Through the Realities of Augmented Reality

Thitirat Siriborvornratanakul(✉)

Graduate School of Applied Statistics, National Institute of Development Administration (NIDA), 118 SeriThai Rd., Bangkapi, Bangkok 10240, Thailand
thitirat@as.nida.ac.th

Abstract. Speaking of Augmented Reality (AR), it is about augmenting an actual world with some virtually generated digital information in order to make the combination of two worlds as seamless as possible. Creating seamless AR effects in real time is non-trivial, requiring interdisciplinary knowledge integration from many fields such as computer vision, signal processing, sensor network, internet of things (IoT), three-dimensional computer graphics, human-computer interaction, and hardware designs. Nevertheless, for the past two decades, it is computer vision that has dominated the field of AR. Hence, common forms of AR that most people are familiar with are about utilizing a hardware device with embedded camera(s) together with a software program powering by computer vision algorithms. Based on our first-hand experiences in AR researches and communities, this paper presents a new summary regarding the world of modern AR from the beginning of the 21st century until now. Our summary divides the modern AR into five major waves based on important trends happening both inside and outside research communities.

Keywords: Augmented Reality · Spatial Augmented Reality · Wearable augmented reality · Artificial intelligence · Deep learning

1 The First Wave: Marker-Based AR

We believed that the first wave of modern AR was dated back in 1999 when an open source AR tracking library named ARToolKit [9] was demonstrated at SIGGRAPH 1999. ARToolKit was a C/C++ open source library that only required simple fiducial markers (i.e. black-and-white square markers) and an off-the-shelf camera to work with. Using ARToolKit's built-in functions, it became very easy for researchers and programmers to jump start in AR and obtain real-time camera's 3D pose estimation (with respect to the ARToolKit marker) regardless of indepth 3D computer vision understanding. During the beginning of the 21st century, this library's ease of use triggered rapid development in diverse applications not only in the field of AR itself but also in other 3D-vision applications. Some example usages of ARToolKit included an ARToolKit-based

C. Stephanidis (Ed.): HCII 2019, LNCS 11786, pp. 253–264, 2019.
https://doi.org/10.1007/978-3-030-30033-3_20

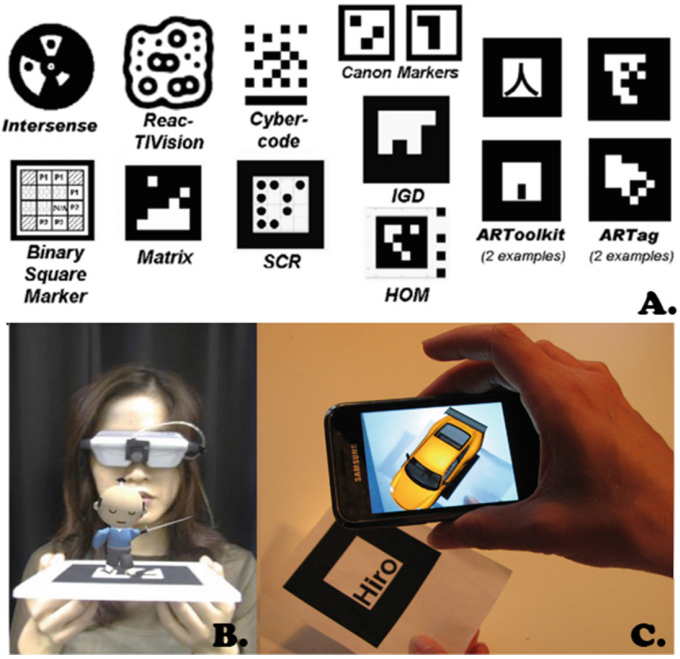

Fig. 1. (A) shows examples of fiducial marker libraries in computer vision; some are specifically designed for AR tasks (e.g., ARToolKit and ARTag) whereas the others are for other tasks in computer vision. (B) and (C) show examples of using ARToolKit markers to generate real-time AR see-through effects. (Image credit: http://campar.in.tum.de/twiki/pub/ISMAR/IarAbstractARTag/IarDetailsFialaSlides.pdf, http://www.hitl.washington.edu/artoolkit/ and https://alternativeto.net/software/artoolk/)

tangible interface for musician [17], ARToolKit for educational exhibitions [28], and ARToolKit as passive markers for a motion capture system [20].

