Live Will Never Be the Same!

How Broadcasting Might Influence the Acceptance and Widespread Usage of Augmented Reality

Wolfgang Broll[1]([⊠]) and Jan Herling[2]

[1] Virtual Worlds and Digital Games Group, TU Ilmenau,
Ehrenbergstr. 29, 98693 Ilmenau, Germany
`wolfgang.broll@tu-ilmenau.de`
`http://www.tu-ilmenau.de/vwdg`
[2] fayteq AG, Erich-Kaestner-Str. 1, 99094 Erfurt, Germany
`jan.herling@fayteq.com`
`http://www.fayteq.com`

Abstract. While anticipated in Hollywood movies and sci-fi literature, a perfect artifical reality – i.e. the ultimate VR environment – will probably not become real within the next couple of years. However, using Augmented Reality (AR) technologies in order to seamlessly embed artificial content into the real environment, might be a much more feasible way to remove the clear border between reality and virtuality. In this paper we will look into lately developed AR technologies and how they relate to recent trends in movie and TV productions. We will further anticipate what this will mean to live broadcasts and which implications this might have for a widespread individual usage of sophisticated AR technology.

Keywords: Augmented Reality (AR) · Diminished Reality · Broadcasting · Postproduction · Compositing · Advertising · Seamless integration

1 Introduction

If you would just ask an arbitrary person on the street what she thinks VR is, she would probably answer: the Matrix or maybe the Holodeck. Another possibility to express this would be to say: what would make a virtual environment undistinguishable from real life? Obviously, there is no easy and no immediate answer to this question. However, if we look at Augmented Reality (AR), it seems that we are already much closer to a situation, where it becomes impossible or at least very difficult for an individual to distinguish between real and artificial (virtual) content. Recent works in the area of Diminished Reality and AR with real lighting reveal that the remaining steps in this area might be much smaller than usually estimated. The main advantage of AR settings compared to VR environments is that reality – perfect as it is – is already there and only rather small parts have to be added or removed. While virtual worlds have to

provide a complete perfect or at least very convincing impression, AR applications may already provide a perfect illusion when still restricted to particular settings. Limitations such as not being capable to deal with correct lighting or occlusions in arbitrarily complex scenes in realtime hence can easily be avoided. Therefore, the current limitation is rather with respect to the overall complexity of the scenario and application then the quality of the augmentation. Adding live virtual content is already well established in the area of broadcasting. Further, we can observe a rapidly growing market for digital product placement, providing the necessary driving force. Thus, perfectly integrated sophisticated (live) virtual content will probably quite soon become a standard element within movies and broadcasts. On the one hand, live transmissions will contain additional unreal content, while on the other hand, real elements are discarded – not recognizable for the observer. Combined with recent advances in see-through display technologies we should not be surprised to already see individually adapted environments undistinguishable from (pure) reality within a couple of years. As already explained, this seems feasible as in contrast to VR most of the content observed will still be real and (selected) virtual content can be adapted to the real one much easier than creating an entire convincible artificial world. As with the Matrix and the Holodeck, this raises the question, how such a development will influence our daily life with respect to communication, interaction, and the reception of our environment.

In this paper we will show how recent technological developments will influence our perception of the environment, in particular regarding the manipulation of content in live broadcasts. The paper is structured as follows: in section two we will provide an overview of the current types of usage of digital video manipulation. In section three we will review recent approaches and developments in Diminished Reality and advanced lighting for AR. In the fourth section the implications of these approaches on broadcasting – in particular live broadcasts – as well as our perception of the environment will be discussed. We will then conclude in the final section of the paper.

2 Digital Video Manipulation

In this section we will review some existing application areas of AR related technologies in broadcasting. While the first application area makes immediate usage of (simple) AR technology, more advanced visual effects are currently still restricted to composting. Compositing is a major step in the postproduction of most movies, allowing for seamlessly combining different video sources, in particular digital content with real video footage. Nowadays, even experts can no longer clearly distinguish the real and the artificial parts of a scene. However, achieving such a perfect illusion is an elaborate and time-consuming process. Actually, a range of types for video editing and manipulation exists. Each of them combines or composes real and virtual content, and requires some kind of camera tracking for registering the artificial content. Further, they all require

some means to adapt lighting to seamlessly integrate the digital content. In order to discriminate between the different types of video effects (VFX) and AR, we would like to introduce the following scheme:

