

Human Abnormal Behavioral Detection for Video Surveillance

QIAN Ying^{1,a}, ZHANG Wenjing^{2,a}

¹Chongqing University of Posts and Telecommunications, College of Software, Chongqing 400065,

²Chongqing University of Posts and Telecommunications, College of Computer Science, Chongqing 400065

^azhangwenjji@163.com

Keywords: Kinect, depth image, skeleton tracking technology, fall detection

Abstract. With the increasing proportion of the elderly, fall will be a serious threat to the health of the elderly, especially elderly people who lives alone. Therefore, how to automatically detect abnormal behavior has become a key problem. Because of image recognition ,which is captured from the general monitoring, is affected by light, shelter and other factors, resulting in performance degradation. This paper introduced the Kinect device as the research platform. Depth images are acquired by analyzing data of human body height change and skeletal points ,which are acquired by skeletal tracking technology. Build an automatic determining abnormal algorithm, which can be alarming for abnormal behavior. The experiment show that the system of strong real-time performance and high detection rate.

Introduction

According to the UN survey predicts that 65 years old and over will account for 18.3% of the population in China until 2040[1]. With the accelerated pace of life, children can't take care of both the older adults and themselves. Therefore, more and more older adults have to live alone. If the older adults who live alone fall, many suffer serious injuries, and can't be found and rescued would have serious consequences[2]. As a result, systems that detect falls have become a hot research topic in recent years.

Fall detection system using wearable sensors in the early years[3]. [4] proposed uses wearable sensors, which is inconvenience and expensive to the older adults. [5] developed a fall detection system which using ordinary camera to detect falls. The system based on motion energy image method. But ordinary camera[6,7] can't work at night and low light conditions, which leads to the system can't be real-time monitoring. Meanwhile, the system is easy to leak user privacy. In order to solve this problem, it introduces Kinect that provides a color image data and 3D depth image data. [8] using the location and an average of 20 key points of speed that Kinect detected to detect falls.

In this paper, fall is defined as abnormal behavior. It presents a new method for detecting abnormal behavior. It is extracted values of height variable sequences, which is acquired by Kinect depth image, and skeleton joints which is acquired by tracking

technology. Using the values to train Support Vector Machine(SVM) classifier. Finally, detect abnormal behaviors and give alarm messages by trained classifier.

Kinect technology

Kinect, which provided by Microsoft company, contains a color camera, an IR camera, an IR emitter and a microphone array. Deep eyes is formed by the junction of an IR emitter and an IR camera. Skeleton tracking technology using machine learning to achieve[9]. The first step is split mask, using edge detection and noise threshold processing, the human body target feature point classification techniques to separate people from the background. The second step is body parts classification, human body contour are classified to the human body, which is using machine learning. The third step is joint point positioning, every possible pixel to judge which joint it is.

The Kinect provides 25 joints of the human body skeleton. All of these joints are defined in 3D coordinate. The origin of coordinate is the IR camera, x axis is horizontal, and the right axis is

positive. Y axis is vertical, and the upward is positive. The positive axis of z is the IR camera optical direction.

Human height measurement based on depth image

In this paper, using triangle knowledge to calculate the body's actual height that shown in Fig.1, which is based on depth image three dimensional pixel information [10]. Horizontal and vertical viewing angles of the camera Kinect respectively 60 ° and 70 °. The Kinect provides approximately 30 frames per second of data. If we assume the Kinect is on a level surface, which is at the distance of 1m above the ground.

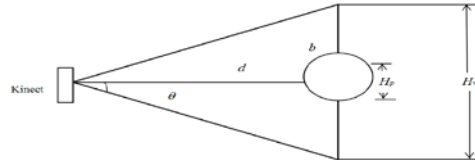


Fig.1 The principle diagram of the height measurement

In Fig.1, the relevant parameters are defined as follows:

θ : the vertical Angle of the camera (35 °);

d : the actual distance from the user to the camera;

b : a half of width of the perspective of the user's effective field

H_p : Body depth pixel height, it is to say that the highest pixel H_{top} and the lowest pixel H_{bottom} of body parts. $H_p = H_{top} - H_{bottom}$

H_w : width of background pixel, the default is 320 mm

H_r : actual height of the user.

Calculation formula is as follows:

$$b = d \times \tan \theta \quad (1)$$

$$\frac{H_p}{H_w} = \frac{H_r}{2b} \Rightarrow H_r = \frac{2b \times H_p}{H_w} \quad (2)$$

Therefore, It is convenient to calculate the actual height of the user by H_p and d . It is the most important clues that the height of human body is varied.

