https://doi.org/10.2352/ISSN.2470-1173.2018.10.IMAWM-454

Learning Enhancement with Mobile Augmented Reality

Xunyu Pan, Joseph Shipway, and Wenjuan Xu
Department of Computer Science and Information Technologies, Frostburg State University, Frostburg, Maryland, USA

Abstract

Traditional text-based instruction may not effectively inspire the motivation for learning especially for those young students attending K-12 schools. However, the booming of mobile devices and multimedia technologies can significantly enhance the effectiveness of the learning process and strengthen student engagement. In this work, we propose a novel mobile knowledge learning system based on Augmented Reality (AR) technology to improve the learning experiences for many users. In our AR system, virtual entities are created and superimposed over real-world images or video streams. Appeared to exist in the real world defined by an image or a video, these virtual entities can directly interact with real-world objects and respond to human activities. Depending on the source of camera input, which can be a static image or a video stream, the proposed mobile AR system supports both the demonstration of physical concepts and the rendering of 3D models. We evaluate the performance of the proposed system on the efficiency and effectiveness of the rendering of virtual AR entities under various conditions. Experimental results demonstrate our system supports real-time AR rendering and provides highly interactive learning experiences for different types of users including K-12 students.

Introduction

The past decade has seen the education system in the United States gradually integrated new technologies such as computers and Internet into classrooms, creating a blended learning environment. However, traditional text-based instruction and tutorial depict prototypical examples that do not represent the diverse examples in the real-world [1]. Thanks to the booming of mobile devices and multimedia technologies [2, 3, 4], educators are now able to utilize various digital learning tools to effectively inspire the motivation for learning. Using these technologies significantly enhances the effectiveness of the learning process and strengthen student engagement especially for those young students attending K-12 schools.

Among many innovative technologies supporting the education for children and the training for adults, *Augmented Reality* (AR) is a technology applied to digital devices to integrate together the reality and virtual worlds by adding a virtual overlay to real world scenarios. AR is a view of the physical world whose elements are augmented by virtual and artificial entities (e.g., images, animations, and videos). Educators are able to use AR technology to enhance students' educational experience by allowing students to interact with 3D objects in their environment. It was shown[5] that students who use AR while learning content, were more likely to retain the information and construct real-world applications of the material, verses students who learned in a more traditional instruction method. Instruction with AR supports student-centered learning as it allows the students to fully grasp the meaning of a subject topic with interactive demonstration with 3D illustrations.

To enhance these advantages, we propose a novel mobile knowledge learning system based on AR technology to improve the learning experiences for different types of users including K-12 students. We incorporate AR technology into the mobile system to create various computer generated entities in a hybrid and interactive environment. The mobile AR system is developed using Java language, *OpenCV* [6] algorithms for webcam input and page determination, *Tess4j* [7] for OCR text recognition, and *JOGL* for model rendering with *OpenGL* [8] library. In our AR system, virtual entities are created and superimposed over real-world images or video streams. Appeared to exist in the real world defined by an image or a video, these virtual entities can directly interact with real-world objects and respond to human activities. For example, when the AR system detects a section of text description about a *Dog* in one page of a book, a 3D virtual dog can be automatically generated and pop up out of that specific page. Users can also rotate the mobile device or even the book for viewing the 3D model from different angles. For another example, users can study the concept of *Reflection* and *Gravity* in physics by observing a virtual ball falling and bouncing with the edges of various objects (e.g., human body, chair, and whiteboard) in a real-world scene. To this purpose, a physics engine is implemented to realize the interaction of the computer generated ball with those real-world edges by estimating the reflection angle when a collision occurs. The virtual ball is able to interact with the environment because of a series of image filters and edge detectors supported by the physics engine.

Starting with the environment of a blank canvas, the proposed mobile AR system can be employed to process two types of camera input. The user can determine whether they would like to use static images or real-time video as system input. In the former case, the system takes a single user-selected image as input, whereas in the latter case it takes input from a webcam allowing virtual objects to react dynamically in real time to its surrounding environment. The real-time option makes the interaction between the users and the AR system to be possible, creating a more interactive learning experience. We evaluate the performance of the proposed system on the efficiency and effectiveness of the rendering of virtual AR entities under various conditions. Survey results also demonstrate positive student perceptions for using the AR based system to study new knowledges. The learning quality is substantially enhanced in this hybrid and interactive environment which provides a better understanding of subject matter than with traditional instruction approaches.

