A Depth Camera Based Fall Recognition System for the Elderly

Rachit Dubey[1], Bingbing Ni[2] and Pierre Moulin[3]

1. Nanyang Technological University, Singapore, 639809
`rach0012@e.ntu.edu.sg`
2. Advanced Digital Sciences Center, Singapore, 138632
`binbing.ni@adsc.com.sg`
3. University of Illinois at Urbana-Champaign, IL 61820-5711
`moulin@ifp.uiuc.edu`

Abstract. Falls are a great risk for elderly people living alone. Falls can result in serious injuries and in some cases even deaths. It is important to recognize them early and provide assistance. In this paper we present a novel computer vision based fall recognition system which combines depth map with normal color information. With this combination it is possible to achieve better results as depth map reduces many errors and gives more information about the scene. We track and extract motion from the depth as well RGB map and then use Support Vector Machines to classify the falls. Our proposed fall recognition system recognizes and classifies falls from other actions with a very high accuracy (greater than 95%).

Keywords: fall detection, SVM, feature extraction, depth image

1 INTRODUCTION

Nowadays, countries have to face the growing population of elderly people. Many of them need to be kept in safe environment with continuous monitoring. It is important to develop technologies to provide a secure living environment and improving the quality of life.

For the older people living alone, one great danger which needs to be monitored is that of falling. Action recognition for the elderly has been an active research topic over the past few years and various efforts have been made over it. Zhang et al. used mobile phone to detect falls by embedding an accelerometer into them [1]. Researchers have also used wearable devices to detect falls [2]. However the problem with wearable devices and mobile phones is that people will often forget to carry or wear them hence sometimes falls will not be monitored and detected.

Computer vision offers a better solution to solve this problem by providing an analysis of actions and activity while doing continuous monitoring. Some researchers extracted motion from the videos using Motion History Image (MHI) and used it to detect falls [3], [4]. However these systems assumed that the motion is very small after a fall. Töreyin et al. combined audio and used Hidden Markov Models (HMM) to detect falls [5]. However, the setup required a silent environment and hence could not be used in all cases. Most of the fall recognition systems have been limited by the sensing device i.e. the color camera restricted them and failed to capture the depth information. Some researchers proposed to use a 3D camera for performing visual fall detection [6] but the camera used had low image resolution and high cost.

Recent emergence of inexpensive depth sensors like the Kinect has made it possible to capture the depth map along with the color images with good resolution and accuracy. The sensing range of the depth sensor is adjustable, and the Kinect software can sense 3D structure of the scene. This allows any cluttered background to be easily segmented out. Thus the combination of the color and depth information can be used to provide comprehensive 3D information of the scene and the person/s involved.

Hence the need to build a fall recognition system which utilizes depth camera and sensor techniques for higher accuracy was felt. Using MHI [7] and Hu moment [8], we propose a high accuracy fall detection system that combines the above techniques with RGB-Depth (RGB-D) information. To further enhance the accuracy and make the system more flexible, we use support vector machines (SVM). This can effectively analyze data and recognize patterns to classify falls from other actions.

The main aim of this paper is to explore the effectiveness of combining depth map with color information and to propose a novel fall recognition approach. An overview of the proposed system is depicted in Figure 1.

The rest of this paper is organized as follows. In Section 2, the fall recognition model is proposed. The results of experiments are presented in Section 3, and Section 4 concludes this paper.

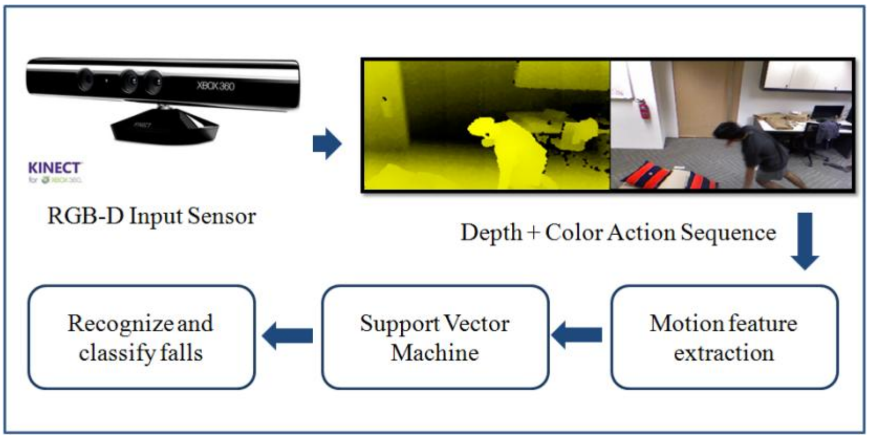

XBOX 360
KINECT for XBOX 360
RGB-D Input Sensor
Depth + Color Action Sequence
Motion feature extraction
Support Vector Machine
Recognize and classify falls

Fig. 1. Overview of the proposed fall recognition approach.

