Decoupling of Real and Digital Content in Projection-based Augmented Reality Systems Using Time Multiplexed Image Capture

Shoaib R. Soomro, Erdem Ulusoy, and Hakan Urey

Koc University, Electrical Engineering Department, Optical Microsystems Laboratory, Istanbul 34450 Turkey
E-mail: ssoomro13@ku.edu.tr

Abstract. *Projection-based augmented reality systems overlay digital information directly on real objects, while at the same time use cameras to capture the scene information. A common problem with such systems is that cameras see the projected image besides the real objects to some degree. This crosstalk reduces the object detection and digital content registration abilities. The authors propose a novel time sharing-based technique that facilitates the real and digital content decoupling in real time without crosstalk. The proposed technique is based on time sequential operation between a MEMS scanner-based mobile projector and rolling shutter image sensor. A MEMS mirror- based projector scans light beam in raster pattern pixel by pixel and completes full frame projection over a refresh period, while a rolling shutter image sensor sequentially collects scene light row by row. In the proposed technique, the image sensor is synchronized with scanning MEMS mirror and precisely follows the display scanner with a half-period lag to make the displayed content completely invisible for camera. An experimental setup consisting of laser pico projector, an image sensor, and a delay and amplifier circuit is developed. The performance of proposed technique is evaluated by measuring the crosstalk in captured content and sensor exposure limit. The results show 0% crosstalk in captured content up to 8 ms sensor exposure. High capture frame rate (up to 45 fps) is achieved by cyclically triggering a 3.2 MP, 60 fps CMOS sensor and using a 60 Hz pico projector. © 2017 Society for Imaging Science and Technology.*

INTRODUCTION

The spatial augmented reality systems provide a unique way of blending real and virtual world by displaying digital content directly on real objects. Such systems use multimedia projectors for display and different sensors (i.e., cameras, inertial sensors) to map environment, detect objects, and enable user interaction. Most of the see-through augmented reality systems and near-to-eye display solutions offer narrow field of view, limited depth of field, and have complex optical configurations.[1–3] On the other hand, projection-based augmented reality systems can provide wide field of view and enhanced depth by directly projecting on real objects. Many types of projection-based augmented reality systems have been developed for different applications. LightSpace[4] and MirageTable[5] from Microsoft use the fixed installations of projectors and

depth cameras to augment interactive graphical objects on walls and tabletop screens. Portable augmented reality systems are developed by using pico projector together with vision sensors and smartphones.[6–10] LightBeam[6] and BeThere[7] use the pico projector and Kinect depth sensor[11] to display content on real objects (i.e., books, wall, table) and track the objects movement to enable user interaction. SideBySide[8] uses hybrid visible/infrared light projectors and near-infrared cameras to register digital content and enable inter-projector interaction. SurfacePhone[9] uses a smartphone paired with a pico projector as single and multi-user display and uses smartphone camera for fingertip and gesture recognition. Many of the augmented reality systems use visible markers to display perspective corrected image and fetch object-related information.[10,12–14] Head mounted projection displays together with retro-reflective screens positioned in the surroundings have also been demonstrated[15–17] to provide bright display content by reflecting projected beam toward projector placed close to the eye.

Most of the augmented reality systems use cameras placed next to projector to capture the real space, detect and register real objects, and enable user interaction. A common problem with projection-based augmented reality systems is the visibility of digital content in the camera frame. The mix of the projected light with actual scene light reduces the ability of the camera to record the real scene, recognize the real objects, and detect the visible markers associated with objects. An alternative technique to capture the scene information is using active infrared cameras or depth sensors but, since essential scene information such as color resides in visible spectrum range, a simultaneous but crosstalk-free image capturing technique is needed to widen the usability of projection-based augmented reality systems.