Inspiring by the success and popularity of ARToolKit, many vision-based fiducial marker libraries were introduced afterwards, mostly for two main reasons—to improve robustness of marker detection and tracking regardless of partial occlusion or difficult lighting situations, and to introduce more visual alternatives of fiducial markers for different tasks. Examples of fiducial marker libraries in computer vision are shown in Fig. 1A. Note that standard barcodes, QR codes and many 2D planar patterns are not suitable as vision-based fiducial markers because they either require some specific camera orientations (relative to the marker) or provide inadequate information for visual computing.

It can be said that in the first wave of modern AR, the key developments heavily relied on computer vision algorithms and applications; the most popular AR features back then were marker-based AR see-through effects using an electronic monitor or a head-mounted device (HMD) as shown in Fig. 1B and C. Nevertheless, during the first wave of modern AR, utilization and popularity of AR outside research laboratories were scarce.

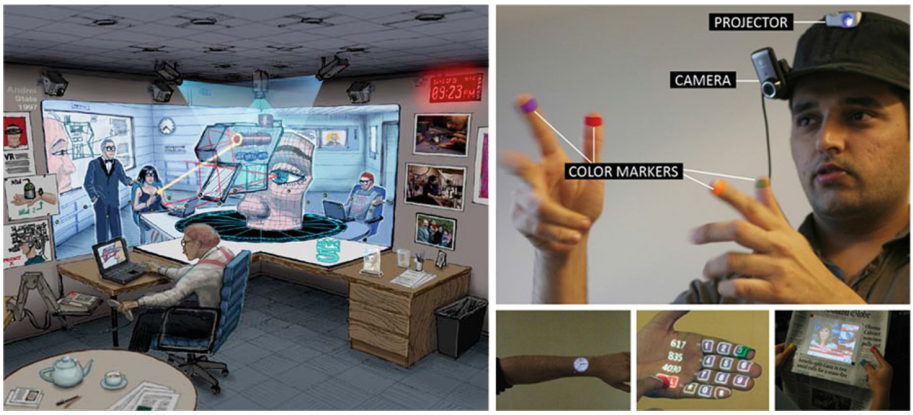

Fig. 2. Examples of Spatial AR for augmenting a physical surface with interactive projected imagery. (Image credit: Office of the future [18] and http://www.pranavmistry.com/projects/sixthsense/)

2 The Second Wave: Spatial AR with Projector

During SIGGRAPH 1998, there was a signal regarding another wave of modern AR. In this conference, Raskar et al. [18] proposed an idea called Office of the Future (Fig. 2 left) where interactive imagery projected from projectors were used to augment arbitrary flat surfaces in a seamless manner; they named this kind of projector-based AR as Spatial Augmented Reality (SAR) [2]. Inspiring by the office-of-the-future proposal as well as the trend of projector miniaturization following afterwards, the first decade of the 21st century was overwhelmed by not only SAR researches using projector-camera devices (a.k.a. pro-cam devices) but also continuous developments for smaller but brighter projection mechanisms. In the following paragraphs, we clarify the rise and fall of SAR in more detail.

Speaking of portable projectors, although they needed not to be firmly fixed on ceiling, their bulky form factors prevented them from being used in truly mobile fashions. After portable projectors became reasonably affordable around the beginning of the 21st century, they were continuously used in diverse researches and applications regarding SAR and computer vision; most of the time, a portable projector was coupled with one or more cameras in order to visually observed the projected results. Some example studies regarding portable projectors included methods to calibrate a pair of projector and camera either geometrically [10] or photometrically [5], methods to calibrate multiple projectors [29], imperceptible [3] or infrared [1] projection techniques, etc.