– **Video Compositing:** this typically includes the usage of chroma keying (green screen, blue screen) video footage, video compositing tools, image editing, and involves plenty of manual editing. The scene composition may be arbitrarily complex. This process often takes several months and by that is feasable for feature films only.
– **Video Manipulation (offline):** this relates to standard video footage enhanced during post production using video compositing or image editing tools, still involving a significant amount of manual editing. The complexity of this kind is fair. The time frame nowadays is from a few days to a few weeks and is used for daily soaps and other TV productions.
– **Realtime Video Manipulation:** this refers to directly adding (or removing) content of video stream immediately before broadcasting them. It requires sophisticated realtime capable tools and often includes special hardware for tracking and/or well-known camera viewpoints. Nevertheless complexity is restricted to rather simple settings for now. The time available is between a few milliseconds up to a few seconds (as even *live* broadcasts are typically delayed by a few seconds). This is currently mainly used for certain types of sport broadcasts.
– **Augmented Reality:** while using the same underlying technologies as direct video manipulations, AR additionally allows the user to freely change her viewpoint, moving around in the augmented scene, and enables him to interact with the real and virtual content. However, existing examples often provide a rather limited tracking and by that registration quality. Additionally they often neglect visual integration.

2.1 Virtual Content in Sports Broadcasts

Sports broadcasts were the first area to establish an extensive usage of AR techniques for overlaying the real video feed by virtual information. Early examples included the overlay of a virtual touchdown line in American football games or the virtual goal distance line for penalty shots in soccer. More advanced examples also included moving representations of all-time records and/or a comparison with competitors in racing, running, cycling, and swimming competitions. However, the virtual content added is still rather simple as it is typically restricted to simple 2D or even 1D objects such as lines, circles, text, and sometimes images (Fig. 1). In fixed settings such as the playing field in soccer or football this was also already occasionally used to add virtual advertising (see also next subsections).

2.2 Digital Product Placement

Product placement has a very long tradition in major Hollywood productions. One of the oldest and best-known examples is probable James Bond driving
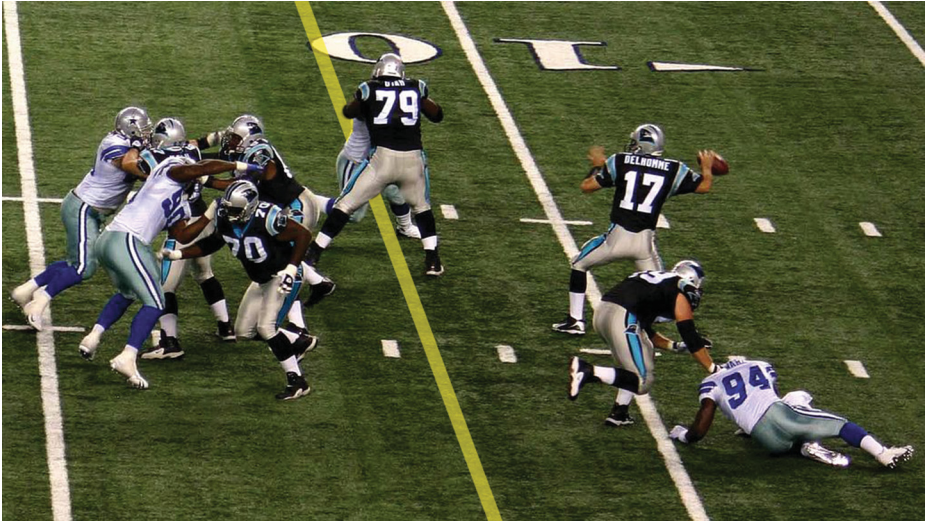
Fig. 1. Using simple augmentation in sport broadcasts (image adapted from [20] for illustration)

an Aston Martin. However product placement was previously often considered as covered advertising and by that has not been allowed as part of TV broadcasts for a long time in many countries including e.g. Germany. After recent law adjustments, product placement has meanwhile become a well-established form of advertisement and is common in most countries. Traditional product placement however, has one big disadvantage compared to standard TV advertising spots: the decision regarding the advertisement has to be made before the actual production of the movie, daily soap, or show. As such productions typically happen several months or at least weeks before their actual broadcast, and in contrast to this, the decision on advertising budgets is a rather short-term decision (a few weeks or even days before), traditional product placements provide a limited flexibility and by that market potential. Digital product placements partially overcome these limitations as they can be added to the scene afterwards. Their origin also trace back to feature films where they were added as part of the video compositing. This work however, is quite time-consuming and elaborate, and by that costly. Thus, it is not suitable or affordable for most productions, and for that reason may not be applied to advertisings for most commercial products. Digital product placement meanwhile has become quite common in certain TV formats such as the CBS production "How I met your mother".