In this article, we present a real and digital content decoupling technique for crosstalk-free imaging using time multiplexed camera–projector approach. A mobile projector displays the content on real objects, while at the same time a camera synchronized with projector captures the scene without having interference from projected light. The proposed technique uses a scanning laser pico projector and a rolling shutter image sensor, sharing the space in time sequential fashion. The projector and camera scanning is precisely delayed to capture and display different parts of
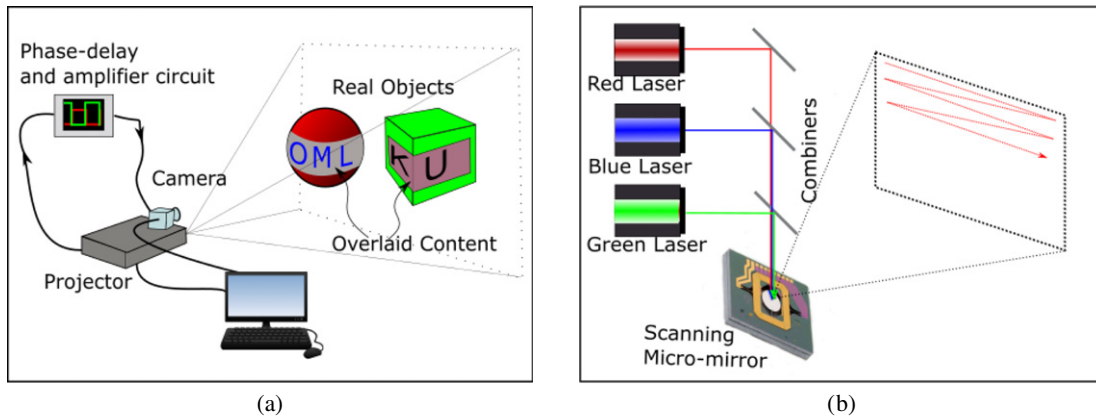
Figure 1. (a) The augmented reality projection system elements for time multiplexed imaging and display. (b) The working principle of MEMS scanner based laser pico projector, where red, green, and blue lasers are combined and reflected off a rotating micro-mirror to project focus-free images.

scene at particular time instant. The proposed technique can be used for a variety of augmented reality applications, where capturing the color information of real object is important (i.e., human face, visible markers). It can also provide real-time scene discovery and facilitate content registration for moving objects in the scene. The use of laser sourced projector provides infinite focus and mobility with no projection optics involved. The section of Decoupling Real and Digital Content discusses the concept of proposed technique and the Results and Discussion section discusses the experimental setup and test results.

DECOUPLING REAL AND DIGITAL CONTENT

The proposed technique uses three different elements as illustrated in Figure 1(a): (1) scanned laser sourced mobile projector, (2) a rolling shutter sensor based camera, and (3) a phase delay and amplifier (PDA) circuit.

The scanned laser projector contains red, green, and blue laser light sources; a MEMS micro-mirror based scanner; and drive electronics. Fig. 1(b) illustrates the layout of projection engine having scanning micro-mirror, laser sources, and optical combiners. The combined RGB laser light is reflected off the rotating micro-mirror to project the focus-free image on screen. The micro-mirror is a coupled 2D electromagnetic actuator that scans the input laser light beam in two directions to project full frame. The laser beam is projected in raster pattern moving from top left to bottom right as shown in Figure 2(a). The projector scans each pixel one by one and row-wise. The particular pixel position in the image is defined by the 2D angular position of the micro-mirror, while the pixel intensity and color value depends on the amount of red, green, and blue components in input light beam at particular time. The vertical scanning period (T) of the scanner is defined by the refresh rate of the projector, while the horizontal scanning period depends on both refresh period and vertical resolution of projector.

