on Mobile and Ubiquitous Multimedia*, MUM '20, p. 179–185. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3428361.3428402

[17] U. Gruenefeld, L. Prädel, and W. Heuten. Locating nearby physical objects in augmented reality. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*, MUM '19. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3365610.3365620

[18] K. Hasegawa and H. Saito. Diminished reality for hiding a pedestrian using hand-held camera. In *2015 IEEE International Symposium on Mixed and Augmented Reality Workshops*, pp. 47–52. IEEE, 2015.

[19] C. Henley, T. Maeda, T. Swedish, and R. Raskar. Imaging behind occluders using two-bounce light. In *European Conference on Computer Vision*, pp. 573–588. Springer, 2020.

[20] IBM Corp. IBM SPSS Statistics for Windows (Version 28.0). `https://www.ibm.com/products/spss-statistics`, 2021.

[21] S. Iizuka, E. Simo-Serra, and H. Ishikawa. Globally and locally consistent image completion. *ACM Transactions on Graphics (ToG)*, 36(4):1–14, 2017.

[22] Y. Kameda, T. Takemasa, and Y. Ohta. Outdoor see-through vision utilizing surveillance cameras. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 151–160, 2004. doi: 10.1109/ISMAR.2004.45

[23] Y. Kameda, T. Takemasa, and Y. Ohta. Outdoor see-through vision utilizing surveillance cameras. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 151–160. IEEE, 2004.

[24] K. Lilija, H. Pohl, S. Boring, and K. Hornbæk. Augmented reality views for occluded interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300676

[25] C. Lin and V. Popescu. Fast intra-frame video splicing for occlusion removal in diminished reality. In *International Conference on Virtual Reality and Mixed Reality*, pp. 111–134. Springer, 2022.

[26] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro. Image inpainting for irregular holes using partial convolutions. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 85–100, 2018.

[27] M. A. Livingston, A. Dey, C. Sandor, and B. H. Thomas. Pursuit of "x-ray vision" for augmented reality. In *Human Factors in Augmented Reality Environments*, pp. 67–107. Springer, 2013.

[28] T. Maeda, Y. Wang, R. Raskar, and A. Kadambi. Thermal non-line-of-sight imaging. In *2019 IEEE International Conference on Computational Photography (ICCP)*, pp. 1–11. IEEE, 2019.

[29] Q. McNemar. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2):153–157, 1947.

[30] C. Mei, V. Popescu, and E. Sacks. The occlusion camera. In *Computer Graphics Forum*, vol. 24, pp. 335–342. Amsterdam: North Holland, 1982-, 2005.

[31] C. Mei, E. Sommerlade, G. Sibley, P. Newman, and I. Reid. Hidden view synthesis using real-time visual slam for simplifying video surveillance analysis. In *2011 IEEE International Conference on Robotics and Automation*, pp. 4240–4245. IEEE, 2011.

[32] S. Mori, S. Ikeda, and H. Saito. A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects. *IPSJ Transactions on Computer Vision and Applications*, 9(1):1–14, 2017.

[33] S. Mori, F. Shibata, A. Kimura, and H. Tamura. Efficient use of textured 3d model for pre-observation-based diminished reality. In *2015 IEEE International Symposium on Mixed and Augmented Reality Workshops*, pp. 32–39. IEEE, 2015.

[34] V. Popescu and D. Aliaga. The depth discontinuity occlusion camera. In *Proceedings of the 2006 symposium on Interactive 3D graphics and games*, pp. 139–143, 2006.

[35] V. Popescu, P. Rosen, and N. Adamo-Villani. The graph camera. *ACM Trans. Graph.*, 28(5):1–8, dec 2009. doi: 10.1145/1618452.1618504

[36] V. Popescu, P. Rosen, and N. Adamo-Villani. The graph camera. *ACM Trans. Graph.*, 28(5):1–8, dec 2009. doi: 10.1145/1618452.1618504

[37] G. Queguiner, M. Fradet, and M. Rouhani. Towards mobile diminished reality. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 226–231. IEEE, 2018.

[38] P. Rosen and V. Popescu. The epipolar occlusion camera. In *Proceedings of the 2008 symposium on Interactive 3D graphics and games*, pp. 115–122, 2008.

[39] P. Rosen and V. Popescu. An evaluation of 3-d scene exploration using a multiperspective image framework. *The Visual Computer*, 27(6):623–632, 2011.

[40] P. Rosen, V. Popescu, K. Hayward, and C. Wyman. Nonpinhole approximations for interactive rendering. *IEEE Computer Graphics and Applications*, 31(6):68–83, 2011.

