Fall Detection With Multiple Cameras: An Occlusion-Resistant Method Based on 3-D Silhouette Vertical Distribution

Edouard Auvinet, Franck Multon, Alain Saint-Arnaud, Jacqueline Rousseau, and Jean Meunier

***Abstract*—According to the demographic evolution in industrialized countries, more and more elderly people will experience falls at home and will require emergency services. The main problem comes from fall-prone elderly living alone at home. To resolve this lack of safety, we propose a new method to detect falls at home, based on a multiple-cameras network for reconstructing the 3-D 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 during a predefined period of time, which implies that a person has fallen on the floor. This method was validated with videos of a healthy subject who performed 24 realistic scenarios showing 22 fall events and 24 cofounding events (11 crouching position, 9 sitting position, and 4 lying on a sofa position) under several camera configurations, and achieved 99.7% sensitivity and specificity or better with four cameras or more. A real-time implementation using a graphic processing unit (GPU) reached 10 frames per second (fps) with 8 cameras, and 16 fps with 3 cameras.**

***Index Terms*—3-D reconstruction, fall detection, multiple cameras, occlusion.**

NOMENCLATURE

$\mathbf{X} = (X, Y, Z)$	Real world coordinates.
$\mathbf{X}_c = (X_c, Y_c, Z_c)$	Camera coordinates.
$\mathbf{f} = (f_x, f_y)$	Focal length (horizontal and vertical).
$\mathbf{c} = (c_x, c_y)$	Optical center coordinates.
$\mathbf{k} = (k_1, k_2, k_3, k_4, k_5)$	Radial distortion parameters.
$\mathbf{T}$	3D translation vector.
$\mathbf{R}$	3D rotation matrix.
(x_n, y_n)	Normalized image projection.
(d_x, d_y)	Tangential distortion vector.
r_n	Radial distance.
(x_d, y_d)	Normalized image coordinates with radial distortion.
(x_p, y_p)	Pixel image coordinates.
α	Skew coefficient.
i_j	Image i of camera j.
b_j	Background model of camera j.
s_j	Binary image of the segmented foreground object for camera j.
z_i	Height of the horizontal plane i.
$S_{i,j}$	Projection of the image provided by camera j on the horizontal plane i.
S_i	Summation of the projection $S_{i,j}$ coming from n cameras.
S_i^*	One slice of the 3D volume reconstructed.
$VVD(i)$	Vertical Volume Distribution of the object at the ith slice.
$VVDR$	Vertical Volume Distribution Ratio.
Th	Segmentation threshold.

I. INTRODUCTION

WHEN approaching 65 years old, the risk of falling is rising. Indeed 30% of people over 65 years of age and living in the community fall each year, and a fifth of fall incidents require medical attention [1]. Hence, falling is the most common cause of injury for elderly people [2]. It was the first cause of death by injury for elderly in 1997 and 1998 [3], [4]. Although, most of the falls result in light injuries, 5%–10% of falls in community dwelling lead to serious injuries such as fractures, head injuries, or serious lacerations [5], [6]. An example of such injuries is hip fracture. Moreover, 25%–75% of "fallers" do not recover their prefracture level of movement and autonomy [7]. Besides, fear of falling appears and/or increases after falling that could increase the risk factor for future falls and reduce the quality of life [5]. This fall problem becomes more important for elderly people living alone because they cannot always call emergency services. Hence, many recent works have tried to develop easy-to-use and automatic techniques to detect falls in elderly people's houses [5], [9]–[15]. The key question is: how to detect that a person has fallen in a house, which contains many objects, and where people can perform a wide range of activities?

Manuscript received February 3, 2010; revised July 1, 2010; accepted October 3, 2010. Date of publication October 14, 2010; date of current version March 4, 2011. This work was supported by the Natural Sciences and Engineering Research Council of Canada and the Mathematics of Information Technology and Complex Systems Network (MITACS).

E. Auvinet is with the Institute of Biomedical Engineering, University of Montreal, Montreal, QC H3T 1J4, Canada, and also with the M2S Laboratory, University of Rennes 2, 35043 Rennes cedex, France (e-mail: auvinet@iro.umontreal.ca).

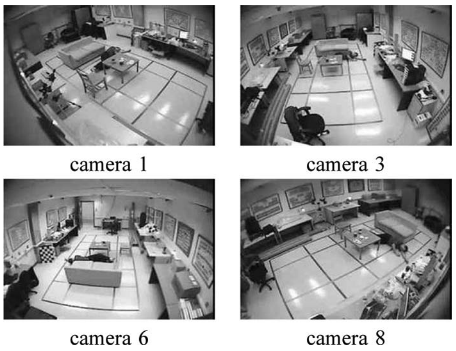
F. Multon is with the M2S Laboratory, University of Rennes 2, 35043 Rennes cedex, France (e-mail: franck.multon@uhb.fr).

A. Saint-Arnaud is with the Lucille Teasdale Health and Social Service Center, Montreal, QC H1W 0A9, Canada (e-mail: alain.starnaud.lteas@ssss.gouv.qc.ca).

J. Rousseau is with the Centre de recherche de l'Institut universitaire de gériatrie de Montreal, University of Montreal, Montreal, QC H3T 1J4, Canada (e-mail: jacqueline.rousseau@umontreal.ca).

J. Meunier is with the Institute of Biomedical Engineering and the Department of Computer Science and Operations Research, University of Montreal, Montreal, QC H3T 1J4, Canada (e-mail: meunier@iro.umontreal.ca).

Digital Object Identifier 10.1109/TITB.2010.2087385



camera 1

camera 3

camera 6

camera 8

Fig. 1. Examples of real occlusions in our experimental setup.

Indeed, one of the key problems is to recognize a fall among all the daily life activities. A description of the various phases of falling has been proposed in previous studies [16], [17]. This classification provides us with physical features proper to fall movement that can be used to detect a fall in daily life. Falling is subdivided into four phases [17]: prefall (linked to daily life motions); critical (loss of balance); postfall (final position after fall); and recovery (return to normal daily life) phases. The critical phase is extremely short (300–500 ms [17], [18]) composed of “free fall” and ”impact with the floor” events. The former is associated with an increase of the body's velocity because of gravity. This velocity reaches abnormal maximal values for vertical and horizontal speeds compared to normal life activities [18]: typically 2–3 times higher values. At the ”impact with the floor” event, the speed decreases down to zero and a sudden inversion of acceleration polarity occurs. During the postfall period, the main features are a horizontal orientation of the body, a proximity to the floor, and commonly, lack of movements.

According to this description, several approaches have been proposed to detect falls. These approaches mainly focus on the critical and postfall phases. Wearable devices composed of accelerometers or gyroscopes directly placed on subjects body parts (mainly chest [5], waist [8], or wrists [8]) enable to capture the high velocities, which occur during the critical phase and the horizontal orientation during the postfall phase. However, these methods are based on the assumption that the subject wears the system at any time (with a warning by the system otherwise), and therefore, if it is uncomfortable, it could bother the user. Additionally, such systems require recharging the battery frequently, which could be a serious limitation for real application.

