

# Detecting Fall Incidents of The Elderly based on Human-Ground Contact Areas

Dao Huu Hung and Hideo Saito

Department of Information and Computer Science  
Keio University  
Yokohama, Japan  
{hungdaohuu,saito}@hvrl.ics.keio.ac.jp

Gee-Sern Hsu

Department of Mechanical Engineering  
National Taiwan University of Science and Technology  
Taipei, Taiwan  
jison@mail.ntust.edu.tw

**Abstract**—People always make a little contact with the ground during usual activities mainly by feet but often lie completely on the ground after accidental falls. Thus, we propose using Human-Ground Contact Areas (HGCA) to classify human states into standing, sitting and lying states. A fall is defined by a fast change of human states from standing or sitting to lying and continuity in lying state for a sufficient duration. Temporal analyzing human-state transitions is used to discriminate falls from usual events. To measure HGCA, we project foreground of monitored person from one view to another by using homography of the ground between them. Overlap regions between projected foreground and foreground in the latter view that only exist in which people are in contact with the ground, due to plane parallax, are measured as HGCA. We generalize a threshold of HGCA to separate lying states from the others from view-invariant distributions of HGCA with respect to human states. We propose using human state simulation in which camera viewpoints are freely changed to capture 3D human models in various states. Hundreds of images are generated from the simulation as training data to build these distributions. We test our method on "multiple camera fall dataset" leading to a competitive performance with other methods tested on the same dataset.

**Keywords**—Fall Detection, The Elderly, Human-Ground Contact Areas

## I. INTRODUCTION

This paper studies a method of detecting fall incidents of the elderly from surveillance cameras. It is critical to help the elderly reach instant treatment after falling for not worsening injury. The healthcare industry has realized the useful application of fall detection in intelligent personal emergency response systems [1]. Falls are sensed for automatically making an instant contact with emergency centers. Therefore, fall detection becomes an emerging research topic.

In general, falls are detected when people change from upright posture to almost lengthened one in a fast pace and subsequently remain relatively immobile in the latter posture for a while due to the shock impact of the fall [2], [3]. It leads to our proposal of using Human-Ground Contact Areas (HGCA) to discern human states and a framework of temporal analyzing human-state transition for fall inference.

Our contributions are three-fold. (1) We propose HGCA, a good feature to discern Lying states from other states,

i.e. Standing, Sitting and Kneeling since people make a little contact with the ground during usual activities but often lie completely on the ground after falls. HGCA is effectively estimated by using homography of the ground between views, facilitating the real-time performance. (2) We present a human state simulation in Google Sketchup environment. A virtual camera is positioned on the surface of a hemisphere to capture images of 3D human models in various states and poses. We generate hundreds of samples under a variety of viewpoints as training data to build distributions of HGCA with respect to human states. Thus, the classification of human states according to HGCA based on these distributions is invariant to camera viewpoints. (3) We present a framework of temporal analyzing human-state transition for reliable fall inference based on above fall definition.

In experiments, we choose "Multiple Camera Fall Dataset" [4], a common benchmark dataset in fall detection. It leads to the fair comparison with state-of-the-art methods [5], [6], [7], tested on the same dataset. Our results are competitive.

We continue the paper with a summary of vision-based fall detection methods, in section II. The proposed method is described in section III. Section IV is dedicated to performance evaluation, comparison and discussions. Finally, section V concludes the paper and delineates future works.

## II. RELATED WORKS

This section only focuses on vision-based techniques for fall detection. Please refer to [2], [3] for interests in wearable device and ambient device-based techniques.

One of the most popular class of approaches is based on the analysis of silhouettes in a single view. Early works [1], [8] extract speed of silhouettes' centroids from top-view cameras and compare it with special thresholds in furniture and non-furniture areas (manually labeled) for detecting falls. Top views are inappropriate for fall detection since vital features of the vertical body motion to recognize a fall are not available. Aspect ratios [9], [10], and its combination with 2D speed [11] are considered with cameras placed either sideways or obliquely. They are prone to false alarms with a fall in parallel to the optical axis and the occlusion

