

# Fall Detection Using Body Volume Reconstruction and Vertical Repartition Analysis

Edouard Auvinet<sup>1,2</sup>, Franck Multon<sup>2</sup>, Alain St-Arnaud<sup>3</sup>,  
Jacqueline Rousseau<sup>4</sup>, and Jean Meunier<sup>1</sup>

<sup>1</sup> University of Montreal - IGB - Montreal, QC, Canada

<sup>2</sup> University of Rennes 2 - M2S - Rennes, France

<sup>3</sup> Health and social service center Lucille-Teasdale - Montreal, QC, Canada

<sup>4</sup> CRIUGM, University of Montreal, Montreal, QC, Canada

**Abstract.** With the life expectancy increase, more and more elderly people risk to fall at home. In order to help them living safely at home by reducing the eventuality of unrescued fall, autonomous systems are developed. In this paper, we propose a new method to detect falls at home, based on a multiple cameras network for reconstructing the 3D shape of people. Fall events are detected by analyzing the volume distribution along the vertical axis, and an alarm is triggered when the major part of this distribution is abnormally near the floor which implies that a person has fallen on the floor. This method is evaluated regarding the number of cameras (from 3 to 8) with 22 fall scenarios. Results show 96% of correct detections with 3 cameras and above 99% with 4 cameras and more.

## 1 Introduction

Falls is a major health security problem which increase with the elderly's population growth. However, since a decade, technological solutions are proposed to reduce the problem of unrescued elderlies who fall on the floor at home and could not raise up by themselves. Those systems use different kind of sensors such as accelerometers and camera-system to detect fall events and then call emergency services. Accelerometer systems are easy to setup and their use is widely experimented [4]. But acceleration measures are sometimes insufficient to discriminate falls from others activities of daily life (ADL) and are totally useless in the case where patients forget to wear it, which is not negligible with patient suffering from dementia or Alzheimer's disease. With camera solutions, it is the contrary situation that prevails. Picture sequences contain too much information, and the major problem here is to identify useful part. Once this segmentation operation is done, it remains to find the descriptor which will define a situation as a fall or not. Some previous works proposed mono-camera descriptors using the movement, such as the reconstructed 3D movement of the head [5] or body silhouette changes [6]. Others consist in using body orientation features such as width and height of a silhouette by comparing a standing and a lying person [7][8][9]. However in case of occlusions, which come easily with

home furnitures, those solutions are disabled by hidden parts of the movement or body. This is why some other descriptors use multi-cameras information in order to overcome this problem. This kind of system permits to reconstruct people volume in space using a visual hull reconstruction method [11]. For instance, Anderson et al. [10] use a descriptor based on principal axis of the reconstructed volume and fuzzy logic analysis to classify human activities, such as fall. But a more simple solution could be possible to detect lying down people on the ground.

Our contribution resides in two points. The first one is the use of an adaptation of visual hull reconstruction method to increase robustness to occlusions [12] which does not need that the silhouette of the person be present in all cameras to reconstruct the volume. The second one is the validation of a descriptor based on a simple volume repartition ratio in order to detect a lying-down-on-the-floor person. Using the fact that the volume of a person in this situation does not exceed 40 centimeters, detection is done by looking at the ratio of the volume under this height threshold divided by the total volume. A previous presentation of this descriptor [1] showed excellent results using 8 cameras. In this paper, we evaluate the performance of this descriptor regarding the number of cameras (from 3 to 8) used to reconstruct the volume. In a following part, a brief description of the method will be given, followed by the result obtained with 22 different simulated fall scenarios. Then a discussion on the method and futur works will be presented in the conclusion.

## 2 Method

Our algorithm can be divided into 3 levels: camera, data fusion and recognition levels. Each level is now described in details.

### 2.1 Camera Level

In order to calculate the volume distribution of a subject in his environment the system must know the relationship between the camera coordinate system and the real 3D space. Thus, preliminary to the fall detection process, the cameras have to be calibrated.

