TOF Cameras in Ambient Assisted Living Applications

Alessandro Leone and Giovanni Diraco

1 Introduction

In the recent years, the phenomenon of population ageing is receiving increasing attention firstly for healthcare and social impacts (rising health-care costs, life-style changes, etc.) and secondly as an opportunity to leverage the full potential of technology in making automated services for lonely elderly people. In this vision, AAL has been introduced as a term describing solutions based on advanced ICT technologies to support conduct of life. Relevant applications in this field relate, for instance, to the prevention and detection of potential dangerous events such as falls in the elderly, integrated in a wider emergency system which may help in saving lives. On the other hand, many AAL applications, especially in the homecare field, exploit the inference of human activities in order to support the everyday living of elderly people. Applications herein are devoted to support a wide range of needs from specific rehabilitation exercises to better insights into how perform the so called Activities of Daily Living (ADLs), helping geriatricians to evaluate the autonomy level of older adults by employing a variety of electronic aids and sensors [1]. The design of such AAL applications is normally based on paradigms of ambient intelligence and context-awareness providing intelligent environments in which various kind of sensors are deployed. Typical adopted sensors are accelerometers, gyroscopes, video cameras, microphones, pressure switches, and so on. Solutions based on these sensors can roughly be grouped on the basis of their operation modality into three main categories: ambient-based, wearable-based and camera-based solutions [2, 3]. The ambient category relates to those sensors that are embedded into appliances or furniture in order to detect

A. Leone (✉) · G. Diraco
Consiglio Nazionale delle Ricerche—Istituto per la Microelettronica ed i Microsistemi,
Via Monteroni, presso Campus Universitario, Palazzina A3, Lecce, Italy
e-mail: alessandro.leone@le.imm.cnr.it

G. Diraco
e-mail: giovanni.diraco@le.imm.cnr.it

F. Remondino and D. Stoppa (eds.), *TOF Range-Imaging Cameras*,
DOI: 10.1007/978-3-642-27523-4_10, © Springer-Verlag Berlin Heidelberg 2013

presence, door open/close, etc. They require typically an ad hoc design or redesign of the home environment. Wearable devices are based essentially on accelerometer and/or gyroscope sensors. This solution does not require any environmental modification since devices are worn by the user; however wearable devices are prone to be forgotten or worn in a wrong body position, exhibiting a low acceptance rate. Camera-based solutions require the installation of at least one camera in each monitored room allowing the capture of the most of the activities performed and avoiding, at the same time, a large number of ambient-based sensors. Furthermore, apart from being non-invasive camera provides a rich and unique set of information that cannot be obtained from other types of sensors.

Aiming to highlight the benefits of TOF-based RIM in relevant AAL contexts, this chapter focuses on two central AAL scenarios, namely the critical event detection and the analysis of human activities. In particular, the fall detection is considered as the main representative application within the first scenario, whereas the problem of posture recognition is faced within the second one since it is a fundamental prerequisite to all kind of human activity inferences. The main principles of the different approaches are discussed with less of a focus on theoretical details, instead methodologies and their integration into practical implementations are suggested, giving realistic hints on how to handle the main technical issues typical of AAL contexts. The presented methodologies are implemented by using a state-of-the-art TOF range camera and a very compact embedded PC. The design of the suggested system takes into account the ethical aspects in order to maximize the user's acceptance rate and minimize the risk of loss of privacy.

The chapter is organized as follows. In Sect. 2, the active vision is compared with the passive vision and the advantages of the first in AAL contexts are highlighted. A full automated system for detection of falls in the elderly by using TOF vision is presented in Sect. 3, in which both methodological and technical issues are considered. In Sect. 4, the presented framework is extended suggesting a TOF-based solution for human posture recognition well suited for AAL contexts. Finally, the Sect. 5 concludes the chapter by discussing the proposed framework and giving some final considerations.