2 SYSTEM OVERVIEW

In this section, we first introduce our feature extraction method. MHI was used to extract the motion from the dataset and then Hu moments were used to extract the features from the MHIs. We then confirm and classify the falls using Support Vector Machine (SVM).

2.1 Motion History Image

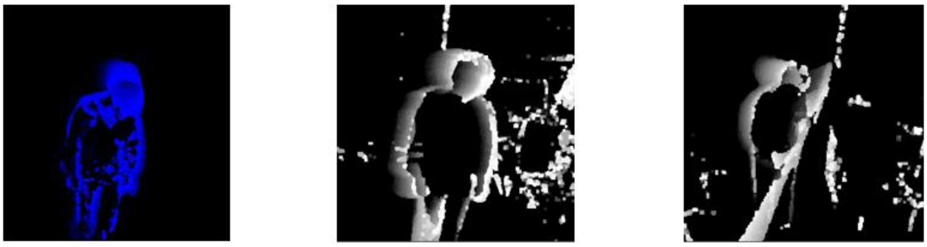
Whenever a person falls, the fall will be associated with a large change in motion. Hence it is important to extract and keep track of the motion information. The result of MHI is an image which stores information on the recency of the motion. The pixel brightness gives information about the most recent motion in the image sequence and thereby tells how the person has moved during the course of an action.

$I(x, y, t)$ is an image sequence and the threshold value for generating the mask is K, t is the fixed duration and $\tau,$ the maximum time window. Then each pixel intensity value in an MHI is a function H_t of the temporal history of motion at that point, namely:

$$H_t(x, y, t) = \begin{cases} \tau, if\big(I(x, y, t) - I(x, y, t-1)\big) > K \\ max(0, H_t(x, y, t-1) - 1)\,, otherwise. \end{cases} \tag{1}$$

2.2 Three- Dimensional Motion History Images (3D-MHIs)

The limitation with conventional MHIs is that they only use the normal RGB camera. In our earlier work [9] we proposed to use depth camera and a new approach - Three-Dimensional Motion History Images (3D-MHIs). The 3D-MHI approach combines the MHIs with two additional channels and extracts motion from depth map as well. For the depth channel we propose two types of Motion History Images named as DMHIs. DMHIs contain forward-DMHIs (fDMHIs) and backward-DMHIs (bDMHIs). To generate the fDMHIs, the same approach as that of MHI is taken except $I(x, y, t)$ i.e. image sequence is replaced by $D(x, y, t)$ which is a depth sequence. Therefore fDMHIs give information about forward motion history i.e. increase of depth. The backward DMHI on the other hand encodes the decrease of depth and the backward motion history. The bDMHIs are generated in a similar fashion except that the thresholding function is replaced by$\big(D(x, y, t) - D(x, y, t-1)\big) < -K$. Hence we combined the conventional MHIs with the 2 DMHIs – fDMHIs and bDMHIs, together they form the 3D-MHIs and we used them for motion extraction. The 3D-MHIs work well in our case as not only they detect change in motion in x-y direction but also change in the depth direction. So actions and falls with similar change in x-y direction but different motion in depth direction which could not be distinguished earlier are now easily classified with the use of 3D-MHIs. Figure 2 shows the 3D-MHIs for an action sequence and clearly show the contrast between them.



MHI – motion history image

fDMHI – forward depth MHI

bDMHI – backward depth

Fig. 2. The 3D-MHIs, showing the difference between them.

2.3 Hu Moment

Hu described a set of 6 moments along with a 7th skew invariant that has been implemented and used in several areas [8]. The 6 moments are that of rotation, scaling and translation invariance. We calculate the 7 hu-moments for each of the 3D-MHIs and use these 21 features for classification.

