

DNN-based Human Activity Recognition by Learning Initial Motion Data for Virtual Multi-Sports

Jong-Sung Kim

Creative Contents Research Division

ETRI, Daejeon, South Korea

js.kim@etri.re.kr

Abstract—This paper proposes a new deep neural network-based approach of human activity recognition for virtual multi-sports. The proposed approach acquires initial motion data of sports balls, not full motion data of human activities, with a single high-speed camera. Then, a deep neural network model is trained to recognize corresponding human activities, such as baseball batting, soccer kicking, golf swing, and so on, by learning those initial motion data. The proposed approach is very efficient and effective for practical applications of virtual multi-sports. In practice, the proposed approach was successfully applied to a unified virtual multi-sport platform. The effectiveness and efficiency of proposed approach was verified with real experimental results.

Keywords— Human activity recognition, deep neural networks, initial motion data, virtual multi-sports, high-speed camera

I. INTRODUCTION

More and more people want to play virtual sports, such as virtual baseball, virtual soccer, virtual golf, and so on, in indoor spaces, because it becomes difficult and uncomfortable to play real sports in outdoor places due to environmental or social changes [1]. For a decade, many computer graphics and vision technologies have been extensively studied to recognize, simulate, and visualize objects and humans for virtual sports. Nowadays, we can play virtual sports in indoor spaces regardless of weather conditions and place limitations.

Recently, virtual multi-sports have been introduced to play many kinds of sports at the same indoor space with the unified multi-sport platform developed for commercial purposes. Actually, the unified multi-sports platform is more attractive than previous virtual sports platforms in that it can supports several sports at the same time and place. However, it necessarily requires many high-tech, high-cost hardware and software, such as multiple high-speed cameras, multi-sports simulators, human activity recognizers, and so on.

In this paper, a new approach of human activity recognition is proposed for virtual multi-sports. The proposed approach acquires the initial motion data of sports balls with a single high-speed camera, not multiple ones. Then, a deep neural network model is trained to recognize human activities, such as baseball batting, soccer kicking, golf swing, and so on, by learning those initial motion data of sports balls instead of full

motion data of human activities. In practice, the proposed approach can be used for virtual multi-sports as shown in Fig. 1. The effectiveness and efficiency of proposed approach were verified with real experimental results.

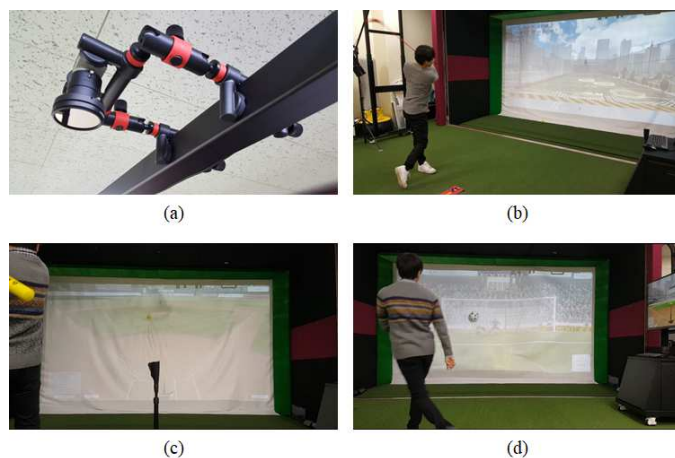


Figure 1. Virtual multi-sports and human activities. (a) High-speed camera equipment. (b) Golf swing. (c) Baseball pitching. (d) Football kicking

II. INITIAL MOTION ACQUISITION

Most of human motion recognition systems adopt either multiple high-speed (HS) cameras [2][3] or multiple inertial-motion-unit (IMU) sensors [4][5] to acquire full motion data of human activities. Although such systems can acquire the full motion data with high accuracy, full motion acquisition systems using either multiple HS cameras, such as Vicon [6] or IMU sensors, such as Xsens [7], are very expensive. In addition, the full motion data of human activities are too complicated to simulate and visualize virtual multi-sports.