2.3   Digital Advertising

Beside this, traditional forms of advertising as used in live events such as soccer games or car races become increasingly interesting for digital enhancements.

This refers for example to the advertising on shirts of players and drivers, or on cars as well as perimeter advertising. When broadcasting such events this local advertising is automatically transmitted as well. Looking for instance at international soccer matches we currently still see the situation that individual ads are used for the perimeters on each side. Having two almost independent camera teams filming the match from each side allows for country-specific ads. However, as such matches are usually also watched in several other countries or regions, a significant amount of the advertising does not reach the intended audience. Therefore, companies such as Supponor have started to enhance perimeters and cameras with appropriate sensor devices to detect these areas within the recorded video stream. This then allows for replacing the image area covered by the perimeter ads by custom-tailored content. Thus, individual ads can be applied and broadcasted to reach a specific audience. While such hardware and labor intensive approaches are currently still the only possibility to apply realtime modifications, cost pressure will eventually quicken the deployment of software solutions.

2.4 Object Removal

An airplane in the sky above Troy, a building visible through the window of the Titanic: original footage often contains undesired content, not matching to the current scene or the anticipated era. Thus, removing undesired objects from video sequences is a major issue in the postproduction of movies and films. Often supporting rigs, dollies, wires, people, buildings, microphones, etc. are either directly or indirectly (in some mirroring surfaces) visible and have to be removed from the original footage for the final cut. Existing tools provide rather limited support for this task. Therefore, objects are typically removed manually by standard image editing tools, which is very cumbersome and time-consuming, and by that a costly process. For this reason, in undesired objects are either just left where they are or the entire scene is removed in productions without the budget of feature films.

3 Related Work

In this section we will have a look at related work. We will focus on approaches and technologies, which may not necessarily be part of the ongoing developments currently, but which will or might influence them significantly in the future.

3.1 Diminished Reality

While Augmented Reality (AR) enhances reality by artificial (virtual) content, Diminished Reality (DR) is exactly the opposite of it. Diminished Reality means removing real content seamlessly and invisible from the reality. In the context of audio this meanwhile represents a standard technology: noise cancelation earphones, allowing for removing environmental audio sources and replacing them

by different (displayed) content are an off-the-shelf commercial product. From a technological point of view, Diminished Reality for video uses a totally different approach. Removing real content from a video stream requires an object removal on each individual frame. We can distinguish between two generally different approaches. The first approach removes objects by revealing the real background covered by the object to be removed. The second approach does not try to discover the real background, but rather tries to generate a coherent and by that convincible image. We will introduce both approaches below.

Regarding the first approach, two different methods exist. One method is to use multiple cameras. Thus, while the background of an occluding object is covered for one camera, it is still visible by another camera. Knowing the transformation between the individual cameras, the transformed real background image can be inserted into the original video frame, effectively removing the undesired real object. This method was e.g., used by Zokai et al. [23], and Enomoto and Saito [5]. Another possibility to remove objects is when either the object or the camera or both move within the video sequence and the occluded background is revealed in one of the other frames. Based on the camera movement the transformed real background can again be inserted into the appropriate video frame. This method originally was introduced by Wexler et al. [21]. Similar methods are meanwhile already available in a couple of video editing resp. compositing tools, e.g. Nuke or Mocha. In contrast to the first method, the second method is not very suitable for real-time usage, as it requires searching for matching image patches in previous or subsequent video frames. In order to allow for real-time or close to real-time processing the search has to be restricted to a small number of preceding frames. Nevertheless processing time remains a critical issue here.

