REVIEW PAPER

Open Access

CrossMark

A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects

Shohei Mori[1*], Sei Ikeda[2] and Hideo Saito[1]

Abstract

In this paper, we review diminished reality (DR) studies that visually remove, hide, and see through real objects from the real world. We systematically analyze and classify publications and present a technology map as a reference for future research. We also discuss future directions, including multimodal diminished reality. We believe that this paper will be useful mainly for students who are interested in DR, beginning DR researchers, and teachers who introduce DR in their classes.

Keywords: Diminished reality, Mediated reality, Mixed reality, Augmented reality, Image-based rendering, Camera pose estimation, Object detection, Object recognition, Image inpainting, Survey

1 Introduction

Diminished reality (DR) is a set of methodologies for concealing, eliminating, and seeing through objects in a perceived environment in real time to diminish the reality. This technique is different from augmented reality (AR) and mixed reality (MR) [1–3] that superimpose virtual objects on the real world to enhance reality (Fig. 1). In AR/MR, virtual objects are newly placed among real objects or real objects are extended with virtual objects. For example, a building is annotated with a virtual billboard or a building is extended with its virtual floors. In the former case, real and virtual objects individually exist in the environment. Therefore, in general, observers do not observe contextual seamlessness between real and virtual objects. In the latter case, however, no apparent gap between real and virtual objects is acceptable; otherwise, real–virtual borders will appear as visual inconsistency. In a similar manner to this, in DR-specific scenarios (e.g., removing real objects), observers assume that there is no apparent gap between the real and virtual scenes because the virtual scene is the reconstruction of real objects unobservable from the observers. That is, most of DR research faces to scenarios in which this consistency at

the real–virtual boundaries is considered important. As with AR/MR, such inconsistencies can be summarized as geometric, photometric, and temporal issues.

The word "diminish" means to "make or become less" or "cause to seem less impressive or valuable," and the meaning recalled from the current DR, such as "removing, alternating, or seeing through" real objects, is not originally included. In the 1990s, DR was defined by Mann in the concept of mediated reality including AR, MR, and DR [4]. In the early DR literature, some examples were shown in which lowering the saturation of some areas to force an observer to face the other regions [4], and virtual objects overwrote undesirable real objects to hide the real information [5]. In the 2000s, Zokai *et al.* proposed a method for removing industrial pipes visually [6]. In this work, they recovered a hidden background at the main view using photos observed at different views to cover the real pipes to remove them visually, and they represented this approach as a DR methodology. Since then, methods for visually removing objects have been considered one DR methodology. From the same point of view, in the 1990s, medical AR/MR attempted to visualize the inside of a patient's body from outside by superimposing computerized tomography (CT) or endoscopic images on AR/MR displays. These techniques can also be categorized as a DR methodology [7, 8]. In the 2010s, a real-time image-inpainting framework was proposed and considered a DR

*Correspondence: mori@hvrl.ics.keio.ac.jp
[1]Keio University, Yokohama, Japan
Full list of author information is available at the end of the article

Mori *et al. IPSJ Transactions on Computer Vision and Applications* (2017) 9:17

Page 2 of 14

Fig. 1 Real, augmented, and diminished scenes. This figure shows differences between real (**a**), augmented reality (**b**), and diminished reality scenes (**c**). In augmented reality, real–virtual inconsistency mainly appears where real and virtual objects contact each other, i.e., geometric registration of virtual objects is one of the most important issues. From this point of view, in DR, real-virtual borders will appear all around the virtual regions since real regions surround the virtual ones

methodology, although image inpainting had been considered a computationally expensive process to be run in real time because the method repetitively searches and composes image patches to fill in missing regions [9, 10].

In this paper, the DR methodologies published during the 20-year period from Mann's publication in the 1990s to the present are reviewed. We especially focused on a comprehensive survey, organization, and analysis of "interactive" methods. However, as real-time performance depends on the machines used, the literature claiming to be DR methods are introduced even if they could not process the method in real time at the time of publication. This paper is a revised version of a paper published in Japanese in 2011 [11]. In this paper, we added discussions of new publications up to 2017 and improved the descriptions of the implementation framework and classification to more general descriptions. To the best of our knowledge, this paper is the first international DR survey paper.

2 Basic functions and usage

DR technology is used to implement *diminish*, *see-through*, *replace*, and *inpaint* functions. Figure 2 shows example DR results from existing publications.

Diminish Degrade visual functions for a certain purpose. For example, the color information of a visual field (i.e., the acquired image) is thinned out or distorted. This corresponds to acquiring, editing, and presenting the light rays jumping into the eyes via see-through head-mounted display (HMD) and reality mediator (i.e., a computer [4]). This representation can be used to help and understand individuals with visual impairments [4].
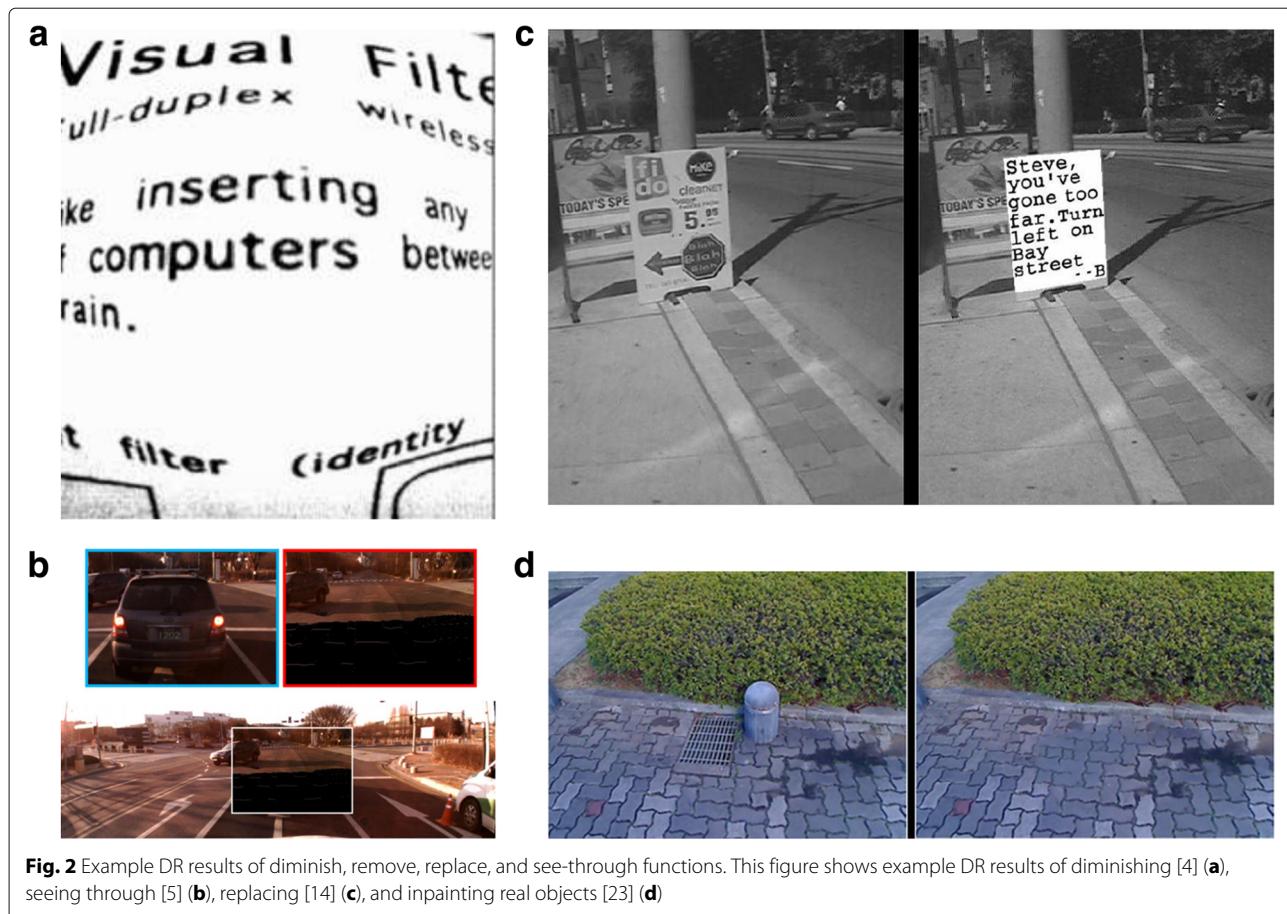
See-through: Cover real objects with images of their occluded background to make the objects virtually invisible in our vision. This process is equivalent to replacing light rays from the real object with light rays from the background to reproduce the scene in which real objects

do not exist in our visual field through HMDs. This process can be implemented in the same manner as the virtual object overlay in AR/MR (e.g., virtual objects of reconstructed backgrounds are overlaid in a perceived vision). This process is used to remove a person from Google Street View pictures to protect his or her privacy [12], to remove a person in a video [13], to remove a vehicle in front of the driver [14], to remove a baseball catcher to visualize the view of the pitcher from a view behind the catcher [15], and to generate a panoramic stroboscopic image [16, 17].

