3-D Interactive Augmented Reality-enhanced Digital Learning Systems for Mobile Devices

Kai-Ten Feng, Po-Hsuan Tseng[a], Pei-Shuan Chiu, Jia-Lin Yang, and Chun-Jie Chiu

Department of Electrical Engineering, National Chiao Tung University, Hsinchu, Taiwan;
[a]Department of Electronic Engineering, National Taipei University of Technology, Taipei, Taiwan

ABSTRACT

With enhanced processing capability of mobile platforms, augmented reality (AR) has been considered a promising technology for achieving enhanced user experiences (UX). Augmented reality is to impose virtual information, e.g., videos and images, onto a live-view digital display. UX on real-world environment via the display can be effectively enhanced with the adoption of interactive AR technology. Enhancement on UX can be beneficial for digital learning systems. There are existing research works based on AR targeting for the design of e-learning systems. However, none of these work focuses on providing three-dimensional (3-D) object modeling for enhanced UX based on interactive AR techniques. In this paper, the 3-D interactive augmented reality-enhanced learning (IARL) systems will be proposed to provide enhanced UX for digital learning. The proposed IARL systems consist of two major components, including the markerless pattern recognition (MPR) for 3-D models and velocity-based object tracking (VOT) algorithms. Realistic implementation of proposed IARL system is conducted on Android-based mobile platforms. UX on digital learning can be greatly improved with the adoption of proposed IARL systems.

1. INTRODUCTION

In the age of information exploration, education and self-learning becomes more and more important in the daily life. Digital learning or e-learning systems with the growth of the Internet make education not confined in the classroom. As the rapid popularization of smartphones, digital learning systems on mobile platform can realize the idea to learn anytime everywhere. It is essential to design digital learning systems with enhanced user experience (UX) to increase learning efficiency. With the advancements of the processing capability of mobile platforms, augmented reality (AR)[1] has become a potential technology for enhanced UX. The virtual information such as videos and images can be imposed onto a digital display through AR technology. Furthermore, integrating AR technology with user interaction capability can draw user interests for digital learning system.

In general, AR on mobile platforms can be divided into two categories. One is to impose geographical information, e.g., place information, onto real-time captured views. This category is suitable for the context-aware digital learning system such as guide system. However, users location is required in this category. Though users location can be acquired In the GPS-equipped smartphone today, position accuracy is still not satisfied for this kind of application especially in the indoor environment. The other is to detect marker or markerless image and replace the image with corresponding three-dimensional (3-D) object on screen. Note that it requires pattern recognition to detect this specific image with high CPU loading. There are existing research works based on AR targeting for the design of e-learning systems.[2–4] However, none of these work focuses on providing 3-D object modeling for enhanced UX based on interactive AR techniques.

In order to interact with e-learning systems for enhanced UX, it is essential to perform real-time object tracking to recognize the commands or orders from the user. There are several research works targeting on

Further author information: (Send correspondence to Kai-Ten Feng)
Kai-Ten Feng: E-mail: ktfeng@mail.nctu.edu.tw
Po-Hsuan Tseng: E-mail: phtseng@ntut.edu.tw
Pei-Shuan Chiu: E-mail: mean11809.cm97@nctu.edu.tw
Jia-Lin Yang: E-mail: brian37.eed98@nctu.edu.tw
Chun-Jie Chiu: E-mail: jack0502801.am98@nctu.edu.tw

The Engineering Reality of Virtual Reality 2013, edited by Margaret Dolinsky, Ian E. McDowall,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 8649, 864908 · © 2013 SPIE-IS&T
CCC code: 0277-786X/13/$18 · doi: 10.1117/12.2003505

real-time vision-based object tracking. Hand tracking with the color glove[5] identifies hand gesture with high precision. However, marker-based tracking is not suitable for mobile platform. Tracking object using background subtraction[6,7] can identify the position of moving object, but steady background is required for subtraction which might not be suitable for moving object. Real-time hand tracking based on depth images identifies foreground object fast and reliable, but depth images[8] can only be acquired from time-of-flight camera or stereo camera, which are not commonly equipped in mobile device. Color based hand tracking[9] can be implemented on different mobile platforms because it only requires a color camera, which is generally equipped in mobile device. The method also allows the mobility of the camera. However, it may encounter mismatches when the object moves in the area where the background has similar color as the object.