Intrinsic parameters were computed using the chess board method[15] to define the focal distance  $\mathbf{f} = (f_x, f_y)$ , the optical center  $\mathbf{c} = (c_x, c_y)$ , the skew coefficient  $\alpha$  and the radial distortion  $\mathbf{k} = (k_1, k_2, k_3, k_4, k_5)$  as presented in [14]. The later parameters are necessary because of non negligible radial distortion due to the large field of view of the camera lenses. External parameters, the rotation matrix  $\mathbf{R}$  and the translation vector  $\mathbf{T}$ , were calculated using feature points manually placed on the floor. Altogether, those parameters define the projective camera model described as follows. Let  $\mathbf{X} = (X, Y, Z)$  be the real world vector of a 3D point and  $\mathbf{X}_c = (X_c, Y_c, Z_c)$  his coordinates in the camera space then :

$$\mathbf{X}_c = \mathbf{R} \mathbf{X} + \mathbf{T}$$

The normalized image projection is defined by:

$$\begin{bmatrix} x_n \\ y_n \end{bmatrix} = \begin{bmatrix} X_c/Z_c \\ Y_c/Z_c \end{bmatrix}$$

The normalized point coordinates with radial distortion become :

$$\begin{bmatrix} x_d \\ y_d \end{bmatrix} = (1 + k_1 r_n^2 + k_2 r_n^4 + k_5 r_n^6) \begin{bmatrix} x_n \\ y_n \end{bmatrix} + \begin{bmatrix} d_x \\ d_y \end{bmatrix}$$

Where the tangential distortion vector is :

$$\begin{bmatrix} d_x \\ d_y \end{bmatrix} = \begin{bmatrix} 2k_3 x_n y_n + k_4 (3x_n^2 + y_n^2) \\ k_3 (x_n^2 + 3y_n^2) + 2k_4 x_n y_n \end{bmatrix}$$

Finally, multiplying the normalized coordinates with the camera matrix gives pixel coordinates

$$\begin{bmatrix} x_p \\ y_p \end{bmatrix} = \begin{bmatrix} f_x & \alpha \cdot f_x & c_x \\ 0 & f_y & c_y \end{bmatrix} \begin{bmatrix} x_d \\ y_d \end{bmatrix}$$

This function can be written as :

$$[x_p, y_p] = \phi(X, Y, Z, \mathbf{f}, \mathbf{c}, \mathbf{k}, \mathbf{R}, \mathbf{T}, \alpha)$$

In order to detect moving objects, each camera image  $i_j$  is subtracted from its own background model  $b_j$  obtained by computing a temporal median image of the sequence [13]. When the absolute difference of a pixel is higher than a previously defined threshold  $Th$ , it is registered as a foreground pixel, otherwise it is considered as a background pixel:

$$s_j(x_p, y_p) = \left\{ \begin{array}{ll} 1 & \text{if } |i_j(x_p, y_p) - b_j(x_p, y_p)| > Th \quad (\text{foreground pixel}) \\ 0 & \text{otherwise} \quad (\text{background pixel}) \end{array} \right\}$$

Finally, in order to reduce noise detection and reinforce large surface detection, an opening morphological operation is done on  $s_j$  to obtain the final 2D silhouette of the person for each camera  $j$ .

## 2.2 Data Fusion Level

This level aims at gathering projections of the 2D silhouette provided by each camera on horizontal slices in order to reconstruct the 3D volume of the subject. Let  $S_{i,j}$  be the projection of the image provided by camera  $j$  on the horizontal plane  $i$  as follow:

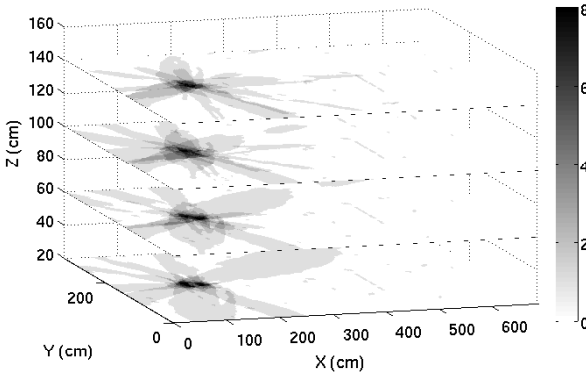
$$S_{i,j}(X, Y) = s_j(\phi(X, Y, Z_i, \mathbf{f}_j, \mathbf{c}_j, \mathbf{k}_j, \mathbf{R}_j, \mathbf{T}_j, \alpha_j))$$

where  $Z_i$  is the height for the horizontal plane  $i$  and  $\mathbf{f}_j, \mathbf{c}_j, \mathbf{k}_j, \mathbf{R}_j, \mathbf{T}_j, \alpha_j$  are the parameters for camera  $j$ .

For each horizontal slice  $i$ ,  $S_i$  is the image corresponding to the summation of projections  $S_{i,j}$  coming from  $n$  cameras:

$$S_i(X, Y) = \sum_{j=1}^n S_{i,j}(X, Y)$$

Where  $n$  is the total number of cameras. Therefore,  $S_i(X, Y)$  takes values between 0 and  $n$  depending on the number of 2D silhouettes (from  $n$  cameras) contributing to the 3D reconstruction at position  $(X, Y)$  and at height  $Z_i$ . The distance between each slice was set arbitrarily to 10 cm in this study. Fig.1 illustrates an example of such kind of fusion. Without occlusion, the person is visible



**Fig. 1.** Representation of 4 slices ( $S_i$ ) where camera views were projected and summed (18 slices were used in practice with a 10 cm vertical interval)

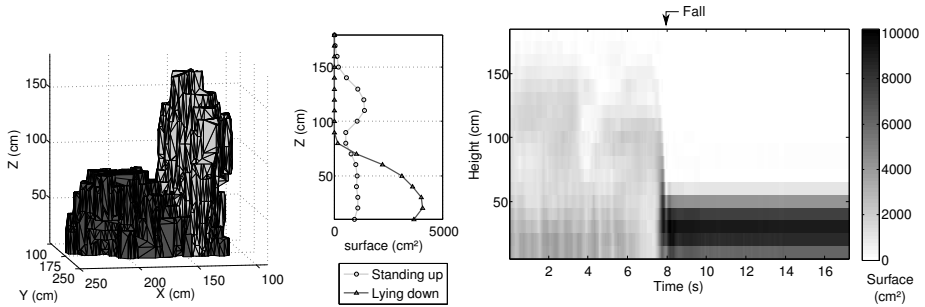
from all cameras and consequently all positions  $(X, Y)$  where  $S_i(X, Y) = n$  define the correct 3D reconstruction (slice by slice). To allow tolerance for one possible occlusion we simply add the positions where  $S_i(X, Y) = n - 1$  at the expense of a slightly larger and coarser reconstruction. Therefore, by thresholding  $S_i$  at  $n - 1$  we obtain the 3D reconstruction as a series of segmented slices  $S_i^*$ :

$$S_i^*(X, Y) = \begin{cases} 1 & \text{if } S_i(X, Y) \geq n - 1 \\ 0 & \text{otherwise} \end{cases}$$

Let  $B$  be the set of pixels in each slice  $S$  belonging to the largest object. The Vertical Volume Distribution of this object at the  $i$ th slice denoted  $VVD(i)$  is given by :

$$VVD(i) = \sum_{(X,Y) \in B} S_i^*(X, Y)$$

Examples of the resulting volume of a standing up (light gray) and lying down (dark gray) positions and their corresponding vertical volume distribution are presented in Fig.2, on left, where the difference is clearly visible. Fig.2, on right, represents the evolution of the vertical volume distribution (displayed with gray levels) of a subject obtained during a fall scenario.



**Fig. 2.** On the left, 3D Reconstruction of a person after fusion of the different points of view and their corresponding vertical volume distribution in the middle graphic. Light gray color is attached to a standing up person, and dark gray for a lying on the ground person. On the right, example of the vertical volume distribution during a fall scenario (displayed with gray levels).