There are two types of shuttering techniques used in image sensors:[18,19] (1) global shutter and (2) rolling shutter. In the global shutter sensors, the shutter is released at
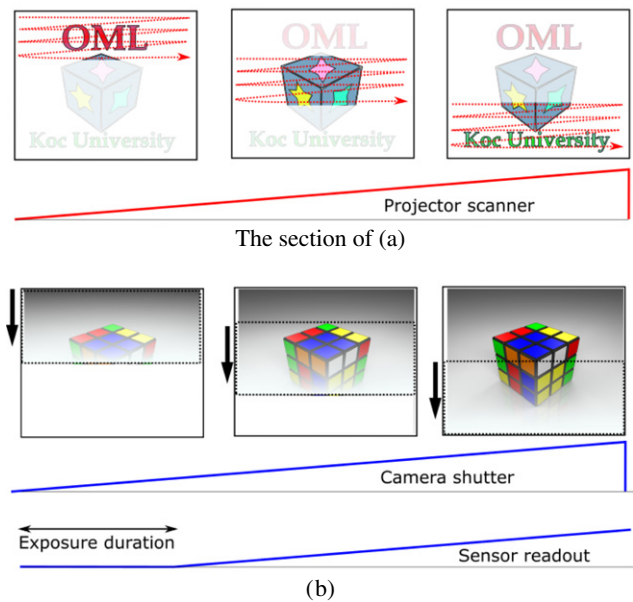

Figure 2. (a) Image projection mechanism of laser pico projector. The image is projected in raster pattern and one pixel at a time. (b) The light collection principle of rolling shutter image sensor. The shutter of each sensor row is released at different times and remains open for the defined exposure duration.

the same time for all sensor pixels, while in the rolling shutter sensors each sensor row opens and closes at slightly different times. The imaging in a rolling shutter sensor starts by releasing the shutter for the first sensor row, and the subsequent sensor rows start light collection at slightly different times as shown in Fig. 2(b). Each sensor row remains active for the predefined exposure duration (E). The number of active rows at a particular time instant depend on exposure duration. The first sensor row finishes the light collection first and whole frame is captured after last sensor row is closed. The total capture duration (D) for the full frame can be expressed as:
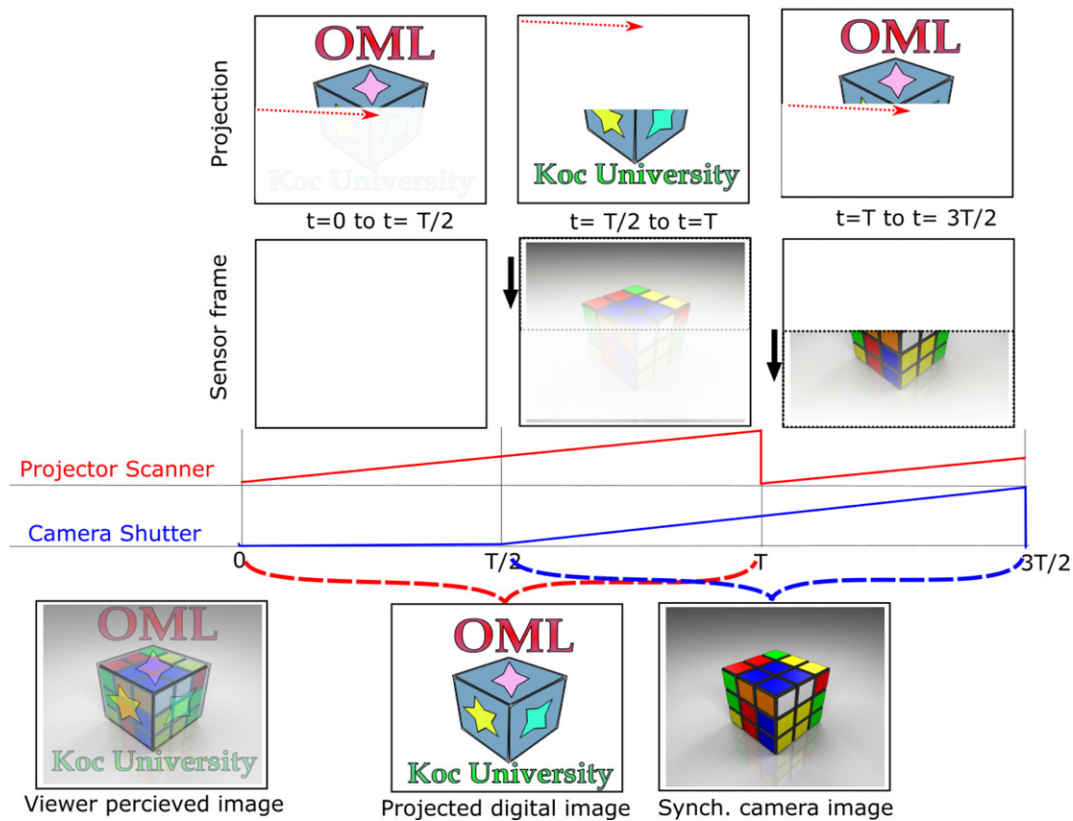
$$D = E + T_s \qquad (1)$$