Related Work

As an important user interface technology, AR has experienced exciting developments during the past few years. People believe that AR has many potential implications and numerous applications in the context of teaching and learning. Currently some popular application fields are AR books, AR gaming, discovery-based learning, object modeling, and skill training [5]. However, the learning enhancement for K-12 students requires educators to engage, stimulate, and motivate students to explore class materials. Hence we are particularly focused on the *Concept Learning* using augmented multimedia content as it helps foster student imagination and creativity [9].

A closely related application field to our research is AR books, where AR technology is utilized in combination with mobile devices to offer students 3D presentations and interactive experiences when they read the book contents. For readers who still like printed books [10], AR books digitally enhance printed books with 3D rendering animation to bridge the gap between the physical and digital world. For example, *MagicBook* is an AR interface system allowing animated or interactive 3D content drawn from any printed book [11]. Children can actively participate in a story as the interface system permits AR content to be produced for a traditional book. Another category of AR books is a pop-up book showing 3D characters when readers wear special glasses such as *Dialogbooks* [5]. Moreover, as a web-based online tool, *Zooburst* [12] allows educators to design their own AR pop-up books. Authors can arrange characters within a 3D world consisting of customized items stored in a built-in database.

While the above described techniques enhance the learning process through 3D illustrations for printed books, little effort has been made to specifically address the direct interaction between virtual AR entities and real-world objects including human activities. Moreover many recent works implement the AR rendering using QR code [13], number[14], and marker [15, 16] without text recognition support for more meaningful AR representation.

Methods

In this section we describe a novel AR based mobile knowledge learning system. The goal of the proposed AR system is to improve the quality of the learning process for most users including K-12 students. The AR technology is integrated into the mobile system to create various 3D entities in a hybrid and interactive environment. In our AR system, virtual entities are fused with real-world images or video streams. Unlike most existing AR systems, these virtual entities can directly interact with those real-world objects and human activities as if they exist in the real world.

The proposed AR system can be deployed on any mobile device and requires no additional hardware equipment. Mobile users interact with the AR system through a control panel as shown in Figure 1. The system environment starts with a blank canvas with two running modes: (a). *3D Model Rendering*. In this mode, a 3D virtual model can be automatically generated on the top of a printed page containing related text description. (b). *Demonstration of Physical Concepts*: In this mode, physical concepts are displayed through the interaction between a virtual entity and various objects in the scene. For these two modes, both static images and video streams can serve as the source of system input. In addition, the ability to process real-time video makes the interaction

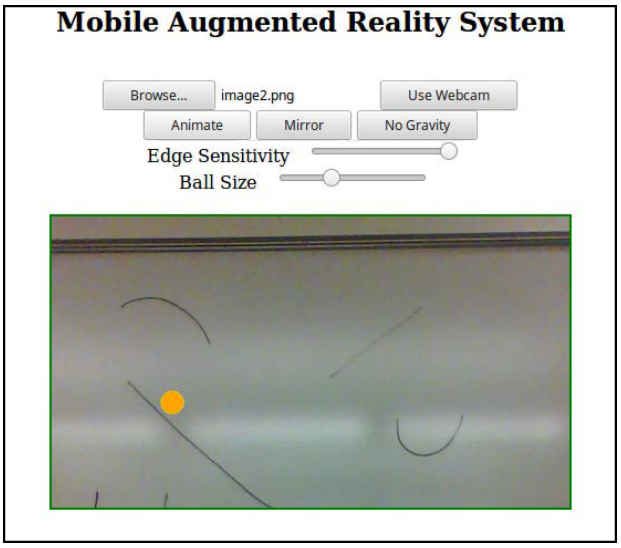


Figure 1. *The Graphical User Interface (GUI) of the proposed mobile AR system.*

between the users and the AR system to be possible, creating a more interactive learning experience.