2 Advantages of Active Vision in AAL Contexts

Generally, the usage of monocular vision for surveillance and monitoring purpose is considerably troublesome since a single camera view can be strongly affected by perspective ambiguity when the viewpoint is unfavorable [4, 5]. The stereo vision (or in general the multiple view vision in which two cameras or more capture the same scene) overcomes perspective problems exploiting 3D geometric representation of human shape. However, stereo/multiple view vision deals with the ill-posed problem of the stereo correspondences strongly affected by poor textured regions and violation of brightness constancy assumption. In addition, the usage of multiple cameras requires both intrinsic and extrinsic camera calibrations that unfortunately are time consuming and error prone activities [6, 7]. Moreover, both monocular and stereo/multiple view vision systems fall within the so called passive vision in which the vision system measures the visible radiation already present in the scene due to natural or artificial illuminations. In general passive vision is well-known to be demoted by many factors such as the presence of shadows, camouflage effects (overlapped regions having similar colors), brightness fluctuations, few surface cues (poor textured regions) and occlusion handling. Recently, the active vision, mainly by using TOF cameras, is increasingly investigated in order to overcome the drawbacks of passive vision systems [8–16]. The manufacture costs of active vision systems in general and TOF cameras in particular are decreasing thanks to a lot of researches in progress especially gained by gaming industry strongly interested in new Natural User Interface: in a near future these devices are likely to be as cheap as webcams are today [11, 17]. Table 1 synthesizes the most important characteristics of TOF sensors in comparison with passive stereo vision systems. The main advantage in the use of TOF is the description of a scene with a more detailed information, since both depth map and intensity image can be used at the same time. In particular, previously mentioned problems of passive vision (foreground camouflage, shadows, partial occlusions, etc.) can be overcome by using depth information that is not affected by illumination conditions and objects appearance. Although the passive stereo vision provides depth information in a less expensive way, this approach presents high computational costs and it fails when the scene is poorly textured or the illumination is insufficient; vice versa, active vision provides depth maps even if appearance information is poor textured and in all illumination conditions [18, 19]. However, it is important to note that both distance and amplitude images delivered by the TOF camera have a number of systematic drawbacks that must be compensated. The main amplitude-related problem comes from the fact that the power of a wave decreases with the square of the distance it covers. For the previous consideration, the light reflected by imaged objects rapidly decreases with the distance between object and camera. In other words, objects with the same reflectance located at different distances from the camera will appear with different amplitudes in the image. Furthermore, in several situations active vision may exhibit unwanted behaviors due to limitations of specific 3D sensing technology (limited depth range due to aliasing, multi path, reflection object properties) [20]. Benefits of TOF sensors in surveillance contexts are summarized in Table 2, whereas drawbacks are reported in Table 3.

In order to understand the advantages in the use of range imaging in surveillance, a qualitative comparison between intensity-based and depth-based segmentation is presented in Fig. 1, using the same well-known segmentation approach (Mixture of Gaussians-based background modelling and Bayesian framework for segmentation) [21]. The two images, intensity and range respectively, are taken by the same TOF camera at the resolution of 176 × 144 pixels. The better segmentation is achieved by using the depth image, whereas the same segmentation approach applied on the intensity image suffers of mimetic effects.

Table 1 Comparison of important characteristics of TOF cameras and stereo vision systems

	TOF sensor	Stereo (passive) vision
Depth resolution	Sub-centimetre (if chromaticity conditions are satisfied)	Sub-millimetre (if images are highly textured)
Spatial resolution	Medium (QCIF, CIF)	High (over 4CIF)
Portability	Dimensions are the same of a normal camera	Two video cameras are needed and also external light source
Computational efforts	On-board FPGA for phase and intensity measurement	High workload (the calibration step and the correspondences search process are hard)
Cost	High for a customizable prototype (1000–3000€)	It depends on the quality of stereo vision system

Table 2 Advantages in the use of TOF sensors in surveillance contexts

	TOF sensor	Passive vision
Illumination conditions	Accurate depth measurement in all illumination conditions	Sensible to illumination variations and artificial lights. Unable to operate in dark environments
Shadows presence	It does not affect principal steps of monitoring applications	Reduced performances in segmentation, recognition, etc
Objects appearance	Camouflage is avoided but appearance could affect depth precision (chromaticity dependence)	Camouflage effects are presented when foreground/background present same appearance properties
Extrinsic calibration	Not needed when only one camera is used	Always needed

