*IEIE Transactions on Smart Processing and Computing, vol. 5, no. 5, October 2016*
http://dx.doi.org/10.5573/IEIESPC.2016.5.5.327


IEIE Transactions on Smart Processing and Computing

# Motion Estimation–based Human Fall Detection for Visual Surveillance

**Heegwang Kim, Jinho Park, Hasil Park, and Joonki Paik***

Graduate School of Advanced Image Science, Multimedia and Film, Chung-Ang University / Seoul 06974, Korea
{heegwang27, dkskzmffps}@gmail.com, {hahaha2470, paikj}@cau.ac.kr

* Corresponding Author: Joonki Paik

*Received October 10, 2016; Accepted October 27, 2016; Published October 30, 2016*

* *Short Paper*

***Abstract***: Currently, the world's elderly population continues to grow at a dramatic rate. As the number of senior citizens increases, detection of someone falling has attracted increasing attention for visual surveillance systems. This paper presents a novel fall-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 an energy map is generated by accumulating the detected human region for a certain period of time. We can then detect a fall using k-nearest neighbor (kNN) classification with the previously estimated motion information and energy map. The experimental results show that the proposed algorithm can effectively detect someone falling in any direction, including at an angle parallel to the camera's optical axis.

***Keywords***: Fall detection, Slip detection, Motion estimation, Integrated spatiotemporal energy map

## 1. Introduction

Globally, the number of persons aged 60 years or more is projected to more than double by 2050, from 901 million in 2015 to 2.1 billion [1]. For that reason, surveillance systems for the elderly have attracted significant attention. Among various human abnormal behaviors, falling down is particularly important since a fall can cause serious injury or death. Falls are the second leading cause of accidental or unintentional injury-related deaths worldwide, and adults older than 65 suffer the greatest number of fatal falls [2]. Therefore, automatic fall detection is an important function in an intelligent visual surveillance system.

Many fall-detection methods have been proposed in the literature. Rougier *et al.* used a combination of motion history and human shape variation [3]. Lin *et al.* used a mixture of Gaussians and applied a motion history image (MHI) to analyze falls [4]. Anderson *et al.* proposed a hidden Markov model–based fall-detection algorithm [5]. However, Anderson's algorithm could not detect a fall when the direction is parallel to the camera's optical axis. To overcome this problem, multi-view–based fall detection algorithms have been proposed. Thome *et al.* classified the silhouette between standing and prone positions, and used a layered hidden Markov model to recognize the poses [6]. Auvinet *et al.* employed a multiple-camera network to reconstruct three-dimensional (3-D) shapes of human objects. Their system detects a fall by analyzing volume distribution along the vertical axis [7]. Yu *et al.* proposed a multiple camera–based fall-detection method that represents a person as a set of 3-D voxels, and classifies the object's behaviors into standing, sitting, and falling [8]. However, these methods cannot distinguish a normal object lying down from an abnormal object falling down, since they do not consider spatiotemporal features.

In order to solve the above-mentioned problems, the proposed method relies on incorporating motion information about the object of interest to generate an energy map by accumulating the motion-compensated foreground object region over a certain period of time. The proposed algorithm consists of four sequential steps: (1) extraction of the foreground object region using a background subtraction method, (2) motion estimation of the object region using an optical flow method, (3) generation of an energy map by accumulating the motion-compensated foreground object region, and (4) detection of a fall using k-nearest neighbor (kNN) classification.

## 2. Foreground Object Region Detection

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

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

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 $\beta$ 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 follows:

$$f_O(x,y) = \begin{cases} 1, & \sum_{c\in\{R,G,B\}} \left| f_B^c(x,y) - f^c(x,y) \right| \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

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

## 3. Motion Estimation

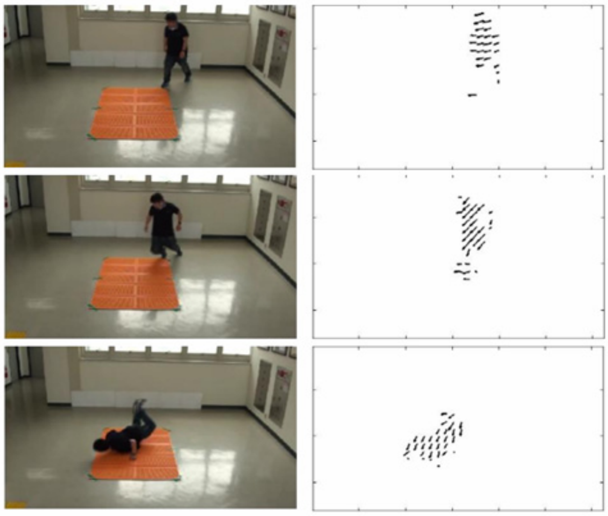
When a human is falling down, the direction of motion vectors in the object region have a downward direction. So, we estimate the motion vector using a combined local–global approach with total variation (CLG-TV) [9], where the motion vector is determined by minimizing the following energy function:

$$E_{CLG-TV} = \int_{\Omega} \left[ \lambda \left( \sum_{P} wr(u,v)^2 \right) + |\nabla u| + |\nabla v| \right] \tag{3}$$

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, defined as

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

Fig. 1 shows the results of optical flow estimation using the CLG-TV method when a human falls in a direction parallel to the camera's optical axis. The estimated motion vectors are quantized into one of nine different directions, as shown in Fig. 2, and then we compute histograms of the nine directions.