Proposed Algorithm

In this paper, a human abnormal behavioral detection system by using Microsoft Kinect sensor as an input device. The reason the Kinect sensor is utilized is because it contains several sensors (RGB camera, depth sensor, and 4 microphone) and it is not affected by variation in illumination. In the proposed algorithm used in this work, height is acquired by depth image information and joints are acquired by skeleton tracking technology. The flow chart of abnormal behavioral detection is shown in Fig.2

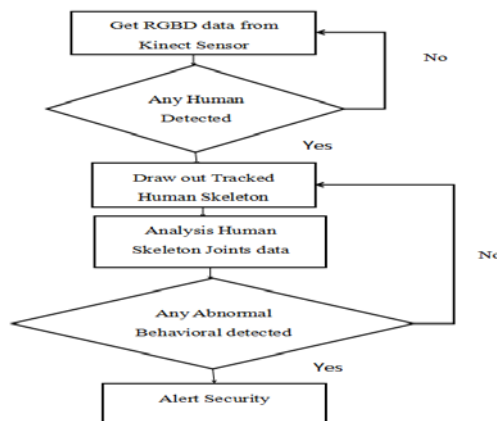


Fig.2 Flow Chart Abnormal Behavioral detection

The following are the steps involved in the analysis of human abnormal behavioral:

1. Get RGB-Depth data from the Microsoft Kinect sensor.

2. Judge whether there are any human. If detected human, the human height is

$$d = H_r \quad (3)$$

3. Human body joints distance data in 30 frames is saved to calculate the velocity. In this paper, velocity of head joint is calculated using the formula(4)(5)

For frame i and $i+1$ the velocity for a particular joint normal to the floor is then:

$$v_i = \frac{d_{i+1} - d_i}{t_{i+1} - t_i} \quad (4)$$

This velocity is averaged over N frames :

$$v_{avg} = \frac{1}{N-1} \sum_{i=1}^{N-1} \frac{d_{i+1} - d_i}{t_{i+1} - t_i} \quad (5)$$

where N =number of frame, d =distance calculated using formula(3)

4. SVM cross-validation algorithm build modeling about characteristic values to detect

abnormal behavioral of fall. In the above steps, it has get height, the head velocity and the minimum distance from ground to head joint and spine center joint characteristic values. It judged target sequence whether it is fall abnormal behavioral by trained model. If the target sequence is fall, the warning information given immediately.

SVM

SVM is a binary classification pattern recognition model. Its main idea is to create a classification hyperplane as a decision surface, so that the edge of isolation between the two categories is maximized. It is based on statistical learning theory, specifically, SVM is a structural risk minimization of approximation. If the training sample is

$$S = \{(x_1, y_1), \dots, (x_m, y_m)\} \in (X \times Y)^m \quad (6)$$

Where $x_i \in X = \mathbb{R}^n$ is input vector, $y_i \in Y = \{1, -1\} (i = 1, 2, \dots, m)$ is category labels, x_i is feature vectors.

1) In the case of linear separable, the training set can be linear by a hyperplane, it is $W \bullet X + b = 0$, where W and b decided the hyperplane position. The problem is transformed into the problem of optimal hyperplane:

$$\min \phi(w, \xi) \quad \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n \xi, c \geq 0, (i = 1, 2, \dots, n) \quad (7)$$

$$s.t. \quad y_i [W \bullet X_i + b] \geq 1 - \xi \quad \xi \geq 0 \quad (8)$$

Where W is coefficient vector of feature space classification hyperplane, ξ is considered the relaxation factor, C is corresponding to penalty factor for samples of incorrect classification samples.

2) In the case of nonlinear separable, the low-dimensional input space \mathbb{R}^n mapped to high-dimensional feature space H in order to linearly separable through a mapping function. First, define a kernel function $K(x_i, x_j)$ and the appropriate parameter C , this problem is converted into

$$\min \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j a_i a_j K(X_i \bullet X_j) - \sum_{i=1}^n a_i, (i = 1, 2, \dots, n) \quad (9)$$

$$s.t. \quad \sum_{i=1}^n y_i a_i = 0, \quad 0 \leq a_i \leq C \quad (10)$$

Results and Analysis

In this paper, the experimental condition is: the 64 bit Windows 10 operating system; Intel (R) Core (TM) i5-4590 CPU @ 3.3GHz; RAM 8.00 GB; The camera device is Kinect for Xbox 360; The camera shooting conditions is on the surface, which is at the distance of 1m above the ground. In order to make the training more effective, the paper collected different person, which are distinguish male and female, high and low differences in different people, action bone data. It

collected the most common six kinds of behavior: walking, jogging, bend, sit down, squat, fall. There are five people in this experiment. There are 60 fall behaviors, 100 bend, sit and squat behaviors.

