Surveillance in Daycare Centers Using 3D Depth Sensor

Hee Jung Yoon and Sang Hyuk Son*
Daegu Gyeongbuk Institute of Science and Technology

Abstract : According to the United States Department of Health and Human Services, more than 3 million reports of child abuse are made in the nation, involving more than 6 million children. This is a huge issue not only in the U.S. but also all around the world, specifically in areas with a high children population such as daycare centers. While daycare centers are considered to be relatively safe, these environments present one of the biggest concerns against of safety of children. In this work, we focus on detecting child abuse that are portrayed by teachers in daycare centers. We use streaming 3D skeleton joint coordinates obtained from Kinect sensors to (1) classify a teacher from a group of children and (2) detect physically abusive behaviors such as such as hitting, kicking, slapping, shaking, and pushing that are performed by the adult. Specifically for age group classification, we analyze the histograms of training samples and implement a bin-based classification method that represents bin-boundaries. For abusive behavior detection, we use a combination of an array of supervised learners to recognize a predefined set of abusive actions. By designing such a system, we aim at opening the possibilities for developing applications that contribute to preserving the safety of children in daycare centers.

Keywords : abusive action, skelatal joint data, Kinect

I. Introduction

Recently, child abuse by teachers in daycare centers has sparked a storm of outrage in Korea [1]. The number of child abuse cases in daycare centers is increasing each year as there were 100 cases in 2010 and 232 cases in 2013 [2]. After series of recent incidences, lawmakers have agreed to enact a policy in an effort to prevent child abuse by requiring at all daycare centers in the nation to deploy surveillance cameras [3]. However, merely deploying CCTV cameras cannot immediately detect nor alert a child's parent if an abusive behaviors is acted onto the child. Particularly for working parents, they will not be able to rely on the recordings during working hours or might not have the time to look through the entire recordings to see if any suspicious act was done. Hence, we introduce a potential application that can contribute to preserving the safety of children in daycare centers.

We focus our work on physical abusive since it is an act of commission that cause actual physical harm or have the potential for harm. For this purpose, we use skeleton joint information from Kinect to (1) classify a teacher from a group of children and (2) detect physically abusive behaviors such as such as hitting, kicking, slapping, shaking, and pushing that are portrayed by the adult. To begin our study, in this paper, we introduce a preliminary design of our system and survey few of recent studies that relate to our work.

* 교신저자(Corresponding Author)
Hee Jung Yoon and Sang Hyuk Son : DGIST
※ This research was supported in part by Global Research Laboratory Program (2013K1A1A2A02078326) through NRF, and the DGIST Research and Development Program (CPS Global Center) funded by the Ministry of Science, ICT & Future Planning.

Ⅱ. System Overview

As shown in Fig. 1, our system is composed of three major architectural components: sensing, classification, and action discovery. This section briefly describes these components in order to provide a high-level overview of the system. We use the streaming skeleton data from the Kinect to detect and classify actions. Kinect recognizes up to 6 users within its field of view, and by default, selects the first two users to track 20 joints on each of their skeletons. Skeleton frames are generated at the rate of 30 frames per second, and each frame consists of the 3D coordinates of 20 body joints along with their tracking states. The 3D coordinates are with respect to a frame of reference centered at Kinect.

To detect skeleton and skeletal movement, our system accumulates a window of w frames and analyzes it. The skeleton tracker and movement detector is a simple threshold-based classifier which acts as an admission controller for the classification phase. It computes the variances of joint coordinates, compares them against a set of predefined and configurable thresholds, and if any of the values are above some threshold, the window is passed on to the bin-based classifiers or expert classifier.

Our system employs two types of classifiers, each of which classifies: (1) child vs. adult and (2) a specific abusive actions. These classifiers are similar in the sense that they classify a window of frames and provide a confidence value of that window's likelihood of belonging to a specific age group or action class.

Ⅲ. Classifiers

We design a system that classifies a human as a child or an adult using a bin-based classifier. Specifically, our system performs the following three steps in order to achieve robust classification: bin creation, frequency detection,

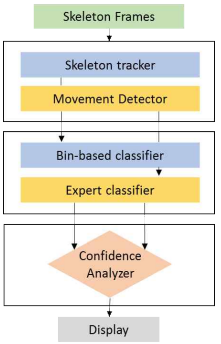

Fig. 1. System overview

weight computation. For detecting abusive actions, five supervised classifiers are used to recognize five common abusive actions: hitting, kicking, slapping, shaking, and pushing. Each of these are binary classifiers recognizing an action from a window of skeleton frames obtained from Kinect. Frames from Kinect are first converted into feature vectors which are invariant to relative position and orientation of the body, and speed of an action. The feature vectors are used to train support vector machine (SVM) classifiers.