Some literature discusses a way of representing occluding objects as semi-transparent in a mixture with alpha blending or the like. This visualization technique is called AR X-ray vision, see-through vision, or ghosted views. These literature discuss the reasonableness of the representations in terms of visibility and depth perception. Semi-transparent representation is useful for seeing through car interiors [18] and walls [19].

Replace Overlap a real object with a virtual object so that the real object appears to be replaced by the virtual object. In other words, light rays from the real object are blocked by the superimposed virtual object in a view [5, 20]. To fully cover the real object, one has to prepare a virtual object with the same or larger size than the real one in the perspective. Alternatively, the virtual object overlay can be performed after the see-through process to replace the real object with a smaller virtual object. For example, a signboard with unnecessary information is hidden with a useful virtual signboard [5]. The removal and then replacement methodology will improve the quality of the AR/MR rebuilding and landscape simulation in which old buildings are completely replaced with new virtual ones.

Inpaint Generate plausible background images based on the surroundings. This technique can provide similar results to see-through but the background image is a

Fig. 2 Example DR results of diminish, remove, replace, and see-through functions. This figure shows example DR results of diminishing [4] (**a**), seeing through [5] (**b**), replacing [14] (**c**), and inpainting real objects [23] (**d**)

synthesis of pixels or image patches in regions surrounding the object to be removed in user perspective. In other words, light rays from the real object are replaced with light rays generated based on surrounding light rays. Thus, there is no guarantee if the background is as it is. On the other hand, this technique can eliminate obstacles with no background (e.g., drawings on a wall and manholes on a street). Most of existing work assumes that the background is planar [9, 10, 21] because image inpainting is potentially a screen space image processing. The recent work can handle curved [22] or multiple surfaces [23].

3 Implementation procedures and terminology

Most existing DR systems are video see-through (VST) systems. Therefore, this chapter describes implementation procedures for diminishing, seeing through, replacing, and inpainting methodologies of VST DR. Table 1 summarizes the relation between each procedure of typical DR and AR/MR methods.

1. **Background observation** Observe backgrounds to acquire background information for see-through. Various types of images are used as follows: Internet photos [13], X-ray images [8, 24], image sets actively

Table 1 Typical diminish, see-through, replace, inpaint, and augmentation (AR/MR) procedures

	Diminish	See-through	Replace	Inpaint	Augmentation
Background observation	×	✓	×	×	×
Scene tracking	×	✓	✓	✓	✓
Detection of region of interest	✓	✓	✓	✓	×
Hidden view generation	×	✓	×	✓	×
Composition	✓	✓	✓	✓	✓

captured in advance [25, 26], streaming image sequences from surveillance cameras [27], multi-view cameras [6, 28–30], and RGB-D cameras [31, 32]. One could store a sufficient number of viewpoints when one carefully captures a scene during enough time [13, 25, 26]. However, one will suffer from the geometric or photometric differences between the real and virtual scenes in the DR results due to the intervals between the preliminary observation and the current DR experience. In the case of real-time observation that can handle dynamic backgrounds [6, 27, 29–32], because the number of installed cameras is physically limited, the number of viewpoints is sometimes insufficient. In addition, temporal inconsistency occurs when camera synchronization is not performed properly. In both observation methods, differences between the camera optics cause photometric inconsistency. These inconsistencies can be modified in the *composition* stage.

2. **Scene tracking** Estimate the camera pose of the current frame or objects of interest. Most DR methods require the camera pose or the relative pose of the camera to the scene as an input to recover the hidden background in the current view. Positioning sensors [33, 34], visual simultaneous localization and mapping (vSLAM) methods [14, 23, 27, 35, 36, 43], pre-calibration [6, 37], and fidicial markers [31, 38] are available.

3. **Region of interest detection** Determine the region of interest (ROI) specifying a target object to be diminished in a view. The ROI is a mask image to be covered with a recovered hidden background image in the next procedure. The ROI is not necessarily the silhouette of the target object and can be a bounding box that convert the object roughly and sufficiently. However, minimizing the ROI will reduce unnecessary artifacts while such rigorous edges may make the hidden area conspicuous. Geometric models of the target objects to be removed [25], simple 3D bounding boxes [6, 26], and image recognition techniques [13, 39] are used to detect the ROI. Several methods place the target objects in a non-reconstruction area not to be appeared in the resulting image [30, 40].

4. **Hidden view generation** Recover a hidden background in the ROI based on *background observation* for see-through or generate plausible images for inpaint. In see-through, image-based methods are preferred for seamless real–virtual composition. Homography transformations of images [13, 19, 35, 36, 41, 42], rendering texture 3D models [6, 31, 43], rendering voxels [37], and

image-based rendering (IBR) [25, 26, 30, 44] are the major approaches. In inpaint, pixels or image patches are synthesized in the ROI to fill in the region with plausible pixels [9, 10, 21, 38].

5. **Composition** Make the real–virtual boundary less noticeable by post-processing or overlaying other effects on the DR image. Poisson blending-based methods [13, 38] and pixel intensity estimation from surrounding pixels [26] are used for seamless real–virtual composition. Similarly, alpha blending at real–virtual boundaries is known to be effective [25, 26]. For the *replacing* process, AR overlay is accomplished [20]. For the see-through process, real–virtual images are blended with an alpha blending [45], and the edges or the saliency map of the real foreground image are overlaid to improve visibility and depth perception [19, 46–48].

4 Classification by procedures

We classify existing methods according to the five procedures described in the previous section. Table 2 shows the procedures and their representative methods.

4.1 Background observation

To diminish a real object from a perceived view, background information hidden from the viewer is necessary to exchange the object with the background information. A direct approach is to observe the background from a different viewpoint or beforehand, and therefore, the observation approaches vary depending on the situation. In this section, we introduce four approaches: pre-observation, active self-observation, real-time observation with additional cameras, and their combinations.

4.1.1 Pre-observation

In some situations, one can create background image datasets before the target objects to be diminished are placed in the environment. In this case, we can expect high-quality DR results because a sufficient number of viewpoint images are carefully captured. However, the geometric and photometric differences between the real and virtual regions must be handled in *composition* stage due to the time intervals between the pre-observation and the DR experience.

Mori et al. restored hidden background images from multi-view perspective images preserved by the user in advance [26]. Cosco et al. removed a haptic device PHANToM on a desk using multi-view photos captured before the device was placed [25, 49].

Takemura et al. proposed a method for restoring the line of sight of the other party wearing an HMD by removing the HMD in the perspective [50, 51]. In this case, the hidden background is the face hidden by the HMD, and therefore, they used pre-captured

Table 2 Procedures and their representative methods

Procedures	Sub-procedures	Representatives
Background observation	Pre-observation	[13, 25, 26, 49–51]
	Active self-observation	[12, 16, 17]
	Real-time observation	[6, 14, 27, 29–33, 40, 41, 46]
	Combination	[19, 27, 45]
Scene tracking	Fixed viewpoint	[6, 52]
	Constrained viewpoint	[16, 17]
	Free viewpoint	[9, 10, 13, 14, 21–23, 25, 26, 35, 36, 38, 41, 43, 49, 107]
Region of interest detection	Overlay without detection	[41]
	Manual detection	[6, 29, 33, 34, 37, 40, 48]
	(Semi-)Automatic detection	[9, 10, 13, 14, 16, 17, 21–23, 25, 26, 31, 38, 39, 46, 47, 49, 59]
Hidden view generation	Homography	[13, 19, 35, 36, 42]
	3D reconstruction	[6, 14, 29, 31, 40, 43]
	Image-based rendering	[25, 26, 30, 49, 66, 67, 107]
	Inpainting	[9, 10, 21–23]
Composition	Seamless blending	[13, 25, 26, 31, 38, 49]
	Semi-transparent representation	[33, 34, 45–48, 71–77]
	AR overlay	[4, 20, 25, 49, 78]

angle-dependent images. Li et al. deleted a person in a video sequence by superimposing an image from Internet photo collections selected based on GPS data [13]. They assumed that a sufficient number of photographs taken close to the current viewpoint exist on the web because the method is used at a sightseeing spot.