(a) (b)

**Fig. 1. Results of optical flow estimation (a) input frame, (b) estimated motion vectors.**


**Fig. 2. Quantization of motion vectors.**

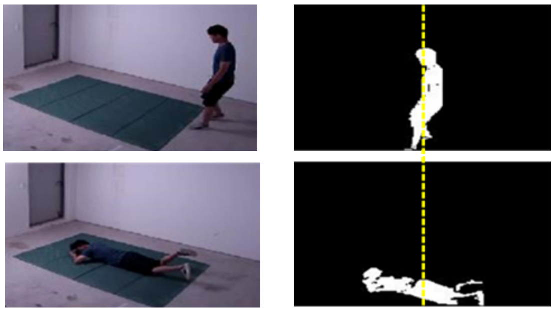
## 4. Integrated Spatiotemporal Energy Map

When a human is walking toward the camera, false detection of a fall may occur. To solve this problem, we additionally use an energy map of the human region. According to Cham and Redfern [10], falls are generally associated with peak slip velocity greater than or equal to 0.8 m/s. For that reason, we assume the time of a fall for an adult of average height at 2 s. An energy map is accumulated from 60 frames in a 30 fps video. The detected human region using background subtraction is represented by a bounding box. The detected human region for a period of time is aligned to the centroid of the x-coordinate of a set of bounding boxes and the bottom of the bounding boxes. The energy map of the human region is generated by accumulating the foreground human region.

Next, we generate a histogram of the horizontal axis of the generated energy map. Fig. 3 shows the results of the aligned human region in each frame. Fig. 4 shows the energy map from accumulating 60 frames. Fig. 4(a) shows the energy map of a walking human, and Fig. 4(b) shows that of a falling human.

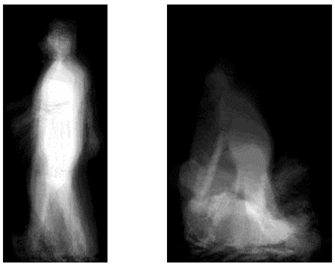
## 5. Classification

For kNN classification, the histogram of quantized motion vectors and the energy map are used as features to



(a) (b)

**Fig. 3. Aligning the object region (a) input frame, (b) object region aligned to the center of the bounding box.**



(a) (b)

**Fig. 4. Energy maps from different behaviors (a) a human walking normally, (b) a human falling down.**

detect the fall. When an object falls, $\theta_1$, $\theta_2$, $\theta_3$, $\theta_4$ of the entire motion histogram becomes larger, and the corresponding energy map is shown in Fig. 4(b). On the other hand, when a human is walking normally or running, the corresponding energy map is shown in Fig. 4(a). Fig. 5 shows the generation process of feature data for a kNN classifier. As a human falls, the integrated energy map of the object region has large values in the bottom region, and the energy maps are generated in a slightly different form, depending on the direction of the fall. Then, we find the maximum values in each row of the energy map and extract these values. The quantized motion vector and the feature vector of the energy map are combined into one set of feature data. In this work, we use the kNN classifier to determine boundaries in the feature containing falling and normal actions. To generate clusters, we take a video containing a human falling down at a 45-degree angle from the camera's optical axis, and train with kNN.

## 6. Experimental Results

The experiments were performed in an 8 m x 7 m x 3 m room, and a surveillance camera was located 2.5 m from the ground. We assumed there are two actions: falling down and normal movement. Figs. 6(a) and (d) show normal movement; Figs. 6(b), (e), (f), (g) and (h) show a human falling down in a direction non-parallel to the camera's optical axis. Figs. 6(c) and (i) show a human falling down parallel to the camera's optical axis. The proposed method can successfully detect a fall in any direction.