2.4 Support Vector Machines (SVM)

To classify the fall from other actions we propose to use SVM. After we calculate the seven hu–moments, we store them in a representation matrix to be used as features for training. SVM finds an optimal way to separate data by finding an optimal hyperplane. This optimal hyperplane maximises the margin of separation between two classes which need to be classified. We want to fix w_o and b_o that define the optimal hyperplane.

$$g(x) = w_o^T x_i + b_o = 0 \tag{2}$$

So a point x_i, in the training data ($1 \leq i \leq N$) is classified as:

$$g(x) = w_o^T x_i + b_o \geq +1 \quad \text{for } d_i = +1 \tag{3}$$

$$g(x) = w_o^T x_i + b_o \leq -1 \quad \text{for } d_i = -1 \tag{4}$$

3 EXPERIMENT AND RESULTS

3.1 Dataset Construction

We use the Microsoft Kinect camera to construct the database for the experiments. All the videos are collected in a lab environment. The horizontal and vertical distances are both 2 metres. This distance is measured from the camera to the centre of the set. The set is such that it is ideal for home and hospital monitoring. The resolutions of both colour image and depth map are 640 × 480 in pixel. The colour image is of 24-bit RGB values; and each depth pixel is a 16-bit integer. Both sequences are synchronized and the frame rates are 30 frames per second.

3.2 Dataset Structure

We are interested in distinguishing falls from other activities. For this purpose a dataset consisting of daily human activities was constructed. Inspired by the guidelines in [10], we have included 12 different kinds of human activities. These include: *make a phone call, mop the floor, enter the room, exit the room, go to bed, get up, eat meal, drink water, sit down, stand up, take off the jacket* and *put on the jacket*. Many different volunteers were asked to perform these actions. Our dataset also consists of 79 *fall down* video sequences, one of which is shown in Figure 3. Our final database consisted of a total of 1198 labeled videos.

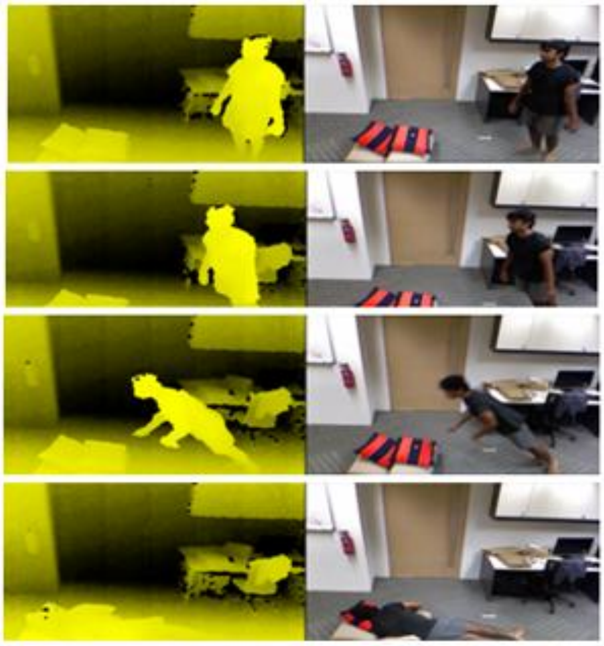


Fig. 3. Figure showing image and depth map of a fall down sequence.

3.3 Experimental Results

To conduct the experiments, we divided our dataset into two parts. We define the *fall down* videos as positive samples and videos of daily activities as negative samples. To train the SVM, we constructed a training set consisting of 160 videos. Out of these 40 of them were positive samples and 120 were negative. In this way the positive samples comprised 1/3rd of the total samples. The testing set consisted of 39 positive samples and the rest 999 videos were negative samples. The performance was measured based on the recognition accuracy and error rate given by:

$$\text{Recognition accuracy} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

$$\text{Error rate} = \frac{\text{False Positive}}{\text{True Negative} + \text{False Positive}}$$

True positive and true negative denote the correctly classified and false positive and false negative cover the wrongly classified videos.

Our proposed system was implemented on Intel's OpenCV library. We used a 64 bit Intel i5 core CPU with 6 GB memory. The results of our proposed fall recognition system are shown in Table-1. Out of 39 *fall down* videos, 38 videos were correctly classified.

RGB-D	Total Videos	Classified falling	Classified non falling
Fall down	39	38	1
Daily activities	999	30	969

Table 1. The experimental results of the proposed fall down recognition system

Furthermore, we also conducted experiments using only the RGB information (only 7 features extracted from MHI). The comparison results of our proposed system with that of only RGB are shown in Table-3.The RGB system gives a recognition accuracy of 87% whereas RGB-D gives an accuracy of 97% thus significantly outperforming the RGB system. This is also evident from the precision/recall curves compared in Figure 4.