In the process of learning of SVM training, artificially divided into training set and test set assembly is affected by the data distribution randomness. So the cross validation method can reduce or eliminate the influence of random distribution of the experimental data on the experimental results. The grid search method is used to determine the optimal parameters of the SVM optimization. Fig. 3 is the optimization result of penalty parameter c (cost) and the kernel function g (gamma) parameter.

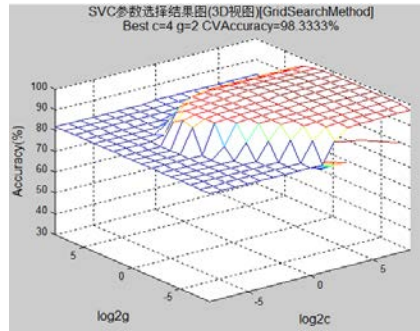


Fig.3 SVM parameter selection results

It shows that the difference between this paper and other paper.[11] the detection rate is 94%, which is based on RGBD monitor video to detect fall. The specific algorithm is to judge the height variation by SVM model. In contrast, it is not affected by light and the detection rate is higher in this paper.[12] Kinect tracking technology is used to detect fall event and the total rate is 93%. Select the person's head, left shoulder, right shoulder, two hip center, left hip, right hip and so on six skeleton point location, relative velocity, moving velocity and left time to judge human fall event.

Conclusion

It is built common six behavior database by the values of highly variable sequences that acquired by Kinect depth image and skeleton joints that acquired by tracking technology. It is located target sequence by sliding window algorithm and extracted height, the head velocity and the minimum distance from ground to head joint and spine center joint. It is detected event by build model of SVM cross validation algorithm, and detection rate is 98.33%. Therefore, the experiment show that SVM cross validation algorithm model can accurate real-time detect fall.

References

- [1]Jing Yuejun,Wang Shengjin.The predicament and countermeasures of population Chinese sustainable development will face in the 21st century[J].Journal of population,2001,1:3-7.
- [2]Heinrich S,Rapp K,Rissmann U,etal.Cost of falls in old age:A systmatic review[J].Osteoporosis International,2010,21:891-902.
- [3]LIU Peng,LU Tancheng,LU Yuanyuan,etal.MEMS tri-axial accelrometer based fall detection[J].Chinese Journal of Sensors and Actuators,2014,27(4):570-574.
- [4]Centers for Disease Control and Prevention. Falls Among Older Adults: An Overview. <http://www.cdc.gov/HomeandRecreationalSafety/Falls/adultfalls.html>
- [5]Li Yibo, Abnormal behavior detection in homes of older adults based on intelligent monitoring[J]. Computer engineering and application,2014,31(2).
- [6]Bulin D, Benezeth Y, Courtial E. Posture recognition based on fuzzy logic for home monitoring of the elderly[J]. Information Technology in Bio-medicine, IEEE Transactions on, 2012,16(5): 974-982.

- [7]Khoshelham K, Elberink S O. Accuracy and resolution of kinect depth data for indoor mapping applications[J]. Sensors,2012, 12(2): 1437-1454.
- [8]Christopher Kawatsu, Jiaxing Li, and CJ Chung. Development of a Fall Detection System with Microsoft Kinect. Lawrence Technological University, USA, April 25, 2012.
- [9]Shotton J, Sharp T, Kipman A, etal. Real-time human pose recognition in parts from single depth images [J].Communications of the ACM, 2013, 56 (1) :116-124.
- [10]Zhou Changshao,Shi Zhuo. Height measurement system design based on the depth image[J]. Electronic science and technology of guilin university journal,2013, 33(3):214-217.
- [11]Wang Junze,Zhu Xiaolong, Qu Chang.Fall automatic detection of human body based on Kinect skeleton tracking[J].Journal of Shanghai jiaotong university,2015,49(9).
- [12] Zhang C, Tian Y, Capezuti E. Privacy preserving automatic fall detection for elderly using RGBD cameras[M]. Springer Berlin Heidelberg, 2012.