This is a preprint of the paper to be presented at the 16th International Conference on Human-Computer Interaction (HCII), Heraklion, Greece, June 22–27, 2014. The final version is available in Springer library http://link.springer.com/chapter/10.1007%2F978-3-319-07854-0_28.

Virtualizing Real-life Lectures with vAcademia, Kinect, and iPad

Andrey Smorkalov[1] and Mikhail Morozov[1]

[1] Multimedia Systems Laboratory, Volga State University of Technology, Russia
{smorkalovay, morozovmn}@volgatech.net

Mikhail Fominykh[2] and Ekaterina Prasolova-Førland[2]

[2] Program for Learning with ICT, Norwegian University of Science and Technology, Norway
{mikhail.fominykh, ekaterip}@ntnu.no

Abstract. In this paper, we present a project aiming at designing a low-cost technological setup for translating real-life lectures into a 3D virtual world. We present the design of the first prototype where we record the voice of a lecturer, presentation slides, and use a motion-capture technique to grasp gestures. Based on these data, we create a scene in a 3D virtual world, display the slides, play the recorded sound, and animate the lecturer's avatar. We discuss evaluation results and discovered limitations, and outline solutions. In addition, we propose the major types of learning scenarios for the use of the designed system.

Keywords: 3D virtual worlds, motion capture, vAcademia, Kinect, iPad.

1 Introduction

Various advanced Virtual Reality (VR) technologies are being developed or adapted for educational purposes, mostly in industry, military, and healthcare [1]. However, a broader deployment requires the development of low-cost and off-the-shelf solutions.

Many studies report the potential of three-dimensional virtual worlds (3D VWs) for educational activities [2]. In this work, we explore new ways of using 3D VWs for creating asynchronous content out of synchronous learning activities. Several techniques were adapted by educators for getting content out of traditional classes, such as video recording of face-to-face lectures and recording of webinars. Such recordings change the context of learning and do not provide a possibility for collaborative work or for further developing the content, except for annotating. 3D VWs are also used for generating educational content. However, activities there are usually recorded as 'flat' 2D video, eliminating many advantages of the technology.

Low-cost motion-sensing technologies such as Microsoft Kinect, Nintendo Wii Remote, and Playstation Move provide researchers and educators with new opportunities for improving learning. Multiple examples include a low-cost alternative for interactive whiteboards and multi-touch teaching stations designed based on Kinect [3]. Tablets, like iPad, and other mobile devices found various applications, among them – augmenting 3D VWs [4]. A tablet provides a more convenient interface for a VR application than mouse and keyboard, especially if the user wants to stand and move.

This paper introduces a project aiming at designing a low-cost technological setup for translating real-life presentations and lectures into a 3D VW. We record sound, use a motion-capture technique to grasp the gestures of a lecturer, and a hand-held tablet device for additional control of the media content. Based on these data, we create a scene in a 3D VW, playing the recorded sound, animating the lecturer's avatar and displaying the media content.

2 Virtualizing real-life lectures

2.1 Technological Setup

We use three available technologies to implement the proposed system, vAcademia, Kinect, and iPad. Kinect is used for capturing the movement of a lecturer, while vAcademia is used for creating and recording the virtual replica of a lecture. In addition, the lecturer can use an iPad to control the environment and media content, such as drawing, switching slides, pointing, viewing the environment, and streaming in 2D.

Microsoft Kinect (http://www.microsoft.com/en-us/kinectforwindows/) is a low-cost motion sensing input device that is able to capture one or two humans. Apple iPad is a hand-help tablet device that has a touch screen can recognize finger gestures (http://www.apple.com/ipad/). vAcademia is an educational 3D VW (http://vacademia.com/). Its most distinctive feature is 3D recording. It allows capturing everything in a given location in the VW in process, including positions of the objects, appearance and movement of the avatars, media contents, text and voice chat messages [5]. 3D recording is often treated as an embedded screen capture mechanism. However, it is conceptually different from the video recording or screen capturing. A replayed 3D recording contains the entire 3D scene with all 3D objects, avatars, and all their actions. In addition to the virtual recording, vAcademia has a special set of tools to work with large amounts of media content [6].

