This is a preprint of the paper to be presented at the 2nd Workshop on Off-The-Shelf Virtual Reality at the International Conference on Virtual Reality (VR), Orlando, FL, USA, 16 March 2013 and published in the IEEE Xplore Digital Library.

Virtualizing Real-life Lectures with vAcademia and Kinect

Andrey Smorkalov[1]

[1] Multimedia Systems Laboratory, Volga State University of Technology, Russia

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

[2] Program for Learning with ICT, Norwegian University of Science and Technology, Norway

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, evaluation results, discovered limitations, and outlined solutions. In addition, we propose the major types of learning scenarios for the use of the designed system. In the first prototype of the system, we record the voice of a lecturer and use a motion-capture technique to grasp his or her gestures. Based on these data, we create a scene in a 3D virtual world, play the recorded sound, and animate the lecturer's avatar. The presentation slides that the lecturer uses are sent directly to the 3D virtual world and displayed on a virtual whiteboard. Such system provides an alternative way of streaming and creating recordings of presentation and lectures.

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

1 INTRODUCTION

Many studies report the potential of three-dimensional virtual worlds (3D VWs) for educational activities [1]. This technology can benefit educational process in many ways. Most of them are considered to exploit advantages of the 3D VWs, such as low cost and high safety, 3D representation of learners and objects, and interaction in simulated contexts with a sense of presence [2, 3]. However, this technology is far from becoming mainstream in education, as there are many challenges in applying it for learning.

One of the most serious challenges in adapting 3D VWs for learning is the lack of features that educators use in everyday teaching [4]. Despite the demand and interest from educators, in most cases, 3D VWs are adopted for educational purposes, but not specially created for them [4]. Cooperation and co-construction in 3D VWs is a complex task, it needs to be supported and requires additional tools [2]. The design of 'learning spaces' within 3D VWs is considered to be important, however, there are no strong guidelines [5]. In addition, 3D VWs are mostly used for synchronous activities and lack support for learning in the asynchronous mode.

Various more advanced Virtual Reality technologies are being developed or adapted for educational purposes. However, the cost is becoming a limiting factor in most cases. Industry, military, and healthcare are named as the only areas where these technologies are starting to be used [4]. Thus, the development of low-cost and off-the-shelf solutions is necessary.

In our research and development projects, we explore new ways of using 3D VWs for learning. One of our particular interests is using 3D VWs for creating asynchronous content out of synchronous learning activities. This paper introduces a project aiming at designing a low-cost technological setup for translating real-life presentations and lectures into a 3D virtual environment. We record sound and use a motion-capture technique to grasp the gestures of a lecturer. Based on these data, we create a scene in a 3D virtual environment, play the recorded sound, and animate the lecturer's avatar. The presentation slides that the lecturer uses are sent directly to the 3D VW and displayed on a virtual whiteboard. Such system provides an alternative way of streaming and creating recordings of presentation and lectures.

2 BACKGROUND

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 web conferences (or webinars). These methods allow creating cheap educational content for asynchronous learning [5, 6]. In many cases, these are the optimal solutions. However, video lectures and web conferences change the context of learning. These technologies do not provide immersion or sense of presence as in 3D VWs. Such recordings do not provide a possibility for collaborative work or a method for further developing the content, except for commenting and annotating it.

3D VWs are also used for generating educational content. Even though this technology allows creating full context of the real-life educational process, it is usually recorded as 'flat' 2D video, which eliminates many advantages of the technology, such as sense of presence – the sense of 'being there' [7, 8]. For example, this approach is used in Machinima – collaborative film making using screen capture in 3D VW and games [9].

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 [10]. Kinect was also used to improve video recording of presentations by designing an automatic camera control system [11].

3 VIRTUALIZING REAL-LIFE LECTURES

3.1 Technological Setup

We use two available technologies to implement the proposed system, Kinect and vAcademia. Kinect is used for capturing the movement of a lecturer, while vAcademia is used for creating and recording the virtual replica of a lecture. The third component of the system is a software plugin for vAcademia that translates the motion data from Kinect, the sound, and the contents of the whiteboard into the 3D VW of vAcademia.

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 [12]. The device

[1] smorkalovAY@volgatech.net
[2] mikhail.fominykh@ntnu.no [3] ekaterip@ntnu.no

consists of a video camera, depth camera, and an IR camera. Kinect SDK (software development kit) is provided on an open-access basis for controlling and getting data from the device [13].