Figure 3. The figure explains the concept of proposed technique, the viewer perceived content (mixture of projection and scene) is captured free of crosstalk by sequentially triggering the image sensor with the phase delayed projector's scanning signal.

where T_s is the scanning period of the image sensor, which can be different from that of projector. The sensor takes additional time to read pixel values, transmit them to other end, and clear the local memory of sensor before the new frame is captured.

The proposed technique utilizes the scanning properties of camera and MEMS mirror based laser projector to decouple real and digital content. The vertical scanning signal from the projector is used to synchronously trigger the image sensor. Figure 3 illustrates the running procedure: as the system starts ($t = 0$), the projector generates the vertical scanning signal, which starts light beam scanning pixel × pixel, while the image sensor remains idle at that time. The projector's scanner signal is fetched and passed through a phase delay and amplifier circuit to introduce half a period delay. The sensor shutter is released when the delayed signal triggers image sensor ($T/2$). By that time, the projection beam reaches the image center and first half of the image is already displayed. The image sensor then synchronously follows the projection beam with constant delay. At the end of the beam scanning period ($t = T$), the projector completes image display and image-sensor scanning reaches the center of the scene. In the following step, the projector's scanner rotates back to display new frame, while the image sensor captures the remaining half of the scene. The projector keeps scanning new frames continuously, while the image sensor takes additional time after each frame capture to allow sensor

readout and memory reset. The process is repeated once the sensor is ready to capture next frame.

To ensure that the camera and projector share same space for image capture and display, the camera is placed close to the projector. The lens of the camera is selected such that the vertical field of view of the camera is equal to that of projector and full projected area is always visible in camera frame. In the current implementation, we used a 50° field of view lens with image sensor. Different camera lenses including wide-angle lens (i.e., fish-eye lens) can also be used for imaging. When using a wide-angle lens, the captured images experience severe barrel distortion and viewing space of camera is expanded compared to that of projector. Due to lens distortion, each straight line in viewing space is imaged as curved line on sensor space and can introduce slight crosstalk from projector. The crosstalk in such a case can be removed by optimizing the trigger delay such that the sensor scanner does not cross the projector scanner at any time instant. A multi-camera structure with single projector can also be used to capture crosstalk-free images from different angles. Each camera in the setup can be either fed with the identical trigger signal having a fixed delay or can have different trigger delays for each camera depending on the geometric configuration of cameras with respect to projector.

Due to the difference in pixel resolution and position of projector and camera, the real object has different pixel locations in captured and projected image and geometric cal-

ibration is required to map the captured image to projected image and facilitate proper registration of digital content with real objects. The mapping between captured and projected images is performed by calculating the perspective geometric transform between camera and projector plane.[20],[21] The perspective transform calculates the scale (zoom), rotation, Translation, and tilt between captured and projected image. The perspective transform is determined by projecting four known points on the screen placed at the working distance (70 cm). All four projected points are captured on the camera and correspondence between captured and projected points is used to calculate the perspective transform. To project the virtual content registered with real object, the real object points are warped from camera plane to projector plane and the resulted image is displayed on projector. The calibration procedure discussed above is performed just once during the initial assembly of the system.