The proposed mobile AR system consists of four major functional components:

1. **Motion Estimation**: The mobile AR system helps users to study physical phenomenon such as mechanical concepts *Reflection* and *Gravity*. A physics engine is employed to estimate the motion of computer generated entities when they interact with various objects (e.g., human body, chair, and whiteboard) in a real-world scene.
2. **Page Determination**: For accurate text recognition and correct model rendering, the precise location and extent of a page in a printed publication (e.g. journal, magazine, and book) is determined by detecting the largest convex quadrilateral in a given image or video frame.
3. **Text Recognition**: The printed text in a detected page is converted to machine-encoded characters using the *Optical Character Recognition* (OCR) technique. For more accurate recognition, the printed page is warped into a new page with standard viewing angle using estimated perspective transformation.
4. **Model Rendering**: Based on the analysis of existing scene, various virtual entities are superimposed over the current real-world image or video stream. If no printed page is detected, a computer generated ball is rendered to fulfill the interaction of this ball with those real-world objects. When a printed page is detected and the corresponding text is recognized by the AR system, a 3D virtual model can be automatically rendered and pop up out of that specific page. Users can further control the mobile device to observe the 3D model from different viewing angles.

Shown in Figure 2 is the high level logic overview of the described mobile AR system. The Page Determination, Text Recognition together with Model Rendering modules serve for the 3D model rendering based on the recognized text on a detected page.