Virtual Academia (vAcademia) is an educational 3D VW developed by Virtual Spaces LLC in cooperation with the Multimedia System Laboratory at the Volga State University of Technology (former Mari State Technical University), Russia. The system is currently under beta testing and free to use (http://vacademia.com/).

The most distinctive feature of vAcademia is 3D recording which allows capturing everything in a given location in the VW in process, including positions of the objects, appearance and movement of the avatars, contents on the whiteboards, text and voice chat messages. 3D recording is often misunderstood and treated as an embedded screen capture mechanism. However, it is conceptually different from the video recording or screen capturing. A replayed 3D recording does not only deliver the image at any virtual camera angle and a synchronized communication messages, but much more. A 3D recording contains the entire 3D scene with all 3D objects and avatars. All the actions of the avatars and in the environments are also saved in a 3D recording. These actions happen again when a 3D recording is replayed or, better to say, visited. It can be visited by a group of avatars that can interact with each other and the recorded objects. Moreover, such a visit can be recorded again. In such a way, this feature allows creating a new type of content that comprises both space and time.

Similar functionalities were realized earlier in few VWs or desktop virtual reality systems [14]. However, 3D recording was never developed into a convenient tool and never adopted for specific use as in vAcademia. In addition, no convenient tools for working with the resultant recordings were developed.

3.2 First Prototype

3.2.1 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. In addition, the scripts provide the Kinect-related information, including the skeleton in 2D and the recognition status of each body part.

The Scripts interact with the *Animation library* of the graphic engine through the *Script Executing Library*. The Animation library interacts with the Kinect plugin, and based on the data from it, controls the lecturer's avatar using Cal3D (Fig. 1).

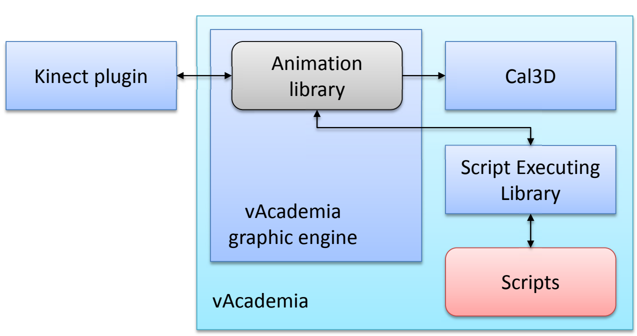

Figure 1: Virtualizing real-life lectures mode interface of vAcademia

vAcademia uses Cal3D library for implementing the skeleton avatar animation [15]. 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. If the lecturer was recognized by Kinect, Cal3D bones of the vAcademia avatar are oriented according to the key points. In the general case, we assumed that the lecturer is standing, which we defined as a 'standing mode'. The system also supported a 'sitting mode', when the motion-capture data were used only for orientating arms and head.

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 when the lecturer is not recognized correctly, we applied the following method. We divided the skeleton into five major parts: left arm, right arm, left leg, right leg, and head. The Kinect data for each of these parts is processed separately to better estimate how adequately they are recognized. We consider a part of the skeleton adequately recognized if the average recognition status of the key points of a specific skeleton part is higher than a certain value.

If a certain part of the skeleton cannot be adequately recognized, the system stops applying Kinect data to this part and uses the last adequately recognized set of key points. If this part of the skeleton remains unrecognized adequately for a certain period of time (0.2–0.5 seconds), the system sets this part into the default state.

3.2.2 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 virtual world 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.

In the general case, vAcademia uses one CPU core, while Kinect uses the other one. The process of animating the lecturer's avatar based on the data from Kinect is not computationally complex. Therefore, the system's performance should be satisfactory on a computer that satisfies the minimum requirements described above, which has been confirmed during the evaluation.

3.2.3 System Evaluation Summary

The system has been evaluated at the Computer Science department and the Volga State University of Technology. Although the evaluation we conduct has not been systematic so far, it allowed us to improve many characteristics of the system. The first versions of the prototype were too unstable for applying them in educational process on normal, not purely experimental, conditions.

For each version of the prototype, we conduct several evaluation sessions in two-three different courses, in several auditoriums of different configurations and lightning, and involving different teachers. The major evaluation data source is a short interview with the lecturer that we conduct while watching together the 3D recording created vAcademia.

Such evaluation data is rather difficult to summarize, as we usually able to improve the prototype slightly after few trials and release the next version. However, 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 Kinect 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 in the 3D virtual environment.