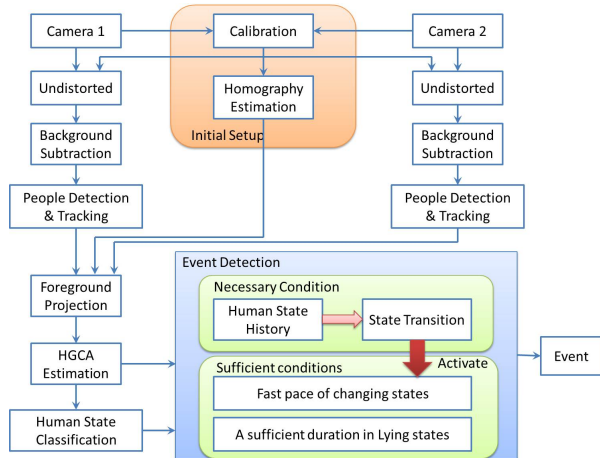


Figure 1: The flowchart of our proposed method

of furniture due to relying on simple features. Several methods quantify the body motion by Motion History Image [12], integrated spatiotemporal energy map [13], and the distance map between two sampling of human skeletons [14] since large motion is likely caused by a fall. Dimensional variations of silhouettes are analyzed upon a large motion to verify falls. Distinction of human shape matching costs during usual activities and falls is presented in [7] at the expense of high computation. Cucchiara *et al.*, [15] train probabilistic projection maps for various postures i.e., standing, crouching, sitting, and lying. Fall events are obtained from a state transition graph.

Multiview approaches are adopted in several works to exploit 3D cues and decision fusion for accuracy enhancement. Cucchiara *et al.*, [16] extend [15] to cope with multiple rooms. Camera handoff is treated by warping human appearance between views based on homography. In [17], 3D angle between vertical line and principal axis of eclipse human models in metric-rectified images is extracted from different views for posture classification. Layer HMM is hand designed to make event inference. A two-level hierarchy of fuzzy logic to infer linguistic summarizations of voxel person's states, reconstructed from two cameras, is introduced for performing activity inference [18]. 3D silhouettes [5] are reconstructed from multiviews since its vertical distributions are different for standing and lying people. Hung and Saito [6] approximate people by a rectangular cuboid by two cameras whose fields of view are relatively orthogonal. A feature space composing of height and bottom area of cuboid is studied to classify human states into standing, sitting and lying for event detection.

We realize that [5], [6], [18] based on 3D human reconstruction have good performance. However, it seems to be not necessary to reconstruct 3D human models. A good estimation of HGCA is also able to discriminate falls from

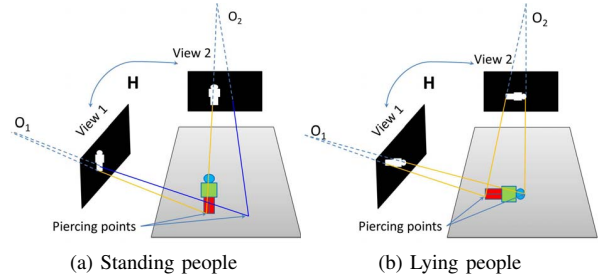


Figure 2: The illustration of using planar homography for fall detection

usual activities with a high accuracy. Hence in this paper, we propose a low-cost and effective scheme to estimate HGCA by using a planar homography constraint and a framework of temporal analyzing human states for reliable fall inference.

### III. PROPOSED METHOD

In multiple-view geometry [19], any two images of a same planar surface (assuming a pinhole camera model) are related by a planar projective transformation, so-called homography. This geometrical relation is successfully applied to tracking people in multiple views [20]. In this paper, we present a low-cost and effective scheme of estimating HGCA based on this geometrical relation. Fig. 1 shows the flowchart of our proposed method.

#### A. Flowchart

Image must be undistorted since the homography is only held under a pinhole camera model. The homography matrix  $H$  of the ground between two views is estimated by simply specifying four point correspondences [19]. The initial setup of our proposed method can be done offline.

Foregrounds are segmented and enhanced by morphological operators before labeled by connected component algorithm, resulting in foreground blobs. These blobs are clustered to form foregrounds of people which are mapped between views by homography of the ground to measure HGCA for event detection.