In this paper, an interactive augmented reality-enhanced learning (IARL) system is proposed to combine both of the technologies. When readers use a mobile phone with a camera to read the picture from a picture book, the proposed IARL system generates a 3-D AR object for this picture on the display. Meanwhile, the design of IARL system also allows user to interact with 3-D AR objects which makes digital learning more attractive, e.g., user's hand moving to roll the 3-D object. The markerless pattern recognition (MPR) is proposed as the AR technology to the IARL system because image recognition can be more useful for e-learning without the place limitation. Considering that IARL system is realized on mobile device, the color-based tracking method which has lower computational complexity is conducted. However, color-based hand tracking may fail to detect the hand when the background has similar color as hand color. The velocity-based object tracking (VOT) is proposed to enhance the hand color tracking with the consideration of the hand moving velocity. Therefore, the VOT algorithm detects hand moving to roll the 3-D object generated by MPR. The design of the IARL system, which transfers static pictures to 3-D AR objects on a mobile phone and allow interaction with the objects, generates a new way to draw attentions from the readers compared to traditional education materials.

The paper consists of four sections. In Section 2, the proposed IARL system is described. Section 3 presents the results and the discussion of the proposed IARL system with examples from picture books that are followed by the conclusion in Section 4.

2. PROPOSED IARL SYSTEM

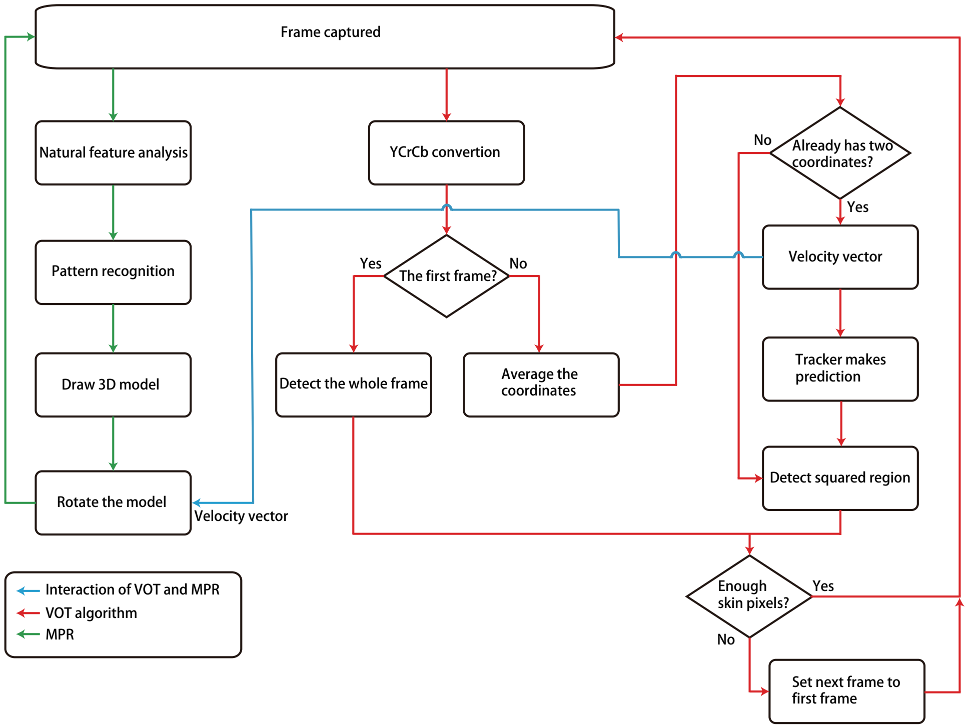


Figure 1. Flowchart of the proposed IARL system

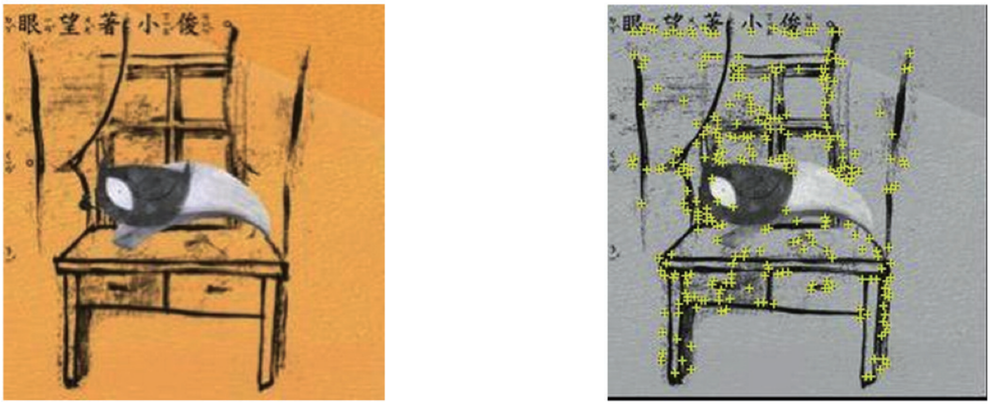
Fig.1 is the flowchart of the proposed IARL system. It comprises two parts - the MPR algorithm in the left part and the VOT algorithm in the right part of Fig.1. The first step of the MPR algorithm is natural feature analysis. Feature points are captured from every incoming frame and stored into an array. The second step is pattern recognition, which compares the feature points in the stored array and the device database, to decide if there is any 3-D object in the database contained in the captured frame. If the decision is true, the system draws the 3D object in the third step. After that, the velocity estimated from the VOT algorithm is utilized to decide whether the object is rotated in the fourth step. In the VOT algorithm, each frame converts to YCrCb space first. In