On the opposite, video systems enable an operator to rapidly check if an alarm is linked to an actual fall or not. Therefore, cameras placed in the subject's environment were used to detect falls by measuring the movement or orientation of the body. A first approach consists in detecting abnormal horizontal and vertical speeds [9] or body silhouette changes [10], [12] associated with the critical phase. Another method consists in using body orientation features, such as width and height of a silhouette by comparing a standing and a lying person [12], [15], [19]. In this case, the detection would be mainly based on postfall shape or orientation features. As these approaches generally use only one camera, they could fail to detect falls in case of occlusions. These occlusions frequently occur in real situations at home because a room contains furniture and objects that could be placed between the subject and the camera (as shown in Fig. 1), contrary to easier and simplified experimental settings in laboratories. Thus, dealing with occlusions is a key issue for using video systems in real situations in order to avoid misdetection and false alarms.

Using multiple cameras could overcome this limitation by offering several different points of view of the subject. It then becomes possible to extract a 3-D silhouette of the subject. Some approaches use homography (a transformation between projective planes) to project silhouette (previously segmented with a foreground/background algorithm) on the ground and parallel planes for gathering information from different cameras and locate the person in the place (e.g., [14], [20]–[22],). Another method, the visual hull [23], consists in back-projecting silhouettes into space using camera models. Contrary to homography, camera models permit to represent more sophisticated situations such as lens distortion. The intersection of all those projections results in the final volume [23]. This method has been applied to fall detection in [13] to detect if the body is vertical or not during the postfall phase. To this end, the method computes the centroid of the volume and its main axis using principal component analysis. The authors did not report any information about the robustness of the system to occlusions. However, since the silhouettes coming from all the cameras were needed, when an occlusion occurs for one camera or more, the reconstructed volume may become unreliable or unusable. To overcome this problem, it is possible to use an occlusion-resistant visual-hull method [24]. This approach is able to reconstruct a volume even if one of the silhouette is not present for one of the cameras (such as when the body is occluded).

In this paper, we propose a method that is capable of dealing with several occlusions that could occur in personal houses. This method is based on two main ideas. First, we use the occlusion-resistant algorithm introduced previously in [23] in order to detect if a person is lying on the ground even if some occlusions occur. Second, we introduce the original and simple idea of vertical volume distribution ratio (denoted VVDR in the remaining of this paper). This ratio is obtained by dividing the volume that is below a given height by the total volume. For people lying on the ground, this ratio is high compared to when they are standing up. We assume that this feature is less sensitive to noise than methods based on the principal axis of the reconstructed volume. VVDR has been successfully tested in a few occlusion-less situations [25]. In the present paper, we tested how this framework is able to manage occlusions in 24 realistic scenarios showing 22 fall events and 24 confounding events (11 crouching position, 9 sitting position, and 4 lying on a sofa position) under several camera configurations. This unique dataset is documented [26] and made available to the scientific community through a website [27]. We also theoretically analyze the robustness of the method to occlusion by identifying the worst occlusion case and testing it with experimental videos. Finally, a real-time implementation using GPU is demonstrated.

The paper is organized as follows. Section II describes the theoretical background and the implementation of the method proposed in this paper in order to detect falls in these

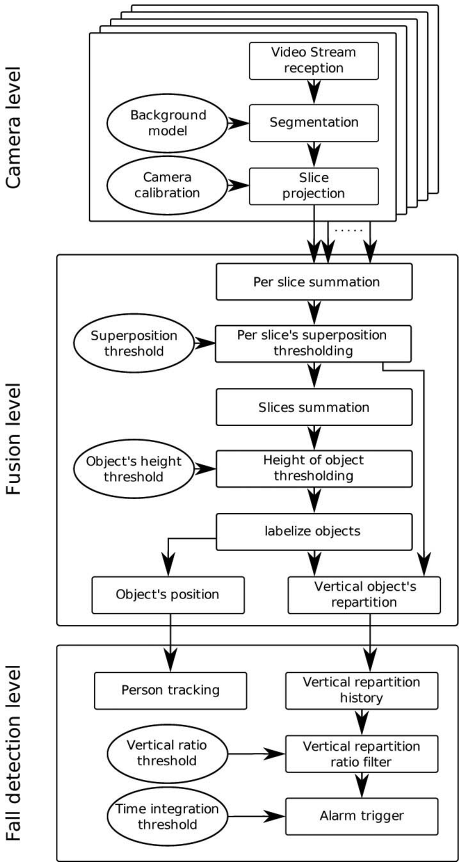


Fig. 2. Schematic representation of the entire process.

scenarios, thanks to the multi-cameras system. Section III describes the experimental setup that was used to generate the 24 scenarios with occlusions to validate the method. Section IV provides some results and discussion about the performance of the method. Finally, Section VI gives the conclusion and perspectives to this paper.

II. Method

Briefly stated, the method essentially involves two main steps. First, with the set of cameras, the 3-D volume of the person is reconstructed with a shape-from-silhouette approach. Second, an index (VVDR) is obtained by dividing the volume that is below 40 cm by the total volume of the person. Then, a simple threshold is used to determine if this index reveals a fall or not. This section describes in more details these steps.

Our algorithm can be divided into three levels: camera and data fusion levels (step 1); and recognition level (step 2), as shown in Fig. 2.

A. 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 3-D space. Thus, preliminary to the fall detection process, the cameras have to be calibrated.

Intrinsic parameters were computed using the chess-board method [28] to define the focal distance $\mathbf{f} = (f_x, f_y)$, the optical center $\mathbf{c} = (c_x, c_y)$, the skew coefficient α, and the radial distortion $\mathbf{k} = (k_1, k_2, k_3, k_4, k_5)$, as presented in [29]. The later parameters are necessary because of nonnegligible 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, these parameters define the projective camera model described as follows. Let $\mathbf{X} = (X, Y, Z)$ be the real world vector of a 3-D 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 (x_n, y_n) 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 (x_d, y_d) with radial distortion become

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

where the tangential distortion vector (d_x, d_y) is

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

and radial distance is: $r_n = \sqrt{x_n^2 + y_n^2}$.

Finally, multiplying the normalized coordinates with the camera matrix gives pixel coordinates (x_p, y_p)

$$\begin{bmatrix} x_p \\ y_p \end{bmatrix} = \begin{bmatrix} f_x & \alpha \cdot f_x & c_x \\ 0 & f_y & c_y \end{bmatrix} \begin{bmatrix} x_d \\ y_d \\ 1 \end{bmatrix}$$

where α is a skew coefficient. This function can be written as follows:

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

In order to detect moving objects, each image of camera j, noted i_j, is subtracted from its own background model b_j obtained by computing a temporal median image of the sequence [30]. 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

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

Finally, in order to reduce noise detection and reinforce large surface detection, an opening morphological operation is done on s_j. An example of this segmentation is given in Fig. 3.