In addition, we received several suggestions on how to increase the educational value of the system. For example, it was suggested to mark the area that is recognized by the system as where the lecturer is pointing on both the physical and virtual screens. The first prototype had such a marker only in the virtual environment. Some teachers suggested using easily recognized gestures for switching presentation slides. Many other suggestions go beyond the current possibilities of Kinect and therefore, they cannot be taken into consideration at this stage.

3.3 Applying Kinect Motion Capture in vAcademia

3.3.1 Challenges

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

1. Kinect does not capture the gestures accurately enough. It does not provide the points of a moving human skeleton. Instead, it is offered to control the averaging ratio of the coordinates, predict the movements, use different recognition states, and other methods. Despite using all available data sources, we could not build a reliable avatar model that can move without unnatural poses. At the same time, we believe that such unnatural poses and other distortions are unacceptable when picturing a lecturer.
2. Kinect does not recognize the turn of the lecturer. The left arm is always on the left, and the right arm is always on the right. If the lecturer turns away from the device, it mixes up the left and the right arms and returns a completely unnatural pose. Kinect cannot recognize the turn of a lecturer despite the possibility to get the quaternion of the turn of the whole body, as the resultant turn is far from the reality.
3. Another 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. This results in additional requirements to the setup and lower recognition accuracy.

3.3.2 Proposals and Solutions

1. We addressed the problem of low accuracy of the motion capture data by introducing several mechanisms. We limited the distance between the Kinect device and the lecturer. Empirical data that we gathered evaluating the first prototype demonstrated that the skeleton can be recognized without significant disturbances at a distance of 1.8 meters or less in the standing mode and 1.3 meters or less in the sitting mode. In addition, we fixed the Kinect device at a 0.5 meters distance from the floor and made a software-based turn into a zero-degree position by the vertical axis to ensure that the lecturer is strictly in the normal plane.

   We introduced an additional filtration mechanism for sorting out unnatural positions of the body parts. We limited the acceptable values of Euler angles between the bones. In addition, we separated hands as distinct body parts.
2. We address the problem of turn recognition 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. If the turn is out of this range, the algorithm cannot calculate the turn correctly.

   We introduced colored markers to the system for better recognition of the turns. Two colored markers are placed on the body of a lecturer on the left and on the right side, facing the Kinect device (Fig. 2). The colors of the markers should be different from the lecturer's clothing and the material should not be shimmering. The video data from Kinect is constantly analyzed in search for markers. If they are recognized, the system considers that the lecturer is in the acceptable turn range. If the markers are not recognized, the system returns the lecturer's avatar in vAcademia to the last correctly recognized state (and later into the default state) and waits for the markers to be recognized.

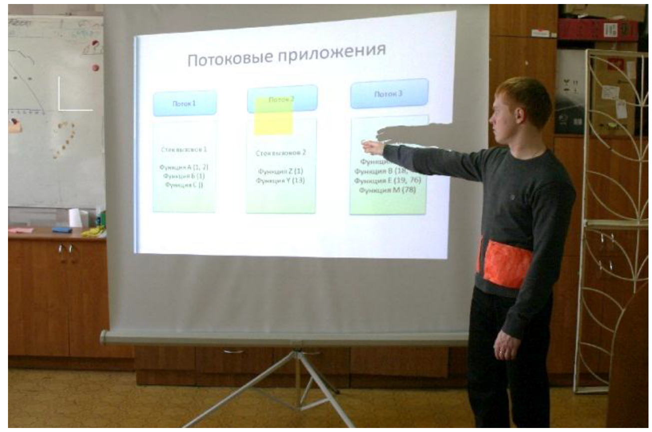

Figure 2: Lecture capturing process

   In order to recognize the markers only on the lecturer's body and ignore the background, the system uses one of the Kinect video data parameters that exists for each pixel and shows if it belongs to a person being captured or not.
3. The challenge of Kinect's inability to capture parts of the body that are covered by other body parts or foreign objects may be addressed by applying multiple Kinect devices, which should improve the accuracy of recognition. The optimal configuration would probably be using three Kinect devices, setting them to the left, to the right, and in front of the lecturer. The whiteboard prevents locating another Kinect device behind the lecturer.

