Industrial Equipment Simulation Training System Based on Kinect

Jun-Feng Yao, Lin Yang, and Lei Lan
Xiamen University, Xiamen, China
Email: yao0010@xmu.edu.cn, liny234@gmail.com

***Abstract*—In carton manufacturing industry, traditionally, new workers are trained by realistic machines. However, training operators in a virtual environment can significantly reduce the budget and avoid unexpected safety accidents. In this paper, we presented a new interactive way in industrial training system based on Kinect. By computing depth map, Kinect sensor can produce the features of human body in real scene. The potential application of Kinect as a gestural interaction tool for equipment simulation training is evaluated in our system. Users can take use of the interactive interface quickly and apply it to navigation through a real-time 3D virtual environment created by Unity3D.**

***Index Terms*—simulation training, Kinect, carton industry, OpenNI middleware**

I. Introduction

In carton factories, the traditional way of training new employees on production lines is inefficient, because some machines are used for training rather than producing, which greatly waste raw materials and result in unnecessary economic expenditure. To make matters worse, many serious security incidents always happen. In this study, interactive 3D technology is mainly applied in training system: Equipment Simulation Training System (ESTS). Based on establishing normative production operation processes and the application model of optimized management processes, the system utilize virtual reality technology to enforce virtualization and multidimensional expression of former processes in the carton production line, such as printing, cutting, notching, nailing, packing etc. ESTS can help trainees understand the production environment and processes of carton packaging industry intuitively. In addition, because all the training processes are taken in 3D scenes, it can completely avoided machine malfunctions and material wastage. Trainees can manipulate equipment comprehensively and accurately and ensure equipments running in safe, so as to improve the production efficiency.

II. Research of Related Kinect Work

Kinect is a motion sensing input device produced by Microsoft for the applications in Xbox 360 video game console and Windows PCs. Based on a RGB camera, depth sensor and multi-array microphone running proprietary software, it enables users to control and interact with the Xbox 360 through a natural user interface using gestures and spoken commands. It is an example of natural interaction devices that capture body movements and sounds to allow a more natural interaction between users and computers in the context of a natural user interface.

But Kinect is not only about entertainment, the options are infinite. Countless people copy the success of Kinect in motion sensing game to other application fields. In the field of education applications, Kinect has the potential to drastically enhance teaching and learning in all learning environments [1]. As revealed by Johnson, L., Smith, R., Willis, H., Levine, A. and Haywood, K. [2], the creation of gesture-based interfaces create promising opportunities for educators to offer students easier and more intuitive ways to interact with the content in multimedia learning environments than ever before. For example using Kinect, users can "become" a musical instrument in music class. Kinect also has great potential in the domain of medical imaging science. The surgeon can't use a mouse or keyboard to retrieve and manipulate virtual medical images in a surgery, because of the risk of infection. Johnson, R., O'Hara, K., Sellen, A., Cousins, C. and Criminisi A. [3] developed a project called touchless interaction in medical imaging, enables doctors to use simple hand gestures to change, move, or zoom in on CT scans, MRIs, and other medical images. It could help make surgery faster, safer, and more accurate.

III. Research in Kinect Gesture Capture

The human movement is actually the body skeleton movement. When we use 3D body skeleton to represent human motion, 3D body need to map onto 2D plane but the relative spatial relationship of each skeleton segments should remain unchanged. Human body skeleton can be regarded as set of rigid body connected by joints, such as upper arm and lower arm is connected by the elbow. After getting joints' space information, we can easily customize several special gestures to OpenNI processing recognition and tracking by comparing the relative space position of the joints.