4.1.2 Active self-observation

When we have no proper background image dataset, we can still observe the backgrounds by moving the viewpoint or waiting for the objects to move so that the initially occluded backgrounds can be observed with a certain time difference but have an advantage that the background observer and the user have the same camera optics. However, the background in contact with the target objects may not be observed and therefore remain unknown.

Flores et al. removed a person from Google Street View images taken with spatial-time differences using the camera attached to a vehicle [12]. Hasegawa and Saito

removed a person from a panoramic image by stitching multiple images recorded during user panning [16, 17]. In this case, a hidden region in a frame is observable in the previous frames.

4.1.3 Real-time observation

Some methods surround the environment with additional cameras to observe occluded backgrounds in the main view. These approaches can acquire background information in real time and provide the current state of the hidden area in the main view. In this case, sharing coordinate systems and time stamps between the cameras is required for providing images.

Zokai et al. used two additional cameras as hidden background observers to erase pipes in a factory in the main view [6]. Kameda et al. [46] and Mei et al. [27] used multiple video cameras behind walls (e.g., surveillance cameras) to see through views occluded by the walls. Enomoto et al. assumed that multiple users with handheld cameras exist in the environment and they observe the backgrounds for the others [41]. In an AR X-ray system for see-through walls [33], a camera-equipped remote control robot was used to observe the hidden background. Rameau et al. used a stereo camera attached to a front vehicle to see through a front vehicle [14]. These methods require all cameras to capture a common area to calculate the relation between the images.

Multi-view cameras are often used for acquiring 3D structure and surface colors of hidden backgrounds [29, 40]. However, multi-view based 3D reconstruction is time-consuming. Thus, Meerits and Saito used an RGB-D observer camera for fast 3D reconstruction of hidden backgrounds as 3D polygon mesh in real time [31]. Ienaga et al. reported an example implementation of a multiple RGB-D camera system to remove the viewer's body in an AR-based mirror system to improve the efficiency of teaching anatomy [32].

Further, multi-view cameras are used for constructing light fields. Mori et al. constructed light fields with a real-time multi-camera system and removed a viewer's hand from the perspective to visualize the viewer's workspace occluded by his or her own hand [30].

4.1.4 Combination

The observation methods described so far can be used together. For example, we can implement a method that removes an object on a plane with an observation-type method and fills the remaining regions with an inpaint method. We can use the current images fetched from a real-time observation-type method to map them on the scene geometry estimated using a pre-observation-type method. This approach will reduce the geometric and photometric inconsistencies.

Mori *et al. IPSJ Transactions on Computer Vision and Applications* (2017) 9:17

Page 6 of 14

Barnum et al. [19] took a similar approach. In their method, an acquired background image is divided into two planes related to a moving object and a non-moving object behind the moving object. Then, they replaced the background image with the pre-observed image. It should be noted as well that they used a relay camera for the main and background observer cameras to create a common region between the two, and therefore, these two cameras can be placed far from each other [19]. Sugimoto et al. switched multi-view RGB-D cameras to cover a wide range of the real-time background observation, and still, unobservable areas are compensated from the past frames [45]. Mei et al. used a 3D map constructed using vSLAM beforehand to localize the current camera and overlaid a real-time video of a surveillance camera in the current view [27].

4.2 Scene tracking
DR methods require calibration of the main camera and the background observer cameras. In addition, one can superimpose virtual objects interactively by estimating the camera pose of the main camera, as in AR/MR.

4.2.1 Fixed viewpoint
Assuming fixed cameras, one can perform calibration to calculate the relative positions and orientations of the cameras in advance (e.g., Zokai et al. [6] and Bayart et al. [52]).

4.2.2 Constrained viewpoint
Allowing some degrees of freedom in camera motion requires tracking the target objects to be diminished or estimating the camera motion (e.g., panning motion [16, 17]).

4.2.3 Free viewpoint
Almost all of the existing DR methods allow six degrees of freedom (6DoF) motion for the main camera using positioning sensors [33, 34], fiducial markers [22, 25, 41, 49], model-based tracking [13, 26], vSLAM [14, 23, 27, 35, 44, 53], etc. Cosco et al. used ARToolKit and ARToolKitPlus markers to track the main camera at pre-observation and run-time for indoor desktop a DR scenario [25, 49]. Enomoto et al. used fiducial markers, ARTag [54], for registering all cameras in a unified coordinate system, and therefore, each camera can move in 6DoF while the marker is visible in the view [41]. Herling and Broll used an object tracker that continuously detects and tracks the ROI to fill in the ROI with pixels using inpaint [9, 10]. If the target object to be diminished is a marker [21, 22, 38] or a marker-attached object [31], then the fiducial markers are used to estimate the camera pose.

For achieving DR in an arbitrary scene, vision-based tracking methods are considered more feasible than artificial markers that should be diminished in the user's perspective too. Mori et al. calculated a camera pose by solving a perspective-n-point problem using 3D–2D correspondences of the features between the current image and the image-based rendering image of a scene [26]. Li et al. used the previous and current homography estimation to smooth the temporal error and thus reduced the jitter [13]. vSLAM systems such as PTAM [55] and KinectFusion [56] are a typical option for estimating the camera poses in an arbitrary environment [35, 36, 43]. For example, Kawai et al. used vSLAM, PTAM, to achieve inpaint under 6DoF camera motion in 3D scenes [23]. Mei et al. used an old vSLAM map to estimate camera motion at run-time in see-through. Rameau et al. proposed a method to match front car local 3D map against the rear image to synthesize the front car image at the rear car perspective [14]. As the 3D local map is updated in real time, the driver can see through roads without the front car in real time.

4.3 Detection of the region of interest
The regions of the target objects to be diminished are determined to fill in the regions with the estimated background image for see-through or plausible image generated using inpaint in the perspective. The occlusions of the target and the other objects must be managed in this step. The removal target area is traced as close to the silhouette of the target object as possible on the screen. Because this removal target area varies due to the viewpoint changes of the main camera and movement or deformation of the removal target object, it is necessary to detect, recognize, and track the area frame by frame. Limiting the ROI can avoid affecting the main camera image with the incomplete reconstruction results, including unnecessary artifacts or can reduce the processing cost. Discontinuities are often observed at the boundaries between the ROI and its surroundings.

4.3.1 Overlay without detection
The method proposed by Enomoto et al. does not estimate a specific ROI for removing target objects because the entire images of the other cameras are projected to the perspective [41]. Although this approach can reduce computational cost of the ROI detection, unnecessary artifacts appear potentially around the target objects.

4.3.2 Manual detection
There are cases where explicit or automatic ROI detection is not required. For example, in the case of [6], the cameras and the target objects are fixed in the environment and the ROI can be set manually. It is not a problem to estimate target areas that are narrower than the actual object area to see part of the wall [33, 34, 48]. Jarusirisawad and Saito ignored their target objects from a projective grid

space (PGS) for their plane sweep algorithm-based 3D reconstruction [29, 40]. They proposed another method for removing a person described by voxels, and they determined the ROI by manually segmenting and labeling the target voxels in a video sequence [37].

4.3.3 (Semi-)Auto detection

When the geometric shapes of the target object are known, the corresponding ROI can be determined by projecting the geometric model on the perspective. Such geometric model data is obtained with manual modeling (e.g., computer-aided design (CAD) data [46, 47]) or automatic structure-from-motion (SfM) modeling software. In the case of a visuo-haptic AR [25, 49], an articulated haptic device, PHANToM, is surrounded by several 3D bounding boxes, and each box position is calculated based on joint angles from the device. In most cases, slightly larger bounding boxes are preferred to sufficiently surround the target objects in the view [25, 26, 31, 49]. Rameau et al. pointed out that vision-based tracking system is still computationally expensive and not robust enough in some cases. Thus, they proposed to use a cuboid to surround a front car of interest assuming that the car pose is estimated at every frame [14].

Lepetit et al. used a semi-automatic segmentation technique [57]. They manually segmented the ROI only in key frames, and the ROIs in the other frames were automatically estimated using a vision technique [39]. Li et al. [13] tracked a person in subsequent frames using comprehensive tracking [58] by which the rectangular region specified at the initial frame was continuously tracked.

