

A Novel Shadow-Assistant Human Fall Detection Scheme Using a Cascade of SVM Classifiers*

Yie-Tarng Chen, You-Rong Lin, and Wen-Hsien Fang

Department of Electronic Engineering,
National Taiwan University of Science and Technology,
Taipei, Taiwan, R.O.C.
{ytchen,whf}@mail.ntust.edu.tw

Abstract. Visual recognition of human fall incidents in video clips has been an active research issue in recent years. However, most published methods cannot effectively differentiate between fall-down and fall-like incidents such as sitting and squatting. In this paper, we present a novel shadow-assistant method for detecting human fall. Normally, complex 3-D models are used to estimate the human height. However, to reduce the high computational cost, only the information of moving shadow is used for this context. Because the system is based on a combination of shadow-assistant height estimation, and a cascade of SVM classifiers, it can distinguish between fall-down and fall-like incidents with a high degree of accuracy from very short sequence of 1-10 frames. Our experimental results demonstrate that under bird's-eye view camera setting, the proposed system still can achieve 100% detect rate and a low false alarm rate, while the detection rate of other fall detection schemes have been dropped dramatically.

Keywords: fall detection, SVM.

1 Introduction

In recent years, visual recognition of human fall incidents in video clips has been an active research issue. In this paper, we consider the problem of using a mono un-calibrated camera to detect if senior citizens fall over, called fall-down incidents hereafter. Such incidents normally occur suddenly and take approximately 0.45 to 0.85 seconds. Both the posture and shape of the victim change rapidly, and he/she usually lies inactive on the floor. Hence, drastic changes in the posture, shape and height of the body are key features in human fall detection. However, modeling those features with low computational complexity is a not a trivial task, especially for accurate human height estimation.

A number of fall detection schemes have been proposed [4-5]. Simple features derived from shape analysis, such as the aspect ratio of the bounding box, the

* This work was supported by National Science Council of R.O.C. under contract NSC 100-222-E-011-134.

angle of the fall and a vertical projection histogram have been used for fall detection. Rougier [4] proposed a fall detection approach based on the Motion History Image (MHI) [3] and changes in body shape. Hidden Markov Models (HMMs) have also been utilized for fall detection. Hiseh [5] developed a triangulation-based skeleton extraction approach to analyze human movements; however, it is not designed specifically for detection fall-down incidents. No approach based on simple features can detect all kinds of human falls. Most video-based fall detection systems based on simple features suffer from high false alarm rates because they do not differentiate between fall-like and fall-down incidents. The high computational cost of human skeleton extraction discourages researchers from using it for real-time human fall detection. Hence, there is need for a reliable fall detection system, a combination of several approaches, which can increase the detection accuracy while still satisfying the real-time constraint.

1.1 A Motivation Example

Figure 1 illustrates the motivation for this paper, which attempted to differentiate the falling posture through the shadow information. We can observe a correlation between the height of standing, sitting down and falling postures and their relevant shadow areas. In particular, the shadow area approaches 0 for a falling posture. Hence, we attempt to investigate the possibility to utilize shadow information for human fall detection. However, shadow is not a stable image cue, especially, it is dependent on the capturing conditions. if a person stands just below a light source, where the projection angle of a light is vertical, the length of a person's shadow is still cannot be detected. Hence, there is need an intelligent combination of the shadow information with other approaches which can increase detection accuracy.

In this paper, we propose a real-time video-based human fall detection system which can support both bird-view and flat-view camera furnishing. The proposed system applies a novel shadow-assistant human height estimation scheme to differentiate between fall-down and fall-like incidents. Normally, complex 3-D models are required to estimate human height in bird-view camera finishing.

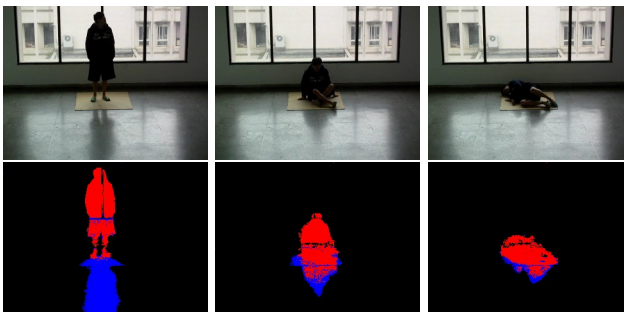


Fig. 1. The shadow (blue) and human foreground (red) for standing, sitting and falling

However, to reduce the high computational costs, the shadow-assistant approach instead, even though it provides less information than 3-D models, but satisfies the requirements of a real-time fall detection system. Furthermore, a fall into any angle can be detected. Specifically, the proposed approach can successfully detect fall toward camera with a low computational cost.

