Motion Estimation-Based Human Falling Detection for Visual Surveillance

Heegwang Kim, Jinho Park, Hasil Park, and Joonki Paik
Department of Image, Chung-Ang University
84 Heuksoek-ro, Dongjak-gu, Seoul 06974, Korea
E-mail : heegwang27@gmail.com, dkskzmffps@gmail.com, hahaha2470@cau.ac.kr, paikj@cau.ac.kr

Abstract: Detection of a human falling event has attracted increasing attention in a visual surveillance system. This paper presents a novel falling event detection algorithm using motion estimation and an integrated spatiotemporal energy map of the object region. The proposed method first extracts a human region using a background subtraction method. Next, we applied an optical flow algorithm to estimate motion vectors, and the energy map is generated by accumulating the detected human region for a certain period of time. We can then detect the falling event using the k-nearest neighbor (kNN) classification with the previously estimated motion information and energy map.

1. Introduction

Automatic detection of an abnormal event is an important function in an intelligent visual surveillance system. Among various human abnormal behaviors, a falling event is particularly important since it causes a serious injury or death. Many falling event detection methods have been proposed in the literature. Anderson *et al.* proposed a hidden Markov model-based falling event detection algorithm [1]. However, Anderson's algorithm could not detect the falling event when the falling direction is parallel to the camera's optical axis. To overcome this problem, Yu *et al.* proposed a multiple camera-based falling event detection method that represents a human object as a set of three-dimensional (3D) voxels [2]. However, Yu's method cannot distinguish a normal lying object from an abnormal falling object since it does not consider spatiotemporal features.

In order to solve the above-mentioned problems, the proposed method relies incorporated motion information of the object-of-interest to generate the energy map by accumulating the motion compensated foreground object region for a certain period of time. The proposed algorithm consists of four sequential steps: i) extraction of the foreground object region using a background subtraction method, ii) motion estimation of the object region using an optical flow method, iii) generation of the energy map by accumulating the motion-compensated foreground object region, and iv) detection of the fall event using the k-nearest neighbor (kNN) classification.

2. Foreground Object Region Detection

In order to detect the object region, we use the background subtraction method after adaptively generating the background image. If the brightness difference between the previous and current frames is greater than a pre-specified threshold, the background image is updated as

$$f_B^t(x,y) = (1-\beta)I_t(x,y) + \beta f_B^{t-1}(x,y). \quad (1)$$

where $f_B^t(x,y)$ and $f_B^{t-1}(x,y)$ respectively represent the background image at time t and $t-1$, and β is the mixing ratio in the range $[0,1]$. The object region in the $t-th$ frame can be detected by using subtraction between the obtained background and input images as

$$f_o(x,y) = \begin{cases} 1, & \sum_{c\in\{R,G,B\}} |f_B^c(x,y) - f^c(x,y)| \\ 0, & otherwise \end{cases}\Bigg\}, \quad (2)$$

where $f_o(x,y)$ represents the binary image of the $t-th$ frame and has the value 1 only in the object region. A morphology filtering is then performed to reduce the noise amplification.

3. Motion Estimation

When a human falling down, the direction of motion vectors in the object region have the downward direction. So, we estimate the motion vector using combined local-global approach with total variation (CLG-TV) [3], where the motion vector is determined by minimizing the following energy function

$$E_{CLG-TV} = \int_\Omega [\lambda(\sum_P wr(u,v)^2) + |\nabla u| + |\nabla v|]. \quad (3)$$

where u and v respectively represent displacements in the x-axis and y-axis directions, and P is the image patch. $r(u,v)$ is the residual between the previous and the current frames that is defined as

$$r(u,v) = (f^t - f^{t-1} + f_x^t u + f_y^t v). \quad (4)$$

Figure 1 shows the result of optical flow estimation using the CLG-TV method when the human falls in the direction parallel to the camera's optical axis.

(a) (b)

Figure 1. Result of optical flow estimation: (a) input frame and (b) estimated motion vectors

The estimated motion vectors are quantized into one of 9 different directions as shown in Figure 2, and then we compute histograms of nine directions.