Yokoi et al. [59] eliminated a lecturer in a lecture video by segmenting the target region using frame–frame differences and the graph cut method [60]. When a target object is a plane marker, the ROI can be automatically determined from the marker information [21, 22, 38]. Hasegawa and Saito used HOG-SVM [61] and Kalman filter [62] to automatically determine an area of a moving person to be eliminated [16, 17].

In inpaint methods, an object of interest is tracked frame-by-frame [9, 10, 23]. Herling and Broll first used classic Snake algorithm [63] to track contours of an object of interest [9] and improved their ROI detection and tracking algorithms for video usages [10]. Their method [10] detects contours of an object of interest based on manually guided footprints in a screen space and tracks them based on homography. In a method proposed by Kawai et al., the user first draws ROI manually in a screen and the ROI determined using 3D scene points fetched from vSLAM [55] is continuously tracked [23]. Kawai et al. present an optional manual ROI cropping procedure because the tracked ROI might include unnecessary part of the scene at the different viewpoint from the initial one.

4.4 Hidden view generation

The synthesized background image must follow the camera motion (i.e., backgrounds must be recovered in 3D). When the backgrounds exist far from the main camera, viewpoint changes do not drastically change the background appearance. In this case, we may approximate this background to a plane or camera motions to rotation-only movement.

4.4.1 Homography

Barnum et al. used two types of homography for handling erratic transformations caused by planar approximation of the objects [19]. Li et al. also used homography to transform images fetched from the web and selected the closest image from the shooting locations [13]. Some tablet-based approaches use *homography* to transform a rear camera image on a tablet to the user's perspective [35, 36, 42] (see the "5.1.3" section for further details).

4.4.2 3D reconstruction

Some methods explicitly extract 3D geometry or depth map of the backgrounds to handle 3D objects in the environment. Zokai et al. used a stereovision technique and approximated the background as a set of multiple planes [6]. Rameau et al. also used a stereovision technique to generate a dense depth map to warp a color image to the main viewpoint [14]. Some methods used multiple cameras to reconstruct backgrounds in PGS using the plane sweep algorithm [29, 40]. Meerits and Saito proposed a graphics processing unit (GPU) processing framework for real-time polygon meshing from depth frames obtained with an RGB-D camera [31]. Baričević et al. reconstructed the 3D scene geometry for transforming rear camera image of a tablet to the user's perspective [43], although the other tablet-based approaches approximate this transformation as *homography* [35, 36, 42].

4.4.3 Image-based rendering

AR/MR methods superimpose arbitrary computer graphics into a real scene while see-through DR methods overlay a synthetic image recovered from observations of a real scene. Therefore, image-based approaches are considered effective. Cosco et al. performed view-dependent texture mapping (VDTM) [64], and for this, they manually built polygon mesh of the environment and a set of pairs of a pre-captured image and its location measured using AR markers in the environment [25, 49]. A pre-observation approach proposed by Mori et al. performed unstructured lumigraph rendering [65] using the structures and images acquired with SfM [26]. There is an example of removing objects using light fields from pre-calibrated multi-view streaming [30]. Synthetic aperture photography (SAP) makes captured foregrounds virtually

Mori *et al. IPSJ Transactions on Computer Vision and Applications* (2017) 9:17

Page 8 of 14

deblurred and invisible by simulating a large aperture camera using regularly arranged cameras [66, 67].

4.4.4 Inpainting

When hidden areas cannot be observed at all, we have no choice except to compensate the background from the surrounding pixels without any background observations. Such methods are referred to as image inpainting, image completion, or video inpainting. Korkalo deleted an AR marker using an inpainting method [68], and this was a pioneering work in this DR area [21].

In general, image-inpainting processing is difficult to achieve in real time, and various attempts have been made (e.g., PatchMatch [69]). Based on the idea of PatchMatch's image patch search method, Herling and Broll implemented a real-time inpainting process that fills in the ROI on a plane with image patches of the surroundings [9]. They proposed a real-time image-inpainting algorithm based on appearance and spatial cost functions, heuristic optimization of the cost functions, and multi-resolution optimization [10]. Kawai et al. proposed a method for simultaneously executing processes related to inpainting and the others using multi-threading [23]. Because image inpainting is an algorithm implemented in the image space, applying the algorithm to a 3D scene is difficult. Kawai et al. proposed a method for segmenting scenes into multiple planes using a point cloud reconstructed with a vSLAM and executed image inpainting on each plane [23]. They also extended an image-inpainting algorithm to inpaint a marker on a deformed surface [22].

4.5 Composition

Inaccurate background recovery results in clear boundaries between the ROI and the other region so that such apparent gaps are absorbed in this step. Applying the AR processes that present illumination and post processes for seamless overlay of virtual objects will improve DR results. On the other hand, many *see-through* DR literature tends to investigate more computationally efficient approaches focusing on compensating real–virtual boundary in screen space (i.e., 2D space) which appears all boundary on ROI. Thereafter, semi-transparent representation is performed to improve user depth perception. In addition, AR/MR objects are overlaid to perform *replace* if necessary.

4.5.1 Seamless blending

To disambiguate the gaps in the regions around the target objects, alpha blending provides a computationally cheap and sufficient solution [25, 26, 49]. Poisson blending-based techniques are computationally expensive but provide a reasonable solution [13, 31, 38]. Li et al. used a mean-value coordinate [70] that limits the ROI to a rectangle to speed up the blending process.

4.5.2 Semi-transparent representation

Some existing work does not remove target objects but renders them semi-transparent to the present foreground and the background at the same time. These techniques are called see-through vision [46, 47], AR X-ray [33, 34, 48, 71–73], and ghosted views [74]. These representation methods will be useful for avoiding the danger of a collision with the diminished objects.

Although these names are different, most of these studies mainly focus on improving and analyzing the depth perception for better spatial understanding. Sugimoto et al. used a simple alpha-blending approach to show their occluding robot arm semi-transparent [45]. Tsuda et al. analyzed various methods for see-through representations, such as wireframes, bird's eye view, and the combinations in a see-through vision framework. They evaluated whether the observer can intuitively grasp the space and reported the best combination [47]. Avery et al. pointed out with a simple transparent representation that information on the wall is lost and that this representation causes problems in the depth perception [34]. Therefore, they proposed a method to show the foreground edges and the background image at the same time to improve the depth perception. Otsuki et al. presented a random dot-based see-through vision, Stereoscopic Pseudo-Transparency, and considered the random dot patterns [75]. Buchmann et al. performed a visibility evaluation of effects for transparency changes on worker's hands at the perspective during block-stacking tasks [76]. Fukiage et al. proposed a framework for optimizing the transparency of each pixel in a superimposed virtual object [77].

4.5.3 AR overlay

For the replace process, the virtual objects are overlaid on a diminished image. There are example reports of covering up a signboard with a virtual one [4] and of overwriting a real object with a larger virtual object [20]. Cosco et al. overlaid virtual tools [25], replaced the user's hand [49] in their visuo-haptic system, and examined effectiveness of their system in terms of user's task performance [49].

In addition, there is a report of a demonstration system for removing furniture with an image-inpainting method and then replacing it with the prepared virtual furniture [78].

5 Classification by devices

We discuss displays and imaging devices in DR. Figure 3 shows example photos of see-through-based, projection-based, and tablet-based DR systems.

Fig. 3 See-through-based, projection-based, and tablet-based DR systems. This figure shows see-through-based [103] (**a**), projection-based [82] (**b**), and tablet-based [35] (**c**) DR systems

5.1 Display device

Most of the display devices used in DR are the see-through based while several publications showed projection-based and tablet-based systems.

5.1.1 See-through based

All of the DR systems introduced so far are VST-type systems. Optical see-through (OST) systems in DR have been lagging the technical difficulties of "shutting off light rays from the real object" that is the preconditioning procedure of DR. In other words, we also expect the emergence of a reasonable solution for occlusion problems of OST display in AR/MR [79].

5.1.2 Projection based

In VST DR, a background image is digitally superimposed on an observer's image to hide target objects, whereas in projection-based DR, the target objects are hidden by literally projecting the hidden background image with the projector to physically diminish the objects.

Seo et al. implemented a projector system to remove textures on a Lambertian plane considering geometric and photometric matching of projector lights and the surface [80]. Bonanni et al. proposed a kitchen system using AR technology, and as one of the functions, they

implemented a mechanism to show the inside of a refrigerator by projecting the image on the door [81]. Iwai et al. proposed Limpid Desk, a system for locating documents of interest from ones stacked and scattered on a desk. When a user touches a certain document in the document group, the documents become transparent to show the target document. They examined interactivity and transparent representations to help users recognize easily the overlap of the documents [82]. Inami et al. proposed a transparent haptic device [83] and suits (optical camouflage suits) [84] using a projector camera (pro-cam), and Yoshida et al. proposed a pro-cam-based transparent cockpit to see through a car body from its inside [18].