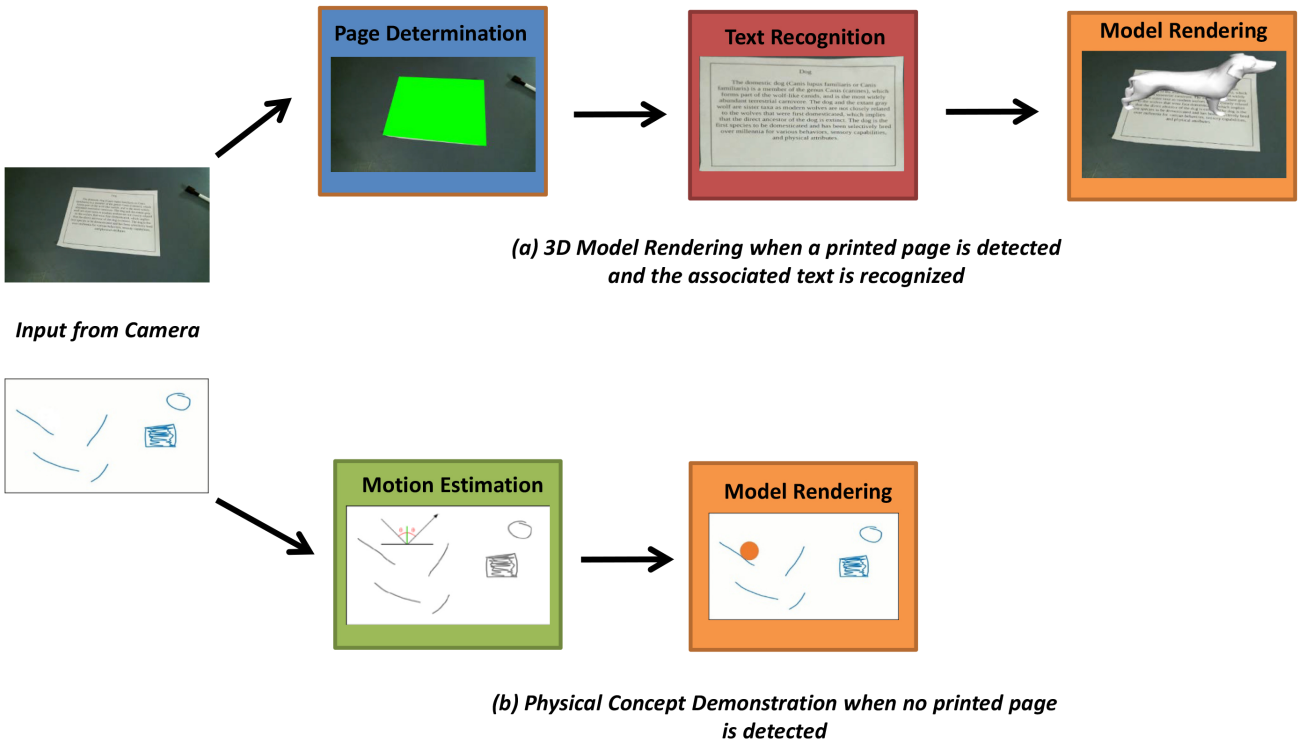


Figure 2. *Our mobile AR system performs differently depending on the input from system camera: (a) 3D virtual Model is rendered when a printed page is detected and the associated text is recognized; (b) Physical concept is demonstrated when no printed page is detected. .*

The Motion Estimation and Model Rendering modules serve for the physical concept demonstration based on the motion estimation for a specific virtual entity.

3D Model Rendering

3D Models are typically rendered over video streams in real-time. Each individual video frame retrieved from a real-time video is processed separately. We describe in detail the entire process of 3D model rendering in a specific video frame with Figure 3 illustrating the main steps of our method.

First, the Page Determination module identifies the largest convex quadrilateral in each video frame aiming to detect one page in a printed book or magazine. More specifically, *Canny* edge detector is used to find all major edges in a video frame. Note that all video frames are compressed into smaller size for rapid processing and noise reduction. We further use mathematical morphological operations to smooth and connect those detected edges. All image regions connected by those edges are analyzed to identify a printed page which is a convex polygon with four corners and the largest area in the current scene. *OpenCV* algorithms are extensively used to perform the above operations.

Next, the Text Recognition module employs OCR technique to retrieve words in the detected page using *Tess4j*, a Java binding for Tesseract OCR software. All retrieved words are sorted based on their Tesseract's confidence level for later usage.

Finally, the Model Rendering module performs the automatic rendering of a 3D virtual model on the top of a printed page detected in the current scene. To precisely locate the 3D model being rendered in the camera coordinate system, *OpenGL* uses a 4×4 *Model View Matrix* to represent the transform from the world coordinate to the camera coordinate. The Model View Matrix $\mathbf{V}$ can be computed as the addition of the *Rotation Matrix* $\mathbf{R}$ and the *Translation Matrix* $\mathbf{T}$, or more explicitly:

$$\boldsymbol{V} = \boldsymbol{R} + \boldsymbol{T} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & 0 \\ r_{21} & r_{22} & r_{23} & 0 \\ r_{31} & r_{32} & r_{33} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & t_x \\ 0 & 0 & 0 & t_y \\ 0 & 0 & 0 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Here the Rotation Matrix represents the rotation from the world coordinate to the camera coordinate, while the Translation Matrix represents the translation from the origin in the world coordinate to the camera coordinate. Since the Y and Z axes of *OpenCV* and those in *OpenGL* are in the opposite direction, the Model View Matrix $\mathbf{V}$ should be inverted in *OpenGL* as:

$$\boldsymbol{V'} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ -r_{21} & -r_{22} & -r_{23} & -t_y \\ -r_{31} & -r_{32} & -r_{33} & -t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

For each individual video frame, a 3D model is rendered at the precise location in the camera coordinate system based on the camera pose estimated in real time. Additionally, a local model database is searched to retrieve the most related 3D model for the recognized word with the highest confidence level in the current video frame.

