Biomedical Signal Processing and Control 44 (2018) 146–153

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Human fall detection using machine vision techniques on RGB–D images

Leila Panahi [a], Vahid Ghods [b,*]

[a] Department of Electrical and Computer Engineering, Semnan Branch, Islamic Azad University, Semnan, Iran
[b] Young Researchers and Elite Club, Semnan Branch, Islamic Azad University, Semnan, Iran

ARTICLE INFO

Article history:
Received 19 September 2017
Received in revised form 5 March 2018
Accepted 15 April 2018

Keywords:
Depth map
Fall detection
Kinect
Machine vision
Telemedicine

ABSTRACT

Falling represents one of the major problems faced by elderly people. In the present research, a machine vision-based system was designed. Depth map images were captured using Microsoft Kinect® camera. They were processed for extracting features and designing the detection algorithm, apply SVM classifier, to distinguish falling pose from normal pose in 70 video samples. Furthermore, another experiment was conducted on the basis of threshold on the feature of distance to the floor, with its outputs replaced SVM responses. In the fall detection algorithm, in order to calculate speed, image features were used rather than accelerometer data. Relying on depth map images and employing Open CV library, the present research outperformed similar works where color images or such devices as accelerometers were used, attaining sensitivity and specificity of 100% and 97.5%, respectively. The use of the distance of the person's centroid to the floor efficiently contributed into better results.

1. Introduction

In today's world, increased life expectancy and reduced fertility rate around the world have made elderly a global phenomenon. As of now, many countries (e.g. Iran) are also experiencing a transition in age structure from youth to elderly [1]. As such, the necessity of prospective planning to control problems related to this group of population have been emphasized. With its objective being to preserve elderly people health in old age, home safety has gained a great deal of [2]. On the other hand, modern life style and the development of electronic communication channels have introduced remote surveillance as a caring approach for elderly people wherein their independency is respected at the same time.

In addition to common elderly diseases, elderly people commonly experience problems with retaining their physical balance due to myasthenia, side-effects of particular drugs, poor eyesight, and lack of sufficient concentration. One of the major risks threatening elderly people is falling. Investigations have shown that, 65% of falls occur at home and 26% on public roads [3]. Falling has been reported as the second major cause of death in elderly people. In USA, the statistical results demonstrate that at least one-third of people aged 65 and over years fall once or more in a year [4]. 47% of non-injured fallers cannot get up without assistance and this period of time spent immobile also affects their health [5], such as failure to take drug on time.

In general, fall detection methods can be classified into two broad categories: those based on machine vision and others. In the methods not based on machine vision, a device (e.g. an accelerometer, cell phone, or smart flooring) is used for fall detection. Considering the limitations faced by these devices such as the need for attaching the device to the elderly's cloth and high rate of error of these devices in distinguishing different poses, the used of cameras and machine vision-based techniques for fall detection in house environment has been further admitted. Some researchers have gone for a combination of wearable devices and machine vision [4]. In the present research, an attempt was made to analyze the images compiled in [4] to design a fall detection system without using the sensors. The first feature used in this work is similar to the obtained inclination angle from ellipse extraction used in [5]; however, rather than color images, the present research goes for ellipse extraction from depth map images. The second feature is the aspect ratio of the bounding rectangle used for example in [4]; however, in order to achieve higher accuracy, the corresponding rectangle to the ellipse was considered. As the third feature, 3D position of the center of the ellipse was employed. This feature is defined as the distance between center of the ellipse and the plane representing floor in 3D space. In most of the works undertaken on depth map images, one can see the feature of distance to floor, but differences

* Corresponding author.
E-mail address: v.ghods@semnaniau.ac.ir (V. Ghods).

https://doi.org/10.1016/j.bspc.2018.04.014

Fig. 1. A Kinect sensor including an RGB camera, an IR projector and an IR camera [10].

come in the form of the choice of the point whose distance to floor is considered. Moreover, in some of researches, occurrence of fast and accelerated movements is assumed as an indication of falling, based on which sensors such as accelerometers were used.

In the present work, it was attempted to use image data to calculate the rate at which the individual lies, so as to distinguish slow lie downs from falling without using any attaching sensors.

The structure of this paper is as follows: Section 2 presents a review on existing literatures. Section 3 explains the proposed method. Experimental results are presented in Section 4. Conclusion and recommendations are given in Section 5.