5.1.3 Tablet based

In MR/AR/DR with tablets as display devices, the tablets themselves are also objects to be diminished with a DR technique, because the camera image displayed on the tablet is not geometrically consistent with the real scene from the user's perspective [3]. In order to decrease the inconsistency, facing planes [42], arbitrary plane [35, 36], non-planar [43, 44], and simplified versions [85, 86] of these methods have been developed. It has also been shown that users prefer this display method [87] and the

Mori *et al. IPSJ Transactions on Computer Vision and Applications* (2017) 9:17

Page 10 of 14

efficiency of the search operation is improved in some cases [88, 89].

5.2 Image sensor

Most DR systems use color cameras only. We introduce DR methods that use special sensors, such as RGB-D (color camera and rangefinder), and medical instruments, such as endoscopy, ultrasound, and X-ray.

5.2.1 RGB-D camera

RGB-D cameras are also used for reconstructing the background in some work [31, 45]. RGB-D cameras help extract scene geometry explicitly. What is important in such systems are the real-time issues of 3D scene reconstruction from RGB-D images, frame-by-frame data transfer of large amounts of geometric information, and perspective transformation of geometric data.

5.2.2 Medical appliance

In the medical field, researchers have used images captured with an endoscope (laparoscope). Fuchs et al. installed a projector that emits structured lights on a body, and they observed the lights with a color camera attached to an endoscope to acquire the 3D surface of the cavity inside the body. The reconstructed image was presented on the body surface [90]. Mourgues et al. proposed a method for reconstructing a 3D cavity inside the body using a stereoendoscope to visually remove the medical device [28]. Similarly, in the medical field, there are many examples of the use of special image sensors. A see-through method of superimposing an ultrasound image on a patient's body [7, 91–97] and a similar method using X-ray images [8, 24] have been proposed and tested.

6 Evaluation method

It is necessary to evaluate the benefits and disadvantages of various DR methods. In this section, we introduce quantitative and qualitative evaluations in DR.

6.1 Quantitative evaluation

If we consider a DR method for generating background images within an ROI according to the current viewpoint, we can evaluate the method separately for each function, such as camera tracking, detecting or recognizing objects, and generating arbitrary viewpoint images. However, to assess whether a resulting image is correct or not, we need the ground truth in DR. For example, the ground truth of the see-through process is the pair of an image sequence of an object to be removed and one without the object in the same spatial-temporal conditions. Focusing on a static background and a static target object, we can acquire such a set of image sequences with a robot arm under a fixed illumination

condition [98]. However, assuming dynamic backgrounds or target objects, it is impossible to acquire such ground truth without guarantees of the repeatability of the backgrounds and targets. Therefore, in such a case, using computer graphics or image sequences of real scenes composited with computer-generated obstacles is one solution.

6.2 Qualitative evaluation

However, one may consider when the correct answer of the DR processing result cannot necessarily be defined or it is not necessary to be a correct answer. For example, in a case of deleting a manhole on a road, the required quality depends on the purpose of the DR process. It is unclear to us whether the pipe system under the manhole should be visualized or asphalt without manholes should be reproduced. Therefore, in such the case, it is necessary to perform a user study to evaluate whether or not the implemented DR method was able to output a visually convincing result. These evaluation methods are often used in literature related to see-through processing [33, 34, 46–48, 71–74], and they are helpful.

The image quality assessment (IQA) method is used in the area of image inpainting that cannot have ground truth images. For example, the method for measuring the gaze amount before and after image processing [99] and the method that uses a saliency map [100, 101] are well-known. Under the same purpose, a method that uses image features has also been proposed [102].

7 Future directions

In this section, we discuss issues that are not limited to vision matters that have not been achieved in the previous DR studies.

7.1 Multi-view calibration

Most DR methods that use multiple viewpoints are premises that cameras are fixed in space. The use of fiducial markers is one solution to this problem although the markers themselves should also be targets to be removed in DR. Allowing cameras to move freely in the 3D space makes a more flexible DR system; we need to tackle challenges known to be difficult in computer graphics and computer vision areas (e.g., artifacts accompanying viewpoint changes in IBR, real-time synchronization, and online multiple camera calibrations).

7.2 Head-mounted displays and binocular stereo

HMDs are actively used in AR/MR, while in DR there are a few use cases only in entertainment systems [103]. OST DR is virtually an unexplored area. Therefore, binocular stereo in DR is one of the unexplored areas although we have an example [104].

7.3 Multimodal DR

In this paper, we focused on extracting basic elements in existing research and classification of the literature only on visual DR which is common in existing DR research. Vision occupies the majority of human senses, but we should discuss similar issues in DR to various modalities that are important in AR/MR [3].

We can filter specific sounds in a frequency domain and therefore will be able to remove a sound from the original sound if digitally recorded sounds are available (cf., noise canceling technique). However, sound waves are felt not only from the ear but also throughout the body as bone conduction and vibration. Therefore, eliminating sound waves is difficult. Likewise, erasing sounds from a specific location is difficult because the sound image localization performance in a 3D space is lower than that of vision. We found a study in which sounds pass through walls after a see-through vision process [105].

In addition, Sawabe et al. proposed a method for diminishing human movement sensing using vection and evaluated its effect [106]. Although augmentation on haptic sensation and taste has been studied, there are no related examples in DR.

8 Conclusion

In this paper, we surveyed diminished reality techniques to visually remove, hide, and see through real objects from the real world. We systematically classified and analyzed publications and presented a technology map as a reference for future research. We also discussed future directions, including multimodal diminished reality. We hope that this paper will be useful mainly for students who are interested in DR, beginning DR researchers, and teachers who introduce DR in their classes.

Authors' contributions
SM collected and categorized publications and built the main structure of this review paper as the primary contributor. SI carried out the categorization, particularly regarding display devices. HS participated in the structural design of this paper and helped to draft the manuscript. All authors read and approved the final manuscript.

Competing interests
As described in the first section, this paper is a revised version of a paper published in Japanese in 2011 [11]. In this paper, we added discussions of new publications up to 2016 and improved the descriptions of the implementation framework and classification to more general descriptions. The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1] Keio University, Yokohama, Japan. [2] Ritsumeikan University, Shiga, Japan.