The real-world object detection and recognition is an important part of projection-based augmented reality systems. There are several object recognition and tracking techniques which rely on captured color images. Visible color leds associated with real objects are used to detect and recognize the real objects.[22] The use of 2D matrix markers of different shapes have also been demonstrated,[23] where printed 2D markers are attached to real object. The markers are captured using visible camera and are processed to retrieve the associated object information, including its position, size, and orientation. A 2D barcode technology have also been demonstrated to recognize the objects in augmented reality environments.[24] To facilitate the outdoor object recognition for augmented reality applications, the Loxel-based visual feature organization is also presented in literature,[25] where the captured input image is matched against a large data of location tagged images and the retrieved digital information is displayed based on the matching results. Any of the above-mentioned technique can be used together with proposed approach to facilitate improved object detection and content registration.

RESULTS AND DISCUSSION

The experimental setup as shown in Figure 4 was built to demonstrate the time multiplexed content decoupling approach. A 15 lumens 60 Hz scanned laser pico projector from Microvision (Show-Wx),[26] a rolling shutter CMOS sensor camera from Point-gray (Fea3, 3.2 MP, 60 fps),[27] and a low-noise preamplifier from Stanford research systems as phase delay and amplifier circuit were used. The capturing rate of up to 45 fps is achieved with zero crosstalk using the 60 Hz projector. If a higher frame rate camera is used, only up to 60 frames in a second can be captured with zero crosstalk, while partial crosstalk at predefined pixel positions can be observed in the remaining frames. Due to synchronized nature of camera and projector, the proposed technique can still provide more scene information compared to an unsynchronized system. The higher capturing rates with zero crosstalk can also be achieved by using the projectors with high refresh rate.
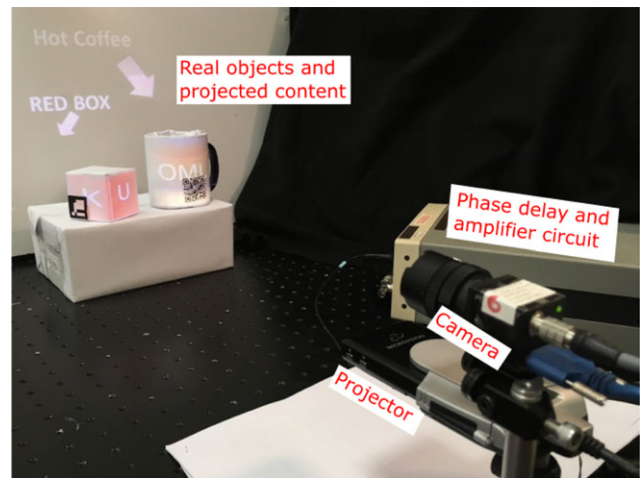

Figure 4. The figure shows the developed experimental setup, where real objects are overlaid with the digital information. The synchronized camera facing toward the scene captures the real objects using the proposed technique and a low-noise preamplifier is used as phase delay and amplifier circuit.

The experimental augmented reality system was tested by overlaying digital content on real-world objects (i.e., cups, boxes). Figure 5(a) shows the photos captured through system camera using the proposed technique; the camera does not capture the light from projector and only scene light is captured. Fig. 5(b) shows the photo captured when the projector and camera are not synchronized, while Fig. 5(c) was obtained by computing the difference between Fig. 5(a) and Fig. 5(b), which shows the digital content only with no scene information. Fig. 5(c) shows that the proposed technique can also be used together with the regular imaging method to subtract the scene light from the projected content. Due to the electronically driven synchronization between camera and projector, the content decoupling capability of the system is not affected by the artifacts inherent in rolling shutter sensors. Our experiments show barely noticeable image bending (skewness) and smearing when fast moving objects are captured. Since there were no sudden illumination changes in the scene, we did not notice the partial exposure artifact in captured images.