As shown in Fig.1, the proposed approach acquires the initial motion data of sports balls with a single HS camera, not multiple ones, which makes it possible to reduce the cost and time for virtual multi-sports. The previous work [8] acquired the initial motion data of sports balls with a single multi-exposure (ME) camera, but the proposed approach uses a single HS camera. As illustrated in Fig. 2, the i -th initial motion datum corresponding to the c -th human activity, where c belongs to the set of human activities, i.e., {baseball batting,

soccer kicking, golf swing, ... }, include initial 3D position, P_i^c , 3D linear velocity, V_i^c , and 3D angular velocity, W_i^c . The ball motion analysis process in [3] was applied to the i -th image sequences in order to compute P_i^c , V_i^c , and W_i^c with a single HS camera, which of all intrinsic and extrinsic parameters were calibrated in a prior.

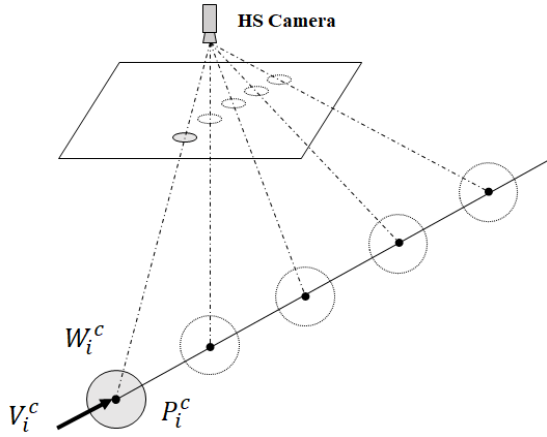


Figure 2. Initial motion acquisition with a single HS camera for virtual multi-sports.

III. HUMAN ACTIVITY RECOGNITION

The proposed approach performs the human activity recognition by learning the initial motion data of sports balls with a deep neural network (DNN) model [9]. The initial motion data of sports balls work as feature vectors, of which the dimension is up to 9, for recognizing human activities, which makes the proposed approach more efficient than the previous approaches using the full motion data of human activities acquired with multiple cameras or sensors. In the proposed approach, a fully-connected DNN model with few layers can be applied to learn the initial motion data of sports balls. Usually, a convolutional neural network (CNN) or recurrent neural network (RNN) model should be used to learn full motion data of human activities [2][3]. The overall procedure of the initial motion acquisition and human activity recognition for virtual multi-sports is illustrated in Fig. 2. This procedure was applied to our unified virtual multi-sports platform in practice.

IV. EXPERIMENTAL RESULTS

The effectiveness and efficiency of proposed approach was verified with real experimental results. The efficiency and effectiveness of the proposed approach was tested with real motion data of baseball batting, soccer kicking, and golf swing.

For experiment, the speed of HS camera was set to 120 frames per second. By using the HS camera, we acquired 300 initial motion data for each activity. Then, we divided those motion data into 120 training data and 180 testing data for each activity. For activity recognition, we used the DNN model with 3 layers and 10 nodes. The AdaGrad algorithm, of which the learning rate and epochs were set to 0.1 and 30, was used to train the DNN model.

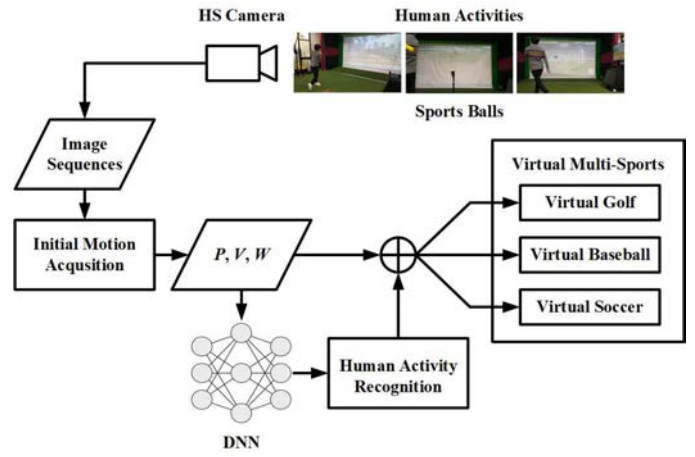


Figure 3. Initial motion acquisition and human activity recognition for virtual multi-sports.

