Denoising 3D Skeleton Frames using Intersection Over Union

Tserenpurev Chuluunsaikhan*, Jeong-Hun Kim*, Jong-Hyeok Choi**, Aziz Nasridinov*
*Dept. of Computer Science, Chungbuk National University
**Bigdata Research Institute, Chungbuk National University
{teo, etyanue, leopard, aziz}@chungbuk.ac.kr

Abstract

The accuracy of real-time video analysis system based on 3D skeleton data highly depends on the quality of data. This study proposes a methodology to distinguish noise in 3D skeleton frames using Intersection Over Union (IOU) method. IOU is metric that tells how similar two rectangles (i.e., boxes). Simply, the method decides a frame as noise or not by comparing the frame with a set of valid frames. Our proposed method distinguished noise in 3D skeleton frames with the accuracy of 99%. According to the result, our proposed method can be used to track noise in 3D skeleton frames.

1. Introduction

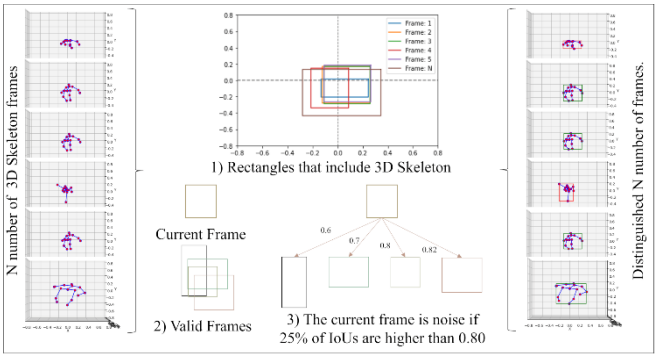
Recently skeleton data has been widely used in the area of action recognition or detecting dangerous situations [1, 2]. Skeleton data brings some advantages, such as not considering personal information privacy, can be to upload to a server quickly, and others. However, obtaining skeleton data properly from a video is still a challenge in computer vision field. Because the quality of skeleton data depends on conditions like barely visible joints, occlusions, clothing, lighting, and many others. Therefore, there are many approaches to improve the skeleton data quality.

In this study, we proposed a methodology to distinguish noise frames in 3D skeleton data. To distinguish noise frames, we used Intersection Over Union (IOU), which is a method for measuring how similarity of two rectangles. Our proposed methodology has the following advantages: a) can detect a frame with too small joints; b) can detect a frame of sudden change; c) can detect a frame with joints' wrong positions.

2. Methodology

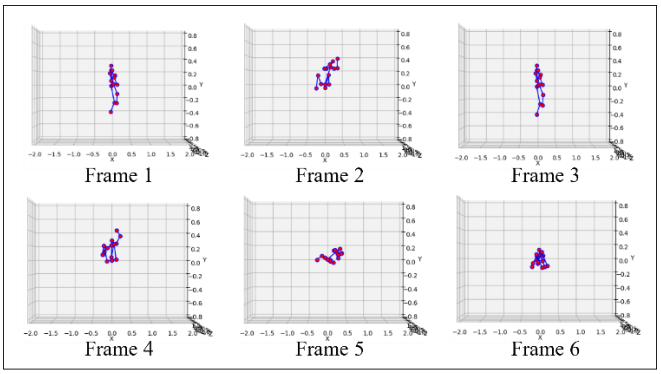
In this study, we propose a method to distinguish noise frames from 3D skeleton frames using IOU. Fig 1 demonstrates the overview of the methodology. The methodology consists of the following steps:

1. Creating a rectangle that includes a 3D skeleton for each frame.
2. Selecting M (i.e., 10), which is the number of initial valid frames based on their width and height ratio.
3. Distinguishing noise frames by comparing a rectangle of each frame with the rectangles of valid frames.
4. When a frame is not noise, updating valid frames by removing the first frame and inserting the frame into the last.



(Fig 1) The overview of the methodology.

A 3D skeleton dataset used in our study was extracted from CCTV videos by using object detection and pose estimation methods [1]. Kim et al. [2] used the 3D skeleton dataset for a study called Abnormal Situation Detection on Surveillance Video Using Object Detection and Action Recognition. The authors mentioned that the study achieved a higher accuracy for detecting abnormal situations than other existed methods. But the 3D skeleton dataset includes noise frames, which affect the model accuracy adversely. Therefore, we aimed to distinguish the noise frames from normal frames using IOU. In this dataset, 3D skeletons have a characteristic that the hip center is located at (0, 0, 0). According to the characteristic, we can apply IOU to distinguish noise in 3D skeleton frames. Fig 2 shows the example of noise frames. The example consists of six frames. Frames 1 and 3 are only normal frames, but not others. Here, the frame 2 represents a sudden change, the frames 4 and 6 represent joints' wrong positions, and the frame 5 represents the example of missing joints. A noise in frames is usually caused by missing joints or joints' wrong position.