B. Data Fusion Level

This level aims at gathering projections of the 2-D silhouette provided by each camera on horizontal slices in order to reconstruct the 3-D volume of the subject. Let S_{ij} be the projection

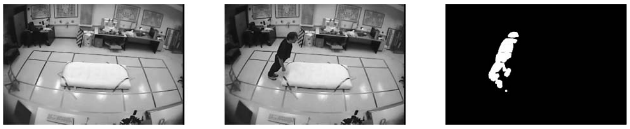


Fig. 3. Result of the moving object segmentation process. From left to right, background median model b_j, current frame i_j, and segmented picture s_j for camera j.

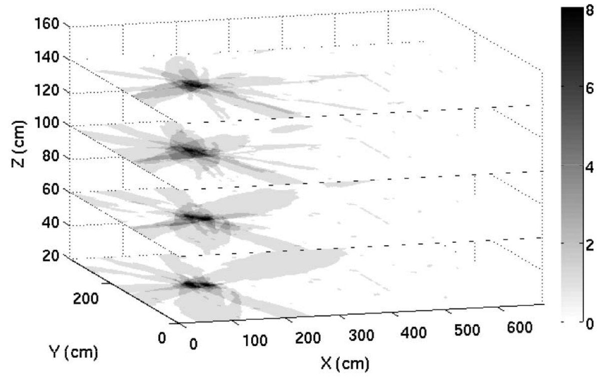


Fig. 4. Representation of four slices (S_i), where camera views were projected and summed (18 slices were used in practice with a 10-cm vertical interval)

of the image provided by camera j on the horizontal plane i as follows:

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

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

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

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

where n is the total number of cameras. Therefore, $S_i\left(X,Y\right)$ takes values between 0 and n, depending on the number of 2-D silhouettes (from n cameras) contributing to the 3-D 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. 4 illustrates an example of such kind of fusion.

Without occlusion, the person is visible from all cameras and consequently all positions (X,Y), where $S_i\left(X,Y\right)=n$ define the correct 3-D reconstruction (slice by slice). To allow tolerance for one possible occlusion, we simply add the positions, where $S_i\left(X,Y\right)=n-1$ at the expense of a slightly larger and coarser reconstruction. Therefore, by thresholding S_i at $n-1$, we obtain the 3-D reconstruction as a series of segmented slices S_i^*

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

As the threshold is applied individually to each position (X,Y), we can also handle the case of multiples partial occlusions in different cameras if they are not affecting the same position (X,Y). Notice that, reducing the threshold to accommodate more occlusions would result in an unacceptable enlargement and innacuracy of the 3-D reconstruction.

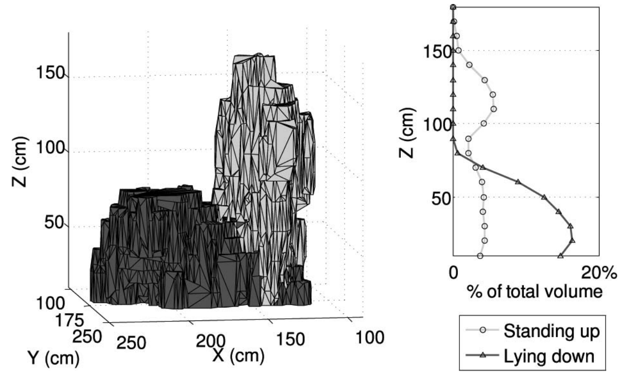


Fig. 5. 3-D reconstruction of a person after fusion of the different points of view and their corresponding VVD on the right. Light gray color is attached to a standing up person, and dark gray for a lying on the ground person.

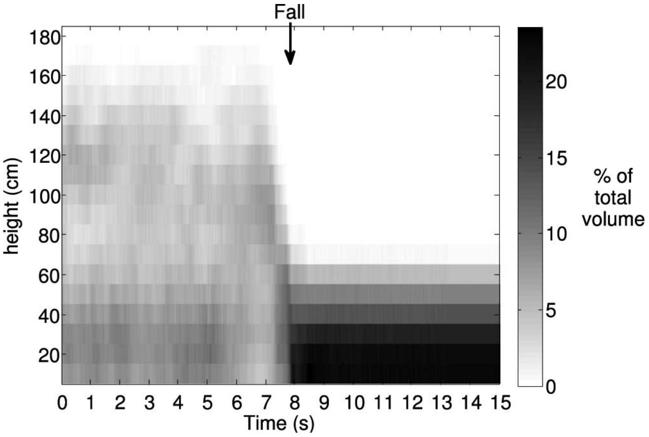


Fig. 6. Example of the VVD during a fall scenario (displayed with gray levels).

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 ith slice denoted VVD (i) is given by

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

Examples of the resulting volume of a standing up (light gray) and lying down (dark gray) positions, and their corresponding VVD are presented in Fig. 5, where the difference is clearly visible. Fig. 6 represents the evolution of the VVD (displayed with gray levels) of a subject obtained during a fall scenario.

C. Fall-Detection Level

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

$$\text{VVDR}=\frac{\sum_{i=1}^{5}\text{VVD}\left(i\right)}{\sum_{i=1}^{m}\text{VVD}\left(i\right)}. \tag{1}$$

This value, 40 cm, is justified by anthropometric data from [31]. In particular, for the 65–80 years old range, the shoulder