### 2.3 Fall Detection Level

To detect a fall, an indicator based on the ratio between the sum of VVD values from the first 40 cm (5 slices starting from the floor) with respect to the whole volume ( $m=18$  slices) is computed as follows:

$$VVDR = \frac{\sum_{i=1}^5 VVD(i)}{\sum_{i=1}^m VVD(i)}$$

A fall is detected if this ratio is above a preselected threshold (see next section) during a predefined period of time (5 seconds in our case).

## 3 Materials

To evaluate the ability of the method to detect falls, 22 simulated fall scenarios were performed by an actor according to many previously published works (e.g. [3]). The room was 7 meters long by 4 meters large. A set of 8 IP cameras GADSPOT 4600 with large field of view lenses was attached to the ceiling for tests.

The returned times of falls detected with this method were compared with the ground true measured visually. If the difference was below 1 second, the fall was considered as detected. In the case of a fall reported outside this time period, it was considered as a false detection.

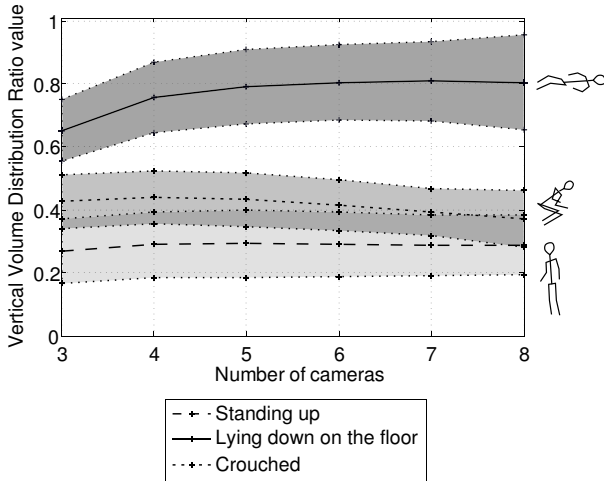
**Table 1.** Number of combinatorial arrangement of cameras possible per number of camera

Number of cameras per combinaison	3	4	5	6	7	8	Total
Number of combinatorial arrangement per scenario	56	70	56	28	8	1	219
Total number of situations	1232	1540	1232	616	176	22	4818

Each scenarios was analysed with all camera combinaisons from 3 to 8 cam-  
 eras, as presented in Tab.1, for a total of 219 camera configurations per scenario.

### 4 Results

The Vertical Volume Distribution Ratio (VVDR) for each body position is pre-  
 sented in Fig.3 as a function of the number of cameras, 95% confidence interval  
 area for the lying-down-on-the-floor position is clearly separated from others.  
 This confirms that fall detection is possible with an appropriate threshold ap-  
 plied to VVDR (e.g. 0.6 for four cameras and more).



**Fig. 3.** Influence of the number of cameras on the capability of the VVDR to discrim-  
 inate body postures. The gray areas correspond to 95% confidence intervals.

**Table 2.** Influence of the number of cameras on correct detection and false alarm  
 rates

Number of cameras	Correct detection rate	False detection rate
$\geq 4$	$\geq 99\%$	$< 1\%$
3	96%	10%

The results, presented in Tab.2, show that falls was correctly detected in 96%  
 of situations with only 3 cameras and above 99% with 4 cameras and more. The  
 false detection rate is 10% with 3 cameras and less than 1% with 4 cameras and  
 more. These results were obtained with a VVDR threshold which was equal to  
 mean point between crouching position and lying-on-the-floor position VVDR  
 value for each number of cameras.

## 5 Conclusion

The results obtained with this method enlighten its efficiency to do its aim : detecting lying people on the floor. However it is based on the assumption that the major part of the fallen body is under 40 centimeters height. When post-fall position differs from this expected position (e.g. fall on a table), it will be misdetected.