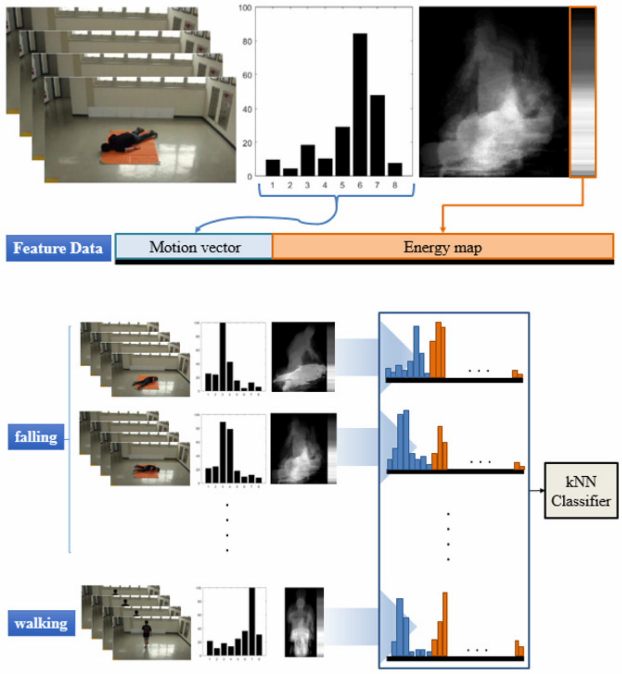
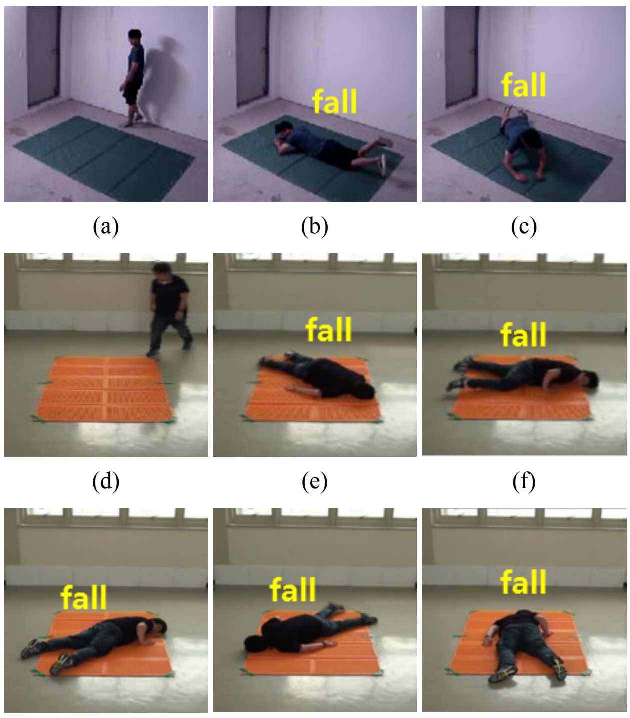



**Fig. 5. Generation of feature data and the kNN classifier.**




(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

**Fig. 6. Fall detection results (a, d) walking, (b, e, f, g, h) falling at a non-parallel angle to the camera's optical axis, (c, i) falling parallel to the optical axis.**

## 7. Conclusion

In this paper, we presented a novel human fall-detection algorithm using motion vectors and an energy map. The proposed method estimates motion vectors when a human is moving, and generates the energy map by accumulating the foreground human region over 60 frames via statistical analysis. The motion vectors are quantized into one of nine different directions. The quantized motion vectors and energy map use the feature data for kNN classification. Experimental results show that the proposed method can successfully detect a fall occurring in any direction. This method can be used in an intelligent surveillance and smart home system.

## Acknowledgement

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) (B0101-16-0525, Development of global multi-target tracking and event prediction techniques based on real-time large-scale video analysis).

## References

[1] United Nations, Department of Economic and Social Affairs, Population Division (2015). *World Population Ageing 2015* (ST/ESA/SER.A/390).

[2] World Health Organization (2007), *WHO Global Report on Falls Prevention in Older Age.* Geneva: WHO.

[3] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Fall Detection from Human Shape and Motion History Using Video Surveillance," *Advanced Information Networking and Applications Workshops, 2007, AINAW '07. 21st International Conference on*, Niagara Falls, Ont., 2007, pp. 875-880. Article (CrosesRef Link)

[4] C. Y. Lin, S. M. Wang, J. W. Hong, L. W. Kang, and C. L. Huang, "Vision-Based Fall Detection through Shape Features," *2016 IEEE Second International Conference on Multimedia Big Data (BigMM)*, Taipei, 2016, pp. 237-240. Article (CrossRef Link)

[5] 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. Article (CrossRef Link)

[6] N. Thome, S. Miguet, and S. Ambellouis, "A Real-Time, Multiview Fall Detection System: A LHHM-Based Approach," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1522-1532, Nov. 2008. Article (CrossRef Link)

[7] E. Auvinet, F. Multon, A. Saint-Arnaud, J. Rousseau, and J. Meunier, "Fall Detection With Multiple Cameras: An Occlusion-Resistant Method Based on 3-D Silhouette Vertical Distribution," in *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 2, pp. 290-300, March 2011. Article (CrossRef Link)

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

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

[10] R. Cham and M. S. Redfern, "Heel contact dynamics during slip events on level and inclined surfaces," *Safety Science*, vol. 40, pp. 559-576, 2002. Article (CrossRef Link)