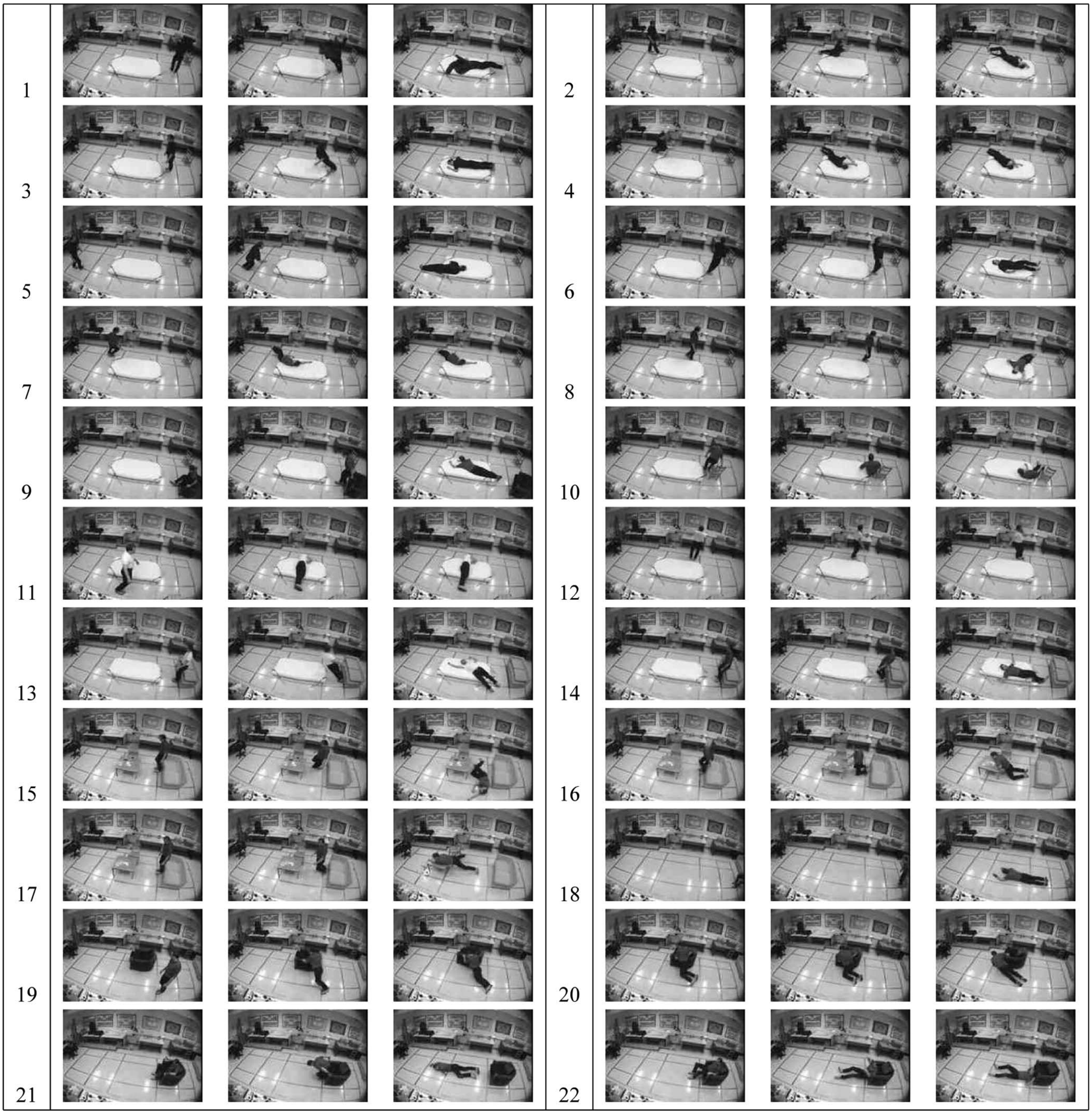


Fig. 7. Description of each scenarios.

width mean is 41.7 cm. This is the highest height to be expected for a lying down body.

A fall is detected if this ratio is above a preselected threshold during a predefined period of time (5 s in our case). This predefined period of time of 5 s is not a sensitive parameter and could be longer if needed. We chose 5 s because after that period, the subject stood up after a fall (we did not ask him to stay on the floor indefinitely) and the confounding events (e.g., crouching down) were lasting shorter periods of time. In practice, this parameter should be chosen by the clinician considering the habits of the elderly person.

III. Materials and Experiments

In order to evaluate the method proposed in this paper we have captured several videos containing a wide set of falls (see Fig. 7). For each situation, we used several synchronized cameras. However, it is impossible to capture real-life situations, where people actually fall. This is why we have designed scenarios that were carried out by an actor who performed the falls in our laboratory with appropriate protection (mattress). One has to notice that the realism of the falling motion is not a key issue here as our approach focuses on the postfall phase.

A. Experimental Setup

The dimension of the area was 7 m per 4 m. A table, a chair, and a sofa were introduced in the capture area in order to reproduce a normal room, where people actually live. Adding such furniture introduces occlusions in the videos for most of the scenarios. We assumed that a commercial system based on our technique would be made up of Internet Protocol (IP) video surveillance cameras with large field of view lenses. For all the scenarios, we thus placed eight such cameras (Gadspot 4600, 110° field of view) all around the area. They were attached to the ceiling at 2.5-m height. Video streams (720 × 480 at 30 fps) were recorded and analyzed on a common desktop PC.

B. Fall Scenarios

We decided to propose a wide range of realistic fall scenarios according to many previously published works (e.g., [32]). Each scenario is defined by a set of characteristics, such as the main falling direction (falling down, forward, backward, and side way) and the departure position (stand up, sit on a chair, or a sofa). Each scenario is depicted in Fig. 7. Some situations, which could lead to false alarms, such as occlusions due to furniture (see Fig. 1), crouching down on the floor, and lying on a sofa (see Fig. 8) are also present to complexify the scenarios. Overall, there were 24 realistic scenarios showing 22 fall events and 24 confounding events (11 crouching position, 9 sitting position, and 4 lying on a sofa position) under several camera configurations. These scenarios captured with eight cameras correspond to a total of 143472 frames (4782.4 s) to be analyzed by the system. This unique dataset is documented in [26] and made available to the scientific community through a website [27].

Each scenario was performed once by one subject and approved by the local Institutional Review Board (IRB) authority. The subject in the videos is one of the author (A. Saint-Arnaud), a clinician, whose research interests are elderly people affected by musculoskeletal and cognitive disorders living in the community. He is well aware of the different features of real falls in elderly people and took care of performing the simulated falls accordingly (e.g., slow motion, falls due to different disorders (loss of balance, blood pressure drop, abrupt sitting due to weakening of the ham-string muscles in elderly people, etc.).

All the cameras were used to capture the fall. However, it was possible to test various camera configurations by using or not some of the video sequences during the analysis process. Hence, we tested configurations using three to eight cameras. It enabled us to evaluate how our method was influenced by the number of the cameras used for the capture. For each scenario, we tested 219 configurations: all the possible combinations when selecting three to eight cameras among eight cameras

$$C_8^3 + C_8^4 + C_8^5 + C_8^6 + C_8^7 + C_8^8 = 219. \qquad (6)$$

Some of these scenarios involved occlusions due to furniture placed in the environment.

In order to test further the ability of the system to tackle the problem of occlusions, we also introduced two artificial occlusions. The first one consists in completely deleting the contribution of one camera, corresponding to a full occlusion of this camera. The second one deletes only the contribution of one camera for the lowest 40 cm of the 3-D volume. This correspond to the worst possible case because the volume of the lowest part becomes underestimated, and consequently, this reduces the VVDR value. This can be explained by the illustrative example in Fig. 9. For a four-camera setup without occlusion, the segmented slice S_i^* is larger (1-2-3-4) than with one occlusion (1-2-3,1-2-4,1-3-4,2-3-4). Therefore, an occlusion may contribute to a higher rate of FNs (failing to detect a real fall).

Fig. 8. Examples of confounding events, from left to right, crouched down, lying on a sofa, and sitted position.

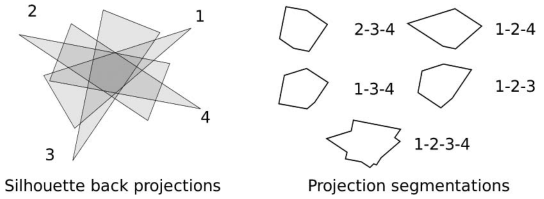


Fig. 9. Illustrative example with four cameras. For a given slice, the segmented surface S_i^* (and the reconstructed volume) is underestimated in case of a camera occlusion.