2.2 First Prototype

System Implementation. The virtualizing real-life lectures mode interface is implemented using vAcademia Scripts. The Scripts initiate the beginning and the end of this mode. They provide Kinect-related information, including the skeleton in 2D and the recognition status of body parts. They interact with the Animation library of the graphic engine through the Script Executing Library. The Animation library controls the lecturer's avatar using Cal3D based on the data from Kinect plugin (Fig. 1).

vAcademia uses Cal3D library for implementing the skeleton avatar animation [7]. It allows to control the skeleton of an avatar by constantly and automatically recalculating its polygonal 3D model to match the current state of the skeleton. In the first prototype of the system, the Animation library of vAcademia requested a set of key points of the lecturer's body from the Kinect-support plugin (Fig. 1). If the lecturer was recognized by Kinect, Cal3D bones of the vAcademia avatar are oriented according to the key points. The system supported a 'sitting mode', when the motion-capture data were used for orientating arms and head and a 'standing mode' for all body parts.

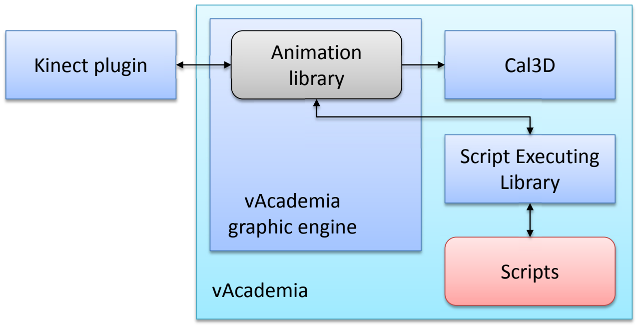

Fig. 1. Virtualizing real-life lectures mode interface of vAcademia

In addition, Kinect provides the status of each key point of the lecturer's body (not recognized, presumably recognized or recognized) as well as indicates if the body is cut in one of four directions by the camera visibility area. In order to eliminate unnatural poses, we divided the skeleton into five major parts (arms, legs, and head) and processed Kinect data for them separately. If a part of the skeleton remains unrecognized adequately for 0.2–0.5 seconds, the system sets it into the default state.

System Performance. vAcademia requires and actively uses one CPU core. Other cores may be also used, but less actively (10–30%), except for the process of the first load of the VW data. Kinect requires a dual-core CPU, but uses only one core, as the second is reserved for the application that uses Kinect data. These settings define the requirements for the computer to run the system designed.

System Evaluation Summary. Although the evaluation has not been systematic so far, it allowed us to improve many characteristics of the system. We conducted several evaluation sessions in several auditoriums with different configurations and lightning. The major evaluation data source is an interview with the lecturer. The following feedback on the technological setup is the most common:

The system applies too many restrictions on the lecturer's movements. First, the lecturer has to stay in the area in front of the Kinect device, and the space between them should be clear. Second, the lecturer has to face the device and not turn. Third, the lecturer has to use only one hand for pointing at the screen. Fourth, the lecturer has to avoid gestures that are difficult to recognize by the users of the 3D VW. In addition, we received suggestions on increasing the educational value of the system.

2.3 Applying Kinect for Motion Capture in vAcademia

The evaluation of the first prototype revealed three major limitations of Kinect in the given context and three corresponding challenges for the system implementation.

- C1: Kinect does not capture the gestures accurately enough. Therefore, we could not build a reliable avatar model and exclude unnatural poses.
- C2: Kinect does not recognize the turn of the lecturer. If the lecturer turns away from the device, it mixes up the left and the right.

- C3: An obvious, but serious challenge for applying Kinect is its inability to capture parts of the body that are covered by other body parts or foreign objects.

We addressed C1 by introducing several mechanisms. We limited the distance between the Kinect device and the lecturer. In addition, we fixed the Kinect device at a 0.5 meters distance from the floor. We introduced an additional filtration mechanism for sorting out unnatural positions, separating hands as distinct body parts. We addressed C2 by implementing the following algorithm. The turn is recognized relatively as a function of the position of the pelvis end points. The resultant value is valid within the range from -110 to 110 degrees against the “facing Kinect device” direction. We introduced colored markers for better recognition of the turns (Fig. 2).