   Applying the multiple-devices approach will definitely improve recognizing the covered body parts. It may also contribute to improving accuracy and turn recognition. However, it might not solve them completely. At the same time, new challenges may appear.
   1. Additional Kinect devices increase the price of the system. This is an important factor, since we position it as a low-cost solution.
   2. The data acquired from the multiple Kinect devices should be adjusted to a single coordinate system. Therefore, the requirements to the accuracy of locating Kinect devices should increase.
   3. Locating Kinect devices to the left and to the right from the lecturer will impose additional requirements to the auditorium. It should be at least seven meters across.
   4. Such approach creates the challenge of merging the data from multiple Kinect devices. In some cases, the data should be averaged. In other cases, the data from a single the most trustable device should be selected. Besides that, it may require to select two the most trustable devices and average their data.

3.4 Supporting Slide Presentations

3.4.1 Challenges

Based on the results of the system's first prototype evaluation, we have identified the following challenges for supporting slide presentations.

1. The position and movement of the lecturer against the whiteboard in the real world should match the position and movement of the lecturer's avatar against the virtual whiteboard.
2. A physical pointer may be an important part of the lecture experience. It needs to be captured and correctly represented in the 3D VW, where it is often even more important. However, Kinect cannot capture a physical pointer.
3. The lecturer needs the slides to be switched, but it requires interaction with the computer, remote control, or the whiteboard (if it is an interactive one). Such actions of the lecturer, when captured by Kinect and translated into the 3D environment, do not convey any meaning.

3.4.2 Proposals and Solutions

We addressed the challenges presented above by introducing the following mechanisms.

1. We introduced a Setup mode for achieving the precise match between the physical whiteboard and the virtual one. The setup should be performed once after installing the physical whiteboard and the Kinect device in the classroom. In the Setup mode, a technician marks the left and the right edges of the physical whiteboard with the left and right hands correspondingly. The edges are captured in Kinect coordinate system. This mechanism allowed us to achieve the precise match between the physical whiteboard and the virtual one horizontally, but the vertical match was achieved by installing the Kinect device and the physical whiteboard on a specified distance from the floor. We further propose to improve the system and make it recognize the borders of the physical whiteboard and create the replica in the 3D VW keeping the proportion automatically.
2. Instead of trying to capture the physical pointer, we direct the virtual pointer based on the position of the lecturer's hand. If the half line that extends from the lecturer's hand towards the physical whiteboard crosses it, the avatar in the 3D VW directs a virtual pointer to the same point (Fig. 3).

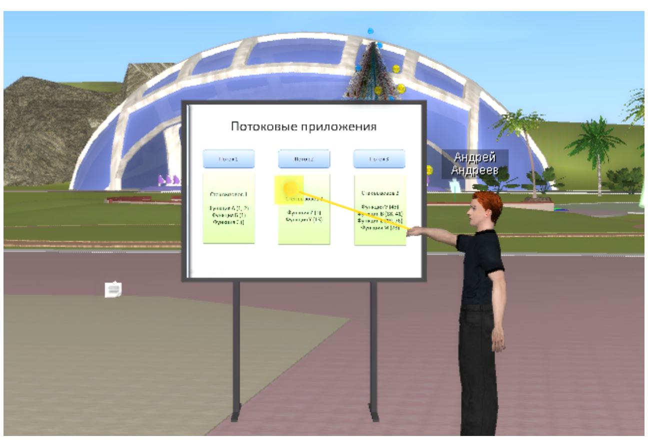

Figure 3: Lecture streaming process

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).In such a way, the lecturer knows if and where the virtual pointer of his or her avatar is directed at any point in time (Fig. 3).

3. We have been developing the switching slides functionality in (currently, only) Microsoft PowerPoint by recognizing standard gestures Swipe Left and Swipe Right.

4 LEARNING SCENARIOS

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

4.1 Scenario 1: Lecturing as a synchronous 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.

- 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 possibilities for interactions between students in the physical and virtual classrooms, as the former might be logged in as avatars in the virtual classroom performing activities together without interrupting the lecture
- The activities of both the students and the lecturer could be recorded, all in the same context.

4.2 Scenario 2: Round-table discussion as a synchronous 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. Multiple Kinect-based systems can be installed in remote locations if necessary. 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 in the non-verbal communication support. The whiteboard can perform rather a supportive role or not used at all.

4.3 Scenario 3: Motion capture as a part of synchronous 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. As in the previous scenario, the advantage of this approach is the possibility to create a shared virtual scene for the role play, as well as the possibility to record it.

4.4 Scenario 4: Creating immersive 3D recordings out of live lectures

The latter consideration motivates another scenario, where 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.

5 Conclusion and Future Work

In this paper, we presented a project aiming at designing a technological setup for translating real-life lectures into a 3D virtual environment. The evaluation of the first prototype demonstrated that the designed system functions as intended and can be useful in educational practice. However, several challenges were outlined.