Another important trigger in the second wave of SAR was probably the SixthSense project (Fig. 2 right) by Mistry et al., published in CHI 2009 [15]. In this project, they built a prototype of wearable pro-cam device where a smartphone-sized laser projector (a.k.a. handheld, mobile or pico projector) and a camera were used to transform any ordinary flat surface into an interactive touch screen.

Their proof-of-concept applications were broadcasted by many news channels, making this project one of the most famous projects in SAR.

When SAR usage scenarios were changed from bulky portable projectors staying (mostly) still on a table to small mobile projectors being moved at all times, problems regarding pro-cam calibration were exaggerated and solutions needed to be provided for dealing with unpredictabilities of projection surfaces. Our previous works started from utilizing a motion sensor for indirect pro-cam geometric calibration [23] and then changed to pro-cam coaxialization using a plate beam splitter in order to completely eliminate the need for geometric calibration [24]; finally, in order to deal with visual crosstalk problems, we decided to maintain projection in the visible light spectrum but change visual analysis tasks into the infrared spectrum [21]. Other popular approaches for dynamic pro-cam geometric calibration included projecting a known pattern on an unknown surface and analyzing the distorted pattern to reconstruct the surface's geometry.

Despite of proof-of-concept demonstrations and few commercial projector-based interactive touchscreen devices, usages of SAR beyond exhibitions and lighting performances were uncommon. Our recent study regarding SAR [22] revealed shared concerns from different experts regarding robustness, price, usability and practicality of SAR in actual usages. Unfortunately, the second wave of modern AR in SAR was not last long. After a decade (approximately) of active development, researchers and manufacturers slowly lost attraction of miniature projector utilization. Our assumption regarding the falling of SAR includes the introduction of Microsoft Kinect Sensor in 2010 and limitations of miniature projectors themselves.

Because of the all-in-one affordable solution provided by Kinect sensor, the needs for developing complicated pro-cam hardwares for geometric calibration or 3D reconstruction were sharply decreased. As for the limitations of miniature projectors, during their golden years, many attempts were pushed towards building brighter, smaller and focus-free projectors. In order to build a focus-free projector (i.e., projected images are always in focus regardless of arbitrary surface depths) with strong brightness, laser projection technologies were once expected as promising solutions. However, because of the international standard of safety for laser devices, the expected solution of laser projection was limited to very small amount of brightness. Even there were other non-laser projection mechanisms that could produce hundreds of lumens in brightness while maintaining their small form factors, without the focus-free ability, our dream of freestyle mobile projection on any desired surface will never come true.

3 The Third Wave: Wearable AR for Corporate

The third wave of modern AR came in the form of professional head-mounted wearable devices as shown in Fig. 3, starting with Google Glass in 2013, Microsoft HoloLens (developer edition) in 2016, MagicLeap One (creator edition) in 2018, and Microsoft HoloLens 2 in 2019; the years specified here are official released years, except for HoloLens 2 that is still in the preorder stage at the moment of writing this paper. Unlike the two previous waves, the third wave of modern

Fig. 3. Leading head-mounted wearable devices for AR from three companies. (A) is Google Glass, (B) is Microsoft HoloLens 1, (C) is MagicLeap One, and (D) is Microsoft HoloLens 2. (Image credit: http://time.com/, https://news.microsoft.com/, https://www.bloomberg.com/ and https://www.microsoft.com/en-us/hololens/buy)

AR has been driven by world famous tech companies whose aims are clearly not just AR prototypes for academic presentations but commercial AR products with (hopefully) mass production. Another uniqueness of this wave is that development of these wearable devices requires high-level interdisciplinary knowledge that goes far beyond computer vision to audio, optics, mechanic, etc. As a result, what most AR researchers do with these wearable AR devices is not to try tweaking their internal mechanisms but to study advantages and disadvantages of using them in each situation. For example, with HoloLens, [12] was able to use the real world geometry as input data and allow a user to define and solve a physical problem by Poisson's equation; [8] conducted experiments using Google Glass for training new scientists in wet laboratory work; [14] created an application on Google Glass that allowed people with Parkinson to monitor their speech volume; [26] discovered that unlike virtual reality headset, users of HoloLens did not suffer from obvious simulation sickness.