The contributions of this paper are two-fold:

1. We have address the use of shadow for visual recognition of human fall. Although shadows have been used to measure the height of a building on aerial images. To our knowledge, shadows are first introduced to human activity recognition in this paper.
2. We develop a fall detection system based on intelligent combination of shadow analysis, shape analysis and motion analysis, achieve 100% detect rate and a low false alarm rate from very short sequence of 1-10 frames while satisfying the real-time constraint.

The remainder of the paper is organized as follows. Section 2 presents foreground objects and moving shadows segmentation. In Section 3, we describe the novel shadow-assistant detection scheme. The results of experiments are detailed in Section 4 and Section 5 concludes the paper.

2 Foreground Objects and Moving Shadows Detection

Foreground object segmentation and moving shadows detection are important preprocessing steps in human fall detection schemes. We use the statistical background subtraction and shadow detection algorithm [1] developed by Horprasert et al. for this context. However, other foreground ground extraction and moving shadow detection algorithms still can be applied to the proposed fall detection scheme.

2.1 Background Modeling

In background modeling, we attempt to obtain a background model and its parameters by several selected images. Each pixel in the background model is assumed to be independent, and it can be represented as a tuple with four parameters. The background is modeled statistically on a pixel by pixel basis. A pixel is modeled by a 4-tuple $\langle E_i, S_i, a_i, b_i \rangle$. where S_i is a standard deviation of color value. It normalizes the pixel color in this work. It is given by

$$S_i = [\sigma_R(i), \sigma_G(i), \sigma_B(i)] \quad (1)$$

where $\sigma_R(i), \sigma_G(i), \sigma_B(i)$ are the standard deviations of the i -th pixel's red, green, blue values over N training frames. The expected color value of pixel i is given by

$$E_i = [\mu_R(i), \mu_G(i), \mu_B(i)] \quad (2)$$

where $\mu_R(i), \mu_G(i), \mu_B(i)$ denote arithmetic means of the i -th pixel's red, green, blue values over N training frames. Let $I_i = [I_R(i), I_G(i), I_B(i)]$ represent the

pixel’s RGB color value in current image. We want to measure the distortion of I_i from E_i by discomposing the distortion measurement into two components: brightness distortion α_i and chromaticity distortion CD_i respectively, which are defined in Equations (3) and (4).

$$\alpha_i = \frac{\left(\frac{I_R(i)\mu_R(i)}{\sigma_R^2(i)} + \frac{I_G(i)\mu_G(i)}{\sigma_G^2(i)} + \frac{I_B(i)\mu_B(i)}{\sigma_B^2(i)}\right)}{\left(\left[\frac{\mu_R(i)}{\sigma_R(i)}\right]^2 + \left[\frac{\mu_G(i)}{\sigma_G(i)}\right]^2 + \left[\frac{\mu_B(i)}{\sigma_B(i)}\right]^2\right)} \tag{3}$$

$$CD_i = \sqrt{\left(\frac{I_R(i) - \alpha_i\mu_R(i)}{\sigma_R(i)}\right)^2 + \left(\frac{I_G(i) - \alpha_i\mu_G(i)}{\sigma_G(i)}\right)^2 + \left(\frac{I_B(i) - \alpha_i\mu_B(i)}{\sigma_B(i)}\right)^2} \tag{4}$$

Furthermore, we consider the variation of the brightness and chromaticity distortions over space and time of the training background images. a_i represents the variation of the brightness distortion of i -th pixel, which is given by

$$a_i = \sqrt{\frac{\sum_{i=0}^N (\alpha_i - 1)^2}{N}} \tag{5}$$

b_i represents the variation of the chromaticity distortion of i -th pixel, which is given by

$$b_i = \sqrt{\frac{\sum_{i=0}^N (CD_i)^2}{N}} \tag{6}$$

2.2 Pixel Classification

Since the different pixels yield different distribution of α_i and CD_i . In order to use a single threshold for all of pixels, we re-scale α_i and CD_i as normalized brightness distortion $\hat{\alpha}_i$ and chromaticity distortion \widehat{CD}_i respectively.