Ⅳ. Related Works

While automatically recognizing characteristics of a person such as identity, emotion, and gender, has been extensively studied, automatic age group estimation has hardly been as explored, despite the fact that it is alone an interesting problem. One of few reasons is that, aging works in a personalized, uncontrollable way that make it a challenging task than that of other characteristics. Of the few existing studies, Basaran et al. propose a system that classify children from adults by using joint information extracted from Kinect [5]. Sandygulova et al. presents a method that gathers 3D body metrics from a robotic system to estimate age and gender [6]. Our work will

similarly classify children from adults by using 3D joint information, however, we will also combine to our system the detection of abusive behaviors that are portrayed by adults.

To the best of our knowledge, there has not been be a study that supports the concerns of child abuse. However, several agitation detection techniques have been proposed in the past. Physiological features, such as the galvanic skin response [7], heart rate [8], and skin temperature [8] have been used to train models of agitating behavior. Studies on the acceleration of the body joints are performed to understand their association with agitation [6]. The weakness of these methods is that, they require an active engagement of the person while collecting data, which is very inconvenient. Compared to these works, the system that we propose is completely non-invasive and passive monitoring system. Kinect has also recently been used in several gesture recognition applications. Skeleton data from Kinect is used to recognize fall detection [9], gait analysis [10], and gesture recognition [11]. However, unlike our study, none of these works consider the issues of abusive behavior.

V. Conclusion

This work presents one of the first efforts to detect abusive actions that are portrayed by teachers in daycare centers. First, we utilize Kinect to collect skeleton joint information of children and adults to classify a teacher from a group of children. Second, we detect abusive actions from adults such as hitting, kicking, slapping, shaking, and pushing. In this paper, we introduce our preliminary system overview and survey few of related works that connects to our proposed work. As of future works, we want combine what we have done in our previous works to develop a system that can be deployed in daycare centers to support the safety of children. We envision that such a system can act as a catalyst to realize applications that can protect children and reduce worries from parent who leave their children in daycare centers.

References

[1] A. Edwards and P. Hancocks, Horrific child abuse at nursery caught on camera is causing outrage, CNN, 2015. http://fox13now.com/2015/01/15/horrific-child-abuse-at-nursery-caught-on-camera-is-causing-outrage

[2] Outrage over abuse at daycare center, The Chosunilbo, 2015. http://english.chosun.com/site/data/html_dir/2015/01/16/2015011601422.html

[3] CCTV to be required at all S. Korean daycare centers, The Korea Times, 2015. http://www.koreatimesus.com/cctv-to-be-required-at-all-s-korean-daycare-centers/

[4] Child Maltreatment 2012, Children's Bureau of the U.S. Dep. of Health & Human Serv., 2012. http://www.acf.hhs.gov/sites/default/files/cb/cm2012.pdf#page=65

[5] Can Basaran, Hee Jung Yoon, Ho Kyung Ra, Sang Hyuk Son, Taejoon Park, and JeongGil Ko, Classifying Children with 3D Depth Cameras for Enabling Children's Safety Applications. ACM UbiComp, 2014.

[6] A. Sandygulova, M. Dragone, and G.M.P. O'Hare, Real-time adaptive child-robot interaction: Age and gender determination of children based on 3d body metrics, IEEE Robot and Human Interactive Communication, pp. 826 - 831, 2014.

[7] W. Liao, W. Zhang, Z. Zhu, and Q. Ji, A real-time human stress monitoring system using dynamic bayesian network, IEEE CVPR, pp. 70 - 70, 2005.

[8] G. E. Sakr, I. H. Elhajj, and H.-S. Huijer. "Support vector machines to define and detect agitation transition", IEEE Trans. on Affective Computing, Vol. 1, No. 2, pp. 98 - 108, 2010.

[9] R. Planinc and M. Kampel, "Introducing the

use of depth data for fall detection", Personal and Ubiq. Computing, Vol. 17, No. 6, pp. 1063-1072, 2012.

[10] M. Gabel, R. Gilad-Bachrach, E. Renshaw, and A. Schuster, Full body gait analysis with kinect, IEEE Engineering in Medicine and Biology Society (EMBC), pp. 1964 - 1967, 2012.

[11] S. Celebi, A. S. Aydin, T. T. Temiz, and T. Arici. Gesture recognition using skeleton data with weighted dynamic time warping, Computer vision theory and applications, 2013.