References
1. Azuma RT (1997) A survey of augmented reality. Presence: Teleoperators and Virtual Environments 6(4):355–385. MIT Press. doi:10.1162/pres.1997.6.4.355
2. Azuma RT (2001) Recent advances in augmented reality. IEEE Comput Graph Appl 21(6):34–47. IEEE, doi:10.1109/38.963459
3. Kruijff E, Swan JE, Feiner S (2010) Perceptual issues in augmented reality revisited. In: Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR). IEEE. pp 3–12. doi:10.1109/ISMAR.2010.5643530
4. Mann S (1994) Mediated reality. M.I.T. Media Lab Perceptual Computing Section, Cambridge, Ma. Technical Report TR 260
5. Mann S, Fung J (2001) Videoorbits on eye tap devices for deliberately diminished reality or altering the visual perception of rigid planar patches of a real world scene. In: Proc. Int. Symp. on Mixed Reality (ISMR). Citeseer. pp 48–55. http://www.eyetap.org/papers/docs/ismr01.ps
6. Zokai S, Esteve J, Genc Y, Navab N (2003) Multiview paraperspective projection model for diminished reality. In: Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR). IEEE. pp 217–226. doi:10.1109/ISMAR.2003.1240705
7. Bajura M, Fuchs H, Ohbuchi R (1992) Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. In: Proc. SIGGRAPH. ACM. pp 203–210. doi:10.1145/133994.134061
8. Navab N, Bani-Kashemi A, Mitschke M (1999) Merging visible and invisible: Two camera-augmented mobile C-arm (CAMC) applications. In: Proc. Int. Workshop on Augmented Reality. IEEE. pp 34–141. doi:10.1109/IWAR.1999.803814
9. Herling J, Broll W (2010) Advanced self-contained object removal for realizing real-time diminished reality in unconstrained environments. In: Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR). IEEE. pp 207–212. doi:10.1109/ISMAR.2010.5643572
10. Herling J, Broll W (2014) High-quality real-time video inpainting with PixMix. IEEE Trans. Visualization and Computer Graphics 20(6). IEEE. doi:10.1109/TVCG.2014.2298016
11. Mori S, Ichikari R, Shibata F, Kimura A, Tamura H (2011) Framework and technical issues of diminished reality: A survey of technologies that can visually diminish the objects in the real world by superimposing, replacing, and seeing-through. Trans Virtual Real Soc Japan 16(2):239–250
12. Flores A, Belongie S (2010) Removing pedestrians from google street view images. In: Proc. Computer Vision and Pattern Recognition (CVPR) Workshop. IEEE. pp 53–58. doi:10.1109/CVPRW.2010.5543255
13. Li Z, Wang Y, Guo J, Cheong LF, Zhou SZ (2013) Diminished reality using appearance and 3D geometry of internet photo collections. In: Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR). IEEE. pp 11–19. doi:10.1109/ISMAR.2013.6671759
14. Rameau F, Ha H, Joo K, Choi J, Park K, Kweon IS (2016) A real-time augmented reality system to see-through cars. IEEE Trans Vis Comput Graph (TVCG) 22(11):2395–2404
15. Hashimoto T, Uematsu Y, Saito H (2010) Generation of see-through baseball movie from multi-camera views. In: Proc. MMSP. IEEE. pp 432–437. doi:10.1109/MMSP.2010.5662060
16. Hasegawa K, Saito H (2015) Diminished reality for hiding a pedestrian using hand-held camera. In: Proc. Int. Symp. on Mixed and Augmented Reality Workshops, Int. Workshop on Diminished Reality as Challenging Issue in Mixed and Augmented Reality (IWDR). IEEE. pp 47–52. doi:10.1109/ISMARW.2015.18
17. Hasegawa K, Saito H (2016) Stroboscopic image synthesis of sports player from hand-held camera sequence. Trans Computational Visual Media 2(3):277–289. IEEE. doi:10.1109/ICCVW.2015.99
18. Yoshida T, Jo K, Minamizawa K, Nii H, Kawakami N, Tachi S (2008) Transparent cockpit: Visual assistance system for vehicle using retro-reflective projection technology. In: Proc. IEEE Virtual Reality. IEEE. pp 185–188. doi:10.1109/VR.2008.4480771
19. Barnum P, Sheikh Y, Datta A, Kanade T (2009) Dynamic seethroughs: Synthesizing hidden views of moving objects. In: Proc. Int. Symp. on

Mori *et al. IPSJ Transactions on Computer Vision and Applications* (2017) 9:17

Page 12 of 14

Mixed and Augmented Reality (ISMAR). IEEE. pp 111–114. doi:10.1109/ISMAR.2009.5336483

20. Richter-Trummer T, Kalkofen D, Park J, Schmalstieg D (2016) Instant mixed reality lighting from casual scanning. In: Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR). IEEE. pp 27–36. doi:10.1109/ISMAR.2016.18

21. Korkalo O, Aittala M, Siltanen S (2010) Light-weight marker hiding for augmented reality. In: Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR). IEEE. pp 247–248. doi:10.1109/ISMAR.2010.5643590

22. Kawai N, Sato T, Nakashima Y, Yokoya N (2015) Augmented reality marker hiding with texture deformation. In: Proc. Trans. on Visualization and Computer Graphics (TVCG). IEEE Vol. PP(99). pp 1–13. doi:10.1109/TVCG.2016.2617325

23. Kawai N, Sato T, Yokoya N (2016) Diminished reality based on image inpainting considering background geometry. IEEE Trans. Visualization and Computer Graphics (TVCG) 22(3):1236–1247. IEEE. doi:10.1109/TVCG.2015.2462368

24. Navab N, Heining SM, Traub J (2010) Camera augmented mobile C-Arm (CAMC): Calibration, accuracy study, and clinical applications. IEEE Trans Med Imaging 29(7):1412–1423. IEEE. doi:10.1109/TMI.2009.2021947

25. Cosco FI, Garre C, Bruno F, Muzzupappa M, Otaduy MA (2009) Augmented touch without visual obtrusion. In: Proc. Int. Symp. on Augmented and Mixed Reality (ISMAR). IEEE. pp 99–102. doi:10.1109/ISMAR.2009.5336492

26. Mori S, Shibata F, Kimura A, Tamura H (2016) Efficient use of textured 3D model for pre-observation-based diminished reality. In: Proc. Int. Symp. on Mixed and Augmented Reality Workshops, Int. Workshop on Diminished Reality as Challenging Issue in Mixed and Augmented Reality (IWDR). IEEE. pp 32–39. doi:10.1109/ISMARW.2015.16

27. Mei C, Sommerlade E, Sibley G, Newman P, Reid I (2011) Hidden view synthesis using real-time visual SLAM for simplifying video surveillance analysis. In: Proc. IEEE Int. Conf. on Robotics and Automation (ICRA). IEEE. pp 4240–4245. doi:10.1109/ICRA.2011.5980093

28. Mourgues F, Deverna F, C-Maniere E (2001) 3D reconstruction of the operating field for image overlay in 3D-endoscopic surgery. In: Proc. Int. Symp. on Augmented Reality (ISAR). IEEE. pp 191–192. doi:10.1109/ISAR.2001.970537

29. Hosokawa T, Jarusirisawad S, Saito H (2009) Online video synthesis for removing occluding objects using multiple uncalibrated cameras via plane sweep algorithm. In: Proc. Int. Conf. on Distributed Smart Cameras (ICDSC). IEEE. pp 1–8. doi:10.1109/ICDSC.2009.5289380

30. Mori S, Maezawa M, Ienaga N, Saito H (2016) Detour light field rendering for diminished reality using unstructured multiple views. In: Proc. Int. Symp. on Mixed and Augmented Reality Workshops, Int. Workshop on Diminished Reality as Challenging Issue in Mixed and Augmented Reality (IWDR). IEEE. pp 292–293. doi:10.1109/ISMAR-Adjunct.2016.0098

31. Meerits S, Saito H (2015) Real-time diminished reality for dynamic scenes. In: Proc. Int. Symp. on Mixed and Augmented Reality Workshops, Int. Workshop on Diminished Reality as Challenging Issue in Mixed and Augmented Reality (IWDR). IEEE. pp 53–59. doi:10.1109/ISMARW.2015.19

32. Ienaga N, Bork F, Meerits S, Mori S, Fallavolita P, Navab N, Saito H (2016) First deployment of diminished reality for anatomy education. In: Proc. Int. Symp. on Mixed and Augmented Reality Workshops, Int. Workshop on Diminished Reality as Challenging Issue in Mixed and Augmented Reality (IWDR). IEEE. pp 294–296. doi:10.1109/ISMAR-10.1109/ISMAR-

33. Avery B, Piekarski W, Thomas BH (2007) Visualizing occluded physical objects in unfamiliar outdoor augmented reality environments. In: Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR). IEEE. pp 285–286. doi:10.1109/ISMAR.2007.4538869

34. Avery B, Sandor C, Thomas BH (2009) Improving spatial perception for augmented reality x-ray vision. In: Proc. IEEE Virtual Reality. IEEE. pp 79–82. doi:10.1109/VR.2009.4811002

35. Tomioka M, Ikeda S, Sato K (2013) Approximated user-perspective rendering in tablet-based augmented reality. In: Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR). IEEE. pp 21–28. doi:10.1109/ISMAR.2013.6671760

36. Tomioka M, Ikeda S, Sato K (2014) Pseudo-transparent tablet based on 3d feature tracking. In: Proc. Augmented Human International Conference (AH). ACM. pp 52:1–52:2. doi:10.1145/2582051.2582103

37. Jarusirisawad S, Saito H (2007) Diminished reality via multiple hand-held cameras. In: Proc. Int. Conf. on Distributed Smart Cameras (ICDSC). IEEE. pp 251–258. doi:10.1109/ICDSC.2007.4357531

38. Kawai N, Yamasaki M, Sato T, Yokoya N (2013) Diminished reality for AR marker hiding based on image inpainting with reflection of luminance changes. ITE Trans Media Technol Appl (MTA) 1(4):343–353. ITE. doi:10.3169/mta.1.343

39. Lepetit V, Berger MO (2001) An intuitive tool for outlining objects in video sequences: Applications to augmented and diminished reality. In: Proc. Int. Symp. on Mixed Reality (ISMR). pp 159–160

40. Jarusirisawad S, Hosokawa T, Saito H (2010) Diminished reality using plane-sweep algorithm with weakly-calibrated cameras. Prog Inform 7:11–20

41. Enomoto A, Saito H (2007) Diminished reality using multiple handheld cameras. In: Proc. Asian Conf. on Computer Vision (ACCV). pp 130–150