$$\hat{\alpha}_i = \frac{\alpha_i - 1}{a_i} \tag{7}$$

$$\widehat{CD}_i = \frac{CD_i}{b_i} \tag{8}$$

Each pixel $M(i)$ is classified into one of the four categories: B (Background), S (Shadow), H (Highlighted background), and F (Foreground object) by the following decision rule:

$$M(i) = \begin{cases} F : \widehat{CD}_i > \tau_{CD} \text{ or } \hat{\alpha}_i < \tau_{\alpha lo}, \text{ else} \\ B : \hat{\alpha}_i < \tau_{\alpha 1} \text{ and } \hat{\alpha}_i > \tau_{\alpha 2}, \text{ else} \\ S : \hat{\alpha}_i < 0 \text{ else} \\ H : \text{ otherwise} \end{cases} \tag{9}$$

where $\tau_{CD}, \tau_{\alpha lo}, \tau_{\alpha 1}$, and $\tau_{\alpha 2}$ are selected threshold values to determine the similarities of the chromaticity and brightness between the background image and the current observed image.

3 Human Fall Detection System

3.1 System Description

After Foreground object segmentation and object tracking, the system begins to perform fall detection. The proposed fall detection scheme as shown in Figure 2 consists of cascaded classifiers which integrates height estimation, posture analysis, shape analysis, and inactivity detection. We assume that a person is in an upright posture when first appearing in the video sequence and becomes immobile on the floor after a fall. All video sequences are captured by a stationary camera, which can be furnished in either a flat-view or a bird view.

Since Histograms of Oriented Gradients (HOG)[2] detectors have shown to give significantly high performance in upright human detection, first, an input image is sent to the HOG-based upright posture detector. Next, any non-upright human foreground is sent to a shadow-assistant falling posture detector as shown in Figure 3, which performs height estimation and shape analysis to detect a falling posture. Finally, we confirm a fall incident by monitoring the inactivity of the person by using a motion history image (MHI)[3].

For classification in the upright posture detection, the falling posture detector and inactivity detector, we use state-of-the-art machine learning techniques—support vector machines, which have been a popular approach for pattern classification and nonlinear regression because of its robustness even in the absence of a rich set of training examples. The virtual height, the HOG vector, and MHI vector are major features in the proposed fall detection scheme, which will be discussed in the following sections.