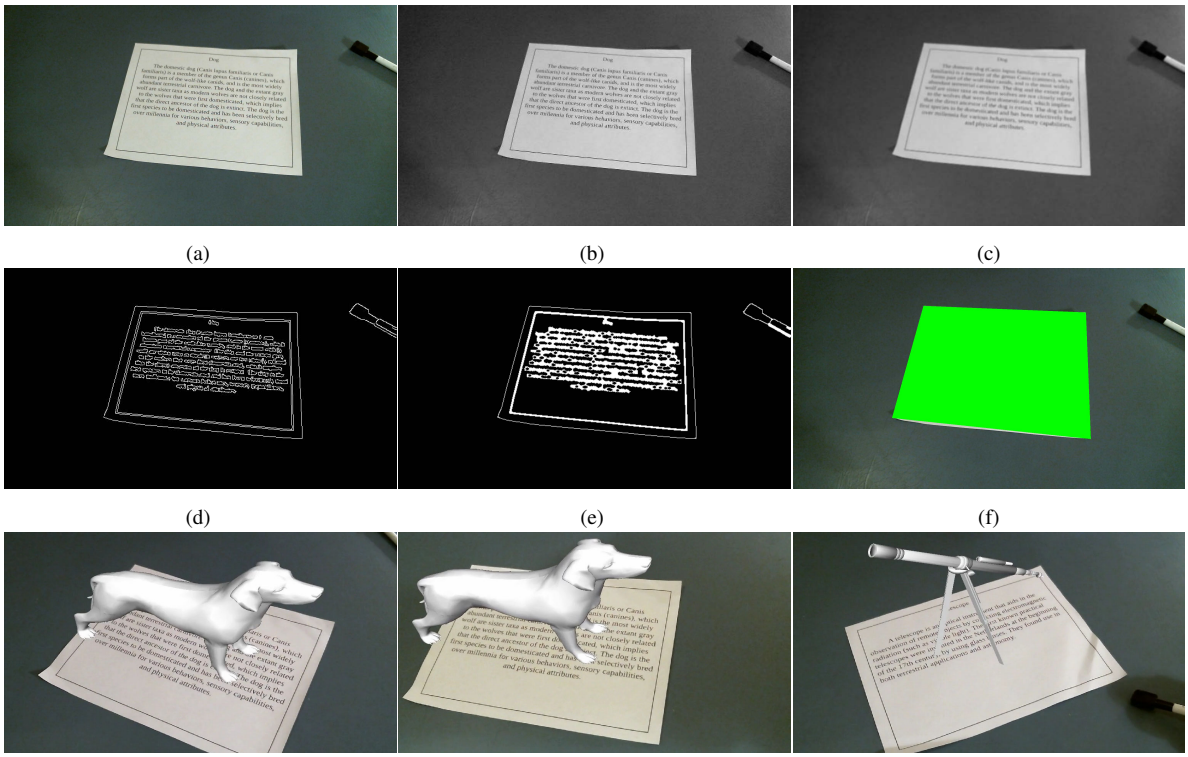


Figure 3. *Main steps of the proposed 3D model rendering method: (a) An original video frame contains a printed page with the text description of the animal* Dog*; (b) The video frame is converted to a grayscale image with smaller size for rapid processing; (c) The image is blurred using bilateral filter to maintain all available edges; (d)* Canny *edge detector is employed to find the edges in the image; (e) Apply morphological operations to smooth and connect those detected edges; (f) The location and extent of the printed page is accurately identified (in green color) by finding the largest convex quadrilateral in the image. (g) Based on the related text recognized by the OCR technique, a 3D dog model is rendered and pops up out of that specific page; (h) Another view of the same dog model when the camera and printed page are rotated (Note: the lighting conditions are also changed); (g) A 3D telescope is rendered when another different printed page about the scientific device* Telescope *is detected.*

Physical Concept Demonstration

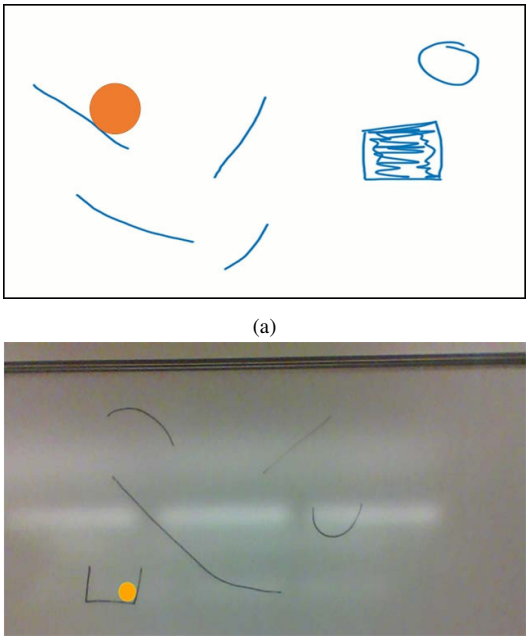
Our mobile AR system helps users understand the physical concepts such as *Reflection* and *Gravity*, which are common learning topics in K-12 education courses. More specifically, the physical interactions between a virtual ball and various objects can be demonstrated in both virtual reality scenes and real-world scenes as shown in Figure 4.