#### B. Projecting foregrounds by using planar homography

Let  $p$  be a 3D point on a plane  $\Pi$  (i.e. the ground), the projections  $p_1 = (x_1, y_1, 1)$  and  $p_2 = (x_2, y_2, 1)$  of  $p$  (in homogeneous coordinates) on two image planes are related by a homography  $H$  of the plane  $\Pi$  between two views [19].

$$p_1 = Hp_2, \quad (1)$$

where  $H$  is a nonsingular  $3 \times 3$  matrix. Let  $H_3$  denote the third row of  $H$ . The point  $p_1$  in the first image is mapped to  $p_m$  in the second one by the homography  $H$  [20].

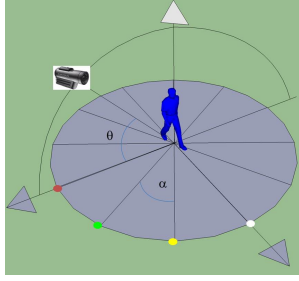


Figure 3: Simulation setup in Google Sketchup. Colorful dots are landmarks for homography calibration

$$p_m = (x_m, y_m, 1) = \frac{H p_1}{H_3 p_1}. \quad (2)$$

If point  $p$  is on the plane  $\Pi$ ,  $p_m$  and  $p_2$  are coincident. But if point  $p$  is not on the plane  $\Pi$ , there exists a misalignment between  $p_m$  and  $p_2$ , reflecting the plane parallax. These observations is elaborated in Fig. 2a. If people are standing or sitting, making contact with the ground by feet, overlap regions between foreground in the second view and projected foreground happen at feet location. But when people lie on the ground, overlap regions cover almost whole body (see Fig. 2).

### C. Estimation of HGCA

Let  $\Psi_1$  and  $\Psi_2$  be sets of foreground in the first and second views, respectively. Let  $\Psi_m$  denote set of mapped foregrounds by homography  $H$  from the first to the second views and let  $X$  be human states, i.e., standing, sitting, kneeling, and lying, etc. We suppose  $\Psi_1 = \Psi_m = \{n\}$  pixels, and  $\Psi_m \cap \Psi_2 = \{m \mid m \leq n\}$ . HGCA is a function of  $X$  and is evaluated by

$$HGCA(X) = \frac{\Psi_m \cap \Psi_2}{\Psi_1} = \frac{m}{n}, \quad (3)$$

### D. Classification of human states

To generalize a threshold of HGCA to separate lying states from the others, training data is needed to build view-invariant distributions of HGCA with respect to human states. This section proposes a human state simulation by using Google Sketchup, as shown in Fig. 3, to generate training images. We consider three typical sitting poses, i.e. sitting on a chair and kneeling on the ground by one or two legs, and three lying poses, corresponding to three typical falls, i.e. falling forward, backward and sideways. A virtual camera is positioned and freely changed on surface of a hemisphere to capture images of 3D human models in various states and poses. The camera viewpoint is modeled in spherical coordinate system by

$$\begin{aligned} P_{camera} &= P(r, \alpha, \theta) \\ r &\approx const \\ \alpha &= [0, 180^\circ], \delta\alpha = 30^\circ \\ \theta &= [30^\circ, 75^\circ], \delta\theta = 15^\circ \end{aligned} \quad (4)$$

where,  $r$  the radial distance,  $\alpha$  the azimuth angle, and  $\theta$  the inclination angle. The angles are measured in degrees. The inclination angle greater than  $75^\circ$  is not taken into account since the camera viewpoints are near the top view which is not appropriate for detecting falls. In addition, indoor surveillance cameras are often positioned obliquely near the ceiling. The spatial constraints make variations of inclination angles in the range of  $[30, 75]$ .

In simulation, both azimuth and inclination angles are changed by steps  $\delta\alpha$  and  $\delta\theta$  in Eq. 4 to generate 196 training images. Fig. 4 illustrates some generated images. Homography of the ground between these views are automatically calibrated by matching colorful dots on the ground. We project foregrounds between every pair of different views  $(\alpha_1, \theta_1)$  and  $(\alpha_2, \theta_2)$  in the training set with  $\Delta\alpha = |\alpha_1 - \alpha_2| > 0$  or  $\Delta\theta = |\theta_1 - \theta_2| > 0$  to estimate HGCA for building distributions. Our aim is to generalize a threshold of HGCA to separate lying states from the others regardless viewpoint variations. It is noted that human models in simulation are stationary and cameras are moving. Both  $\Delta\alpha$  and  $\Delta\theta$  are determined by positions of cameras. But in reality, cameras are fixed and people are moving. When people travel in the fields of views of a fixed-camera pair, both azimuth and inclination angles  $(\Delta\alpha, \Delta\theta)$  vary freely which are determined by positions of both people and cameras, not by the positions of cameras.