The method is clearly efficient even with only three cameras. Although the false detection rate is still not negligible in this case. Fortunately with four cameras or more, the false detection rate becomes negligible.

The foreground segmentation algorithm used was a very basic one. This is possible because the volume reconstruction method, being robust to occlusions, is also robust to segmentation errors.

For future work, more experimentations on occlusion resistance would have to be done in order to better assess the limit of this method. This could be done with real life and simulated occlusions. Some improvements could be worked on, such as the optimisation of the segmentation and back projection processings which could be done on a GPU, or a better foreground/background segmentation algorithm.

The major contribution of this paper is the validation of the VVDR efficiency with 3 cameras and more with the aim to detect lying down person after a fall. But this reconstruction method could be applied to other types of applications, such as quantifying daily life activities which is a key issue of our modern society.

## Acknowledgment

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and MITACS.

## References

1. Auvinet, E., Reveret, L., St-Arnaud, A., Rousseau, J., Meunier, J.: Fall detection using multiple cameras. In: EMBS 2008, 30th Annual International Conference of the IEEE (2008)
2. Xinguo, Y.: Approches and Principles of Fall Detection for Elderly and Patient. In: 10th IEEE Intl. Conf on e-Health Networking, Application and Service (2008)
3. Noury, N., Fleury, A., Rumeau, P., Bourke, A., Laighin, G., Rialle, V., Lundy, J.: Fall detection - principles and methods. In: 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS 2007), pp. 1663–1666 (2007)
4. Kangas, M., Konttila, A., Lindgren, P., Winblad, I., Jämsä, T.: Comparison of low-complexity fall detection algorithms for body attached accelerometers. *Gait & Posture* 28, 285–291 (2008)
5. Rougier, C., Meunier, J., St-Arnaud, A., Rousseau, J.: Monocular 3D Head Tracking to Detect Falls of Elderly People. In: Conference of the IEEE EMBS, August 30-September 3 (2006)

6. Rougier, C., Meunier, J., St-Arnaud, A., Rousseau, J.: Fall Detection from human Shape and Motion History using Video Surveillance. In: IEEE First International Workshop on Smart Homes for Tele-Health, Niagara Falls, May 2007, pp. 875–880 (2007)
7. Tao, J., Turjo, M., Wong, M.F., Wang, M., Tan, Y.-P.: Fall Incidents Detection for Intelligent Video Surveillance. In: Conference on Information, Communications and Signal Processing, December 6-9 (2005)
8. Lee, T., Mihailidis, A.: An intelligent emergency response system: preliminary development and testing of automated fall detection. *Journal of telemedicine and telecare* 11(4), 194–198 (2005)
9. Miaou, S.G., Shung, P.H., Huang, C.Y.: A customized human fall detection system using omni-camera images and personal information. In: D2H2 2006, 1st Transdisciplinary Conf. on Distributed Diagnosis and Home Care, pp. 39–42 (2006)
10. Anderson, D., Keller, J.M., Skubic, M., Chen, X., He, Z.: Recognizing Falls from Silhouettes. In: Conference of the IEEE EMBS, August 30-September 3 (2006)
11. Laurentini, A.: The visual Hull concept for silhouette-based image understanding. *IEEE Trans. Pattern Anal. Mach. Intell.* 16(2), 150–162 (1994)
12. Kim, H., Sakamoto, R., Kitahara, I., Orman, N., Toriyama, T., Kogure, K.: Compensated Visual Hull for Defective Segmentation and Occlusion. In: International Conference on Artificial Reality and Telexistence (2007)
13. Piccardi, M.: Background subtraction techniques: a review. In: IEEE International Conference on Systems, Man and Cybernetics (2004)
14. Heikkila, Silven: A Four-step Camera Calibration Procedure with Implicit Image Correction. In: Conference Vision Pattern Recognition (1997)
15. Bouguet, J.Y.: Camera Calibration Toolbox for Matlab (2007), [http://www.vision.caltech.edu/bouguetj/calib\\_doc/](http://www.vision.caltech.edu/bouguetj/calib_doc/)