The crosstalk between imaging and display is measured by calculating the amount of projected light in captured frame. A white image was displayed on a regular projection screen (white paper) in dark environment and captured through image sensor. Three different photos were captured: (1) when the system is in operation as per the proposed technique, (2) when camera and projector work independently, and (3) when the projector is off. Furthermore, the crosstalk in captured content is calculated as:

$$Crosstalk = \frac{P_{sys} - P_{off}}{P_{reg} - P_{off}} 100\% \qquad (2)$$

where P_{sys} and P_{reg} are respectively the average pixel values when the proposed technique is used and when camera projector are not synchronized, while P_{off} is the average pixel value when projector is off. The proposed technique
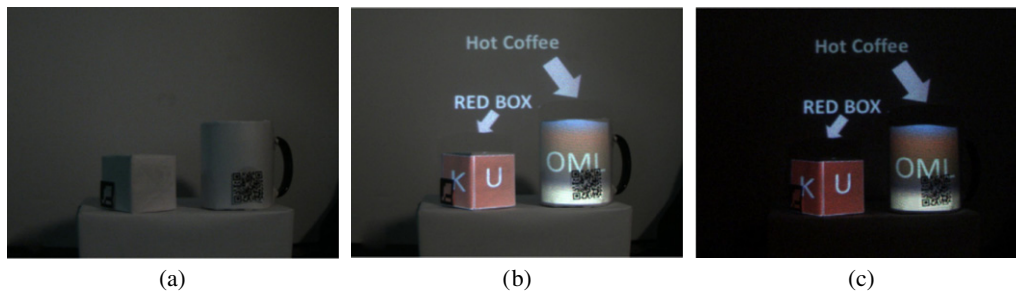
Figure 5. A projection-based augmented reality demonstration, where real objects (white box and cup with the QR code and AR marker) are overlaid with projected digital content (the texts, arrows, colors on the box and cup). (a) The image captured using proposed technique. (b) The image captured using regular image capture. (c) The image obtained by computing the difference of (a) and (b).
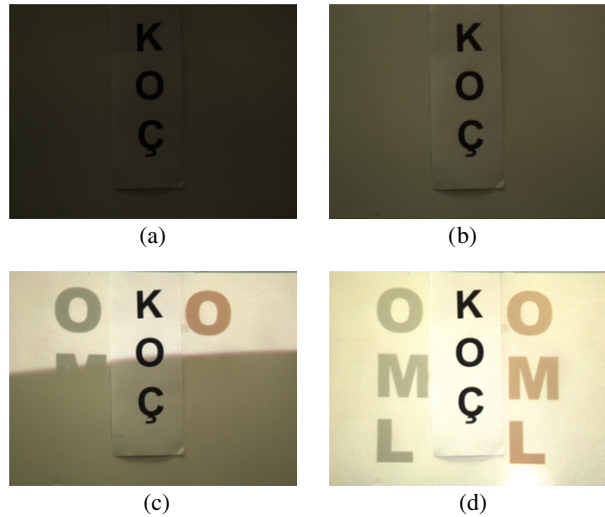

Figure 6. The images captured by system camera with different exposure durations (a) 4 ms, (b) 8 ms, (c) 12 ms, and (d) 16 ms. The camera shows no crosstalk up to 8 ms exposure. Longer exposure duration (> 8 ms) results in proportional increase in crosstalk.

completely skips the projected light in captured photo and provides same values for P_{sys} and P_{off} (as long as the exposure duration is within specified limit as discussed below), which results 0% crosstalk between captured and displayed content.

The maximum allowable sensor exposure depends on the vertical scanning period of projector. Each sensor row remains open for the defined exposure duration. Due to half-period scanning delay between image sensor and projection beam, the sensor shutter is released when projection beam has traveled to half of the scanning cycle. The projection scanner takes another half period to complete frame projection and scan new frame. For crosstalk-free imaging, it is important for a particular sensor row shutter to be closed before the scanning beam reaches that row. Considering this fact, the maximum allowable exposure duration can be defined as:

$$E \leq \frac{T}{2}. \tag{3}$$

To evaluate the effect of varying sensor exposure on system performance, four photos of a scene having printed text "KOC" (real object) on screen (white paper) and the projected text "OML" (digital content) were captured using the system camera at different exposure values (from 4 ms to 16 ms) as shown in Figure 6. Using 60 Hz pico projector, the camera captures real objects (printed text) only up to the exposure duration of 8 ms. The shorter exposure value (4 ms) as in Figure 6(a) results in reduced amount of captured light, which can be improved by using the cameras with highly sensitive sensors and large lens aperture. As the exposure duration is increased to 12 ms (Figure 6(c)), the camera starts to see the projected image, which results in about 50% crosstalk. The further increase in exposure duration (16 ms in Figure 6(d)) shows the full projected image seen in captured frame besides the actual scene and results in 100% crosstalk between imaging and display.

CONCLUSION
In this article, a simultaneous imaging and display technique for projection-based augmented reality systems was proposed. A MEMS scanner based laser projector and a rolling shutter image sensor were used to demonstrate crosstalk-free imaging and display. A signal delay and amplifier circuit was used to synchronously release sensor shutter with constant delay. The crosstalk between imaging and display

was calculated by measuring the presence of projected light in captured images.

The proposed technique can be used for a variety of spatial augmented reality applications, where capturing the decoupled augmented scene provides significant information about real objects. The proposed technique can also be used to capture the scene through transparent projection screens seamlessly. Our result shows 0% crosstalk between imaging and display and enough exposure limit (up to 8 ms) to facilitate real-life scene capture with decent quality. The proposed technique is valid for changing camera–projector configurations, provided that the camera and projector are geometrically calibrated.

REFERENCES
[1] H. Hua and B. Javidi, "Augmented reality: easy on the eyes," Opt. Photonics News **26**, 26–33 (2015).

[2] Microsoft HoloLens https://www.microsoft.com/microsoft-hololens/en-us, accessed June-2016.

[3] M. I. Olsson, M. J. Heinrich, D. Kelly, and J. Lapetina, "Wearable device with input and output structures", US Patent 20,130,044,042 (2013).

[4] A. D. Wilson and H. Benko, "LightSpace: Combining multiple depth cameras and projectors for interactions on, above and between surfaces," *Proc. 23rd Annual ACM Symposium on User Interface Software and Technology* (ACM, New York, NY, 2010), pp. 273–282.

[5] H. Benko, R. Jota, and A. Wilson, "MirageTable: freehand interaction on a projected augmented reality tabletop," *Proc. ACM SIGCHI Conf. Human Factors in Computing Systems* (ACM, New York, NY, 2012), pp. 199–208.

[6] J. Huber, J. Steimle, C. Liao, Q. Liu, and M. Mühlhäuser, "LightBeam: interacting with augmented real-world objects in pico projections," *Proc. ACM 11th Int'l. Conf. Mobile and Ubiquitous Multimedia* (ACM, New York, NY, 2012), pp. 2513–2518.

[7] R. S. Sodhi, B. R. Jones, D. Forsyth, B. P. Bailey, and G. Maciocci, "BeThere: 3D mobile collaboration with spatial input," *Proc. ACM SIGCHI Conf. on Human Factors in Computing Systems* (ACM, New York, NY, 2013), pp. 179–188.

[8] K. D. Willis, I. Poupyrev, S. E. Hudson, and M. Mahler, "SideBySide: ad-hoc multi-user interaction with handheld projectors," *Proc. 24th Annual ACM Symposium on User Interface Software and Technology* (ACM, New York, NY, 2011), pp. 431–440.

[9] C. Winkler, M. Löchtefeld, D. Dobbelstein, A. Krüger, and E. Rukzio, "SurfacePhone: a mobile projection device for single-and multiuser everywhere tabletop interaction," *Proc. 32nd Annual ACM Conf. on Human Factors in Computing Systems* (ACM, New York, NY, 2014), pp. 3513–3522.