The future development of the system will include the support for interactive whiteboards that can replace the projector in the current setup. It can significantly extend the variety of presentation contents. In addition, we plan to implement emotion capture functionality using Kinect and Face Tracking SDK. The latter features will be especially important for supporting role plays and creation of low-cost simulations for training in the areas such as counseling, conflict management, negotiation and teacher education.

A full-featured evaluation of the system will be conducted as soon as the quality of the resultant 3D recordings of the lectures is satisfactory. We plan to conduct a study on the asynchronous use of lectures, observing and comparing five different modes: video-recording classroom lectures, recording webinars, screen capturing lectures in a 3D VW, 3D recording lectures in vAcademia, and Virtualizing real-life lectures with vAcademia and Kinect.

Acknowledgements

The authors would like to thank Mikhail Morozov from the Volga State Technical University and Alexey Gerasimov from the Virtual Spaces LLC for their leadership allowing this project to develop.

References

[1] S. de Freitas, G. Rebolledo-Mendez, F. Liarokapis, G. Magoulas, and A. Poulovassilis, "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, 2009, pp. 43–50.

[2] S. Warburton, "Second Life in higher education: Assessing the potential for and the barriers to deploying virtual worlds in learning and teaching," British Journal of Educational Technology, vol. 40(3), 2009, pp. 414–426.

[3] R. Mckerlich, M. Riis, T. Anderson, and B. Eastman, "Student Perceptions of Teaching Presence, Social Presence, and Cognitive Presence in a Virtual World," Journal of Online Learning and Teaching, vol. 7(3), 2011, pp. 324–336.

[4] H. Rajaei and A. Aldhalaan, "Advances in virtual learning environments and classrooms," in 14th Communications and Networking Symposium (CNS), Boston, MA, USA, 2011, pp. 133–142.

[5] P. Brusilovsky, "Web Lectures: Electronic Presentations in Web-based Instruction," Syllabus, vol. 13(5), 2000, pp. 18–23.

[6] S. M. Engstrand and S. Hall, "The use of streamed lecture recordings: patterns of use, student experience and effects on learning outcomes," Practitioner Research in Higher Education (PRHE), vol. 5(1), 2011, pp. 9–15.

[7] J. V. Draper, D. B. Kaber, and J. M. Usher, "Telepresence," Human Factors: The Journal of the Human Factors and Ergonomics Society, vol. 40(3), September 1, 1998 1998, pp. 354-375.

[8] M. Slater, "Measuring Presence: A Response to the Witmer and Singer Presence Questionnaire," Presence: Teleoperators and Virtual Environments, vol. 8(5), 1999, pp. 560-565.

[9] G. Barwell, C. Moore, and R. Walker, "Marking machinima: A case study in assessing student use of a Web 2.0 technology," Australasian Journal of Educational Technology, vol. 27(Special issue, 5), 2011, pp. 765–780.

[10] S. N. Cheong, W. J. Yap, R. Logeswaran, and I. Chai, "Design and Development of Kinect-Based Technology-Enhanced Teaching Classroom," in Embedded and Multimedia Computing Technology and Service, J. J. Park, Y.-S. Jeong, S. O. Park, and H.-C. Chen, Eds., 2012, pp. 179–186.

[11] M. B. Winkler, K. M. Hover, A. Hadjakos, and M. Muhlhauser, "Automatic camera control for tracking a presenter during a talk," in International Symposium on Multimedia (ISM), Irvine, CA, USA, 2012, pp. 471–476.

[12] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-Time Human Pose Recognition in Parts from Single Depth Images," in Machine Learning for Computer Vision, R. Cipolla, S. Battiato, and G. Maria Farinella, Eds. Berlin Heidelberg: Springer, 2013, pp. 119–135.

[13] Microsoft, Kinect, http://www.microsoft.com/en-us/kinectforwindows/.

[14] M. Morozov, A. Gerasimov, and M. Fominykh, "vAcademia – Educational Virtual World with 3D Recording," in 12th International Conference on Cyberworlds (CW), Darmstadt, Germany, 2012, pp. 199–206.

[15] L. Jingtang, Z. Gang, T. Dunming, and X. Junjie "Research of skeletal-based virtual human animation using Cal3D," in 3rd International Conference on System Science, Engineering Design and Manufacturing Informatization (ICSEM) Chengdu, China, 2012, pp. 269–273.