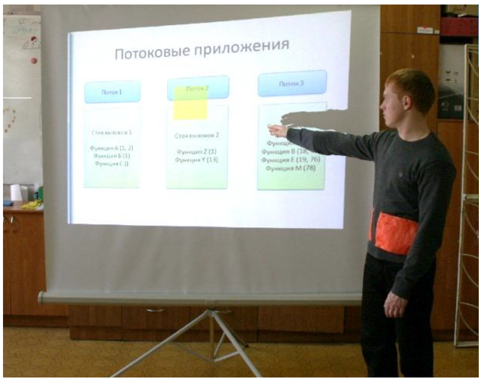

Fig. 2. Lecture capturing process

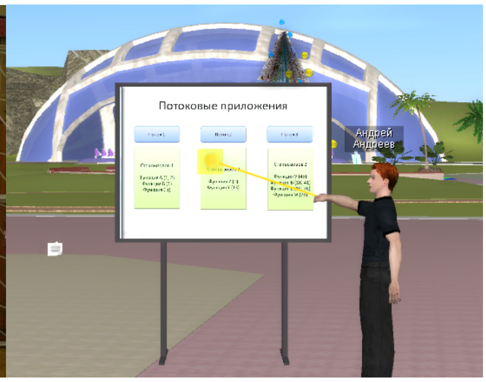

Fig. 3. Lecture streaming process

Challenge C3 may be addressed by applying multiple Kinect devices. At the same time, new challenges may appear, such as the increase of the price, complexity in setting up the system, and complexity in merging the data from multiple sources.

2.4 Applying Kinect and iPad for Controlling Whiteboards in vAcademia

Based on the results of the system’s first prototype evaluation, we have identified the following challenges for controlling virtual whiteboards and their contents.

- C4: The position and movement of the lecturer against the physical whiteboard should match the position and movement of his/her avatar against the virtual one.
- C5: A physical pointer needs to be captured and represented in the 3D VW, where it is often even more important. However, Kinect cannot capture a physical pointer.
- C6: Switching the slides requires interaction with the computer or a remote control, which does not convey any meaning when captured and translated into the 3D VW.

Addressing C4, we introduced a Setup mode for achieving the precise match between the physical and the virtual whiteboard (horizontally) and installing Kinect at 0.5 meters from the floor (vertically). Instead of trying to capture the physical pointer, we direct the virtual pointer based on the position of the lecturer’s hand, addressing C5. If the half line that extends from the lecturer’s hand crosses the physical whiteboard, the avatar in the 3D VW directs a virtual pointer to the same point (Fig. 3). In order to keep the lecturer aware of his or her hand being captured, we display a semi-transparent yellow area on the physical whiteboard on top of the slides (Fig. 2).

Addressing C6, we have been developing the switching slides functionality by recognizing standard Kinect gestures Swipe Left and Swipe Right. In addition, we decided to employ iPad for extending the possibilities of controlling media contents. Using vAcademia Presentation Remote app, the lecturer can stream handwriting and drawing to a virtual whiteboard and control other content on it without going back to the computer. The tablet is connected to the vAcademia client software through the vAcademia communication server using access code.

3 Learning Scenarios

We can identify several major learning scenarios for the systems we have been designing based on the evaluation results and user feedback.

Scenario 1: Lecturing as a Mixed Reality Activity. In this scenario, a live lecture could be attended simultaneously by the learners in real and virtual classrooms. As opposed to some existing solutions where distant learners watch the video of the lecturer, this setting has a number of advantages. This includes the possibility to design the virtual classroom recreating the physical one, thus providing distant learners with a greater sense of presence at the lecture. Additional advantages include possibilities for interactions between students in the physical and virtual classrooms (as the former might be logged in as avatars in the virtual classroom and perform activities there during the lecture) and recording student and lecturer activities in the same context.