We weight all viewpoints equally, that is, all  $(\Delta\alpha, \Delta\theta)$ <sup>1</sup> are allowed to happen with a uniform probability. The overall histograms of HGCA with respect to human states are built and shown in Fig. 5a, together with distribution fits. Threshold of HGCA to separate lying states from the other ones is determined at the intersection of distributions of HGCA with respect to lying states and sitting states.

We realize from the simulation that measuring HGCA by using Eq. 3 is usually inaccurate for a pair of viewpoints with  $\Delta\alpha < 90^\circ$  as shown in Fig. 4 since overlapping regions usually cover body parts which are not in contact with the ground. This phenomenon does not happen for a pair of viewpoints with  $\Delta\alpha \geq 90^\circ$ . Thus, using a pair of cameras positioned at least  $90^\circ$  apart (in terms of azimuth angles) will lead to fewer situations occurring  $\Delta\alpha < 90^\circ$  when people travel in their fields of views. The constraints of indoor spaces and camera placement make some viewpoints frequently happen and some rarely happen. Thus, viewpoints can be weighted unequally to adapt to specific contexts. Fig.

<sup>1</sup>One of them must be non-zero

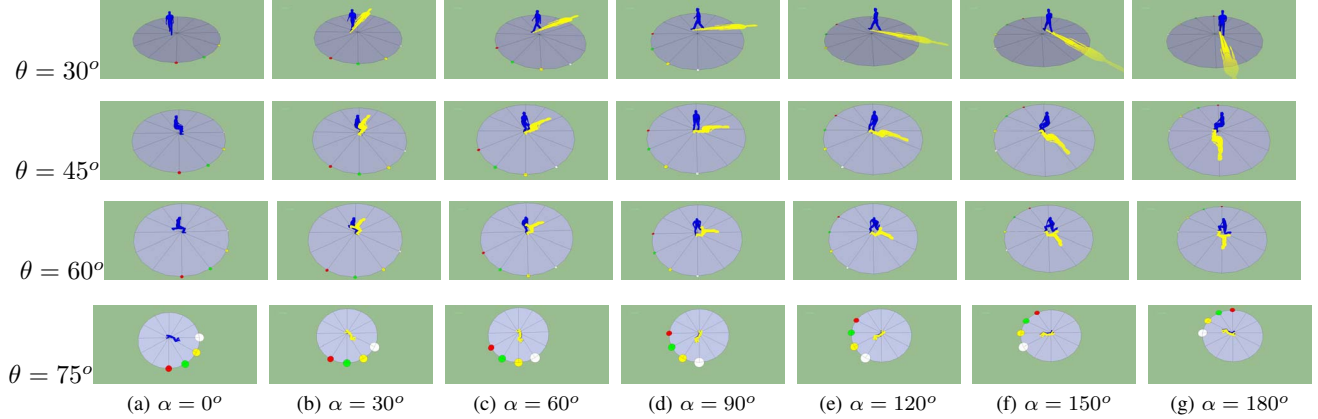
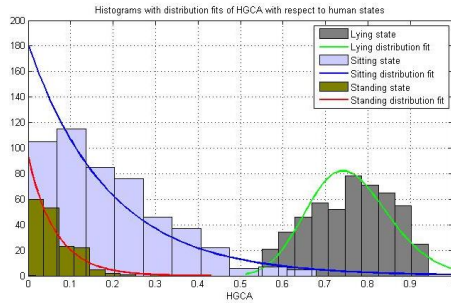
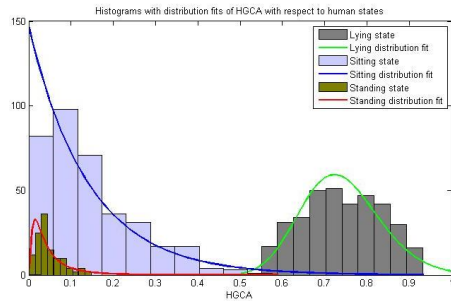


Figure 4: Some generated image samples from the simulation. We show standing, sitting, kneeling and lying people in rows. Foregrounds in first-column images are projected and overlaid by yellow in images in other columns



(a) All viewpoints



(b) All viewpoints except  $\Delta\alpha = 30^\circ$

Figure 5: Histograms and distribution fits of HGCA with respect to human states. Exponential and normal distributions fits are for standing, sitting and lying states, respectively.