The second approach is based on image inpainting methods. Image inpainting aka content-aware fill is a technique where a certain (masked) area of an image is filled with synthetic content based on the remainder of the image. While this meanwhile has become a standard tool in image editing software, its application to video editing is still in its infancy. While image inpainting approaches such as one by Simakov et al. [19] or Barnes et al. [2] usually search for suitable image patches in the remainder of the image, filling the mask with patches ensuring coherency throughout the entire image, video inpainting additionally requires coherency among subsequent video frames. In order to apply video inpainting to a live video stream, the inpainting process additionally has to be real-time capable. A real-time capable video inpainting approach was first presented in our previous work [9] and further enhanced in our more recent approaches [10,12]. It allows for generating an artificial content, coherent with the surrounding image throughout the entire video sequence (Fig. 2). This even allows for removing objects, which could not be removed in reality.

One challenging task in video inpainting approaches is the identification or specification of the undesired image content. The specification can be done either in a pre-precessing step training the application for detecting the undesired object automatically or by a user-interaction in the moment the undesired image content becomes visible (see [10,13]). An automatic detection can be realized by

Fig. 2. Real-time removal of objects from a video sequence in real-time while generating a coherent stream.
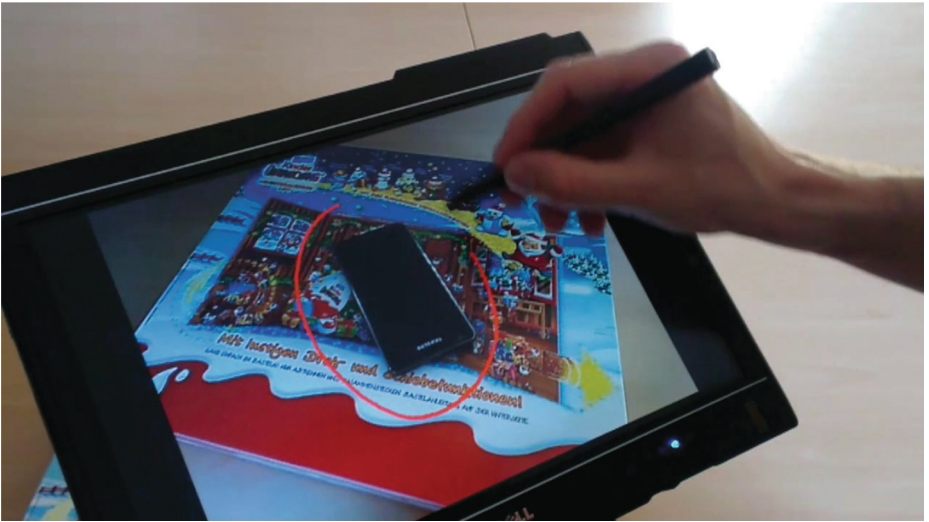

Fig. 3. User-defined selection of undesired image content by definition of a rough object contour on a tablet device.

applying visual patterns as used by Gordon and Lowe [7], our descriptorless approach [11], or by the application of geometric CAD models as used by Reitmayr and Drummond [17] or Wuest and Stricker [22]. However, if the undesired object is not known in advance the object needs to be selected by the user manually. Several individual interaction techniques are conceivable. The selection of the object's contour is one of the most intuitive ways defining the undesired object (Fig. 3). Nevertheless, masking of the entire undesired image content with a brush or mask tool may also be appropriate.

All approaches to video inpainting require some kind of tracking. Either the location of the object to be removed has to be tracked or the movement of the camera, or both. Traditional tracking techniques as used e.g., for AR are only of limited usability for this task. While putting traditional black and white markers on objects to be removed does not represent a feasible solution, most

other (feature-based) tracking approaches such as those introduced by Gordon and Lowe [7] or Bay et al. [3] will fail as well. One reason is, that objects to be removed often do not provide suitable features, or features change due to camera or object movements, and new features cannot clearly be assigned either to belong to the object or the background. Further, besides detecting the movement of the camera in relation to the scene and/or the object, it is also necessary to clearly separate the object to be removed from the background. This image segmentation is the by far most difficult task of DR-related tracking. An approach applying fingerprints may be found in our previous work [10] allowing for the segmentation of undesired object in real-time. However, if more accurate segmentation is required (e.g., applying sub-pixel accuracy) more time consuming segmentation approaches like those used by Rother et al. [18] or Arbelaez et al. [1] may be more appropriate. An additional difficulty is the fact that objects to be removed may partially or even temporarily completely become out of sight at the video image border. Even worse, other scene content, such as moving people, cars, etc. may partially or temporarily completely occlude the objects to be removed, making a proper segmentation difficult up to impossible.