2. Literature review

In the research works performed so far to achieve this objective, major approaches have been to use smart bathroom flooring [6], [7], attaching sensors [8], smart phones [9], various types of cameras, or a combination of these devices. Smart phones have been used to detect falling of the individual in outdoor environment. Smart floors could be only used to monitor bathroom environment. Even though the slippery nature of bathroom flooring increases the chance of falling, but an elderly may spend short times in bathroom, necessitating short-term monitoring for bathroom environment. In the present research, remote monitoring was considered for a place where an elderly spends most of his/her time. The use of attaching sensors has been poorly welcomed by elderly people. Moreover, considering the high rate of distinguishing error incurred by these sensors, researchers deviated to machine vision techniques.

In machine vision-based research on fall detection, two types of video images have been processed: RGB (red/green/blue) images and depth map images acquired by a set of infrared (IR) projector and camera called Kinect sensor and return a Grayscale image indicating the distance of different objects to the camera [10]. Some researchers used two or more RGB cameras to find depth map. They used stereovision method to calculate the third dimension of each pixel and perform 3D feature extraction.

Obtained depth map images from Kinect sensor not only are free of color and shadows, but also provide us with the third dimension of each pixel using the two types of cameras incorporated into the sensor. Kinect sensor has its own noises and disturbances [11]. For example, it encounters an undefined data error in direct sunlight. Pixels exposed to direct sunlight have a certain amount meaning that no depth information is available. A Kinect camera can be observed in Fig. 1.

In [4], researchers combined the information obtained from a special sensor called Inertial Measurement Unit (IMU) which was attached to the individual, replacing the accelerometer. In order to reduce rate of error, they employed Kinect camera and machine vision as a complementary tool in drawing the final conclusion. Among the strengths of their research was the compilation of a large database of movement information of various individuals. In this research, the authors not only extracted shape features, but also determined the equation of floor plane and used to calculate the distance between individual's center of gravity and floor; the distance was then used as a feature. Once the individual was confirmed to be on move based using IMU, depth map images were processed and in case abnormal state was detected by the IMU sensor, SVM was applied to extract features of the images, so as to detect laid pose. Used features by these researchers were aspect ratio of the individual's bounding rectangle, the ratio expressing the height to the individual's surrounding rectangle in the current frame to the natural height of the person, and standard deviations along the horizontal and vertical axes. The research ended up with a specificity and sensitivity of 99.67% and 100%, respectively. The authors of [12] took a similar approach of [4] with the difference that they used the fuzzy system instead of SVM to detect fall conditions from non-fall. They achieved specificity and sensitivity of 95% and 100%, respectively. In [3], two analyzing methods were used to select the more effective features from twenty one extracted features from depth images. They showed the feature that the genetic algorithm recommended was more useful for detecting a fall, than PCA. They yielded specificity and sensitivity of 95.45% and 97.13%, respectively.

In [13], a fall detection system based on tracking different parts of human body using Kinect camera was presented. Authors used modified random decision tree (RDT) algorithm to track head movements, so as to extract body joints in a 3D space. Once finished with finding the path through which head movements occurred, SVM classifier was used for fall detection and fed by 3D path of the points as input. Once an abnormal pose was reported by SVM, final step was to calculate the time spent in falling pose and send notifications in case the time was prolonged. In their experiments, they achieved specificity, sensitivity, and accuracy of 100%, 95.3%, and 97.6%, respectively, along with an error rate of 2.4%. In [14], methodologies were introduced to track body skeleton using Kinect camera. Individual's fitted 3D cube was extracted from depth map images and used for fall detection in [15]. In [16], the researchers applied an optical flow feedback convolutional neural network to the video stream in a home environment to detect human posture as a key factor for fall detection. They represented feature feedback mechanism scheme (FFMS) and achieved the accuracy of % 92.65. In [17], the position and velocity of person extracted from Kinect sensor was used to detect fall. They yielded specificity and sensitivity of 91.3% and 100%, respectively.

In the present research, an attempt was made to design a machine vision-based system where shape features similar to those in [5] and 3D position feature were extracted to detect human fall. In [5], due to the use of RGB images, not only complicated methods were used for background elimination, but also further shape features should be extracted as the individual's 3D position was lacking; these necessitated more complicated calculations. One of the features they used was a polar histogram that obtained from dozens of mathematical operations [5]. In the present work, we used simple shape features and employed 3D information contained in Kinect-captured images to obtain the distance feature from less than 10 mathematical operations.

3. Methodology

3.1. Database

In the present research, the database compiled in [4] was used. This database contains video and imagery files acquired using