Scenario 2: Round-table Discussion as a Mixed Reality Activity. The system designed can be used for supporting round-table discussions. Some of the participants can join such an activity through the 3D VW, while some other can be captured from the real world through multiple Kinect-based systems. Each of them can capture two participants. Gestures are a very important component of round-table discussion, and the designed system provides a significant advantage over pure 3D VWs.

Scenario 3: Mixed Reality Educational Role Plays. In this scenario, the participants from the physical and the virtual classroom could engage in an educational role play, e.g. simulating a conflict management situation. The limitation in this context is that only one or two users could be captured by Kinect at a time, which could be dealt with by people from the physical classroom taking turns or by letting the users captured by Kinect play the roles of facilitators. The advantage of this approach is the possibility to create and record a shared virtual scene for the role play.

Scenario 4: Immersive 3D Recordings of Lectures. In this scenario motion capture is used for easy and low-cost creation of educational content for later (asynchronous) use, such as lectures and simulations. Using the vAcademia 3D recording functionality, any activity, including streaming Kinect-captured lectures, in the 3D VW can be easily saved and revisited later. The realization of this scenario can provide an advanced alternative to the video-recorded lectures, as the resultant 3D recordings combine the convenience of video and immersive qualities of 3D VWs. In this case, the limitation of 1–2 users being captured at a time is less important as several captured sessions with different avatars could be superimposed into one recording.

4 Conclusion and Future Work

The evaluation of the system for translating real-life lectures into a 3D VW demonstrated that the system functions as intended and has practical value. We also present the challenges we outlined for the system, including ways to address them.

Future development of the system will include support for interactive whiteboards that can replace the projector in the current setup. We plan to implement and evaluate emotion capture functionality using Kinect and Face Tracking SDK that will be important for supporting role-plays and creation of training simulations.

References

1. Rajaei H., Aldhalaan A.: Advances in virtual learning environments and classrooms. In: Abhari A (ed.) 14th Communications and Networking Symposium (CNS), Boston, MA, USA, April 3–7, pp. 133–142. Curran Associates, Inc., (2011)
2. de Freitas S., Rebolledo-Mendez G., Liarokapis F., Magoulas G., Poulovassilis A.: Developing an Evaluation Methodology for Immersive Learning Experiences in a Virtual World. In: 1st International Conference in Games and Virtual Worlds for Serious Applications (VS-GAMES), Coventry, UK, March 23–24, pp. 43–50. IEEE, (2009)
3. Cheong S.N., Yap W.J., Logeswaran R., Chai I.: Design and Development of Kinect-Based Technology-Enhanced Teaching Classroom. In: Park JJ, Jeong Y-S, Park SO, Chen H-C (eds.) Embedded and Multimedia Computing Technology and Service. Lecture Notes in Electrical Engineering, vol 181, pp. 179–186. (2012)
4. Pakanen M., Arhippainen L., Hickey S.: Studying Four 3D GUI Metaphors in Virtual Environment in Tablet Context. Visual Design and Early Phase User Experience Evaluation. In: 6th International Conference on Advances in Computer-Human Interactions (ACHI), Nice, France, February 24–March 1, pp. 41–46. (2013)
5. Morozov M., Gerasimov A., Fominykh M., Smorkalov A.: Asynchronous Immersive Classes in a 3D Virtual World: Extended Description of vAcademia. In: Gavrilova M, Tan CJK, Kuijper A (eds.) Transactions on Computational Science XVIII, vol 7848. Lecture Notes in Computer Science, pp. 81–100. Springer Berlin Heidelberg (2013)
6. Smorkalov A., Fominykh M., Morozov M.: Stream Processors Texture Generation Model for 3D Virtual Worlds: Learning Tools in vAcademia. In: Lin Q, Muhlhauser M, Sheu P (eds.) 9th International Symposium on Multimedia (ISM), Anaheim, CA, USA, December 9–11, pp. 17–24. IEEE, (2013)
7. Jingtang L., Gang Z., Dunming T., Junjie X.: Research of skeletal-based virtual human animation using Cal3D. In: 3rd International Conference on System Science, Engineering Design and Manufacturing Informatization (ICSEM) Chengdu, China, October 20–21, pp. 269–273. IEEE, (2012)