YCrCb space, Y represents brightness information, while Cr and Cb represent color information. In the first frame detection, it scans the whole frame and estimates the hand coordinates as the initial value. When the second frame is captured, the system starts to scan pixels around the hand coordinates in previous frame. After consecutive coordinates of user's hand are captured, hand moving velocity can be estimated and serves as an input to the MPR algorithm to rotate the object. In the following subsections, the MPR and VOT algorithm are introduced in details.

2.1 Proposed MPR Algorithm for 3-D Models

In order to implement the markerless pattern recognition, Qualcomm Vuforia[10] is adopted to identify the real-world target instead of the marker which only comprises black and white shapes. The process of the markerless pattern recognition is divided into two sections including natural feature analysis and pattern recognition.

2.1.1 Natural Feature Analysis

Identifying a real-word target is a challenging task because large amount of pixels computation are required. To speed up the recognition process while sacrificing some sort of precision is acceptable, only special and explicit points of images are extracted. Vuforia adopts a technique called natural feature to simplify the recognition process. Natural feature is a sharp, spiked, chiseled detail in the image. It is defined as a sharp turning point at the edge of two contrast colors. The natural features can be utilized to describe the features of the target image simply because they are the most apparent parts of an image for computers to recognize. For example, a frame with a black rectangular on white background, the natural features would at four corners of rectangular since they are the sharp turning points at the edge and their colors, i.e., black and white, are contrast to each other. Another example, a black circle on white background, there would be no natural feature because there is no sharp turning point at the edge. Vuforia SDK offers an online target management system for user to analysis its own image target. After natural features analysis, those feature points will be stored into an array and output to a pattern file. In Fig.2(b), yellow stars represent the feature points of Fig.2(a) to serve as an example for natural feature analysis. In the proposed IARL system, the natural feature analysis is applied to detect the feature points. The natural feature points of the images in the picture books are stored in advance to offer a database for online pattern recognition.