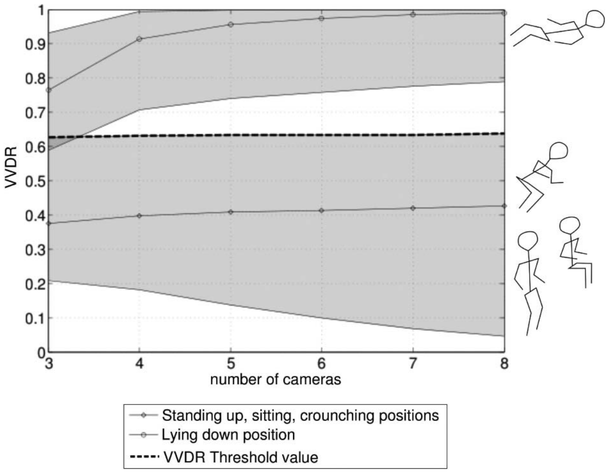


Fig. 10. Influence of the number of cameras on the capability of the VVDR to discriminate body postures obtained without artificial occlusions. The gray areas correspond to 95% confidence intervals and the solid lines are the medians.

C. Data Analysis

In this paper, we wish to evaluate the ability of the VVDR to discriminate lying-on-the-floor position (corresponding to a fall) from others. To this end, we computed the VVDR for all the images coming from the sequences. As shown in Fig. 10, VVDR for lying down positions is clearly different than others, such as standing up, sitting down, or crouched positions. This statement is true for whatever the number of cameras and even with only three cameras, where the separation remains acceptable. The actual time, where a fall occurs (denoted t_{fall}) is manually measured in the video sequences. This time is defined as the beginning of the postfall period when the body hits the

TABLE I
SENSIBILITY AND SPECIFICITY OBTAINED WITH VVDR THRESHOLD SET AT 97.5 PERCENTILE OF THE NO-FALL REGION

Number of camera	original video	
	sensitivity	specificity
3	0.806 (+- 0.021)	1.000 (+- 0.000)
4	0.997 (+- 0.001)	0.998 (+- 0.000)
5	0.999 (+- 0.000)	1.000 (+- 0.000)
≥6	1.000 (+- 0.000)	1.000 (+- 0.000)
	worst occlusion	
	sensitivity	specificity
3	0.550 (+-0.022)	1.000 (+- 0.000)
4	0.895 (+- 0.019)	1.000 (+- 0.000)
5	0.933 (+- 0.016)	1.000 (+- 0.000)
≥ 6	0.954 (+- 0.011)	1.000 (+- 0.000)
	total occlusion	
	sensitivity	specificity
3	0.947 (+-0.007)	0.995 (+- 0.001)
4	0.999 (+- 0.000)	0.990 (+- 0.002)
5	1.000 (+- 0.000)	0.984 (+- 0.003)
≥ 6	1.000 (+- 0.000)	0.956 (+- 0.008)

Mean +- standard deviation of leave-one-out.

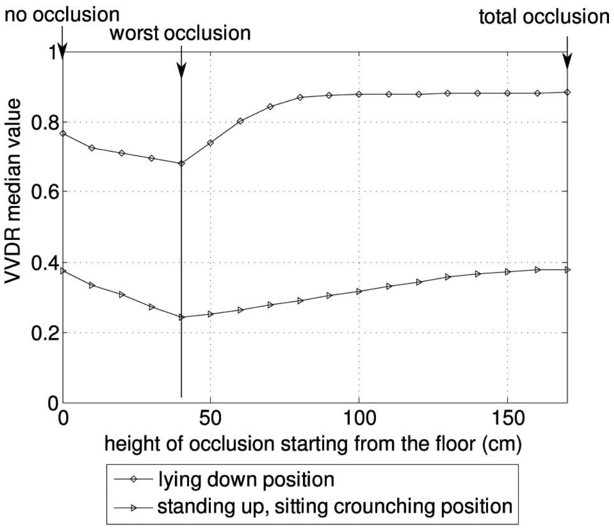


Fig. 11. VVDR for a standing up and lying down situations, where one camera is gradually occluded (from the ground to the head of the subject) in a three-cameras setup with 95% confidence interval in light gray.

ground. If our method is able to detect a fall event after t_{fall}, the detection is supposed to be correct (true positive). If the method does not detect any fall, it is supposed to have failed (false negative). If it detects a fall event before t_{fall} this time interval, it is supposed to have generated a false detection (false positive). If no fall is detected before t_{fall}, it is then considered as TN.

D. Statistical Analysis

The VVDR threshold to set a fall detection was simply taken as the 97.5% percentile of the no-fall region in Fig. 10. This means that with VVDR alone, 2.5% of false positives (FPs) will occur, but we will get most, if not all, the true positives (TPs). This bias toward TP is reasonable since we prefer a few more FPs to avoid some miss-detections of fall (risk minimization). Moreover, the predefined period of inactivity (5 s) after a potential fall will remove several other FPs.

We have tested this threshold with an unbiased leave-one-out strategy to compute the sensitivity and specificity of the complete system (including the period of inactivity of 5 s) in Table I. This means that for each scenario tested, we have computed the VVDR threshold corresponding to the 97.5% percentile of the no-fall region obtained from all the remaining (training) scenarios.

To analyze our recognition results, we compute the sensitivity and the specificity, as follows:

1) *Sensitivity:* $\text{Se} = \dfrac{\text{TP}}{(\text{TP} + \text{FN})}$
2) *Specificity:* $\text{Sp} = \dfrac{\text{TN}}{(\text{TN} + \text{FP})}$

where

1) *True Positives (TP):* number of falls correctly detected (among the 22 fall events multiplied by the total number of camera configurations).
2) *False Negatives (FN):* number of falls not detected.
3) *False Positives (FP):* number of normal activities detected as a fall (among the 24 normal segments in each scenario multiplied by the total number of camera configurations).
4) *True Negatives (TN):* number of normal activities not detected as a fall.

IV. RESULTS

This section presents results obtained from experimentation with the dataset previously presented. In the first part, the ability of the VVDR to detect a fall is examinated. Then, the real-time constrain is tested with respect to the number of camera.

A. VVDR Behavior

Results shown in Fig. 10 prove that VVDR enables to discriminate lying-on-the-floor from others positions. Indeed, the 95% confidence intervals (gray areas) around the mean value of VVDR for lying-on-the-ground and others positions are very well separated with four cameras or more. The separation remains acceptable for the three-camera setup, although, some overlap appears between the confidence intervals.

Moreover, the distance between the two confidence intervals increases with the number of cameras, which tends to show that the ability to detect lying positions increases with the number of cameras.