Both pre-stored images and real-time video streams can be handled by the proposed system. In the latter case, all individual frames are retrieved from a real-time video stream which is captured by the webcam available on most mobile devices. The Motion Estimation module employs a physics engine to simulate most physical phenomenon in the real world. For example, the physics engine uses *Sobel Filter* to detect the collision between a computer generated ball and the edges of various objects in the current scene. Based on the shape and the location of an object, the moving direction and velocity of the virtual ball can be accurately estimated using reflection and trajectory physics. The Model Rendering module is then initialized to superimpose the virtual ball and its corresponding motion over a specific image or video stream. The Motion Estimation and Model Rendering modules work together to support the demonstration of many physical concepts covered in K-12 courses.

Results

We develop the proposed mobile AR system on Windows 10 (x64) operating system. The standard Java 8 JDK is used and all programming is performed within the Eclipse Development Environment. In addition, the camera input processing and page determination are implemented using *OpenCV* algorithms. Meanwhile, the OCR API *Tess4j* recognizes the text on the detected page and further determines the corresponding 3D model to be retrieved from the local database. Finally *JOGL* is employed as a Java wrapper to access the *OpenGL* library aiming to render 3D models on the top of a printed page. To make it a more practical application, the mobile AR system is executed on a common Dell Inspiron 1545 laptop which has an Intel Pentium Dual-Core CPU running at 2.1 GHz with 4 GB of memory. The machine also comes with an Integrated Graphics Controller with Mobile Intel 4 Series Express Chipset.

We evaluate the system performance from two perspectives: *Text Recognition Rate* and *Model Rendering Time*. The Text Recognition Rate measures the average OCR confidence level for successful word detection. The Model Rendering Time measures the average amount of time required for one single 3D model to be rendered, which consists of both camera pose estimation time and graphical model rendering time.



(a)

(b)

Figure 4. *A virtual ball of orange color interacts with various objects in two distinct scenes: (a) A virtual reality scene with computer generated objects; (b) A real-world scene with lines drawn on a whiteboard where the virtual ball is contained in a "Square Cup".*

Text Recognition Rate highly relies on the font size of text. For experimental purposes, a set of English words collected from *Wikipedia* are printed with different font sizes on paper. We compute the average OCR confidence level for each font size. As shown in Figure 5, these words are printed from top to bottom with font sizes 14pt, 16pt, 18pt, 22pt, 24pt, 26pt, 28pt, and 32pt respectively. Liberation Serif is the font used in these experiments. The camera of the mobile system is located 13 inches above the paper. Shown in the Table 1 is the performance comparison of the proposed system on OCR text recognition rate for different font sizes, measured in confidence level. As indicated in the table, when the font size reduces, the mean and median values of OCR confidence level decrease. Moreover, the standard deviation and range values become larger as the font size decreases. Any word with font size equal or greater than 26pt can be easily detected by the OCR without any errors. However, the detection results deteriorate when the font size is less than 26pt. Generally, the OCR confidence level of 80% is the threshold for successful text recognition. It was observed that a word printed at font size 18pt with OCR confidence level as low as 59.41% can still be accurately detected, though it is not quite common.

We also measure the Model Rendering Time for a set of 3D models. On average, the amount of time for camera pose estimation which involves the addition of the Rotation Matrix and the Translation Matrix is around 10ms. On the other hand, the amount of time for graphical model rendering which includes the searching time for local model database is around 20ms. Totally, the average time for rendering a 3D model in each individual video frame is around 30ms. Note that the time for OCR text recognition is not included in the Model Rendering Time.