(a) Original picture

(b) Feature points of the picture

Figure 2. Natural feature analysis

2.1.2 Pattern Recognition

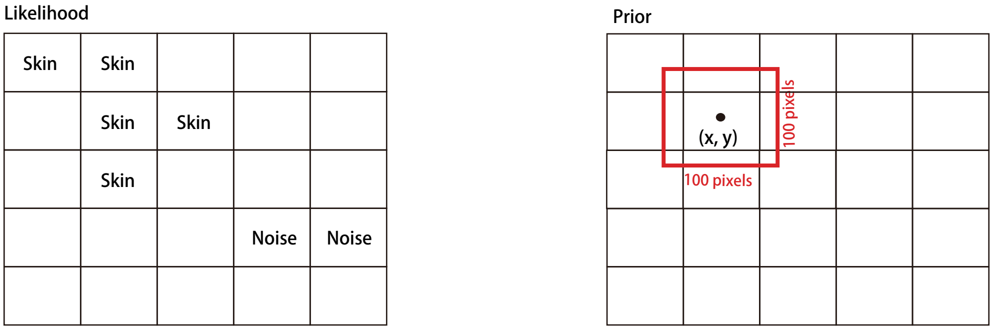
In the proposed IARL system, pattern recognition detects if any specific targets from the picture books present in the captured frame. If so, the 3-D object is displayed and the size of the 3-D object can be determined by the distance between the book and the camera. A database is required for the pattern recognition to decide if the current frame contains any pictures from the picture book. There are two ways to build the pattern database, i.e., device database and cloud database. Device database is more common for database building and the pattern files are imported in the device for online pattern recognition. Cloud database is the most popular tendency recently because of the progress of the Internet. With high speed Internet network, the target image can be transferred through the network and the pattern recognition can be achieved with the database on cloud. In the proposed IARL system, device database is introduced for pattern recognition. The proposed IARL system establishes mathematic model for the 3-D object. Those mathematic models can be used to compute the 3-D coordinates for the recognized target. The target position will be returned so it can establish the coordinate on the image to build the 3-D object and the rotation of the object can be calculated real-time based on the mathematic models.

2.2 Proposed VOT Algorithm

The concept of Bayesian estimation is utilized in the proposed VOT algorithm to estimate the hand location for the interaction purpose. The camera captures a frame and skin color is detected pixel-by-pixel. To estimate the hand location, the pixels with skin color are defined as the hand location. However, determining hand position by skin color may be interfered by the other skin color objects or other people's hands in the background. To enhance the performance, Bayesian estimation is introduced in this algorithm as,

$$P(\mathbf{x} \mid \mathbf{y}) \propto P(\mathbf{y} \mid \mathbf{x}) P(\mathbf{x}) \tag{1}$$

where $\mathbf{x}$ is the hand position and $\mathbf{y}$ is the color frame captured from the camera. $P(\mathbf{x} \mid \mathbf{y})$ is the posterior probability of the hand position given the evidence of the captured pixels. $P(\mathbf{y} \mid \mathbf{x})$ is the likelihood function, which is defined as the skin distribution of total frame captured by camera. $P(\mathbf{x})$ is the prior probability, which is the information of hand region in previous frame.



(a) Likelihood of the frame

(b) Prior of the frame

Figure 3. Bayesian concept in the VOT algorithm

Through the Bayesian estimation, the posterior probability considers both the prior information from the estimated hand region in previous frame and the likelihood function based on the captured frame. The hand tracking performance can be enhanced with the knowledge of posterior probability. When total frame is utilized to detect the hand position, other skin-color objects which are not of interests may be included. As shown in Fig.3(a), the hand position is at the left upper corner. However, there are other skin-color objects at the right corner denoted as noise. With estimated hand position in previous frame, as shown in Fig.3(b), the prior information of the hand position is at the left upper corner. The prior region as the red box in Fig.3(b) represents the prior information which is centered at the previous estimated hand position. By multiplying the prior information and the likelihood function in (3), it removes the possibility of the hand location in the right corner region. Therefore, Bayesian estimation enhances the performance by considering hand tracking in the proposed algorithm.