[41] M. Rosenthal, A. State, J. Lee, G. Hirota, J. Ackerman, K. Keller, E. D. Pisano, M. Jiroutek, K. Muller, and H. Fuchs. Augmented reality guidance for needle biopsies: An initial randomized, controlled trial in phantoms††a preliminary version of this paper was presented at the medical image computing and computer-assisted intervention

(miccai) 2001 conference in utrecht. the netherlands (rosenthal et al., 2001). *Medical Image Analysis*, 6(3):313–320, 2002. Special Issue on Medical Image Computing and Computer-Assisted Intervention - MICCAI 2001. doi: 10.1016/S1361-8415(02)00088-9

[42] A. State, M. A. Livingston, W. F. Garrett, G. Hirota, M. C. Whitton, E. D. Pisano, and H. Fuchs. Technologies for augmented reality systems: Realizing ultrasound-guided needle biopsies. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '96, p. 439–446. Association for Computing Machinery, New York, NY, USA, 1996. doi: 10.1145/237170.237283

[43] J. Swan, M. Livingston, H. Smallman, D. Brown, Y. Baillot, J. Gabbard, and D. Hix. A perceptual matching technique for depth judgments in optical, see-through augmented reality. In *IEEE Virtual Reality Conference (VR 2006)*, pp. 19–26, 2006. doi: 10.1109/VR.2006.13

[44] Unity Software, Inc. Unity Real-Time Development Platform (Version 2020.3.25). `https://unity.com/`, 2021.

[45] P. Vávra, J. Roman, P. Zonča, P. Ihnát, M. Němec, J. Kumar, N. Habib, and A. El-Gendi. Recent development of augmented reality in surgery: a review. *Journal of healthcare engineering*, 2017, 2017.

[46] E. Veas, R. Grasset, E. Kruijff, and D. Schmalstieg. Extended overview techniques for outdoor augmented reality. *IEEE transactions on visualization and computer graphics*, 18(4):565–572, 2012.

[47] J. D. Velazco-Garcia, N. V. Navkar, S. Balakrishnan, G. Younes, J. Abi-Nahed, K. Al-Rumaihi, A. Darweesh, M. S. M. Elakkad, A. Al-Ansari, E. G. Christoforou, M. Karkoub, E. L. Leiss, P. Tsiamyrtzis, and N. V. Tsekos. Evaluation of how users interface with holographic augmented reality surgical scenes: Interactive planning mr-guided prostate biopsies. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 17(5):e2290, 2021. doi: 10.1002/rcs.2290

[48] A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. G. Bawendi, and R. Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature communications*, 3(1):1–8, 2012.

[49] L. Wang, J. Chen, Q. Ma, and V. Popescu. Disocclusion headlight for selection assistance in vr. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 216–225. IEEE, 2021.

[50] L. Wang, H. Jin, R. Yang, and M. Gong. Stereoscopic inpainting: Joint color and depth completion from stereo images. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8. IEEE, 2008.

[51] L. Wang, J. Wu, X. Yang, and V. Popescu. Vr exploration assistance through automatic occlusion removal. *IEEE transactions on visualization and computer graphics*, 25(5):2083–2092, 2019.

[52] L. Wang, W. Wu, Z. Zhou, and V. Popescu. View splicing for effective vr collaboration. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 509–519. IEEE, 2020.

[53] L. Wang, H. Zhao, Z. Wang, J. Wu, B. Li, Z. He, and V. Popescu. Occlusion management in vr: A comparative study. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 708–716. IEEE, 2019.

[54] R. F. Woolson. Wilcoxon signed-rank test. *Wiley encyclopedia of clinical trials*, pp. 1–3, 2007.

[55] J. Wu, L. Wang, H. Zhang, and V. Popescu. Quantifiable fine-grain occlusion removal assistance for efficient vr exploration. *IEEE Transactions on Visualization and Computer Graphics*, 2021.

[56] J. Wu, L. Wang, H. Zhang, and V. Popescu. Quantifiable fine-grain occlusion removal assistance for efficient vr exploration. *IEEE Transactions on Visualization and Computer Graphics*, 28(9):3154–3167, 2022. doi: 10.1109/TVCG.2021.3053287

[57] M.-L. Wu and V. Popescu. Multiperspective focus+ context visualization. *IEEE Transactions on Visualization and Computer Graphics*, 22(5):1555–1567, 2015.

[58] M.-L. Wu and V. Popescu. Efficient vr and ar navigation through multiperspective occlusion management. *IEEE transactions on visualization and computer graphics*, 24(12):3069–3080, 2017.

[59] M.-L. Wu and V. Popescu. Efficient vr and ar navigation through multiperspective occlusion management. *IEEE transactions on visualization and computer graphics*, 24(12):3069–3080, 2017.

[60] M.-L. Wu and V. Popescu. Anchored multiperspective visualization for efficient vr navigation. In *International Conference on Virtual Reality and Augmented Reality*, pp. 240–259. Springer, 2018.

[61] M.-L. Wu and V. Popescu. Rgbd temporal resampling for real-time occlusion removal. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, pp. 1–9, 2019.

[62] M.-L. Wu and V. Popescu. Rgbd temporal resampling for real-time occlusion removal. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, pp. 1–9, 2019.

[63] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang. Generative image inpainting with contextual attention. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5505–5514, 2018.

[64] S. Zollmann, R. Grasset, G. Reitmayr, and T. Langlotz. Image-based x-ray visualization techniques for spatial understanding in outdoor augmented reality. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design*, pp. 194–203, 2014.