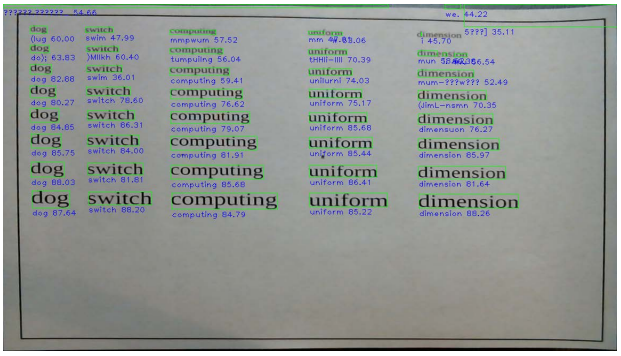


Figure 5. *Text recognition for different words printed from top to bottom with font sizes 14pt, 16pt, 18pt, 22pt, 24pt, 26pt, 28pt, and 32pt respectively.*

Conclusions

The world revolves around technology today. With the wide use of mobile devices and multimedia technologies, the use of AR technology in classrooms is currently on the rise. In this work, we introduce a new AR based knowledge learning system to enhance the learning process and student engagement. Our mobile AR system can automatically generate virtual entities and superimpose them over real-world images or video streams. These virtual entities can interact with real-world objects and respond to human activities in various real-time situations. For the input of static images or video streams, the mobile platform supports the demonstration of physical concepts and the rendering of 3D models. The system performance is assessed on the efficiency and effectiveness of the AR rendering process under various environmental conditions. Experimental results demonstrate our mobile AR system provides high accuracy for page determination and text recognition. In addition, the technique fulfills the real-time AR rendering requirement and supports the interactive learning experiences for users with various background including K-12 students.

The developed AR system has shown to be appealing to many users due to its robust 3D presentation of abstract concepts in an interactive learning environment. Some potential improvements are achievable for the proposed system in the near future: (a). Instead of detecting a printed page for each individual video frame, the system should track the detected page and hence improve the system efficiency; (b). Improve OCR text recognition performance by integrating the clearest part of the same word from multiple video frames; (c). In addition to text recognition, support the recognition of various real scene objects; (d). Enhance user participation through the control of AR models from the GUI; (e). Introduce associated audio and animation for 3D models for better user experiences. The mobile AR system is expected to be ultimately integrated into the K-12 education system to help learners explore and discover our exciting world.

Acknowledgements

This work was partially supported by Al and Dale Boxley Faculty Research Award and by a Frostburg State University Foundation Opportunity Grant (# 34035).

Table 1. Performance comparison of the proposed system on text recognition confidence (in percentage) for different font sizes.

	Font Size							
	14pt	16pt	18pt	22pt	24pt	26pt	28pt	32pt
Mean	54.84	60.78	64.89	76.00	80.96	85.71	86.09	88.08
Median	53.31	60.38	64.68	76.82	81.76	85.65	85.69	88.23
Standard Deviation	10.26	12.43	12.49	8.10	6.05	3.95	2.57	2.88
Range	35.14	45.13	46.87	32.72	22.18	14.12	8.04	11.02

References

[1] H. Crompton, M. R. Grant, and K. Y. H. Shraim, "Technologies to enhance and extend children's understanding of geometry: A configurative thematic synthesis of the literature.," *Journal of Educational Technology & Society*, vol. 21, no. 1, pp. 59–69, 2018.

[2] X. Pan, J. Wilson, M. Balukoff, A. Liu, and W. Xu, "Musical instruments simulation on mobile platform," in *IS&T Symposium on Electronic Imaging (IS&T-EI)*, (San Francisco, CA), 2016.

[3] X. Pan, T. Cross, L. Xiao, and X. Hei, "Musical examination and generation of audio data," in *SPIE Symposium on Electronic Imaging (SPIE-EI)*, (San Francisco, CA), 2015.

[4] X. Pan and S. Lyu, "Region duplication detection using image feature matching," *IEEE Transactions on Information Forensics and Security (TIFS)*, vol. 5, no. 4, pp. 857–867, 2010.

[5] S. C.-Y. Yuen, G. Yaoyuneyong, and E. Johnson, "Augmented reality: An overview and five directions for ar in education," *Journal of Educational Technology Development and Exchange*, vol. 4, no. 1, pp. 119–140, 2011.