Table 3 Drawbacks in the use of TOF sensors in surveillance contexts

	Drawback description
Aliasing	It affects the non-ambiguity range i.e., the maximum achieved depth is reduced (up to 7.5 m)
Multi-path effects	Depth measurement is strongly corrupted when the target surface presents corners
Objects reflection properties	Materials having different colors exhibit dissimilar reflection properties that affect reflected light intensity and, therefore, depth resolution
Field of view	Usually it is limited so that an accurate positioning of the sensor is needed. A pan-tilt architecture could be useful

Moreover, the use of only depth images for measuring allows to improve the preprocessing process and, at the same time, to guarantee the person's privacy since chromatic information is not acquired: only depth measurements are sufficient to detect body movements and postures.

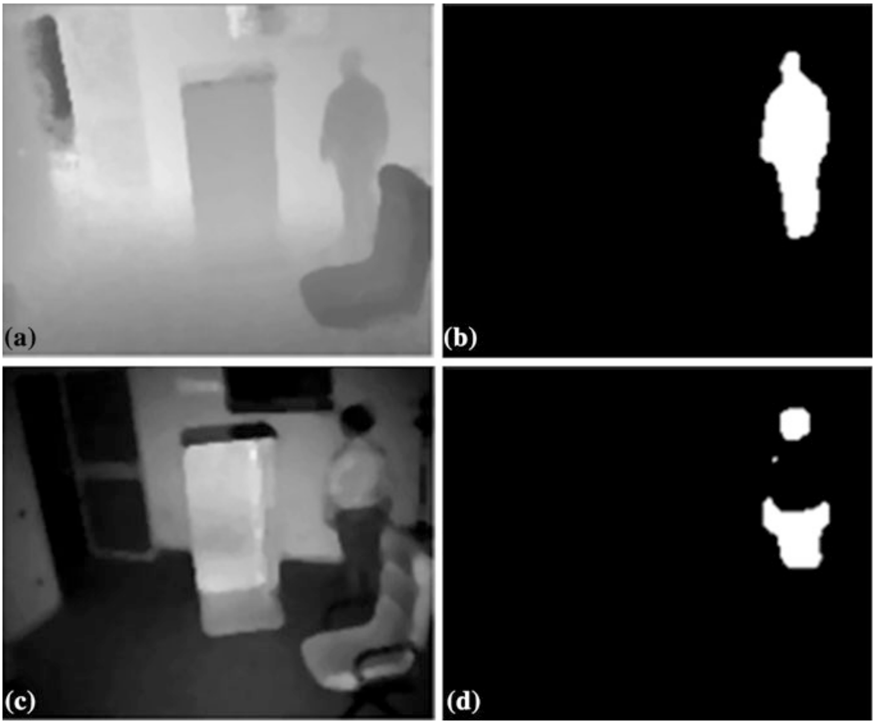


Fig. 1 Segmentation results are shown (**b**, **d**) when the segmentation approach in [39] is applied on depth (**a**) and intensity (**c**) information, respectively. The better segmentation is achieved starting from the depth image, whereas the same segmentation approach applied on the intensity image suffers of mimetic effects (sweater and wall present the same brightness)

3 A TOF Camera-Based Framework for Fall Detection

Actually, the problem of falls in the elderly has become a healthcare priority in all industrialized countries around the world due to the related high social and economic costs [22]. The consequences of falls in elderly may lead to psychological trauma [23], physical injuries [24], hospitalization and death in the worst case [25]. The medical importance of automatic fall-detection is apparent if the two following aspects are taken into account: (a) the involuntarily remaining on the floor for a long period after a fall is related with the morbidity/mortality rate [26]; (b) the elderly may not be able to activate a Personal Emergency Response Systems (PERS) due to the potential loss of consciousness [27]. The most investigated camera-based approach is the monocular one in which a camera alone captures the image frames. A monocular approach was investigated by Shaou-Gang et al. [28], detecting falls by measuring the aspect ratio of the bounding box of the body. Instead, Jansen and Deklerck [29] used depth maps obtained by a stereo camera system to detect inactivity by estimating the orientation change of the body.