This method can be applied to interaction of interactive system. Suma, E., Lange, B., Rizzo, A., Krum, D. and Bolas, M. [4] developed a middleware FAAST, which can

Manuscript received July 25, 2013; revised September 10, 2013.

doi: 10.12720/ijoee.1.4.291-294

identify and track the user's action and control a virtual avatar on the network. Ok-Hue Cho and Won-Hyung Lee [5] presented a new way of interaction in interactive media art. It implemented 5 gestures with movie clip events and other 5 gestures with image events. Rogge, S., Amtsfeld, P., Hentschel, C., Bestle, D. and Meyer, M. [6] developed a software prototype for interactively modify turbine blades with different hand gestures in a virtual environment, intended to help engineers to gain a better understanding of design modifications by receiving a direct feedback, and to create improved design solutions faster during the development process. Chun-Yen Chang, Yu-Ta Chien, Cheng-Yu Chiang, Ming-Chao Lin and Hsin-Chih Lai [7] presented the theoretical underpinnings for adopting gesture-based multimedia learning and described how they used Kinect to embody the most common type of multimedia learning in classroom. In order to obtain a more friendly interaction experience, Zhenxing Gao [8] used Kinect developed actions related to somatosensory interaction and approaches for achieving action recognition. Experimental results demonstrated that real-time navigation was available under 3D virtual environments for the somatosensory interaction actions with Kinect. It is very similar to our work.

IV. Research of Kinect Gestural Interaction in ESTS

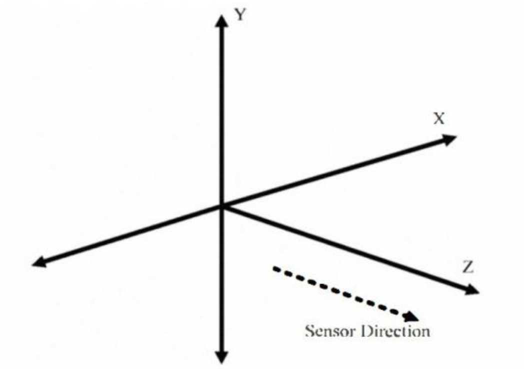

Figure 1. The 3D space sensed by Kinect.

The Kinect contains a RGB image camera, an IR camera, and a laser-based IR projector that emits a known structured light pattern of light and dark speckles at 830 nm. The IR camera and the IR projector form a stereo pair with a baseline of approximately 75 mm. Depth for each pixel can be calculated by triangulation against the known pattern from the IR projector [9]. Based on the cameras and projector, Kinect reacts the 3D space in front of it from 850 mm to 4,000 mm and the center of which is IR camera, while, facing behind of the depth camera's left hand is positive direction of X-axis, above of it is the positive direction of Y-axis, and in front of it is the positive direction of Z-axis, as shown in Fig. 1.

OpenNI provides a set of open source APIs and these APIs provide support for hand gestures and body motion tracking. It defined 24 joints, but actually we can only use 15 joints (Fig. 2) to analyze skeleton by middleware. It indicates the joint space information with position and orientation.

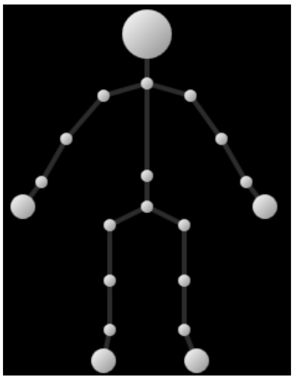

Figure 2. Skeleton and 15 joints we can used defined by OpenNI.

Six special gestures are used for our work (cf. Fig. 3). The method of comparing the relative space position of joints after getting their space data is adopted to process recognition and body motion track. The character's neck, elbows, hands and hips are main coordinates, as listed in Table I. Additionally, the character's skeleton is bound to a virtual avatar. The process of gesture interaction is shown in Fig. 4. Data of a character are collected with Kinect. Next avatar reproduces every movement of the character's at the same time and operates machine models in real time.