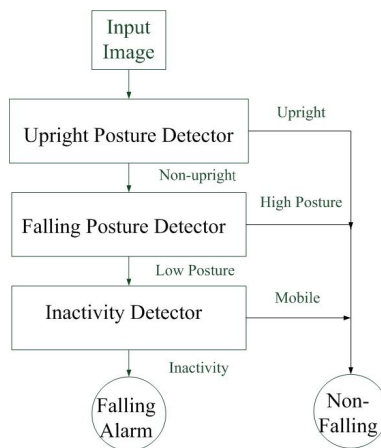


Fig. 2. The flowchart of the shadow-assistant fall detector

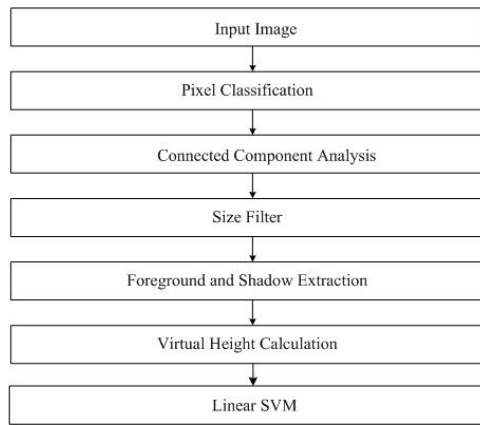


Fig. 3. The flowchart of the falling posture detection scheme

3.2 Upright Posture Detection

The Histogram of Oriented Gradient (HoG) human detector [2] is applied to detect the upright human posture. This detector combined locally normalized histogram of oriented Gradient (HoG) descriptor and a linear support vector machine classifier. The major idea behind the Histogram of Oriented Gradient descriptors is to describe local object appearance and shape by the distribution of intensity gradients or edge direction. The HOG upright posture detector consists of the following steps: gradient computation, orientation binning, block normalization and SVM classifiers. The input to the SVM upright pose classifier comes from a HOG vector in the foreground object. However, instead of using all pixels in an image to compute HOG, only pixels within a bounding box enclosed the foreground object. Similar to the SVM height classifier, The sign +1 of the SVM output was assigned to the upright posture and -1 to the non- upright posture. The details of the Histogram of Oriented Gradient (HoG) human detector can be found in [2].

3.3 Shadow-Assistant Falling Posture Detector

The virtual height, denoted as V_h , is defined as

$$V_h(B) = \begin{cases} \frac{HS_a}{WF_a} & \text{if } B \neq 0 \\ \frac{H}{W} & \text{if } B = 0 \end{cases} \tag{10}$$

where H and W are the height and the width of a bounding box respectively, S_a represents the areas of a shadow, and F_a represent area of a foreground object, which are given by:

$$S_a = \sum_{i=1}^n \delta(M(i) - S) \tag{11}$$

$$F_a = \sum_{i=1}^n \delta(M(i) - F) \quad (12)$$

If the shadow is occurring, both shadow analysis and shape analysis, such as the aspect ratio of the bounding box, are used to detect the falling posture, whereas, if the shadow is not available, only the shape analysis is considered to detect the falling posture.

We assume that a person first appears in the scene with an upright posture. We can use the first few images in the video sequence to detect the presence of the shadow. If the upright detector is flagged as positive and the shadow area is approximate to zero, the shadow information cannot be used for human falling detection. However, if the upright detector is flagged as positive, and the shadow area is larger than a threshold, the shadow information cannot be used to the human fall.

B is the index of the shadow-assistant function in the human fall detection system. If $B = 1$, the shadow-assistant function operates. On the other hand, if $B = 0$ the shadow information will not be considered in the proposed human fall detection scheme.

$$B = \begin{cases} 1 & \text{if } U_r(t) = 1 \text{ and } S(t) = 1 \\ 0 & \text{if } U_r(t) = 1 \text{ and } S(t) = 0 \end{cases} \quad (13)$$

where $U_r(t)$ is the upright detector for frame(t) and $U_r(t) = 1$ if the upright detector classifies frame(t) as an upright human posture and $U_r(t) = 0$ otherwise. $S(t)$ is the shadow index for frame(t) and $S(t) = 1$ if the area of the human shadow in frame(t) is larger than a predefined threshold and $S(t) = 0$ otherwise.

3.4 Inactivity Detector

Motion History Image (MHI) [3] is defined by a simple replacement and decay operators:

$$H_\tau(x, y, t) = \begin{cases} \tau & \text{if } D(x, y, t) = 1 \\ \max(H_\tau(x, y, t-1) - 1) & \text{otherwise} \end{cases}$$

where $D(x, y, t)$ is a binary image sequence indicating region of motion, generated by image differencing.

Motion History Image (MHI) is a scalar-value image where more recently moving pixels are brighter. The Motion History Image can be used to represent how motion the image is moving.

We check the following condition to confirm the occurrence of a fall-down incident: the Motion History Image (MHI) of the human object is lower than the threshold T_p . This condition confirms that the person is inactive for a period of time after a fall.

4 Experiments and Results

4.1 Experimental Settings

In order to validate the efficacy of the shadow-assistant shape analysis for human fall detection, we perform experiments on five datasets: NTUST bird-view dataset, NTUST outdoor dataset, NTUST Lab dataset, ETHZ dataset, and Shoaib dataset. Since the page is limited, only the performance of NTUST bird-view dataset and Shoaib dataset are reported in this paper. We implemented the proposed shadow-based fall detection system on Intel's OpenCV library. The experiments were run on a PC with Windows XP, an Intel Pentium D 3.2GHz CPU and a 2 GB RAM. The performance metrics used in the fall detection experiments are the detection rate P_d and the false alarm rate P_n . The detection rate, i.e., the fraction of fall events detected correctly, is calculated as follows:

$$P_d = \frac{TP}{(TP + FN)} \quad P_n = \frac{FP}{(TN + FP)}$$

where true positive (TP) and true negative (TN) are the counts of correct detection while false positive (FP) and false negative (FN) are the counts of incorrect prediction.