3.2   Advanced Lighting for Seamless Integration

In order to seamlessly integrate artificial virtual content into a real-life environment or video footage, proper lighting is an important issue. In order to achieve a good integration, mutual influence of lighting has to be considered. On the one hand, environmental light influences the appearance of augmented virtual content. On the other hand, this virtual content might also influence its (real) environment. In order to model the environmental lighting influence on artificial objects the light distribution in the scene has to be estimated. Traditional methods applied sphere maps or cube maps received from fisheye lenses or multiple cameras, or used spheres placed in the real scene in order to detect highlights on them [4]. By applying for instance shadow maps and irradiance volumes [8] or light propagation volumes [6] such methods provide surprisingly realistic renderings of virtual objects within real environments. However, they allow for simulating the captured light sources only and become less accurate when virtual objects are rendered at other locations. While placing spheres or cameras within a scene is often not feasible, they provide live information of the lighting situation. Spheres however have the disadvantage that they can only detect light sources shining more or less in same direction as the observer is looking.

For an appropriate modeling of the global illumination of the real scene, the actual geometry resp. the surface normal of each surface is required. The acquisition of the scene geometry can for instance be done using 3D laser scanners. This however, restricts the information to static scenes. As soon as the environment contains any dynamic items, the scene information and by that the resulting lighting will become incorrect. Another possibility to gather the required scene information are SLAM (simultaneous localization and mapping) approaches [14], which allow to create a 3D model of the environment while simultaneously tracking the movement of the camera within it. Finally, depth cameras provide another

Fig. 4. Influence of light reflected from virtual objects on real environment.

possibility to capture the scene information. While typically providing a rather rough resolution, their RGBD image allows for reconstructing surfaces and by that surface normals. In order to do so, the depth image has to be smoothed, outliers have to be removed and holes have to be filled. Depth cameras typically use infrared light and are either based on regular patterns projected onto the environment (e.g. Kinect) or use the time-of-flight (ToF) principle (e.g. Mesa or Kinect II). Approaches applying such information in order to seamlessly integrate virtual content into a real environment include Kinect Fusion [15,16].

As stated above, another issue is the influence of the artificial content on the real environment. This includes direct and indirect lighting of the real environment by artificial (virtual) lights, casting of shadows of virtual objects onto real objects, and real lighting reflected from virtual objects back to real objects (Fig. 4). These aspects of mutual lighting always require a 3D model of the real environment to be applied realistically. If neither information about the light sources nor the geometry of the environment from scanners or depth sensors is available, e.g. in a (2D) video stream, a proper estimation of the real lighting is difficult up to impossible. However, when replacing flat surfaces in a scene (i.e. billboards), the lighting (including its changes throughout multiple frames) may (roughly) be extracted and applied to a 2D replacement. If the camera is moved, SLAM approaches, as introduced above, may also be used.

4 Applications and Implications

4.1 Adding Desired Content

We may now apply AR technologies for frame-to-frame tracking of objects such as billboards or perimeters. By overlaying the real content by artificial digital

content, ads, or other types of information can be added right into the scene as if they already have been there during the original shoot. Suddenly this will no longer be restricted to feature films. Integration of such technologies into postproduction tools will offer a new generation of powerful mechanisms for billboard replacement, digital product placement, etc. This will allow adding digital content even in low budget productions as this may then be done as part of the regular video editing rather than a separate composting step during the postproduction. However, this will also require looking into additional aspects, typically not considered in standard AR applications. Those include aspects such as blurring due to objects being out of focus, or due to fast moving objects or cameras. In contrast to traditional AR environments, the digitally added virtual content will have to be adapted according to the visual parameters in the original footage. Further, lighting aspects will not be limited to those described above. In particular dynamic shadows and highlights may frequently occur and will have to be transferred to added digital content in order to achieve a believable output.

4.2   Removing Undesired Content

By applying DR technology for the removal of undesired content in videos, a much wider usage than with traditional postproduction techniques can be achieved. Using DR technology allows for easily removing undesired content with minimal or even without manual effort. The technology not only allows for removing undesired objects, but also provides means for removing content due to legal demands. Thus, e.g. product placement has to be discarded for broadcasts on children channels (at least in some countries), certain content has to be removed before a movie may be shown aboard of airplanes (airplane editing). Further, this approach may also be used to remove station logos and captions from material received from other stations.