To analysis the accuracy of human activity recognition for different data configuration, we tested $\{P, W\}$, $\{P, V\}$, $\{V, W\}$, and $\{P, V, W\}$. Table 1 shows the accuracy of human activity recognition with the proposed deep learning-based approach for virtual multi-sports. As shown in Table 1, the proposed approach using all initial motion, $\{P, V, W\}$, outperformed the other cases.

TABLE 1. ACCURACY OF HUMAN ACTIVITY RECOGNITION

Initial Motion Data	$\{P, W\}$	$\{V, W\}$	$\{P, V\}$	$\{P, V, W\}$
Batting	96.67	64.17	95.00	93.33
Kicking	95.83	80.83	90.83	92.50
Swing	1.67	76.67	67.50	98.33
Average	64.72	73.89	84.44	94.72

V. CONCLUSIONS

This paper presented a new DNN-based approach of human activity recognition for virtual multi-sports. The proposed approach acquired the initial motion data of sports balls, including baseballs, soccer balls, and golf balls, with a single HS camera. Then, a DNN model was trained to recognize human activities, such as batting, kicking, and swing, by learning those initial motion data of sports balls, not full motion data of human activities, contrary to previous approaches. The proposed approach was successfully applied to the unified virtual multi-sport platform that can be used to play virtual baseball, virtual soccer, and virtual golf in practice. The effectiveness and efficiency of proposed approach was verified with real experimental results.

ACKNOWLEDGMENT

This research was supported by Ministry of Culture, Sports and Tourism (MCST) and Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program R2020040036.

REFERENCES

- [1] C. Choi, C. Greenwell, M. A. Hums, and M. E. Hambrick, "Understanding consumer behaviors in virtual golf: Differences in leisure constraints," *Sport Marketing Quarterly*, vol. 28, no. 1, pp. 46–57, 2019.
- [2] A. Sundaresan and R. Chellappa, "Markerless motion capture using multiple cameras," in *Proc. IEEE CVPR'05*, 2005, USA, pp. 15-26.
- [3] P. U. Putra, K. Shiman, and K. Shimatani, "Markerless activity recognition method based on deep neural network model using multiple cameras," in *Proc. IEEE CoDIT'18*, 2018, Greece, pp. 13-18.
- [4] A. Filippeshi, M. Schmitz, M. Miezal, G. Bleser, E. Ruffaldi, and D. Stricker, "Survey of motion tracking methods based on inertial sensors: A focus on upper limb human motion," *Sensors*, vol. 17, 6 1257, Jun. 2017.
- [5] F. M. Rueda, R. Grzeszick, G. A. Fink, S. Feldhorst, and M. ten Hompel, "Convolutional neural networks for human activity recognition using body-worn sensors," *Informatics*, vol. 5, no. 1, pp. 1–7, May 2018.
- [6] Vicon motion capture system, <http://www.vicon.com>
- [7] Xsens motion tracking system, <http://www.xsens.com>
- [8] J. Kim and M. Kim, "A new single camera-based ball motion analysis system for virtual sports," in *Proc. ACM ICVIP'17*, 2017, Singapore, pp. 212-217.
- [9] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, Jan. 2015.



Jong-Sung Kim received his BS degree in radio sciences and engineering in 2000 from Korea University, Korea, his MS degree in electrical and electronic engineering in 2002, and his PhD degree in 2008 from POSTECH, Korea. In 2002 he worked at Microsoft Research Asia in Beijing, China, as a visiting student. Currently, he is a senior researcher with the S/W & Content Research Laboratory, ETRI, Daejeon, Korea. His research interests include image processing, computer vision and virtual reality.