With four cameras or more, the system achieved almost 100% sensitivity and specificity, as presented in the first part of Table I. The less favorable results were obtained with three cameras, for which the sensitivity decreased down to 80.6%. Whatever the scenario was, simulating a partial occlusion of the lowest 40 cm above the ground (worst occlusions) in one camera led to an artificial decrease of VVDR (see Fig. 11) resulting in a lower detection rate (55% sensitivity with three cameras), but also in the same way, a lower false detection rate (100% specificity), as shown in the second part of Table I. On the contrary, simulating total occlusion of one camera increased the VVDR (see Fig. 11) resulting in a higher detection rate (94.7% sensitivity with three cameras) at the expense of a higher FP rate (95.6% specificity for six cameras and more), as shown in Table I.

TABLE II
INFLUENCE ON THE IMPLEMENTATION ON COMPUTATION TIME

Segmentation	CPU	CPU	GPU
Projection	CPU	GPU	GPU
Number of cameras	time (msec/frame)	time (msec/frame)	time (msec/frame)
3	1140	98	63
4	1516	100	72
5	1888	111	79
6	2258	122	88
7	2613	133	96
8	2980	145	105

The fact that the inflexion point is located at 40 cm, when all contribution of one camera for the lower part of the body are clearly occluded, demonstrates that this is the worst occlusion case, as explained by Fig. 9. In this case, the numerator of (5) is the most underestimated.

B. Real-Time Implementation

Three different implementations of this method have been tested. The first one used only the processor to deal with all the computations. The second one used the GPU [35] for reconstructing the voxels, while the remaining of the computations were performed by the processor. The last one used the GPU for reconstructing the voxels and segmenting the image. Computation times for these three methods are reported in Table II. The main result is that the algorithm, which used the GPU can go 18 times faster than the one with only the processor, for three cameras. This ratio increased up to 28 when using eight cameras.

V. DISCUSSION

Our results compare very favorably with those reported in the literature. For instance, Rougier *et al.* [10] have developed a fall-detection system with a single camera based on silhouette deformation of the subject with the same dataset used here. Their results gave a sensitivity and specificity of 95.5% and 96.4%, respectively, that are lower than those obtained with our method, although, this comparison is somewhat unfair because they used only one camera. Anderson *et al.* [13] used fuzzy logic with a multicamera setup on a dataset containing 14 falls and 32 no-falls events. They obtained 100% TP detection and 6,25% false detection (sensitivity = 100% and specificity = 93.75%). However, they did not address the problem of occlusions and their approach requires the manual adjustment of several parameters. Cucchiara *et al.* [14] proposed a posture-classification system that was able to achieve 97.23% accuracy with some occlusions and for four types of postures including lay down position. Notice that [13] and [14] used datasets with different (unspecified number of) subjects, while we used one (experimented) subject.

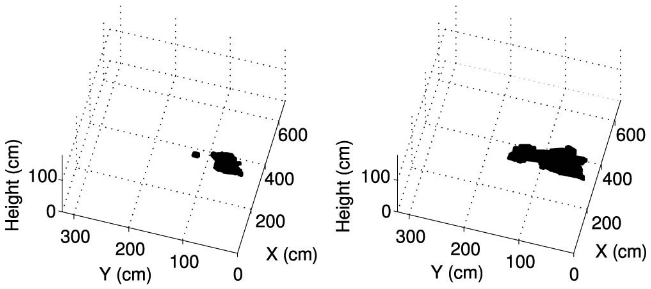


Fig. 12. Volume reconstructed in case of real occlusions with pictures shown in Fig. 1. Left: classical method. Right: occlusion-resistant method presented in this paper.

The method presented in this paper is able to deal with an occlusion of one camera without significantly decreasing the detection rate. In real life, situations, where several cameras are occluded generally occur. However, the reconstruction algorithm is applied independently for each 3-D position. Hence, the algorithm is able to deal with several occlusions, except if there is more than one occlusion for the same 3-D position. In some real situations, such as the one depicted in Fig. 1, all the cameras may be partially occluded. A classical reconstruction method [13] may fail in recognizing a lying person in that case, as shown in Fig. 12, whereas, the method presented in this paper is able to reconstruct the volume of the actor. The quality of the resulting reconstructed volume is sufficient to compute VVDR, and thus determine if the actor is lying down or not.

Another important feature of the method is the fact that it works without considering the speed or motion of the person. Indeed, by simply looking for abnormal volume distribution along the vertical axis, i.e., when the major part of the body is near the floor, fall detection is made possible. This point is important because motion is generally difficult to measure and usually needs more computer resources, and a high and fixed frame rate to be accurate; these requirements add complexity and could impair real-time processing.

Regarding the possibility that people just happen to have their bodies close to the floor for a long period of time (maybe to pick up something or to tie their shoelaces). This problem is usually tackled by the computation of the VVDR itself because a large part of the body remains above 40 cm, but in the unusual case, where the elderly is very near the floor, the predefined limit of time (5 s in our case) is sufficient to avoid FP. The limit could be increased to a higher value if necessary depending on the habits of the elderly person and is not a sensitive parameter. In this paper, the 11 crouching-down events were correctly identified as TN (except for the three-camera configuration). However, in the case, where the subject finishes the fall onto an object (e.g., table, wall, or other furniture), a large part of the body could remain above 40 cm, and thus, our system could fail. Adding some knowledge about the environment could help to detect these difficult cases. Similarly, a fall ending in a sitting posture (on the floor) could cause a miss detection.

Notice that another moving object entering/leaving the room, like a cat or dog would be ignored because the system analyses only the largest object (human) in the scene (see Section II). Very big dogs are out of the question because of the additional risk of fall for an elderly person living alone.

To bring the system to a multiroom setup, a set of cameras needs to be installed in every room. Fortunately, this does not require much more computer resources. Actually, the computer simply needs to know in what room the elderly person is and then process the data coming from this room only for fall detection.

The presence of a person in a room can be easily and quickly monitored with simple background subtraction for all cameras (in all rooms) checked one after the other at a low frame rate per camera (e.g., 1 frame/sec). Only the cameras involved in the identified room would be processed at a higher frame rate for fall detection. More sophisticated alternatives are also possible and will be investigated in the future.

One shortcoming of a multiroom setup of our video system is the requirement for installation of adequate infrastructures that may cause a significant modification of the subjects home environment. Although, this means certainly much work (e.g., compared to wearable devices), with the current miniaturization of cameras and reliable Wi-Fi technology, we believe that such system is nevertheless realizable for real application. The cost of such system could be another problem for the user or provider; however, we think that the economic advantages will be noticeable when compared with traditional intervention, i.e., placing the elderly in a specialized establishment (instead of the home setting). Another concern about video systems could be the intimacy and privacy of the user. For this reason, this system should use a closed circuit: the system will be activated to send an alarm signal toward an outside resource (e.g., via a phone or internet) if and only if a fall is detected, then the images for that event could be accessed (with a password) by the designated persons (e.g., the main care giver or an emergency call centre). Moreover, in some areas (e.g., bathroom), the images could be processed (blurring, pixelization, or silhouette extraction) to ensure some privacy. Finally, it is worth mentioning that in a recent study on the perception of intelligent videomonitoring system by elderly people [33], 96% of participants were favorable or partially favorable to such system for fall detection at home.