RGB	Total Videos	Classified falling	Classified non falling
Fall down	39	34	5
Daily activities	999	50	969

Table 2. The experimental results of the RGB system

	RGB-D	RGB
Recognition Accuracy	97%	87%
Error rate	0.03%	0.05%

Table 3. Comparison of the proposed system with the RGB system in terms of recognition accuracy and error rate

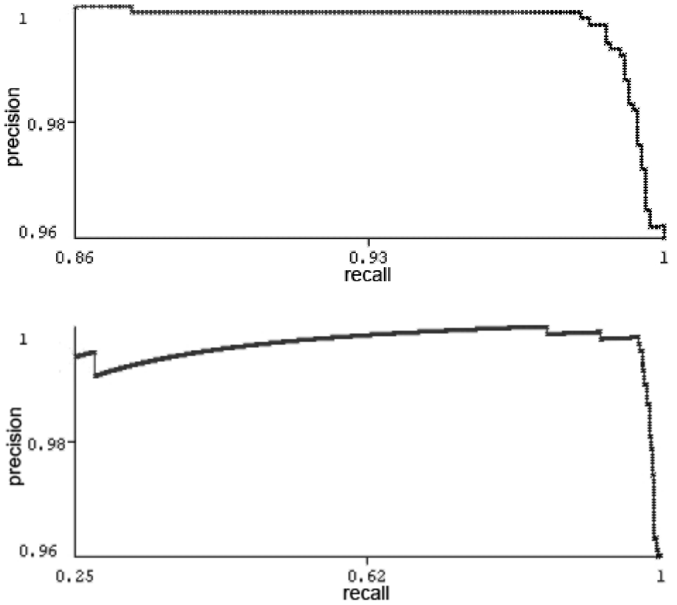


Fig. 4. Comparison of the precision/recall curves. From top to bottom: RGB-D and RGB respectively

4 CONCLUSION AND FUTURE WORK

We have presented a novel approach to detect falls and help the elderly people using an inexpensive depth sensor. Our system gives good results and can recognize falls with over 95% accuracy. We have also compared and evaluated the advantages of using depth information over commonly used RGB. Our future work will consist of extracting more features to make a system capable of recognising all kind of actions and which can be used for many other purposes.

5 REFERENCES

1. Fall Detection by Embedding an Accelerometer in Cellphone and Using KFD Algorithm *Tong Zhang, Jue Wang, Ping Liu and Jing Hou, in* IJCSNS International Journal of Computer Science and Network Security, VOL.6 No.10, October 2006
2. Fall Detection by Wearable Sensor and One-Class SVM Algorithm Tong Zhang, Jue Wang, Liang Xu and Ping Liu in Intelligent Computing in Signal Processing and Pattern Recognition
3. Video Analytic for Fall Detection from Shape Features and Motion Gradients *Muhammad Jamil Khan and Hafiz Adnan Habib in* Proceedings of the World Congress on Engineering and Computer Science 2009 Vol II WCECS 2009, October 20-22, 2009, San Francisco, USA
4. Fall Detection from Human Shape and Motion History Using Video Surveillance *Rougier, C.; Meunier, J.; St-Arnaud, A.; Rousseau, J.; Dept. d'Inf. et de Rech. Operationnelle, Univ. de Montreal, Montreal, QC in* Advanced Information Networking and Applications Workshops, 2007, AINAW '07.
5. Hmm based falling person detection using both audio and video *B. T¨oreyin, Y. Dedeoglu, and A. C¸ et in* IEEE International Workshop on Human-Computer Interaction, Beijing, China, 2005.
6. 3D human pose recognition for home monitoring of elderly *Bart Jansen, Frederik Temmermans and Rudi Deklerck in* Proceedings of the 29th Annual International Conference of the IEEE EMBS Cité Internationale, Lyon, France August 23-26, 2007.
7. The representation and recognition of action using temporal templates *A. Bobick and J. Davis in* IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(3):257–267, 2001.
8. M. Hu. Visual pattern recognition by moment invariants. IRE Transactions on Information Theory, 8(2):179–187, 1962.
9. B. NI, G. Wang and P. Moulin, RGBD-HuDaAct: A Color-Depth Video Database For Human Daily Activity Recognition, ICCV workshop on consumer depth camera, 2011.
10. K. Krapp. Activities of daily living evaluation. Encyclopedia of Nursing and Allied Health, 2002.