4.3   Real-Time Application for Broadcasting

As the technologies presented in section two already showed, their application is generally not restricted to offline tasks as part of the postproduction process. In fact they provide the basis for applying the above scenarios even to live broadcast. Similar to the usage within sports broadcasts today, the addition of digital (virtual) content, and the removal or the replacement of real content will become common even in live broadcasts in the near future. Software based technologies currently do not provide the necessary robustness to be used to this kind of scenario. In order to be used for live broadcasts, they have to be absolutely fail-safe. As discussed before, partial and/or temporal occlusions lead to complex tracking issues, which require to be solved before. Very dynamic scenes (imagine for instance a car race) imply further challenges due to the very small number of frames where an object to be modified might be visible. Applications will include the removal of unlawful content. Broadcast of tobacco ads for example is illegal in Europe, while ads for alcoholic beverages might not be allowed in some Islamic countries. As those advertisements might still

be allowed in the real setting, real-time DR and AR approaches will provide a convenient and efficient solution preventing the broadcast of this content by removing or replacing undesired ads on player shirts. However, while anticipating those opportunities, a clear labeling of modified content (as it is currently already required for (digital) product placement in some countries), will be important to make people aware of the fact that every content they observe might be manipulated.

4.4 Implications on Usage and Acceptance of AR

A general availability of a certain technology is often based on general developments not directly related to the original product. 3D graphics adapters for example were very expensive and available in high-end workstations only unless they were put at reasonable prices in every personal computer to enable it for gaming. The same applies to AR technology when it comes to smartphones and tablets. While the processing power, the graphics, the camera, and the sensors required made creating an AR system a cumbersome and costly adventure, almost everything necessary nowadays is available within a standard smartphone or tablet. What currently is missing from the device side is an appropriate head-worn display. However, with the currently ongoing developments (such as Google's Glass, Epson's Moverio, and Laster's SeeThru) it is quite likely that affordable wearable displays will be available shortly. Although their first generation might still lack the required field-of-view or brightness required for full outside AR and DR, their general availability and usage will lead to frequent updates and improvements (similar to the other developments mentioned here). When mobile phones became widespread, it took some time until one got used to people walking around and talking (in particular when using a head set). The same process will probably take place with head-worn displays. However, being used to the fact that even a live transmission may be enhanced by additional content, while other content might have been discarded, will achieve a higher acceptance than confronting people directly with the possibilities of the new technologies. Nevertheless, this will also open up a new field for empirical studies in the area of social sciences to study the implications of a widespread usage of such technologies. Looking at the implications of smartphones one might assume that arising social implications shall be pretty serious. With the broadcasting and advertising industry driving the development of convincing life-like real-time AR and DR, and the general availability of the necessary hardware for a significant fraction of the population, it seems reasonable to assume that highly sophisticated AR/DR undistinguishable from pure reality - might become common within a couple of years.

5 Conclusion and Future Work

In this paper we presented our vision towards a fundamental change of the reception of our environment. Due to the rapid development of new technologies for

seamlessly integrating virtual content into real scenes as well as removing real content from them, our perception of live broadcast transmissions will fundamentally change compared to what we are currently used to. Moreover, once established for live broadcasts, the technology will also quickly become affordable and sufficiently robust for individual usage. In combination with current and upcoming hardware this will even change the perception of our real surrounding. Thus, a live view will truly no longer be the same as we were used to. However, in order to achieve this step there are still a couple of problems to solve, in particular when it comes to automatic segmentation, tracking, and analysis of lighting conditions. Further, such scenarios come along with a whole bunch of ethical and social implications, to be investigated by social scientists. Further, security issues in particular due to differences between the individual reception and the physical reality represent an important issue requiring further attention.