Finally, the quality of the images was rather poor here due to large field of view lenses and compression artifacts of low-cost cameras, resulting in noise on segmented pictures. Such noise may lead to errors in the silhouette of moving objects. However, missing part of the silhouette could be considered as partial occlusions that our method is able to overcome. Imperfect segmentation are thus partially compensated by the method, but improvement in the segmentation algorithm would certainly be desirable in the future to compensate for the limitations associated with low-cost cameras. Today's higher end cameras will certainly become more affordable in the future and could also contribute to better performance of the system.

VI. Conclusion

The results presented in this paper had shown that a multi-camera system is reliable in order to detect falls even if some occlusions occur. This result is valid even with only three cameras, but four or more cameras will offer better performances.

This research has led to five contributions:

1) VVDR, a simple and robust feature for fall detection; of occlusion-resistant volumetric reconstruction to fall detection; of a unique dataset that is now documented [26] and made available to the scientific community through a website [27]; analysis of the robustness of the method to occlusion by identifying the worst occlusion case and testing it with experimental videos; and
2) real-time implementation with GPU.

One of the major contributions of this paper is the design of a simple index, the VVDR, which focuses on the change of shape vertical distribution of the subject (from standing up to lying down on the ground). VVDR is robust to some inaccuracies that could occur for a few images (because of too multiple occlusions or segmentation errors). It also means that using a lower frame rate could be considered without affecting the performance of the system, since only the shape distribution at each frame is considered. Hence, we could imagine that a unique system could be used to monitor several rooms at a low frame rate for each camera. In this way, an entire home for autonomous people or multiple resident in a community dwelling could be monitored, thanks to a unique system composed of a network of cameras and only one computer. In this paper, we have also shown that this type of detection process could be real time if necessary by simply using a GPU.

The reconstruction method proposed in this paper could be applied to other types of applications, such as quantifying daily life activities, which is a key issue of our modern society. As for detecting falls, real situations involve many occlusions and classical methods based on multiple cameras (e.g., [13]) generally fail in solving this problem. More generally, this method should be useful for applications involving spatial location and activity classification, depending on shape of subjects. Contrary to approaches mainly based on image analysis, dealing with 3-D volumes in space brings richer information that should be useful to address complex monitoring processes.

References

[1] L. Gillespie, W. Gillespie, M. Robertson, S. Lamb, R. Cumming, and B. Rowe, "Interventions for preventing falls in elderly people," Cochrane Database Syst. Rev., no. 3, 2003.

[2] P. Raina, S. Dukeshire, L. Chambers, and J. Lindsay, "Sensory impairments among canadians 55 years and older: An analysis of 1986 and 1991 health activities limitation survey," McMaster Univ., Hamilton, Canada, Tech. Rep., 1997.

[3] D. L. Hoyert, K. D. Kochanek, and S. L. Murphy, "Deaths: Final data for 1997," *Nat. Vital Statist. Rep.*, vol. 47, no. 19, pp. 1–104, 1999.

[4] S. L. Murphy, "Deaths: Final data for 1998," *Nat. Vital Statist. Rep.*, vol. 48, no. 11, pp. 1–105, 2000.

[5] S. M. Friedman, B. Munoz, S. K. West, G. S. Rubin, and L. P. Fried, "Falls and fear of falling: Which come first? A longitudinal prediction model suggests strategies for primary and secondary prevention," *J. Amer. Geriatric Soc.*, vol. 50, pp. 1329–1335, 2002.

[6] J. J. Hindmarsh and E. H. Estes Jr., "Falls in older persons: Causes and interventions," *Arch. Internal Med.*, vol. 149, no. 10, pp. 2217–2222, 1989.

[7] J. Magaziner, E. M. Simonsick, T. M. Kashner, J. R. LHebel, and J. E. Kenzora, "Predictors of functional recovery one year following hospital discharge for hip fracture: A prospective study," *J. Gerontol.*, vol. 45, no. 3, pp. M101–M107, 1990.

[8] A. K. Bourke, J. V. Brien, and G. M. Lyons, "Evaluation of a threshold-based tri-axial accelerometer fall detection," *Gait Posture*, vol. 26, pp. 194–199. 2007.

[9] M. Kangas, A. Konttila, P. Lindgren, I. Winblad, and T. Jms, "Comparison of low-complexity fall detection algorithms for body attached accelerometers," *Gait Posture*, vol. 28, pp. 285–291, Aug. 2008.

[10] C. Rougier, J. Meunier, A. Saint-Arnaud, and J. Rousseau, "Monocular 3D head tracking to detect falls of elderly people," in *Proc. Conf. IEEE Eng. Med. Biol. Soc. (EMBS)*, Aug. 30–Sep. 3. 2006, pp. 6384–6387.

[11] C. Rougier, J. Meunier, A. Saint-Arnaud, and J. Rousseau, "Procrustes shape analysis for fall detection," presented at the 8th Int. Workshop Vis. Surveillance, Marseille, France, 2008.

[12] C. Rougier, J. Meunier, A. Saint-Arnaud, and J. Rousseau, "Fall detection from human shape and motion history using video surveillance," in *IEEE 1st Int. Workshop Smart Homes Tele-Health*, Niagara Falls, May 2007, pp. 875–880.

[13] J. Tao, M. Turjo, M. F. Wong, M. Wang, and Y.-P. Tan, "Fall incidents detection for intelligent video surveillance," in *Proc. Conf. Inf., Commun. Signal Process.*, Dec. 06–09, 2005, pp. 1590–1594.

[14] 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," *Comput. Vis. Image Understand.*, vol. 113, pp. 80–89, 2009.

[15] R. Cucchiara, A. Prati, and R. Vezzani, "A multi-camera vision system for fall detection and alarm generation," *Expert Syst.*, vol. 24, no. 5, pp. 334–345, 2007.

[16] S.-G. Miaou, P.-H. Shung, and C.-Y. Huang, "A customized human fall detection system using omni-camera images and personal information," in *Proc. 1st Transdisciplinary Conf. Distrib. Diagnosis Home Health Care* (D2H2 '2006), Arlington, USA, Apr. 2–4, 2006, pp. 39–42.

[17] Y. Xinguo, "Approaches and principles of fall detection for elderly and patient," in *Proc. 10th IEEE Int. Conf. e-Health Netw., Appl. Serv.*, 2008, pp. 42–47.

[18] N. Noury, P. Rumeau, A. K. Bourke, G. Laighin, and J. E. Lundy, "A proposal for the classification and evaluation of fall detectors," *Ingénierie et recherche biomédicale (IRBM)*, vol. 29, pp. 340–349, 2008.

[19] G. Wu, "Distinguishing fall activities from normal activities by velocity characteristics," *J. Biomech.*, vol. 33, pp. 1497–1500, 2000.

[20] T. Lee and A. Mihailidis, "An intelligent emergency response system: preliminary development and testing of automated fall detection," *J. Telemed. Telecare*, vol. 11, no. 4, pp. 194–198, 2005.