5b show distributions of HGCA excluding viewpoints of  $\Delta\alpha = 30^\circ$ . However, these distributions in Fig. 5a and 5b are quite similar. In this paper, we use the distributions in Fig. 5a in experiments for various camera settings.

### E. Framework for Fall Inference

The fall definition in Section I leads to a typical fall characteristics based on HGCA, as shown in Fig. 6. When a fall happens at  $T_{fall}$ , HGCA increases by  $\Delta HGCA$  to make a state transition from Usual states composing of Standing, Sitting, and Kneeling to Lying states in a so-called falling period of  $\Delta T_1$ . Subsequently, the fallen person is relatively immobile in a period of  $\Delta T_2$ . In this section, we present a framework for fall inference based on the fall definition that can be broken into *necessary* and *sufficient conditions*.

*Necessary condition* is the change of human postures from upright to lengthened. In our framework, the change of human postures is described by a state transition from Usual states to Lying states. Such state transitions are caused not only by fall events but also by lying-down events (people lying on a sofa or a bed). Fall events are associated with a fast movement of human body. In contrast, lying-down events are performed in a leisure manner by the elderly. Therefore to claim a state transition as a fall event, we must verify the following *sufficient conditions*.

*Sufficient conditions* compose of a fast pace of changing states and an observation of Lying states in a sufficient duration after the state transition. Without satisfying both sufficient conditions, the state transition is definitely not caused by a fall event. The proposed framework for fall inference is described in the module of Event Detection in Fig. 1.

To this end, we keep both human states  $X$  and HGCA in the buffer of  $N$  frames. Upon a state transition at  $T_{trans}$ , we start verifying both sufficient conditions to whether claim a fall event. Fast pace of changing states is characterized by  $\Delta T_1$  and  $\Delta HGCA$  which are evaluated by

$$\begin{aligned} \Delta HGCA &= HGCA[T_{trans}] - HGCA[T_{fall}] \\ \Delta T_1 &= T_{trans} - T_{fall} \end{aligned} \quad (5)$$

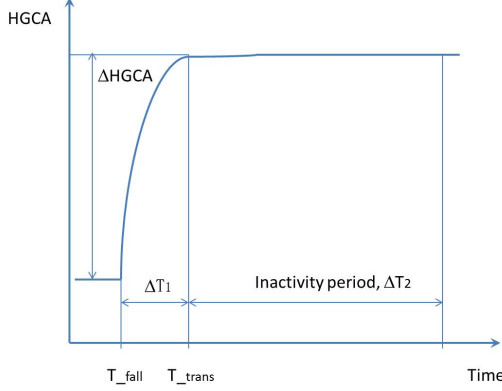


Figure 6: Typical fall characteristics based on HGCA

where  $\Delta HGCA$  the increment of HGCA in falling period. From distributions in Fig. 5,  $\Delta HGCA$  should be at least 0.3 to prevent from trivial state transitions, likely caused by noise.

$$\Delta HGCA \geq 0.3 = Min_{\Delta HGCA} \quad (6)$$

Combine Eqs. 5 and 6,

$$HGCA[T_{fall}] \leq HGCA[T_{trans}] - Min_{\Delta HGCA} \quad (7)$$

$T_{fall}$  and  $\Delta T_1$  are calculated by using Eqs. 5 and 7. A state transition is considered as fast if  $\Delta T_1 \leq 1$  second [2], [3]. Finally, extracting  $\Delta T_2$  is very straightforward to verify the second sufficient condition. In practice, 5 seconds are considered to be long enough for experiment conditions.