In the past, the first Google Glass and HoloLens were promoted with their AR capabilities. However, MagicLeap One and HoloLens 2 are now being advertised as mixed reality (MR). According to the long understanding in reality-virtuality continuum, MR refers to everything where real and virtual worlds are mixed up; this literally means that AR is a subset of MR. Nevertheless, this perspective has slightly been changed since MagicLeap has positioned their wearable device as non-AR but MR where the mixing between two worlds is indistinguishable in a 3D hologram manner. But for the sake of this AR review paper, we will stick with the term AR.

Until now, Google Glass, Microsoft HoloLens and MagicLeap are three big names that have worldly represented the future of wearable AR where the overlaid virtual information is more controllable than the previous wave of SAR as discussed in our previous work [22]. These wearable devices from the three companies share many things in common. They all are packed with sophisticated hardwares and algorithms developed by great engineers. They all are famous not only among AR researchers but also among technologists around the world; this is in particular for Google Glass, the first AR product that successfully popularized AR to end users. Despite of good things in common, they all are struggling the same problem of very high and unaffordable price tags. Their introductory prices are 1,500 USD for Google Glass, 3,000 USD (developer price) or 5,000 USD (commercial price) for HoloLens 1, 2,295 USD for MagicLeap One, and 3,500 USD for commercial HoloLens 2. This problem alone has made utilizations of these wearable AR devices being limited to small groups of researchers or big organizations who can afford (e.g., United State Army, NASA, DHL, General Motors).

For this third ongoing wave of modern AR, there are important lessons learnt from the original Google Glass whose initial aim of being a consumer-grade gadget failed due to privacy laws, driving regulations and social disapproval [7]. This means that the reasons consumers refrain from Google Glasses are not only their unaffordable prices but also their unsuitability regarding consumer life styles. Despite of the previous failure, Google Glass has already come back and this time Google as well as Microsoft have directed their attention to enterprise customers whose personal development and corporate training can take great advantages from these hi-tech and pricey AR headsets.

4 The Fourth Wave: Markerless AR in Smartphone

During the first and second waves of marker-based AR and SAR, one of the most popular techniques is utilizing known visual markers to pinpoint the virtually generated AR contents in world coordinates. Some visual markers are very obvious and not blending to the working environment like those in Fig. 1A. In many researches and AR creator platforms, to avoid using markers with intrusive visibility, natural objects are used as visual markers by help from vision-based feature point matching techniques; this allows us to use something like companies' logos as AR markers.

In our previous work [23], we addressed problems of marker based interactions and proposed a multi-target tracking solution in order to include non-marker objects into AR calculation. Multi-target tracking is a good start to taggle this problem, but in the long run, unless we have a proper map of objects in the environment, augmenting the environment with interactive AR contents remains difficult. To incorporate everything in the environment into AR systems, SLAM (Simultaneous Localization And Mapping) is a promising solution that

has become popular recently. Using SLAM techniques, we are able to use camera images in conjunction with other information in order to reconstruct and update a map of an unknown environment in real time. SLAM is especially popular for interactive systems that deal with unknown environments; this includes usage situations of wearable AR devices (Sect. 3) and smartphone AR (Sect. 4).