[6] Itseez, "Open source computer vision library." https://github.com/itseez/opencv, 2015.

[7] Tess4j, "Java jna wrapper for tesseract ocr api." https://github.com/nguyenq/tess4j.

[8] J. Kessenich, G. Sellers, and D. Shreiner, *OpenGL®Programming Guide: The Official Guide to Learning OpenGL®, Version 4.5 with SPIR-V*. Addison-Wesley Professional, 9 ed., 2016.

[9] A. Dünser and E. Hornecker, "An observational study of children interacting with an augmented story book," in *Proceedings of the 2nd International Conference on Technologies for e-Learning and Digital Entertainment*, Edutainment'07, (Berlin, Heidelberg), pp. 305–315, Springer-Verlag, 2007.

[10] C. C. Marshall, "Reading and Interactivity in the Digital Library: Creating an experience that transcends paper," in *Proceedings of the CLIR/Kanazawa Institute of Technology Roundtable*, pp. 1–20, July 2003.

[11] M. Billinghurst, H. Kato, and I. Poupyrev, "The magicbook: a transitional ar interface," *Computers & Graphics*, vol. 25, pp. 745–753, 2001.

[12] C. Kapp, www.zooburst.com, *"ZooBurst, Augmented Reality 3D Pop-up Books*.

[13] T.-W. Kan, C.-H. Teng, and W.-S. Chou, "Applying qr code in augmented reality applications," in *Proceedings of the 8th International Conference on Virtual Reality Continuum and Its Applications in Industry*, VRCAI '09, (New York, NY, USA), pp. 253–257, ACM, 2009.

[14] M. T. Qadri and M. Asif, "Automatic number plate recognition system for vehicle identification using optical character recognition," in *Proceedings of the 2009 International Conference on Education Technology and Computer*, ICETC '09, (Washington, DC, USA), pp. 335–338, IEEE Computer Society, 2009.

[15] J. Li, H. Aghajan, J. R. Casar, and W. Philips, "Camera pose estimation by vision-inertial sensor fusion: An application to augmented reality books," in *IS&T International Symposium on Electronic Imaging 2016*, vol. 2016, (San Francisco, US), pp. 1–6, February 2016.

[16] H. S. Yang, K. Cho, J. Soh, J. Jung, and J. Lee, "Hybrid visual tracking for augmented books," in *Entertainment Computing - ICEC 2008* (S. M. Stevens and S. J. Saldamarco, eds.), (Berlin, Heidelberg), pp. 161–166, Springer Berlin Heidelberg, 2009.

Author Biography

Xunyu Pan received the B.S. degree in Computer Science from Nanjing University, China, in 2000, and the M.S. degree in Artificial Intelligence from the University of Georgia in 2004. He received the Ph.D. degree in Computer Science from the State University of New York at Albany (SUNY Albany) in 2011. From 2000 to 2002, he was an instructor with Department of Computer Science and Technology, Nanjing University, China. In August 2012, he joined the faculty of Frostburg State University (FSU), Maryland, where he is currently an Associate Professor of Computer Science and the Director of Laboratory for Multimedia Communications and Security. Dr. Pan is the recipient of 2011∼2012 SUNY Albany Distinguished Dissertation Award and 2016 FSU Faculty Achievement Award in Teaching. His publications span peer-reviewed conferences, journals, and book chapters in the research fields of multimedia security, image analysis, medical imaging, communication networks, computer vision and machine learning. He is a member of the ACM, IEEE, and SPIE. (Corresponding Author: xpan@frostburg.edu)

Joseph Shipway received B.S. degree in Computer Science with Honor from Frostburg State University (FSU) in 2017. He is currently working toward the M.S. degree in Computer Science at FSU. He is also a member of Upsilon Pi Epsilon Computer Honor Society.

Wenjuan Xu received the Ph.D. degree in Information Technology from the University of North Carolina at Charlotte. She is currently an Associate Professor of the Department of Computer Science and Information Technologies at Frostburg State University, Maryland.