A manually calibrated multiple camera approach was used by Cucchiara et al. [30] in order to detect a fall by inspection of the 3D body shape. Since passive vision (both monocular and stereo based) systems suffer of previously discussed drawbacks, recently authors start to investigate the problem of detection of falls by using active vision [12, 16] also in conjunction with other kind of sensors [14]. The suggested methodology for fall detection is discussed in the following subsections starting with the description of the hardware platform.

3.1 The Hardware Platform

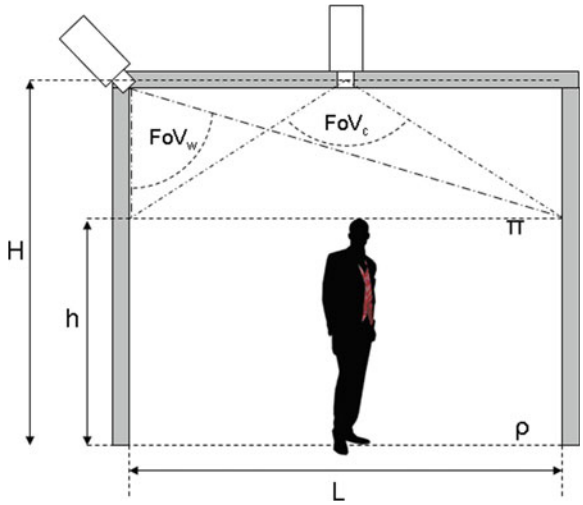
The hardware platform used in the fall-detection framework includes two main components: an embedded PC equipped with an Intel® Atom™ Processor and managed by a Linux-based OS, and the MESA SR3000 [31] TOF camera installed in a wall mounting static setup as discussed in the following subsection. The extrinsic camera calibration is performed in a fully automated way by using a self-calibration procedure (see Sect. 3.2) in order to meet the easy-to-install requirement, whereas the intrinsic calibration is not required since the camera comes intrinsically calibrated by manufacturer.

3.2 Camera Mounting Setup

In this subsection the mounting setup of a TOF camera is discussed. The best camera mounting setup can be defined taking into account the following constraints: (1) the camera is static to limit the computational cost of a pan/tilt handling algorithm; (2) a people height of 1.75 ± 0.20 m is assumed. The two camera mounting configurations investigated were both ceiling and wall mounting setup. The Fields-of-View FoVw (for wall mounting) and FoVc (for ceiling mounting) can be quantitatively compared assuming that the following quantities are given: the covered room length L, the room height H, the average people height h. The two planes ρ and π are considered in order to evaluate the effective camera Field-of-View (Fig. 2): the first plane is referred to the ground floor, whereas the second one is referred to the people head position. In particular FoVw and FoVc are constrained in order to capture the whole π plane. Defined the ratio between the room length L and the distance H–h from the people head position to the ceiling, that is $L' = L/(H–h)$, the FoVw and FoVc are computed by using the following relations:

$$FoV_c(L') = 2\tan^{-1}\left(\frac{L'}{2}\right), \qquad FoV_w(L') = \frac{\pi}{2} - \tan^{-1}\left(\frac{1}{L'}\right). \tag{1}$$

Fig. 2 Two possible camera mounting setups were considered: ceiling mounting and wall mounting



In Fig. 3 FoVc and FoVw are plotted by using Eq. 1 for three typical room having L dimensions of 3, 5 and 7 m. The distance H–h in indoor environments ranges typically from 1 to 2 m, hence a wall mounted camera requires a narrower FoV than a ceiling mounted one. The ceiling mounting configuration is less sensitive to occlusion issues, multiple reflections and flickering effects due to high reflectivity surfaces (windows, mirrors, etc.). On the other hand, in the wall mounting configuration the maximum achievable distance from the camera is greater than that achievable in the ceiling mounting configuration. However, the wall mounting configuration is more sensitive to occlusion problems and spikes may appear in the depth map due to high-reflectivity surfaces. Although ceiling mounting configuration offers many advantages, it does not allow to monitor a wide area, especially when the active sensor is positioned at a limited height from the floor plane. The previous considerations and the narrow FoV typical of TOF cameras motivate the preference of wall mounting setups in AAL contexts.