While the third wave of AR in professional wearable devices is still ongoing, the fourth wave of modern AR in smartphone has already touched the ground with the official releases of Apple's ARKit in 2017 and Android's ARCore in 2018. Like Google's Project Tango proposed back in 2014, ARKit and ARCore utilize SLAM techniques to create markerless AR effects on smartphones. But unlike the discontinued Project Tango relying on specially designed cameras and specific computational modules, ARKit and ARCore use smartphones' built-in motion sensors and cameras to perform SLAM, enabling sustainable smartphone AR in the long run.

Similar to the third wave of wearable AR headsets, the wave of smartphone AR has been driven mainly by big tech companies. However, while wearable AR headsets are very expensive and aim for corporate customers, smartphone AR is mostly free and involves diverse applications for arbitrary smartphone users. Examples regarding smartphone AR include AR DeepCalorieCam [25] that uses ARKit to measure the actual size of the meal (in order to estimate the total calories); interactive AR coding environments where N. Dass et al. [4] show that participant satisfaction is better with smartphone AR (using ARKit) than a traditional tablet or Microsoft HoloLens; a mixed-reality mobile remote collaboration system [6] using ARCore position tracking.

5 The Fifth Wave: AR Underneath Artificial Intelligence

During the years of 2014–2016 (approximately), extended reality technologies (including virtual, augmented and mixed realities) made the headlines that excited many technologists and researchers around the world. But in the past couple of years, majority of the world has turned their interest to Artificial Intelligence (AI) driven by machine learning, particularly the field of computer vision that has been disrupted significantly by deep learning. The popularity in AI (in comparison to AR) during the past decade is illustrated in Figs. 4 and 5. In Fig. 4, it can be seen that the increase in numbers of AI papers is the most obvious in IEEE Xplore Digital Library during 2017 and 2018. As for Google Trends (worldwide) in Fig. 5, popularity in AI keyword has totally beat AR keyword since October 2016.

Because computer vision always plays important roles in AR, disruption in computer vision results in disruption in AR as well. Hence, under the huge umbrella of machine learning and deep learning, there is also AR underneath. For vision-based tasks of recognizing and annotating objects with AR virtual contents, using state-of-the-art pre-trained convolutional neural networks gives researchers and developers a huge jump start with promising image recognition and annotation results. For example, AR DeepCalorieCam [25] uses Inception-v3