[21] E. Auvinet, E. Grossmann, C. Rougier, M. Dahmane, and J. Meunier, "Left-luggage detection using homographies and simple heuristics," in *Proc. 9th IEEE Int. Workshop Performance Eval Tracking Surveillance (PETS)*, New York, 2006, pages 51–58.

[22] D. Arsic, M. Hofmann, B. Schuller, and G. Rigoll, "Multi-camera person tracking and left luggage detection applying homographic transformation," in *Proc. 10th IEEE Int. Workshop Performance Eval Tracking Surveillance (PETS)*, Rio de Janeiro, Brazil, Oct. 14, 2007, pp. 1–8.

[23] A. Williams, D. Ganesan, and A. Hanso, "Aging in place: Fall detection and localization in a distributed smart camera network," in *Proc. 15th Int. Conf. Multimedia*, 2007, pp. 892–901.

[24] A. Laurentini, "The visual Hull concept for silhouette-based image understanding," *IEEE Trans. Pattern Annal. Mach. Intell.*, vol. 16, no. 2, pp. 150–162, Feb. 1994.

[25] H. Kim, R. Sakamoto, I. Kitahara, N. Orman, T. Toriyama, and K. Kogure, "Compensated visual hull for defective segmentation and occlusion," in *Proc. Int. Conf. Artif. Reality Telexistence*, 2007, pp. 210–2107.

[26] E. Auvinet, L. Reveret, A. Saint-Arnaud, J. Rousseau, and J. Meunier, "Fall detection using multiple cameras," in *Proc. IEEE 30th Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBS)*, 2008, pp. 2554–2557.

[27] E. Auvinet, C. Rougier, A. Saint-Arnaud, J. Rousseau, and J. Meunier, "Multi camera fall dataset," Université de Montréal, DIRO, Montreal, Quebec, Canada, Tech. Rep. #1350, 2010.

[28] Multi camera fall dataset. (2010). [Online]. Available: http://vision3d.iro.umontreal.ca/fall-dataset/.

[29] J. Y. Bouguet. (2007). "Camera Calibration Toolbox for Matlab," [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc/.

[30] J. Heikkila and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *Proc. Vis. Pattern Recognit.*, 1997, pp. 1106–1112.

[31] M. Piccardi, "Background subtraction techniques: A review," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, 2004, pp. 3099–3104.

[32] R. Motmans and E. Ceriez, "Anthropometry table," Ergonomie RC, Leuven, Belgium, 2005 Available: www.dinbelg.be.

[33] N. Noury, A. Fleury, P. Rumeau, A. Bourke, G. Laighin, V. Rialle, and J. Lundy, "Fall detection: Principles and methods," in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS 2007)*, pp. 1663–1666.

[34] S. T. Londei, J. Rousseau, F. Ducharme, A. Saint-Arnaud, J. Meunier, J. Saint-Arnaud, and F. Giroux, "An intelligent videomonitoring system for fall detection at home: Perceptions of elderly people," *J. Telemed. Telecare*, vol. 15, pp. 383–390, 2009.

[35] NVIDIA Cuda Library. (2010). [Online]. Available: http://www.nvidia.com/object/cuda_home.html.

Edouard Auvinet received M.Eng. degree in signal analysis and telecommunication from the Louis de Broglie Engineering School, Rennes, France, in 2006, and the M.Sc. degree in electronic and signal analysis from Rennes 1 university, Rennes. He is currently working toward the Ph.D. degree in biomedical engineering at the University of Montreal, QC, Canada, and in biomecanic at M2S Laboratory, university of Rennes 2.

His major research interests include signal and picture analysis in biomedical domain. he has been involved with Interscience for bacterial colonies counting with picture of petry box in 2003, with public establishment as a cardiologist at the Institut of Montreal for measuring cholesterol layer in rabbit artery in 2006, with the National Veterinary School of Nantes for measuring and analysing of the field-breathing parameters of horses athlete in 2005, and with Pégase Mayenne for analysis of accelerometric signal of walk movement in 2004.

Franck Multon received the Ph.D. degree in motion control of virtual humans from the Institut National de Recherche en Automatique et Informatique (INRIA), Rennes, France, in 1998.

He is currently a Professor at University Rennes 2, Rennes, where he is involved in research in biomechanics in the M2S Laboratory and in character simulation in Bunraku/INRIA Rennes. His research interests include biomechanics, character simulation, and interaction between real and virtual humans.

Since 1999, he has been an Assistant Professor at University Rennes 2, where he has defended his "authorization to supervise research" in 2006, and is currently a Full Professor since 2008. He the author or coauthor of 20 journal papers and 23 conference papers, and has reviewed papers in several conferences and journals in several domains including computer animation, robotics, virtual reality, biomechanics, neurosciences and anthropology.

Dr. Multon is a member of Association for Computing Machinery's Special Interest Group on Computer Graphics and Interactive Techniques (ACM SIGGRAPH), and the European Society of Biomechanics, and was a member of the international program committee of ACM SIGGRAPH Symposium on Computer Animation (SCA), Computer Animation and Social Agent (CASA), IEEE-Virtual Reality, Conference on Graphics theory and Applications, and ACM SIGGRAPH ACM Symposium on Virtual Reality Software and Technology (VRST).

Alain Saint-Arnaud received the Master's degree in psychology from the Université du Québec à Trois-Rivières (UQTR), QC, Canada, in 1987.

He pursued his training in neuropsychology at the Université de Montréal. He is currently a Clinician at the Health and Social Service Center, Lucille-Teasdale (health and social care system), QC, where he practices in a psychogeriatric team . He is also the Research Coordinator of the psychogeriatric team. His clinical and research interests include the elderly people affected by cognitive and mental health disorders living in the community.

Mr. Saint-Arnaud is a member of the Quebec Rehabilitation Research Network.

Jacqueline Rousseau received the M.Sc. degree and the Ph.D degree in biomedical sciences (rehabilitation) from the Université de Montréal, QC, Canada, in 1992, and 1997.

She practiced as a clinician from 1981 to 1989 mainly into a community integration program. She is currently a Full Professor at the School of Rehabilitation, and a Researcher at the Research Center of Institut universitaire de gériatrie de Montréal, Université de Montréal, QC. Her research interests include home adaptation and community integration for people living with permanent disabilities (motor, cognitive, visual, and mental health) focusing on the development of assessment tools and technology to facilitate their social participation.

Jean Meunier received the B.Sc. degree in physics from the Université de Montréal, Quebec, Canada, in 1981, and the M.Sc.A. degree in applied mathematics, and the Ph.D. in biomedical engineering from Ecole Polytechnique de Montréal, Quebec, in 1983 and 1989, respectively.

In 1989, after being a Postdoctoral fellow at the Montreal Heart Institute, he joined the Department of Computer Science, Université de Montréal, QC, where he is currently a Full Professor and Chair. His research interests include computer vision and its applications to medical imaging and health care.

Dr. Meunier is a regular member of the Biomedical Engineering Institute, Université de Montréal.