3.3 Self-Calibration of Extrinsic Parameters

Despite TOF cameras are normally intrinsically calibrated by manufacturers, the external calibration parameters must be estimated. In this section a camera self-calibration algorithm is presented, allowing to achieve a very simple installation process, in agreement with the easy-to-use feature typically required in AAL contexts. The external calibration refers to the estimation of the camera position and orientation (i.e. the camera pose) with respect to a world reference frame fixed at floor level. Both world reference frame (Ow, X, Y, Z) and camera reference frame (Oc, x, y, z) are represented in Fig. 4a in which the camera is accommodated in a wall mounting static configuration at height H from the floor plane.

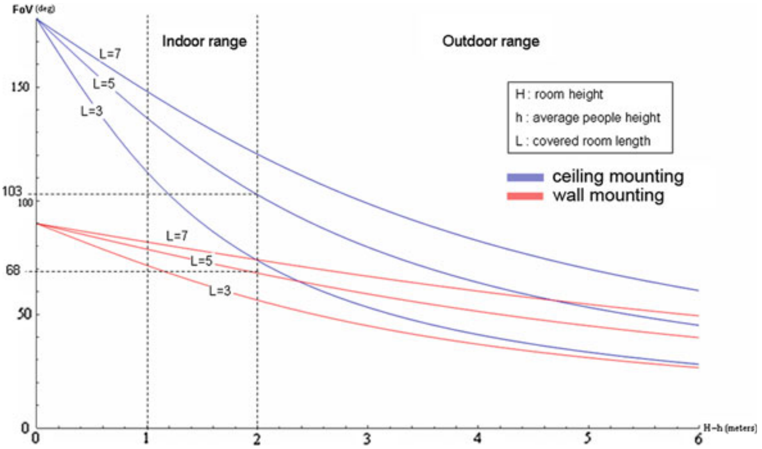


Fig. 3 The fields-of-view FoVc and FoVw are plotted by using Eq. 1 for three typical room dimensions (L=3 meters, L=5 meters and L=7 meters), in function of the distance H-h between person's head and ceiling. In indoor applications the wall mounting setup requires a narrower FoV than the ceiling mounting one

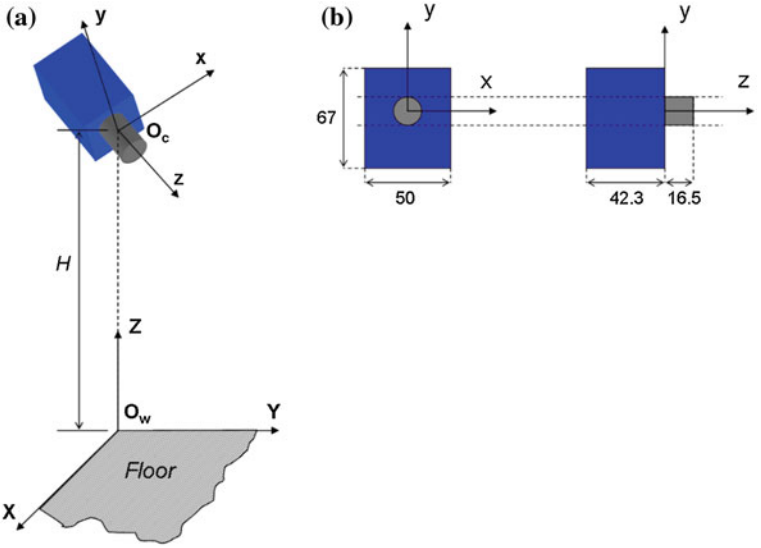


Fig. 4 (**a**) (Ow,X,Y,Z) and (Oc,x,y,z) are world and camera reference frames respectively. The camera is accommodated in a wall-mounting static setup. (**b**) Also camera dimensions are provided

In order to define the camera calibration algorithm, the following assumptions seem to be reasonable for indoor environments:

(A.1) the camera is oriented to capture a relatively large floor plane surface;