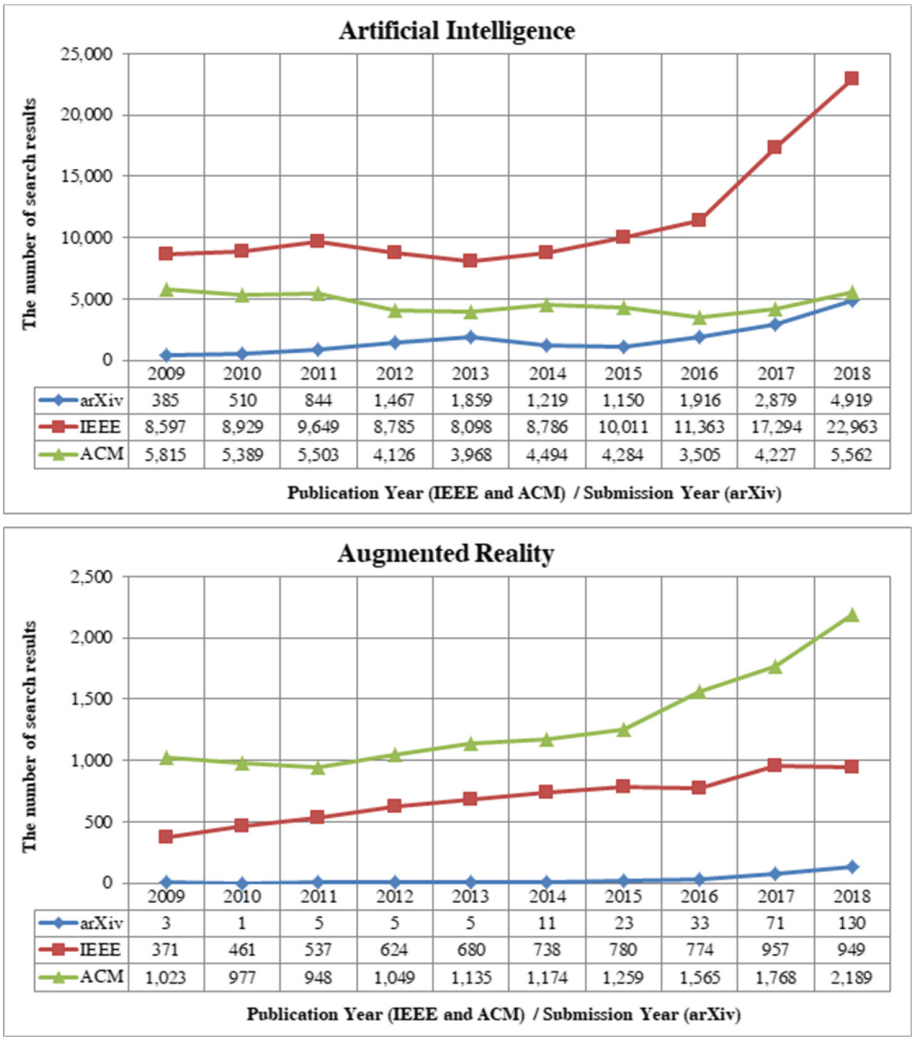

Fig. 4. The number of search results during 2009 to 2018 regarding two keywords (i.e., 'artificial intelligence' and 'augmented reality') from three research paper platforms—arXiv.org, IEEE Xplore Digital Library and ACM Digital Library. (Data retrieved on 19 April 2019)

(pre-trained on ImageNet dataset and fined tune on UEC-FOOD100 dataset) to recognize food category from an image before applying ARKit to measure the actual size of each food. Recognizing and annotating humans in images is another task that has received lot of attention recently. The latest proposal from Wang et al. [27] uses a self-supervised deep learning technique to predict human 3D poses from 2D image inputs. This kind of human 3D pose estimation mechanisms

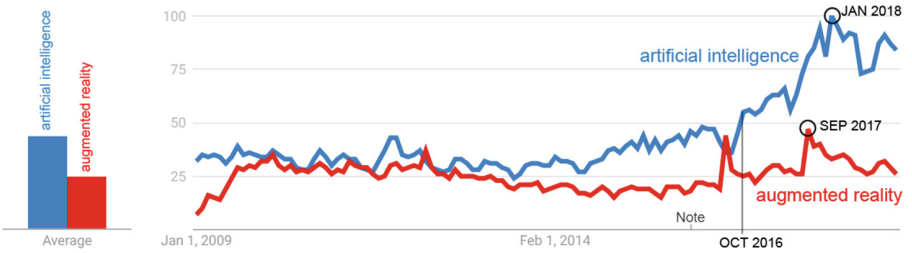

Fig. 5. Google Trends (worldwide) comparing the two keywords of 'artificial intelligence' and 'augmented reality' during 1 January 2009 to 31 December 2018. The vertical axis represents interest over time where 100 is the peak popularity of the corresponding keyword during the specified time interval. (Data retrieved on 19 April 2019)

can help leverage any AR systems that require indepth understanding of human's real-time actions/behaviors.

For interactive virtual makeup applications, it becomes a lot easier to overlay virtual makeups on a moving face as facial landmarks can be precisely located with many free machine learning libraries. An example is in the automatic virtual makeup system of [16] that uses Dlib for extracting 2D facial landmarks from a face image; Dlib is a machine learning library that originally utilizes a combination of Histogram of Oriented Gradients (HoG) and linear Support Vector Machine (SVM). In BeautyGAN [11], more advance technique of deep learning is used to simply transfer a makeup style from a reference makeup face to another non-makeup face. This growth in deep learning techniques helps reform traditional AR virtual makeup systems and strengthen them to become the next generation of artificial intelligence based AR systems.