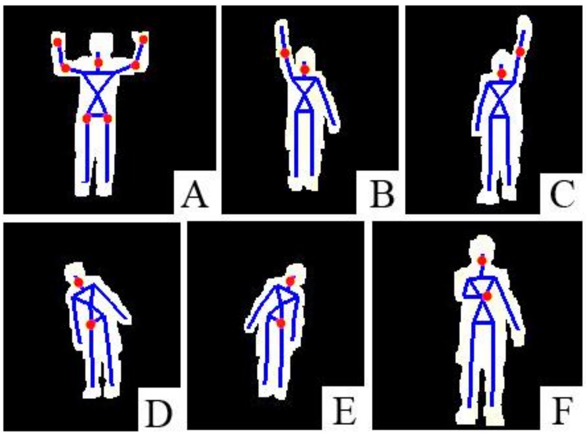

Figure 3. Our custom gestures.

The first gesture is to control an on-screen mouse cursor with user's right hand. Moving right hand to simulate the movement of mouse cursor and holding on the button for 0.5s to trigger the click events of the button. Using this gesture, users could achieve functions as follows: Zoom in a machine model, rotate a model, play a model's animation, switch scenes and cameras, and exit system.

The second gesture (Fig. 3, B) is used to automatically switch to the next scene by the following procedures: raising up the left hand and move upward, when the user's left elbow joint is greater than neck along Y axis the scene changes.

TABLE I. RECOGNIZABLE JOINTS LIST OF ESTS

1	**2**	**3**	**4**
Neck	Left Elbow	Right Elbow	Left Hand
5	**6**	**7**	
Right Hand	Left Hip	Right Hip	

The third gesture (Fig. 3, C) would switch to the last scene by using right hand to repeat the same procedures performed in the second gesture.

The fourth one (Fig. 3, D) would make the machine model rotate to the right by tilt body to the left until the left hip joint exceeds the neck joint along X axis, while the fifth gesture(Fig. 3, E) would do the opposite to make the target model rotate to the left.

Figure 4. Process of gesture recognition system.

The last gesture is shown in Fig. 3, F, to translate the left hand rightward, when the left hand joint is over the neck joint along X axis, the program would switch to the next camera.

V. ESTS SYSTEM DEMONSTRATION

In our system, 4 carton industrial machines including packaging machine, notching machine, grinding machine and binding machine (cf. Fig. 5) are created for training. Different combinations of these gestures we have defined to achieve functions as follows: zoom in machine models, rotate models or view cameras, switch cameras or machines, trigger some virtual buttons, quit system and demonstrate different animated segments about how the machines work.

In the functions, the training system mainly includes three modules:

A. 3D Panoramic Simulation Module

By roaming in the virtual manufacturing environment, users can achieve the results that they get personally in a realistic scenario.

B. Work Instruction Training Module

It can automatically simulate process of actual production, and provide standardized work instruction. According to their own needs, users can arbitrarily choose a process animation demo. Interactive operations are provided for manipulating virtual equipments. Users can view real-time states of internal mechanical motions of equipments in the section view.

C. Virtual Equipment Maintenance Module

Virtual equipment maintenance process animations provide standardized work instruction. When the user selects one kind of maintenance point, the module will automatically indicate its position, and illustrate the corresponding maintenance message.

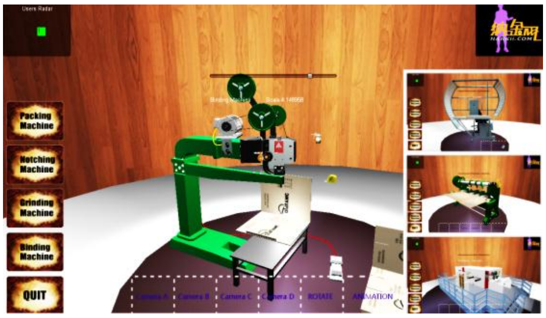

Figure 5. The four machines in the system.

VI. CONCLUDES AND FUTURE WORK

We presented a new way of interaction in interactive industrial training system. It is a high effective tool for training new workers. Additionally, it is interesting, impressive and easy to learn as well. This system has been applied to a carton factory in China for half a year and more than 120 workers have been trained. About 200 new workers are trained each year in the factory.