In addition, the constant velocity hand movement model is also adopted as the state update model to serve as the prior information. By assuming user's hand moves with a constant velocity over different time instants, next hand position can be predicted based on the estimated velocity. The prior region as the red box in Fig.3(b) moves to the location of the predicted position. With the information of velocity, the proposed VOT algorithm can adapt to the hand movement faster than the scheme without the information of velocity, because the prior region can move to the location where user's hand may appear in the next frame first.

Next, the VOT algorithm is illustrated in detail. In the beginning of the algorithm, the frame is transformed into YCrCb color space and further divided into 16×16 blocks, where the pixel number of each block is relative to the resolution of the display. Note that Cr and Cb represent color information with the full 8-bit range of $0-255$. In the VOT algorithm, the skin pixel is detected if the Cr and Cb color information of the pixel satisfies the following equation,

$$\text{skin color} : \begin{cases} 77 \leq \text{Cb} \leq 127 \\ 133 \leq \text{Cr} \leq 173 \end{cases} \tag{2}$$

If the number of skin pixels in a block is over the pre-defined threshold, e.g., 60% of pixels in a block are detected as skin color, this block is regarded as a skin block. After every blocks in a frame is detected, the hand coordinates are estimated as the average coordinates of the skin block as the following equation,

$$(x, y) = \frac{\sum_{i=1}^{N}(x_i, y_i)}{N} \tag{3}$$

where (x, y) denotes the hand coordinates and (x_i, y_i) represents the coordinates of each skin block. N represents total skin block number.

After detecting hand coordinates in the first frame, the region near the hand coordinates is served as the prior information for next time instant. The prior region as the red box in Fig.3(b) is chosen as 100×100 pixels squared region center at previous hand coordinates, where the region size can be determined according to different scenarios. As the prior information shown in Fig.3(b), the region outside this prior region has zero prior probability. When detecting hand's coordinates in the second frame, the skin color detection only focus on this prior region. Other skin-color objects appear outside the squared region will be filtered out because the VOT algorithm only detects the skin color inside the prior region. Consequently, the tracking error and the computational complexity can both be reduced.

After hand positions at the two consecutive time instants are estimated, the velocity information is obtained through the following equation,

$$(v_x^n, v_y^n) \propto [(x^{n-1}, y^{n-1}) - (x^{n-2}, y^{n-2})] \tag{4}$$

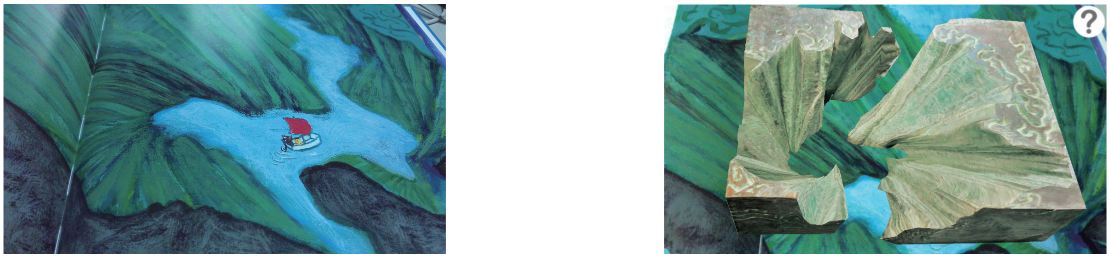
where v_x^n represents the velocity along x-axis in nth frame, v_y^n represents the velocity along y-axis in nth frame. Note that the velocity is measured as the hand position difference per sample time between two consecutive frames. x^n and y^n represent the coordinates of user hand in nth frame. From (4), the hand coordinates can be predicted by the velocity information starting from the third frame as the following equation,