Superimposing an AR content over a live video of the actual world cannot be more indistinguishable when there is a neural style transfer technique from deep learning to help blend two different image styles together; [13] demonstrates this concept using the neural style transfer and ARKit. Even for AR tasks involving 3D reconstruction for AR headsets (mentioned in Sect. 3), there is a recent proposal from Rematas et al. [19] that uses deep learning to convert a typical youtube video of soccer game into dynamic 3D information; this means that all players in the game are dynamically 3D reconstructed in a way that we can wear a 3D AR headset to see this soccer game in 3D AR style.

It can be said that with recent disruption in artificial intelligence, we can expect AR systems (regarding all four previous waves mentioned earlier) to become more intelligent, more seamless and more interactive in the near future. Once AR can overcome their long-standing technical difficulties, what remain unsolved are the true problems of AR in the long run—problems regarding affordability, user experience and practical usage scenario.

6 Conclusion

For the past two decades, it can be said that modern AR has been through a lot of good and bad times. This paper reviews these two decades and summarizes it into five waves of modern AR. The first wave of fiducial marker based AR is classic and can still be seen until now, especially in AR explorer apps and other proof-of-concept AR systems. The second wave of projector-based AR was once popular; but due to many limitations, it has become slow recently. The third and fourth waves of modern AR have been both driven by big tech companies. The third wave of expensive AR headsets has aimed for corporate training whereas the fourth wave of smartphone AR has directed their attention to mass consumers. Finally, all four waves of modern AR have been elevated by the fifth wave of artificial intelligence disruption where solutions regarding long-standing AR technical difficulties have been proposed one after another.

References

1. Akasaka, K., Sagawa, R., Yagi, Y.: A sensor for simultaneously capturing texture and shape by projecting structured infrared light. In: Proceedings of the 6th International Conference on 3-D Digital Imaging and Modeling, pp. 375–381 (2007)
2. Bimber, O., Raskar, R.: Spatial Augmented Reality: Merging Real and Virtual Worlds. A. K. Peters Ltd., Natick (2005)
3. Cotting, D., Naef, M., Gross, M., Fuchs, H.: Embedding imperceptible patterns into projected images for simultaneous acquisition and display. In: Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR 2004), pp. 100–109 (2004)
4. Dass, N., Kim, J., Ford, S., Agarwal, S., Chau, D.: Augmenting coding: augmented reality for learning programming. In: Proceedings of the International Symposium of Chinese CHI (ChineseCHI 2018), pp. 156–159 (2018)
5. Fujii, K., Grossberg, M., Nayar, S.: A projector-camera system with real-time photometric adaptation for dynamic environments. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), vol. 1 (2005)
6. Gao, L., Bai, H., He, W., Billinghurst, M., Lindeman, R.: Real-time visual representations for mobile mixed reality remote collaboration. In: Proceedings of SIGGRAPH Asia 2018 Virtual and Augmented Reality (SA 2018) (2018)
7. Hong, J.: Considering privacy issues in the context of Google Glass. Commun. ACM **56**(11), 10–11 (2013)
8. Hu, G., Chen, L., Okerlund, J., Shaer, O.: Exploring the use of Google Glass in wet laboratories. In: Proceedings of the ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA 2015), pp. 2103–2108 (2015)
9. Kato, H., Billinghurst, M.: Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In: Proceedings of the 2nd International Workshop on Augmented Reality (IWAR 1999) (1999)
10. Lee, J., Dietz, P., Maynes-Aminzade, D., Raskar, R., Hudson, S.: Automatic projector calibration with embedded light sensors. In: Proceedings of the ACM Symposium on User Interface Software and Technology (UIST 2004), pp. 123–126 (2004)