42. Hill A, Schiefer J, Wilson J, Davidson B, Gandy M, MacIntyre B (2011) Virtual transparency: Introducing parallax view into video see-through AR. In: Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR). pp 239–240. doi:10.1109/ISMAR.2011.6092395

43. Baričević D, Lee C, Turk M, Hollerer T, Bowman DA (2012) A hand-held AR magic lens with user-perspective rendering. In: Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR). IEEE. pp 197–206. doi:10.1109/ISMAR.2012.6402557

44. Baricevic D, Hollerer T, Sen P, Turk M (2016) User-perspective AR magic lens from gradient-based IBR and semi-dense stereo. IEEE Trans Vis Comput Graph (TVCG) 23(7):1–1. doi:10.1109/tvcg.2016.2559483

45. Sugimoto K, Fujii H, Yamashita A, Asama H (2014) Half-diminished reality image using three RGB-D sensors for remote control. In: Int. Symp. on Safety, Security, and Rescue Robotics (SSRR). IEEE. pp 1–6. doi:10.1109/SSRR.2014.7017676

46. Kameda Y, Takemasa T, Ohta Y (2004) Outdoor see-through vision utilizing surveillance cameras. In: Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR). IEEE. pp 151–160. doi:10.1109/ISMAR.2004.45

47. Tsuda T, Yamamoto H, Kameda Y, Ohta Y (2006) Visualization methods for outdoor see-through vision. IEICE Trans Inf Syst E89D(6):1781–1789. IEICE. doi:10.1093/ietisy/e89-d.6.1781

48. Sandor C, Cunningham A, Dey A, Mattila VV (2010) An augmented reality x-ray system based on visual saliency. In: Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR). IEEE. pp 27–36. doi:10.1109/ISMAR.2010.5643547

49. Cosco F, Garre C, Bruno F, Muzzupappa M, Otaduy MA (2013) Visuo-haptic mixed reality with unobstructed tool-hand integration. IEEE Trans Vis Comput Graph (TVCG) 19(1):159–172

50. Takemura M, Ohta Y (2002) Diminished head-mounted display for shared mixed reality. In: Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR). IEEE. pp 149–156. doi:10.1109/ISMAR.2002.1115084

51. Takemura M, Kitahara I, Ohta Y (2006) Photometric inconsistency on a mixed-reality face. In: Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR). IEEE. pp 129–138. doi:10.1109/ISMAR.2006.297804

52. Bayart B, Didier JY, Kheddar A (2008) Force feedback virtual painting on real objects: A paradigm of augmented reality haptics. In: Proc. EuroHaptics. Springer, Berlin, Heidelberg. pp 776–785. doi:10.1007/978-3-540-69057-3_99

53. Makoto T, Ikeda S, Sato K (2013) Rectification of real images for on-board camera tablet-based augmented reality. In: Technical Report of Institute of Electronics, Information and Communication Engineers (IEICE). IEICE, in Japanese. pp 347–352

54. Fiala M (2005) ARTag, a fiducial marker system using digital techniques. In: Proc. Computer Vision and Pattern Recognition (CVPR), vol. 2. IEEE. pp 590–596. doi:10.1109/CVPR.2005.74

55. Klein G, Murray D (2007) Parallel tracking and mapping for small AR workspaces. In: Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR). IEEE. pp 225–234. doi:10.1109/ISMAR.2007.4538852

56. Newcombe RA, Davison AJ, Izadi S, Kohli P, Hilliges O, Shotton J, Molyneaux D, Hodges S, Kim D, Fitzgibbon A (2011) KinectFusion: Real-time dense surface mapping and tracking. In: Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR). IEEE. pp 127–136. doi:10.1109/ISMAR.2011.6092378

Mori *et al. IPSJ Transactions on Computer Vision and Applications* (2017) 9:17

Page 13 of 14

57. Lepetit V, Berger MO (2000) A semi-automatic method for resolving occlusion in augmented reality. In: Proc. Computer Vision and Pattern Recognition (CVPR). IEEE. pp 225–230. doi:10.1109/CVPR.2000.854794

58. Zhang K, Zhang L, Yang MH (2012) Real-time compressive tracking. In: Proc. European Conference on Computer Vision. Springer, Berlin, Heidelberg. pp 866–879. doi:10.1007/978-3-642-33712-3_62

59. Yokoi T, Fujiyoshi H (2006) Generating a time shrunk lecture video by event detection. In: Proc. Int. Conf. on Multimedia and Expo (ICME). IEEE. pp 641–644. doi:10.1109/ICME.2006.262527

60. Boykov YY, Jolly MP (2001) Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images. In: Proc. Int. Conf. on Computer Vision (ICCV), vol. 1. IEEE. pp 105–112. doi:10.1109/ICCV.2001. 937505

61. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: Proc. Computer Vision and Pattern Recognition (CVPR), vol. 1. IEEE. pp 886–893. doi:10.1109/CVPR.2005.177

62. Kalman RE (1960) New approach to linear filtering and prediction problems. J. Basic Engineering 82(1):35–45

63. Kass M, Witkin A, Terzopoulos D (1988) Snakes: Active contour models. Int. J. Computer Vision (IJCV) 1(4):321–331. doi:10.1007/BF00133570

64. Debevec PE, Taylor CJ, Malik J (1996) Modeling and rendering architecture from photographs: A hybrid geometryand image-based approach. In: Proc. SIGGRAPH. ACM. pp 11–20. doi:10.1145/237170. 237191

65. Buehler C, Bosse M, McMillan L, Gortler SJ, Cohen MF (2001) Unstructured lumigraph rendering. In: Proc. SIGGRAPH. ACM. pp 425–432. doi:10.1145/383259.383309

66. Vaish V, Wilburn B, Joshi N, Levoy M (2004) Using plane + parallax for calibrating dense camera arrays. In: Proc. Computer Vision and Pattern Recognition (CVPR). IEEE. pp 2–9. doi:10.1109/CVPR.2004.1315006

67. Levoy M (2006) Light fields and computational imaging. IEEE Comput 39(8):46–55. IEEE. doi:10.1109/MC.2006.270

68. Siltanen S (2006) Texture generation over the marker area. In: Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR). IEEE. pp 253–254. doi:10.1109/ISMAR.2006.297831

69. Barnes C, Shechtman E, A Finkelstein DBG (2009) PatchMatch: A randomized correspondence algorithm for structural image editing. ACM Trans Graph (TOG) 28(3). ACM. doi:10.1145/1531326.1531330

70. Farbman Z, Hoffer G, Lipman Y, Cohen-Or D, Lischinski D (2009) Coordinates for instant image cloning. ACM Trans Graph (TOG) 28(3). ACM. doi:10.1145/1531326.1531373

71. Bane R, Hollerer T (2004) Interactive tools for virtual x-ray vision in mobile augmented reality. In: Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR). IEEE. pp 231–239. doi:10.1109/ISMAR.2004.36

72. White WW (2004) X-ray window: Portable visualization on the international space station. In: Proc. SIGGRAPH Sketches. ACM. p 100. doi:10.1145/1186223.1186348

73. Santos M, Souza I, Yamamoto G, Taketomi T, Sandor C, Kato H (2015) Exploring legibility of augmented reality x-ray. Multimedia Tools and Appl 75(16):9563–9585. Springer US. doi:10.1007/s11042-015-2954-1

74. Kalkofen D, Veas E, Zollmann S, Steinberger M, Schmalstieg D (2013) Adaptive ghosted views for augmented reality. In: Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR). IEEE. pp 1–9. doi:10.1109/ISMAR.2013.6671758

75. Otsuki M, Kuzuoka H, Milgram P (2015) Analysis of depth perception with virtual mask in stereoscopic AR. In: Proc. Int. Conf. on Artificial Reality and Telexistence and Eurographics Symp. on Virtual Environments (ICAT-EGVE). Eurographics Association. pp 45–52. doi:10.2312/egve.20151309

76. Buchmann V, Nilsen T, Billinghurst M (2005) Interaction with partially transparent hands and objects. In: Proc. Australasian Conf. on User Interface. Australian Computer Society, Inc. Vol. 40. pp 17–20. http://dl. acm.org/citation.cfm?id=1082246

77. Fukiage T, Oishi T, Ikeuchi K (2014) Visibility-based blending for real-time applications. In: Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR). IEEE. pp 63–72. doi:10.1109/ISMAR.2014.6948410

78. Siltanen S (2017) Diminished reality for augmented reality interior design. Vis Comput 33(2):193–208. Springer-Verlag New York, Inc. doi:10.1007/s00371-015-1174-z