$$(x^n, y^n) \propto (x^{n-1}, y^{n-1}) + (v_x^n, v_y^n) \tag{5}$$

The prior region moves to the predicted coordinates and the proposed VOT algorithm starts to detect skin pixels in this squared region which center at the predicted position. The prior information combines the previous estimated hand position with the velocity information. Therefore, hand location can be tracked with the constant velocity movement model. By sequentially repeating these steps with the incoming frames, the user's hand can be estimated with more reliable precision and lower computational complexity. Providing that the hand tracking lose track, i.e., the pixels in the prior region does not reach the skin color threshold, the VOT algorithm detect the whole frame automatically and then find a new prior region to detect user hand position in next frame.

3. RESULT AND DISCUSSION

First, the results of MPR algorithms are presented to discuss the pattern recognition and 3-D object generation. Note that the proposed IARL system including the MPR and VOT algorithm is implemented on the Android-based mobile phone. Fig.4(a) is an two-dimension (2-D) image of a picture book. Fig.4(b) shows the same picture and the 3D object of the sculpture by MPR technique. When mobile camera captures a frame with a specific pattern in the picture book, the 3-D object is generated with the pre-specified 3-D model. The result shows that the 3-D AR-enabled canyon sculpture is generated at the position where the camera of mobile device captures the 2-D version of canyon sculpture. As shown in Fig.4(b), the 3-D object is generated and displayed on the screen to overlaid with the 2-D model. The coordinates of the 3-D object would be calculated real-time according to the distance from the camera to the picture book or the rotation of the mobile camera. For example, different view angle of the 3-D object can be achieved by rotating the mobile camera.

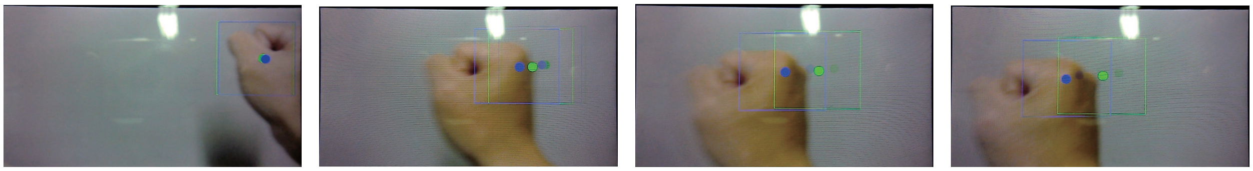


(a) 2D image in picture book

(b) 3D model of the image

Figure 4. Results of the proposed MPR algorithm

To illustrate the effectiveness of hand tracking in the proposed VOT algorithm, Bayesian method with and without velocity information are compared in Fig.5. In the example, user's hand moves fast from right to left as shown in the frames captured consecutively from Figs.5(a) to 5(d). The green color denotes the algorithm without velocity information, while the blue color represents the proposed VOT algorithm. When user's hand is static, blue point and green point locate in the same coordinates as shown in Fig.5(a). As hand starts to move, blue point can predict the next hand coordinates and moves first in Figs.5(b), 5(c) and 5(d). The VOT algorithm adapts to the hand movement faster than Bayesian algorithm without velocity information. This example demonstrates that the prediction in the VOT algorithm with constant velocity movement model improves the tracking performance.



(a)

(b)

(c)

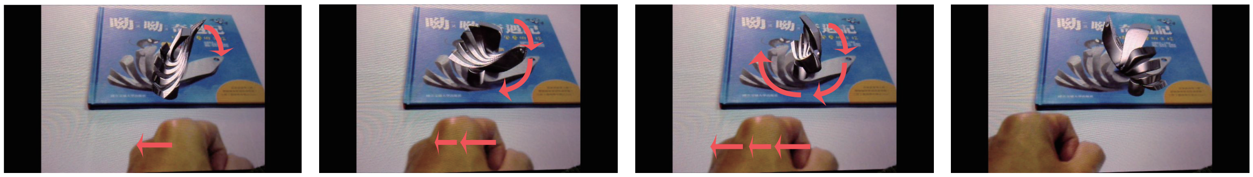
(d)