(Fig 2) The example of noise frames.

IOU is a method to calculate the area of overlapped sections of two rectangles. Specifically, it is the ratio of overlapped sections and non-overlapped sections. The IOU of two rectangles is found by Equation 1. Here, the area of intersection represents overlapped sections, and the area of union represents non-overlapped sections. The IOU of two rectangles returns values between 0 and 1. We can assume that if the value is 1, the rectangles completely overlapped.

$$IOU = \frac{Area\ of\ Intersection}{Area\ of\ Union} \tag{1}$$

To distinguish noise frames, first, we created a rectangle (i.e., a box included a 3D skeleton fully) for each frame. The rectangles were created by Equation 2. Then, we selected ten initial valid frames based on their ratio of width and height. The valid frames were selected by Equation 3. After that, each frame rectangle was compared with the valid frame rectangles by Equation 1. We assumed a frame as noise if the frame is not similar to 75% of IOUs. The IOUs of each frame are calculated by Equation 4.

$$box_i = [\max(x), \max(y), \min(x), \min(y)]$$
$$boxes = \sum_{i=1}^{N} [box_i] \tag{2}$$

$$validBoxes = \begin{cases} if\ \ i < 10\ \ and\ \ width_i \leq height_i * 0.8\ \ add \\ else\ \ not\ \ add \end{cases} \tag{3}$$

$$IOUs = \sum_{j=1}^{M} IOU_{i,j} \tag{4}$$

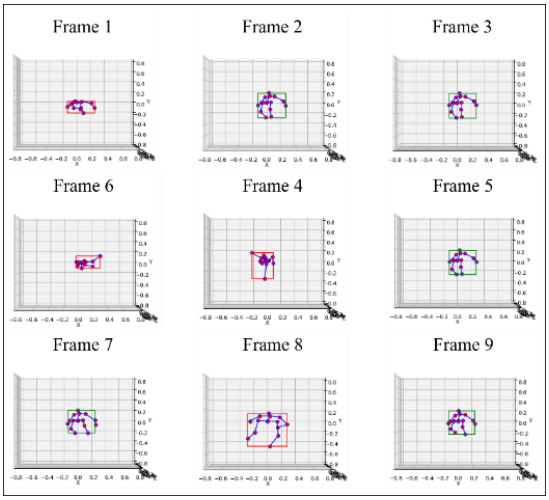
3. Results

To evaluate the proposed method, skeletons obtained from five videos of 734 frames were used. Each frame was labeled manually as noise or not. Then, we compared it with the result of our methodology. Table 1 shows the classification report of the methodology. Here, precision represents the rate of all frames that labeled as normal or noise, how many actually normal or noise; recall represents the rate of all frames that actually normal or noise, how many are labeled as normal or noise; F1-score represents the average of precision and recall; support represents how much data is used for the test. Lastly, accuracy is the ratio of correctly labeled data and the total data.

<Table 1> The classification report of the methodology.

	Precision	Recall	F1-score	Support
Normal	0.99	1.00	0.99	707
Noise	0.87	0.74	0.80	27
Accuracy			0.99	734

Fig 3 shows the visualization example of the method. Here, red box represents the noise frame and green color represents normal frame. Out methodology can distinguish the following noise frames: a) too small skeleton; b) suddenly big changes between frames; and c) the chaotic position of joints.



(Fig 3) The visualization example of the methodology.

4. Conclusion

In this study, we distinguished the noise frames in 3D skeleton frames using IOU. The results have shown that the method can detect noise in 3D skeleton data. The model distinguished the following types of noises like sudden changes and joints' wrong positions, and others. Moreover, our methodology does not require much time, since it is not machine learning or deep learning-based method because execution time is important in a real-time video analysis system. In the feature, we will apply the proposed method to the existing model [2] and compare the results with and without our methodology.

Acknowledgement

This work was supported by Institute for Information communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (2016-0-00406, SIAT CCTV Cloud Platform).

References

[1] K. Lee, I. Lee, and S. Lee. "Propagating LSTM: 3D Pose Estimation based on Joint Interdependency," Proceeding of the European Conference on Computer Vision (ECCV), pp. 119-135, 2018.

[2] J.-H. Kim, J.-H. Choi, Y.-H. Park, and A. Nasridinov, "Abnormal Situation Detection on Surveillance Video Using Object Detection and Action Recognition," Journal of Korea Multimedia Society, vol. 24(2), pp. 186-198, 2021.