79. Kiyokawa K, Billinghurst M, Campbell B, Woods E (2003) An occlusion-capable optical see-through head mount display for

80. supporting co-located collaboration. In: Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR). IEEE. pp 133–141. doi:10.1109/ISMAR. 2003.1240696

80. Seo BK, Lee MH, Park H, Park JI (2008) Projection-based diminished reality system. In: Proc. Int. Symp. on Ubiquitous Virtual Reality (SUVR). IEEE. pp 25–28. doi:10.1109/ISUVR.2008.21

81. Bonanni L, Lee CH, Selker T (2005) CounterIntelligence: Augmented reality kitchen. In: Proc. Conf. on Human Factors in Computing Systems (CHI). pp 2239–2245

82. Iwai D, Hanatani S, Horii C, Sato K (2006) Limpid desk: Transparentizing documents on real desk in projection-based mixed reality. In: Proc. IEEE Virtual Reality. IEEE. pp 30–31. doi:10.1109/VR.2006.95

83. Inami M, Kawakami N, Sekiguchi D, Yanagida Y, Maeda T, Tachi S (2000) Visuo-haptic display using head-mounted projector. In: Proc. IEEE Virtual Reality. IEEE. pp 233–240. doi:10.1109/VR.2000.840503

84. Inami M, Kawakami N, Tachi S (2003) Optical camouflage using retro-reflective projection technology. In: Proc. IEEE Int. Symp. on Mixed and Augmented Reality (ISMAR). IEEE. pp 348–349. doi:10.1109/ISMAR. 2003.1240754

85. Pucihar KC, Coulton P (2014) [Poster] Contact-view: A magic-lens paradigm designed to solve the dual-view problem. In: Proc. IEEE Int. Symp. on Mixed and Augmented Reality (ISMAR). IEEE. doi:10.1109/ISMAR.2014.6948458

86. Hincapié-Ramos JD, Roscher S, Büschel W, Kister U, Dachselt R, Irani P (2014) cAR: Contact augmented reality with transparent-display mobile devices. In: Proc. of Int. Symp. on Pervasive Displays (PerDis). ACM. doi:10.1145/2611009.2611014

87. Samini A, Palmerius KL (2016) A user study on touch interaction for user-perspective rendering in hand-held video see-through augmented reality. In: Lecture Notes in Computer Science. Springer, Cham. pp 304–317. doi:10.1007/978-3-319-40651-0_25

88. Čopič Pucihar K, Coulton P, Alexander J (2013) Evaluating dual-view perceptual issues in handheld augmented reality: Device vs. user perspective rendering. In: Proc. Int. Conf. on Multimodal Interaction. In: Proc. ACM on Int. Conf. on Multimodal Interaction. ACM. pp 381–388. doi:10.1145/2522848.2522885

89. Čopič Pucihar K, Coulton P, Alexander J (2014) The use of surrounding visual context in handheld AR: Device vs. user perspective rendering. In: Proc. Conf. on Human Factors in Computing Systems (CHI). ACM. pp 197–206. doi:10.1145/2556288.2557125

90. Fuchs H, Livingston MA, Raskar R, Colucci D, Keller K, State A, Crawford JR, Rademacher P, Dranke SH, Meyer AA (1998) Augmented reality visualization for laparoscopic surgery. In: Proc. Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, Berlin, Heidelberg. pp 934–943. doi:10.1007/BFb0056282

91. State A, Chen DT, Tector C, Brandt A, Chen H, Ohbuchi R, Bajura M, Fuchs H (1994) Case study: Observing a volume rendered fetus within a pregnant patient. In: Proc. Conf. on Visualization (VIS). IEEE. pp 364–368. doi:10.1109/VISUAL.1994.346295

92. Fuchs H, State A, Pisano ED, Garrett WF, Hirota G, Livingston MA, Whitton MC, Pizer SM (1996) Towards performing ultrasound-guided needle biopsies from within a head mounted display. In: Proc. Int. Conf. Visualization in Biomedical Computing. Springer, Berlin, Heidelberg. pp 591–600. doi:10.1007/BFb0047002

93. Garrett WF, Fuchs H, Whitton MC, State A (1996) Real-time incremental visualization of dynamic ultrasound volumes using parallel BSP trees. In: Proc. Conf. on Visualization (VIS). IEEE. pp 235–240. doi:10.1109/VISUAL. 1996.568114

94. State A, Livingston MA, Garrett WF, Hirota G, Whitton MC, Pisano ED, Fuchs H (1996) Technologies for augmented reality systems: Realizing ultrasound-guided needle biopsies. In: Proc. SIGGRAPH. ACM. pp 439–446. doi:10.1145/237170.237283

95. State A, Ackerman J, Hirota G, Lee J, Fuchs H (2001) Dynamic virtual convergence for video see-through head-mounted displays: Maintaining maximum stereo overlap throughout a close-range work space. In: Proc. Int. Symp. on Augmented Reality (ISAR). IEEE. pp 137–146. doi:10.1109/ISAR.2001.970523

96. Rosenthal M, State A, Lee J, Hirota G, Ackerman J, Keller K, Pisano ED, Jiroutek M, Muller K, Fuchs H (2001) Augmented reality guidance for needle biopsies: A randomized, controlled trail in phantoms. In: Proc. Int. Conf. on Medical Image Computing and Computer-Assisted

Mori *et al. IPSJ Transactions on Computer Vision and Applications* (2017) 9:17

Page 14 of 14

Intervention (MICCAI). Springer, Berlin, Heidelberg. pp 240–248. doi:10.1007/3-540-45468-3_29

97. Rosenthal M, State A, Lee J, Hirota G, Ackerman J, Keller K, Pisano ED, Jiroutek M, Muller K, Fuchs H (2001) Augmented reality guidance for needle biopsies: An initial randomized, controlled trial in phantoms. In: Proc. Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer-Verlag. pp 313–320. doi:10.1007/3-540-45468-3_29

98. Mori S, Eguchi Y, Ikeda S, Shibata F, Kimura A, Tamura H (2016) Design and construction of data acquisition facilities for diminished reality research. ITE Trans Media Technol Appl (MTA) 4(3):259–268. ITE. doi:10.3169/mta.4.259

99. Venkatesh MV, Cheung S-CS (2010) Eye tracking based perceptual image inpainting quality analysis. In: Proc. Int. Conf. on Image Processing (ICIP). IEEE. pp 1109–1112. doi:10.1109/ICIP.2010.5653640

100. Ardis PA, Singhal A (2009) Visual salience metrics for image inpainting. In: Proc. IS&T/SPIE Electronic Imaging. pp 72571W–72571W. doi:10.1117/12.808942

101. Oncu AI, Deger F, Hardeberg JY (2012) Evaluation of digital inpainting quality in the context of artwork restoration. In: Proc. European Conf. on Computer Vision (ECCV). Springer, Berlin, Heidelberg. pp 561–570. doi:10.1007/978-3-642-33863-2_58

102. Isogawa M, Mikami D, Takahashi K, Kojima A (2016) Eye gaze analysis and learning-to-rank to obtain the most preferred result in image inpainting. In: Proc. Int. Conf. on Image Processing (ICIP). IEEE. pp 3538–3542. doi:10.1109/ICIP.2016.7533018

103. Sakauchi D, Matsumi Y, Mori S, Shibata F, Kimura A, Tamura H (2015) Magical mystery room, 2nd stage. In: Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR) Demo

104. Matsuki H, Mori S, Ikeda S, Shibata F, Kimura A, Tamura H (2016) Considerations on binocular mismatching in observation-based diminished reality. In: Proc. Symp. on 3D User Interface (3DUI). IEEE. pp 259–260. doi:10.1109/3DUI.2016.7460070

105. Hoshino J (2001) See-through representation of video. In: Proc. Int. Symp. on Mixed Reality (ISMR). pp 155–156

106. Sawabe T, Kanbara M, Hagita N (2016) Diminished reality for acceleration - Motion sickness reduction with vection for autonomous driving. In: Proc. Int. Symp. on Mixed and Augmented Reality Workshops, Int. Workshop on Diminished Reality as Challenging Issue in Mixed and Augmented Reality (IWDR). IEEE. pp 277–278. doi:10.1109/ISMAR-Adjunct.2016.0100

107. Baričević D, Höllerer T, Sen P, Turk M (2014) User-perspective augmented reality magic lens from gradients. In: Proc. Symp. on Virtual Reality Software and Technology (VRST). ACM. pp 87–96. doi:10.1145/2671015.2671027