Figure 5. Results of the proposed VOT algorithm

Finally, the example by utilizing the IARL system which combines the MPR and the VOT algorithms is illustrated in Fig.6. The interaction with the 3-D AR object is initiated by the hand movement. As shown in Fig.6, the user hand movement makes the 3-D AR object recognized from a 2-D picture rotate clockwise or counterclockwise. When user's hand moves from left to right, the model rotates counterclockwise. From Figs.6(a) to 6(d), a phoenix rotates clockwise when user's hand moves from right to left. The example demonstrates that the proposed IARL system can interact with 3-D AR objects on the mobile devices.

4. CONCLUSION

Implementation of the proposed interactive augmented reality-enhanced learning (IARL) system is conducted on Android mobile platform. The markerless pattern recognition (MPR) algorithm recognizes the specific target



(a) (b) (c) (d)

Figure 6. Results of the proposed IARL system

from the camera captured two-dimensional (2-D) frame to generate the three-dimensional (3-D) augmented reality (AR) object. The proposed velocity-based object tracking (VOT) algorithm tracks the hand movement by combining the skin color detection and the velocity information. Therefore, user not only reads the picture book but also interact with the 3-D AR object on mobile phone through the hand movement. The proposed IARL system enhances user experiences, which are beneficial for digital learning systems. The interaction with 3-D AR objects on mobile phone can increase the learning efficiency such as draw the reader's attention or create a new way for people to learn. However, there is still a room to reduce computational complexity. To further reduce the computation complexity of the IARL system for more accurate pattern recognition or hand tracking is considered as the future work.

ACKNOWLEDGMENTS

This work was in part funded by NSC 101-2218-E-027-006, NSC 101-2628-E-009-003, NSC 101-2219-E-009-027, Aiming for the Top University and Elite Research Center Development Plan, National Chiao Tung University (NCTU) Press, and Digital Content Production Center at NCTU.

REFERENCES

[1] J. Bolter and B. Macintyre, "Is It Live or Is It AR?," *IEEE Spectrum* **44**, pp. 30–35, Nov. 2007.

[2] C. Hughes, C. Stapleton, D. Hughes, and E. Smith, "Mixed Reality in Education, Entertainment, and Training," *IEEE Computer Graphics and Applications* **25**, pp. 24–30, Nov. 2005.

[3] S. Lee, J. Choi, and J. Park, "Interactive e-learning System using Pattern Recognition and Augmented Reality," *IEEE Transactions on Consumer Electronics* **55**, pp. 883–890, Nov. 2009.

[4] M. Sugimoto, "A Mobile Mixed-Reality Environment for Children's Storytelling Using a Handheld Projector and a Robot," *IEEE Transactions on Learning Technologies* **4**, pp. 249–260, Nov. 2011.

[5] A. A. Argyros and M. I. A. Lourakis, "Real-Time Tracking of Multiple Skin-Colored Objects with a Possibly Moving Camera," in *Proc. of 8th European Conference on Computer Vision (ECCV)*, pp. 368–379, 2004.

[6] H. Ribeiro and A. Gonzaga, "Hand Image Segmentation in Video Sequence by GMM: A Comparative Analysis," in *Proc. of 19th Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI 2006)*, pp. 357–364, Oct. 2006.

[7] C.-P. Chen, Y.-T. Chen, P.-H. Lee, Y.-P. Tsai, and S. Lei, "Real-time hand tracking on depth images," in *Proc. of IEEE Visual Communications and Image Processing (VCIP)*, pp. 1–4, Nov. 2011.

[8] R. Y. Wang and J. Popović, "Real-time Hand-tracking with a Color Glove," *ACM Transactions on Graphics* **28**(3), 2009.

[9] S. Malik, "Real-time Hand Tracking and Finger Tracking for Interaction," tech. rep., Department of Computer Science, University of Toronto, 2003.

[10] *Vuforia Developer*, Qualcomm.