11. Li, T., et al.: BeautyGAN: instance-level facial makeup transfer with deep generative adversarial network. In: Proceedings of the ACM International Conference on Multimedia (MM 2018), pp. 645–653 (2018)
12. Logg, A., Lundholm, C., Nordaas, M.: Solving Poisson's equation on the Microsoft HoloLens. In: Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST 2017) (2017)
13. MadeWithARKit: Realistic AR brush texture paintings with deep learning style transfer by the ever so creative @Laan Labs. http://www.madewitharkit.com/post/166751998274/realistic-ar-brush-texture-paintings-with-deep. Accessed 28 Feb 2019
14. McNaney, R., Poliakov, I., Vines, J., Balaam, M., Zhang, P., Olivier, P.: LApp: a speech loudness application for people with Parkinson's on Google Glass. In: Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2015), pp. 497–500 (2015)
15. Mistry, P., Maes, P., Chang, L.: WUW - wear ur world: a wearable gestural interface. In: Proceedings of the CHI Extended Abstracts on Human Factors in Computing Systems (CHI 2009), pp. 4111–4116 (2009)
16. Park, J., Kim, H., Ji, S., Hwang, E.: An automatic virtual makeup scheme based on personal color analysis. In: Proceedings of the International Conference on Ubiquitous Information Management and Communication (IMCOM 2018) (2018)
17. Poupyrev, I., Berry, R., Kurumisawa, J., Billinghurst, M., Airola, C., Kato, H.: Augmented groove: collaborative jamming in augmented reality. In: SIGGRAPH 2000, Emerging Technologies (2000)
18. Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L., Fuchs, H.: The office of the future: a unified approach to image-based modeling and spatially immersive displays. In: Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 1998), pp. 179–188 (1998)
19. Rematas, K., Kemelmacher-Shlizerman, I., Curless, B., Seitz, S.: Soccer on your tabletop. In: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
20. Sementille, A., Lourenco, L., Brega, J., Rodello, I.: A motion capture system using passive markers. In: Proceedings of the ACM SIGGRAPH International Conference on Virtual Reality Continuum and Its Applications in Industry (VRCAI 2004), pp. 440–447 (2004)
21. Siriborvornratanakul, T.: Vision-based smart mobile projection: a study of infrared projection and sensing in a ubiquitous environment. Int. J. Digit. Content Technol. Appl. (JDCTA) **8**(2), 1–12 (2014)
22. Siriborvornratanakul, T.: Enhancing user experiences of mobile-based augmented reality via spatial augmented reality: designs and architectures of projector-camera devices. Adv. Multimedia **2018** (2018)
23. Siriborvornratanakul, T., Sugimoto, M.: A portable projector extended for object-centered real-time interactions. In: Proceedings of the European Conference for Visual Media Production (CVMP 2009), pp. 118–126 (2009)
24. Siriborvornratanakul, T., Sugimoto, M.: Multiscale visual object detection for unsupervised ubiquitous projection based on a portable projector-camera system. In: Proceedings of the IEEE International Conference on Digital Image Computing: Techniques and Applications (DICTA 2010), pp. 623–628 (2010)
25. Tanno, R., Ege, T., Yanai, K.: AR DeepCalorieCam V2: food calorie estimation with CNN and AR-based actual size estimation. In: Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST 2018) (2018)

26. Vovk, A., Wild, F., Guest, W., Kuula, T.: Simulator sickness in augmented reality training using the Microsoft HoloLens. In: Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI 2018) (2018)
27. Wang, K., Lin, L., Jiang, C., Qian, C., Wei, P.: 3D human pose machines with self-supervised learning. IEEE Trans. Pattern Anal. Mach. Intell. (T-PAMI 2019) (2019, to appear). https://arxiv.org/abs/1901.03798
28. Woods, E., et al.: Augmenting the science centre and museum experience. In: Proceedings of the International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia (GRAPHITE 2004), pp. 230–236 (2004)
29. Zhou, J., Wang, L., Akbarzadeh, A., Yang, R.: Multi-projector display with continuous self-calibration. In: Proceedings of the ACM/IEEE International Workshop on Projector Camera Systems (PROCAM 2008) (2008)