#### IV. PERFORMANCE EVALUATION AND DISCUSSION

##### A. Performance evaluation and comparison

We adopt the common metric used in state-of-the-art methods [5], [6], [7] for a fair performance comparison by computing *sensitivity* and *specificity*,

$$SE = \frac{TP}{TP + FN}$$

$$SP = \frac{TN}{TN + FP}, \quad (8)$$

where  $SE$  the sensitivity,  $SP$  the specificity,  $TP$  True Positive,  $FN$  False Negative,  $TN$  True Negative,  $FP$  False Positive, In each video sequence, the actual time occurring a fall (denoted  $t_{fall}$ ) is manually annotated. This time is defined by the first moment of human body hitting the ground after a fall. A fall is detected after or before  $t_{fall}$ , resulting in TP or FP, respectively. A fall is not detected after or before  $t_{fall}$ , resulting in FN or TN, respectively.

We test our method with two pairs of cameras, i.e. cameras 2 and 7 (positioned  $180^\circ$  apart), cameras 2 and

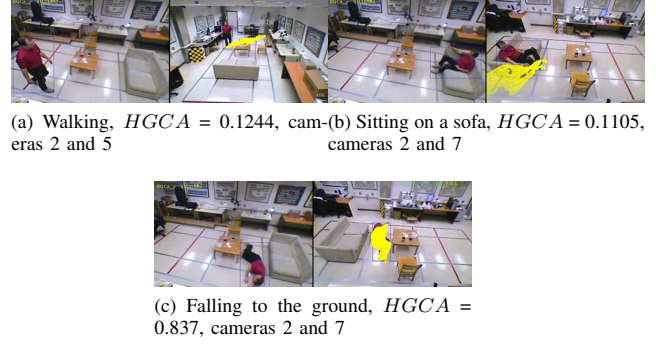


Figure 7: Results of our method in scene 15. The yellow dots are projected foregrounds, overlaid in the right images

5 (positioned  $90^\circ$  apart) and with three of them. Performance of three experiments are compared with that of other methods, tested on the same dataset, in Table I. Although results in [6] seem to be better than ours, it does not consider a fall from sitting to lying states that frequently happens with the elderly. Our performance is competitive with low computational cost. Average frame rate of our non-optimized OpenCV and C code is about 22 fps (2 cameras), running on a common PC with chipset Intel core i7 3820QM, in comparison with approximate 15 fps in [6] (2 cameras), 5 fps in [7] (one camera), and 0.9 and 15 fps in [5] by CPU and GPU implementation (three cameras), respectively. However, the comparison is not quite fair since they are implemented on different hardware. In next section, we discuss several issues, which may affect the performance of our method.

##### B. Discussions

1) *Lighting Conditions*: We believe that the lighting conditions in the dataset are typical in indoor surveillance. Shadow and reflection happen frequently but not severely. They may cause errors in estimating HGCA of standing and sitting states. Although the performance of our method without using shadow removal algorithms is not sensitive by lighting variations, we will continue to investigate further the influence of severe lighting variations as well as integrate shadow removal algorithms to make the method robust to a variety of lighting conditions.

2) *Furniture Occlusion*: Occlusion by furniture frequently happens in indoor surveillance that make HGCA estimation inaccurate, degrading performance. Our method fails to detect the fall events in scenes 15 and 22 by a pair of cameras 2 and 5 but correctly detects them by a pair of cameras 2 and 7. The reason is that the sofa covers a part of human body when he lies on the ground after falling. Similarly, our method fails to detect the fall event in scene 19 by a pair of cameras 2 and 7 but correctly detect it by



Table I: Performance comparison between our method and three state-of-the-art methods [5], [6], [7], tested on the same dataset. Results in [5] are with 3 cameras.

	Sensitivity	Specificity
Our method (cameras 2, 5 and 7)	95.8 %	96 %
Our method (cameras 2 and 7)	88 %	100 %
Our method (cameras 2 and 5)	88 %	96 %
Hung and Saito, 2012 [6]	95.8 %	100 %
Rougier <i>et al.</i> , 2011 [7]	95.4 %	95.8 %
Auvinet <i>et al.</i> , 2011 [5]	80.6 %	100 %

a pair of cameras 2 and 5 due to the same reason. Thus, we combine 3 these cameras to make the method robust to furniture occlusion since we believe that people are occluded in this view but are likely visible in the other ones. Decisions made independently by these camera pairs are fused simply by OR rules to enhance the event detection accuracy as shown in Table I.