There are three main future lines to follow from this point on. Firstly, further research on how the system is developed will allow us to use more interactive methods to interact with it. Kinect's audio device and motor device technology being developed could be applied to improve its performance. That provides a wider range of operating area and fantastic spoken commands interaction. The second future work line covers the development of being able to identify and track more operators, and richer interactive contents. Thirdly, we would combine Kinect with augmented reality glasses to create a more immersive training experience.

ACKNOWLEDGMENT

This work supported by the National Natural Science Foundation of China (Grant No. 61174161, the second collaborator) and the 2011 Fujian Provincial Science and Technology Foundation (Grant No. 2011H0031).

We are grateful to Narkii (China) Information Technology for financial support on practical application research.

REFERENCES

[1] H. M. J. Hsu, "The potential of kinect as interactive educational technology," IPED. IACSIT Press, 2011, vol. 13, pp. 334-338.

[2] L. Johnson, R. Smith, H. Willis, A. Levine, and K. Haywood, "The new media consortium," *The 2011 Horizon Report* , Austin, TX, 2011.

[3] R. Johnson, K. O'Hara, A. Sellen, C. Cousins, and A. Criminisi, "Exploring the potential for touchless interaction in image-guided interventional radiology," in *Proc. Computer Human Interaction Conference*, 2011, pp. 3323-3332.

[4] E. Suma, B. Lange, A. Rizzo, D. Krum, and M. Bolas, "FAAST: The flexible action and articulated skeleton toolkit," in *Proc. IEEE Virtual Reality Conference*, 2011, pp. 247-248.

[5] O. H. Cho and W. H. Lee, "Gesture recognition using simple-openNI for implement interactive contents," *Lecture Notes, in Electrical Engineering*, vol. 179, pp. 141-146, 2012.

[6] S. Rogge, P. Amtsfeld, C. Hentschel, D. Bestle, and M. Meyer, "Using gestures to interactively modify turbine blades in a virtual environment," *Emerging Signal Processing Applications* (ESPA), pp. 115-158, 2012.

[7] C. Y. Chang, Y. T. Chien, C. Y. Chiang, M. C. Lin, and H. C. Lai, "Embodying gesture-based multimedia to improve learning," *British Journal of Educational Technology*, vol. 44, pp. E5–E9, 2013.

[8] Z. X. Gao "Study and implementation on somatosensory interaction in 3D virtual environments," in *Proc. 2012 IEEE Symposium on Electrical & Electronics Engineering*, 2012.

[9] K. D. Mankoff and T. A. Russo, "The Kinect: A low-cost, high-resolution, short-range 3D camera," *Earth Surface Processes and Landforms*, vol. 38, no. 9, pp. 926-936, 2013.

Jun-Feng Yao was born in Shanxi, China, in 1973. He received the B.S., M.S. and Ph.D. degrees in data modeling engineering from Central South University, Hunan, China, in 1995, 1998, and 2001, respectively, and the Post.D. degree in simulation and control engineering from Tsinghua University, Beijing, China, in 2003. From 2004 to 2005 he was a Visiting Researcher with University of SUSSEX, Brighton, UK. Between 2009 and 2010 he was a Visiting Researcher with Southern Polytechnic State University, Ga, USA. He is currently a Professor with Xiamen University, Xiamen, China, and is engaged in research on virtual reality, simulation and optimization of complex industry process.

Lin Yang was born in Chongqing, China, in 1990. He received the B.S. in electric science and technology from Hunan University, Hunan, China, in 2011. He is currently studying for his M.E. degree in electronics and communication engineering from Xiamen University, Xiamen, China. His major research interests are somatosensory interaction and virtual reality.

Lan Lei is a postgraduate student at the Software School of Xiamen University, Fujian, China. He received his B.E. in Artificial Intelligence in 2011, Xi'an University of Posts and Telecommunications. His major research interests are virtual reality and simulation.