We compared the performance of the proposed fall detection system with that of existing approaches: approach 1: the proposed shadow assistant scheme, approach 2: combination of shape analysis and MHI [4], approach 3: Skeleton matching [5] and, approach 4: Shape analysis. shape analysis uses the aspect ratio of the bounding box as a feature and Support Vector Machine as a classifier. In skeleton matching, a skeletons is classified by near-neighbor scheme and the distance map is used as the distance function.

4.2 Experiments on NTUST Bird-View and Shoaib Data-Sets

NTUST bird-view dataset as shown in Figure 4 contains 6 activities performed by one actor in the NTUST design square, an outdoor environment and captured in a bird-view by a single stationary and un-calibrated camera furnished on 8 meters height. The dataset consists of 54 actions in 1800 frames, including 27 falling actions, 7 sitting-down on the ground, 5 sitting on a chair, 4 squat actions, 8 bend actions and 12 walking actions. Shoaib data-set contains one actor in a cluttered home environment.

The proposed shadow-assistant scheme outperforms other fall-detection approaches in both the NTUST bird-view dataset and Shoaib dataset as shown in Table 1 and Table 2 since the only light source in the NTUST bird-view dataset is the sunlight, a simple point light source, which can cast simple shadow contours. The detection rate of other human fall-down detection schemes drop significantly in bird-view dataset in compassion with they did in flat-view datasets since the actual height of a human posture in a bird-view image by cannot be estimated by conventional two-dimensional fall detection schemes. The average execution time of the shadow-assistant scheme takes 0.156 second per frame, which satisfies the real-time constraint.



Fig. 4. Parts of the test videos:(left three) walking, sitting-down, and falling down from bird-view dataset, and (right two) bending and falling down from Shoaib dataset

Table 1. Comparison of fall detection schemes for outdoor bird view data sets

approach	TP	FN	TN	FP	P_d	P_n
our approach	27	0	34	2	1.00	0.055
approach 2	20	7	25	11	0.74	0.31
approach 3	24	3	33	3	0.889	0.08
approach 4	25	2	31	5	0.926	0.138

Table 2. Comparison of fall detection schemes for Shoaib data sets

approach	TP	FN	TN	FP	P_d	P_n
our approach	10	0	24	4	1.00	0.166
approach 2	8	2	23	5	0.80	0.217
approach 3	4	6	26	2	0.40	0.077
approach 4	10	0	14	14	1.00	0.500

5 Conclusion

We have presented a novel shadow-assistant human fall detection system which can support different viewpoints. The robust human fall detection scheme relies on shadow and shape analysis to differentiate fall-down and fall-like incidents under different viewpoints. Our experiment results demonstrate that the proposed system can achieve a high detection rate and low false alarm rate while satisfying real-time constraints.

References

1. Horprasert, T., Harwood, D., Davis, L.: A Statistical Approach for Real-Time Robust Background Subtraction and Shadow Detection. In: IEEE International Conference on Computer Vision, ICCV 1999, Frame-Rate Workshop, pp. 1–19. IEEE Press, New York (1999)
2. Dalal, N., Triggs, B.: Histograms of Oriented Gradients for Human Detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, pp. 886–893. IEEE Press, New York (2005)
3. Bobick, A., Davis, J.: The Recognition of Human Movement Using Temporal Template. IEEE Transactions on Pattern Analysis and Machine Intelligence 23, 257–267 (2001)
4. Rougier, C., Meunier, J., St-Arnaud, A., Rousseau, J.: Fall Detection from Human Shape and Motion History Using Video Surveillance. In: the 21st International Conference on Advanced Information Networking and Applications, Niagara Falls, Canada, pp. 875–880. IEEE Press, New York (2007)
5. Hsu, Y.T., Liao, H.Y., Chen, M.C.C., Hsieh, J.W.: Video-based Human Movement Analysis and Its Application to Surveillance Systems. IEEE Transactions on Multimedia 10, 372–392 (2008)