4. Integrated Spatio-temporal Energy Map

When a human is walking toward the camera, a false detection of the falling event may occur. To solve this problem, we additionally use an energy map of the human region. The detected human region using the background subtraction is represented by a bounding box. The detected human regions for a

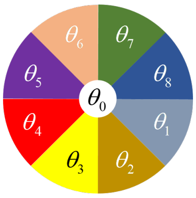
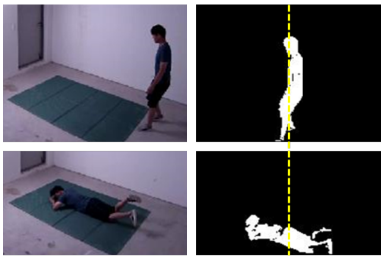


Figure 2. Quantization of motion vectors

period of time are aligned to the centroids of the x-coordinate of a set of bounding boxes and bottom of the bounding boxes. The energy map of the human region is generated by accumulating the foreground human regions. Next, we generate the histogram of the horizontal axis of the generated energy map. Figure 3 shows results of the aligned the human region in each frame.



(a) (b)

Figure 3. Aligning the object region; (a) input frame and (b) aligned object region to the center of bounding box

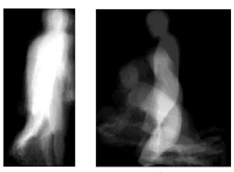
Figure 4 shows the energy map by accumulating for a certain period of time. Figure 4(a) shows the energy map of a walking human, and Figure 4(b) shows that of a falling human.

5. Classification

The histogram of quantized motion vectors and the energy map are used as a feature to detect the falling event. When an object falls down, $\theta_1, \theta_2, \theta_3, \theta_4$ of the entire motion histogram becomes larger, and the corresponding energy map is shown in Figure 4(b). On the other hand, when a human is normally walking or running, the corresponding energy map is shown in Figure 4(a). In this work, we use the kNN classifier to determine boundaries in the feature containing the falling and normal actions. To generate clusters, we take a video containing a human falling down in the direction of angle 45 degrees from the camera's optical axis and training with kNN.

6. Experimental Results

The experiments were performed in a 8 m x 7 m x 3 m room, and a surveillance camera is located at 2.5 m from the ground. We assume that there are two actions: falling down and normal moving. Figure 5 shows the experiment results.



(a) (b)

Figure 4. Energy maps of different behavior: (a) a normally walking human and (b) a falling human.

(a) (b) (c)

Figure 5. Detection of a falling event: (a) walking, (b) falling in the right angle of the optical axis, and (c) falling parallel to the optical axis.

Figure 5(a) shows a normal moving, Figures 5(b) and (c) respectively show the human falling down in the right angle and parallel to the camera's optical axis. The proposed method can successfully detect the falling event in the any directions.

7. Conclusion

In this paper, we presented a novel human falling detection algorithm using motion vectors and the energy map. The proposed method estimates motion vectors when a human is moving, and generates the energy map by accumulating the foreground human region for certain number of frames. Experimental results show that the proposed method can successfully detects the falling event in any direction. This method can be used in an intelligent surveillance system.

8. Acknowledgments

This work was supported in part by the Technology Innovation Program (Development of Smart Video/Audio Surveillance SoC & Core Component for Onsite Decision Security System) under Grant 10047788, and by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government(MSIP) (B0101-15-0525, Development of global multi-target tracking and event prediction techniques based on real-time large-scale video analysis), and by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2016-H8501-16-1018) supervised by the IITP(Institute for Information & communications Technology Promotion).

References

[1] D. Anderson, J. M. Keller, M. Skubic, X. Chen, and Z. He, "Recognizing falls from silhouettes," in *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*. IEEE, 2006, pp. 6388–6391.

[2] M.-C. Yu, S. M. Naqvi, A. Rhuma, and J. Chambers, "One class boundary method classifiers for application in a video-based fall detection system," *Computer Vision, IET*, vol. 6, no. 2, pp. 90–100, 2012.

[3] M. Drulea and S. Nedevschi, "Total variation regularization of local-global optical flow," in *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*. IEEE, 2011, pp. 318–323.