References

1. Arbeláez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. IEEE Trans. Pattern Anal. Mach Intell. **33**(5), 898–916 (2011)
2. Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D.B.: Patchmatch: a randomized correspondence algorithm for structural image editing. In: SIGGRAPH, pp. 1–11. ACM, New York (2009)
3. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (surf). Comput. Vis. Image Underst. **110**(3), 346–359 (2008)
4. Debevec, P.: Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. In: Computer Graphics, pp. 189–198. ACM (1998)
5. Enomoto, A., Saito, H.: Diminished reality using multiple handheld cameras. In: ACCV 2007 Workshop on Multidimensional and Multi-view Image Processing (2007)
6. Franke, T.A.: Delta light propagation volumes for mixed reality. In: Proceedings of 2013 IEEE International Symposium on Mixed and Augmented Reality (IEEE ISMAR 2013). IEEE Computer Society, Piscataway (2013)
7. Gordon, I., Lowe, D.G.: What and where: 3D object recognition with accurate pose. In: Ponce, J., Hebert, M., Schmid, C., Zisserman, A. (eds.) Toward Category-Level Object Recognition. LNCS, vol. 4170, pp. 67–82. Springer, Heidelberg (2006)
8. Grosch, T., Eble, T., Mueller, S.: Consistent interactive augmentation of live camera images with correct near-field illumination. In: Proceedings of the 2007 ACM Symposium on Virtual Reality Software and Technology (ACM VRST 2007), pp. 125–132. ACM, New York (2007)
9. Herling, J., Broll, W.: The ocean framework: providing the basis for next-gen MR applications. In: Proceedings of the IEEE Virtual Reality 2010 Workshop, 3rd Workshop on Software Engineering and Architectures for Realtime Interactive Systems, SEARIS 2010 (2010)
10. Herling, J., Broll, W.: PixMix: a real-time approach to high-quality diminished reality. In: 11th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2012, USA, pp. 141–150. IEEE, November 2012

11. Herling, J., Broll, W.: Random model variation for universal feature tracking. In: Proceedings of the 18th ACM Symposium on Virtual Reality Software and Technology, VRST 2012, pp. 169–176. ACM, New York (2012)
12. Herling, J., Broll, W.: High-quality real-time video inpainting with pixmix. IEEE Trans. Vis. Comput. Graph. **20**(6), 866–879 (2014)
13. Kawai, N., Sato, T., Yokoya, N.: Diminished reality considering background structures. In: 2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 259–260, October 2013
14. Klein, G., Murray, D.: Parallel tracking and mapping for small AR workspaces. In: Proceedings of ISMAR 2007, Washington, DC, USA, pp. 1–10 (2007)
15. Lensing, P., Broll, W.: Instant indirect illumination for dynamic mixed reality scenes. In: Proceedings of the 11th IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE Computer Society, Piscataway (2012)
16. Newcombe, R.A., Davison, A.J., Izadi, S., Kohli, P., Hilliges, O., Shotton, J., Molyneaux, D., Hodges, S., Kim, D., Fitzgibbon, A.: Kinectfusion: real-time dense surface mapping and tracking. In: Proceedings of the 10th IEEE International Symposium on Mixed and Augmented Reality, pp. 127–136. IEEE Computer Society, Piscataway, October 2011
17. Reitmayr, G., Drummond, T.: Going out: robust model-based tracking for outdoor augmented reality. In: Proceedings of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR 2006, pp. 109–118. IEEE Computer Society, Washington, DC (2006)
18. Rother, C., Kolmogorov, V., Blake, A.: "grabcut": interactive foreground extraction using iterated graph cuts. In: SIGGRAPH, pp. 309–314. ACM, New York (2004)
19. Simakov, D., Caspi, Y., Shechtman, E., Irani, M.: Summarizing visual data using bidirectional similarity. In: CVPR, IEEE (2008)
20. Tomasland: American football sample scene "jake delhomme goes deep" (2009). https://www.flickr.com/photos/tomasland/3965511917/, available under cc 2.0. https://creativecommons.org/licenses/by/2.0
21. Wexler, Y., Shechtman, E., Irani, M.: Space-time completion of video. IEEE Trans. Pattern Anal. Mach. Intell. **29**(3), 463–476 (2007)
22. Wuest, H., Stricker, D.: Tracking of industrial objects by using cad models. JVRB J. Virtual Reality Broadcast. **4**(1) (2007). URN: urn:nbn:de:0009-6-11595
23. Zokai, S., Esteve, J., Genc, Y. Navab, N.: Multiview paraperspective projection model for diminished reality. In: Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR 2003, USA, pp. 217–226. IEEE (2003)