[10] J. Kim, J. Seo, and T. D. Han, "AR Lamp: interactions on projection-based augmented reality for interactive learning," *Proc. ACM 19th Int'l. Conf. on Intelligent User Interfaces* (ACM, New York, NY, 2014), pp. 353–358.

[11] Z. Zhang, "Microsoft kinect sensor and its effect," IEEE Multimedia **19**, 4–10 (2012).

[12] A. Sand and R. Ismo, "Mixed reality with multimodal head-mounted pico projector," *Proc. ACM Virtual Reality Int'l. Conf.* (ACM, New York, NY, 2013), pp. 14–15.

[13] S. Asahi, Y. Manabe, N. Yata, and Y. Uranishi, "Overlayable and rotation-free transmissive circular color marker for augmented reality," *Proc. IS&T CGIV2012 Sixth European Conf. on color in Graphics, Imaging, and Vision* (IS&T, Springfield, VA, 2012), pp. 115–120.

[14] J. Park and K. Myoung-Hee, "Smart pico-projected AR marker," *Proc. ACM SIGGRAPH Asia Posters* (ACM, New York, NY, 2012).

[15] D. M. Krum, E. A. Suma, and M. Bolas, "Augmented reality using personal projection and retroreflection," *Personal and Ubiquitous Computing* (ACM, New York, NY, 2012), 16, pp. 17–26.

[16] D. Kade, K. Akşit, H. Ürey, and O. Özcan, "Head-mounted mixed reality projection display for games production and entertainment," Per. Ubiquitous Comput. **19**, 509–521 (2015).

[17] C. Genç, S. Soomro, Y. Duyan, S. Ölçer, F. Balcı, H. Ürey, and O. Özcan, "Head mounted projection display & visual attention: Visual attentional processing of head referenced static and dynamic displays while in motion and standing," *Proc. ACM SIGCHI Conf. on Human Factors in Computing Systems* (ACM, New York, NY, 2016), pp. 1538–1547.

[18] Global and Rolling shutters, http://www.red.com/learn/red-101/global-rolling-shutter, accessed in June-2016.

[19] D. Bradley, B. Atcheson, I. Ihrke, and W. Heidrich., "Synchronization and rolling shutter compensation for consumer video camera arrays," *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops* (IEEE, Piscataway, NJ, 2009), pp. 1–8.

[20] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," *Proc. Seventh IEEE Int'l. Conf. Computer Vision* (IEEE, Piscataway, NJ, 1999), pp. 666–673.

[21] S. Fernandez and S. Joaquim, "Planar-based camera-projector calibration," *Proc. 7th IEEE Int'l. Symposium on Image and Signal Processing and Analysis (ISPA)* (IEEE, Piscataway, NJ, 2011), pp. 633–638.

[22] M. Bajura and N. Ulrich, "Dynamic registration correction in augmented-reality systems," *Proc. IEEE Virtual Reality Annual Int'l. Symposium* (IEEE, Piscataway, NJ, 1995), pp. 189–196.

[23] J. Rekimoto, "Matrix: A realtime object identification and registration method for augmented reality," *Proc. 3rd IEEE Asia Pacific Computer Human Interaction* (IEEE, Piscataway, NJ, 1998), pp. 63–68.

[24] J. Rekimoto and A. Yuji, "CyberCode: designing augmented reality environments with visual tags," *Proc. ACM Designing Augmented Reality Environments* (ACM, New York, NY, 2000), pp. 1–10.

[25] G. Takacs, C. Vijay, G. Natasha, X. Yingen, W.-C. Chen, B. Thanos, G. Radek, P. Kari, and G. Bernd, "Outdoors augmented reality on mobile phone using loxel-based visual feature organization," *Proc. 1st ACM Int'l. Conf. Multimedia Information Retrieval* (ACM, New York, NY, 2008), pp. 427–434.

[26] Microvision Laser Pico-projector ShowWX http://www.microvision.com/product-support/showwx/.

[27] Point Grey Flea-3 camera, https://www.ptgrey.com/flea3-32-mp-color-usb3-vision-sony-imx036-camera.