## V. CONCLUSIONS AND FUTURE WORKS

We have presented a low-cost scheme to estimate HGCA efficiently for detecting fall incidents of the elderly. Although a good performance is reported, it is the results evaluated merely on the common dataset containing limited challenges in the real world. In the future, we will continue this work by testing it against real falls of the elderly in real home environments, instead of simulated falls in the dataset. We will study more about fusion rules to combine decisions made independently from various pairs of cameras to enhance further the accuracy. Shadow removal algorithms will be integrated to reduce segmentation errors caused by severe lighting variations.

## REFERENCES

- [1] T. Lee and A. Mihailidis, "An intelligent emergency response system: preliminary development and testing of automated fall detection," *J. of Telemedicine and Telecare*, vol. 11, no. 4, pp. 194–198, 2005.
- [2] N. Noury, A. Fleury, P. Rumeau, A. Bourke, G. Laighin, V. Rialle, and J. Lundy, "Fall detection - principles and methods," *EMBS*, pp. 1663–1666, 2007.
- [3] X. Yu, "Approaches and principles of fall detection for elderly and patient," *Int'l Conf. on ENAS*, pp. 42–47, 2008.
- [4] E. Auvinet, C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Multiple cameras fall dataset," *Technical Report 1350, DIRO Université de Montréal*, pp. 1–24, 2010.
- [5] 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," *IEEE Trans. on ITB*, vol. 15, no. 2, pp. 290–300, 2011.
- [6] D. H. Hung and H. Saito, "Estimation of heights and occupied areas of humans from two orthogonal views for fall detection," *IEEJ Trans. on Electro., Info., and Sys.*, vol. 113, pp. 117–127, 2013.
- [7] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Procrustes shape analysis for fall detection," *Int'l Workshop on Visual Surveillance*, 2007.
- [8] H. Nait-Charif and S. McKenna, "Activity summarisation and fall detection in a supportive home environment," *ICPR*, vol. 4, pp. 323–326, 2004.
- [9] D. Anderson, J. Keller, M. Skubic, X. Chen, and Z. He, "Recognizing falls from silhouettes," *EMBS*, pp. 6388–6391, 2006.
- [10] C.-L. Liu, C.-H. Lee, and P.-M. Lin, "A fall detection system using k-nearest neighbor classifier," *ESA*, vol. 37, no. 10, pp. 7174–7181, 2010.
- [11] B. Huang, G. H. Tian, and X. L. Li, "A method for fast fall detection," *World Cong. on ICA*, pp. 3619–3623, 2008.
- [12] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Fall detection from human shape and motion history using video surveillance," *Int'l Conf. AINA*, vol. 2, pp. 875–880, 2007.
- [13] Y. T. Liao, C. L. Huang, and S. C. Hsu, "Slip and fall event detection using bayesian belief network," *PR*, vol. 45, no. 1, pp. 24–32, 2012.
- [14] Y.-T. Chen, Y.-C. Lin, and W.-H. Fang, "A hybrid human fall detection scheme," pp. 3485–3488, 2010.
- [15] R. Cucchiara, C. Grana, A. Prati, and R. Vezzani, "Probabilistic posture classification for human-behavior analysis," *IEEE Trans. on SMC Part A: Sys. and Hum.*, vol. 35, no. 1, pp. 42–54, 2005.
- [16] R. Cucchiara, A. Prati, and R. Vezzani, "A multi-camera vision system for fall detection and alarm generation," *ES*, vol. 24, no. 5, pp. 334–345, 2007.
- [17] N. Thome, S. Miguët, and S. Ambellouis, "A real-time, multiview fall detection system: A lhmm-based approach," *IEEE Trans. on CSVT*, vol. 18, no. 11, pp. 1522–1532, 2008.
- [18] D. Anderson, R. H. Luke, J. M. Keller, M. Skubic, M. Rantz, and M. Aud, "Linguistic summarization of video for fall detection using voxel person and fuzzy logic," *CVIU*, vol. 113, no. 1, pp. 80–89, 2009.
- [19] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.
- [20] S. Khan and M. Shah, "A multiview approach to tracking people in crowded scenes using a planar homography constraint," *ECCV*, vol. 4, pp. 133–146, 2006.