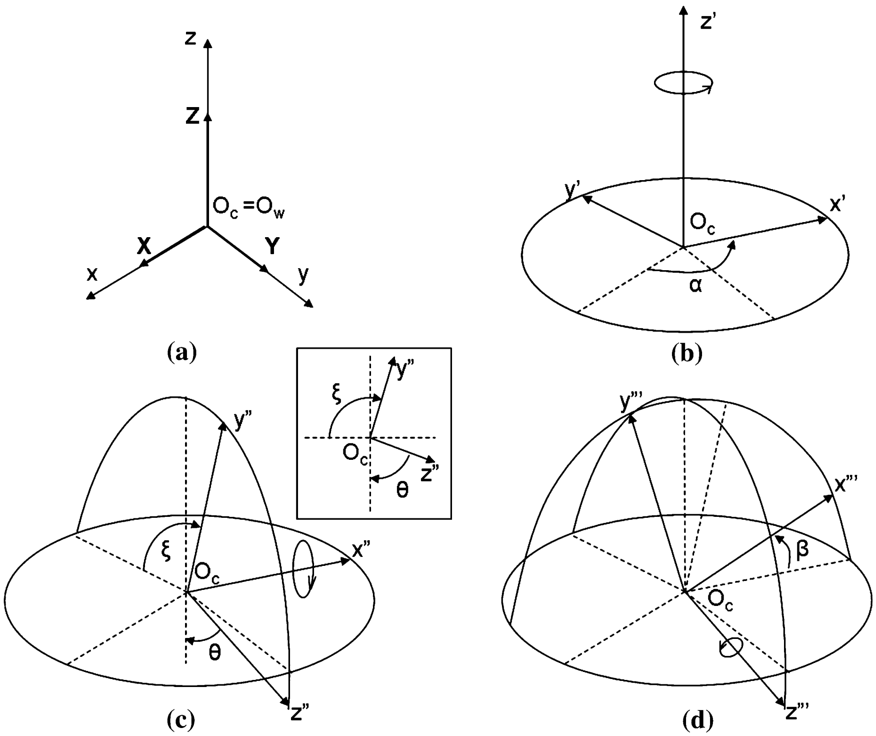


Fig. 5 The camera orientation is defined in terms of Pan (α), Tilt (θ) and Roll (β) angles by using the well known z-x-z convention. Starting from the camera reference frame aligned with the world reference one (**a**), the first rotation is performed around the z-axis of an angle α (**b**); the second rotation around the x-axis of on angle $\xi = \pi\text{-}\theta$ (**c**); and finally the third rotation around the z-axis of a β angle

(A.2) the floor plane could be covered by carpet-like surfaces;
(A.3) the presence of little objects like poufs, boxes, etc., is very limited: the floor is not entirely covered by little objects;
(A.4) the camera could capture other planar surfaces (tables, walls, etc.).

Given the previous A.1, A.2, A.3 and A.4 assumptions, a camera calibration procedure based on floor plane detection is defined. The camera orientation can be defined in terms of pan (α), tilt (θ) and roll (β) angles with respect to a world reference frame as represented in Fig. 5. Following the well known z–x–z convention the camera orientation can be represented as a composition of three rotations, starting from the world coordinated axes (Fig. 5a) and performing: (1) a rotation around the z-axis of α (Fig. 5b), (2) a rotation around the x-axis of $\xi = \pi - \theta$ (Fig. 5c), and finally (3) a rotation around the z-axis of β (Fig. 5d). In homogeneous coordinates the transformation matrix from the camera reference frame into the world reference frame can be written as follows:

$$\mathbf{M} = \begin{pmatrix} \mathbf{R} & \bar{T} \\ \bar{0}^T & 1 \end{pmatrix}, \text{ where } \bar{T} = \begin{pmatrix} 0 \\ 0 \\ H \end{pmatrix} \bar{0} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \text{ and} \tag{2}$$

$$\mathbf{R} = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\xi & -\sin\xi \\ 0 & \sin\xi & \cos\xi \end{pmatrix} \cdot \begin{pmatrix} \cos\beta & -\sin\beta & 0 \\ \sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3}$$

Defining the camera orientation with respect the world reference frame (that is fixed at the floor level) is the same as defining the floor plane orientation in camera coordinates. Hence, the floor plane can be written in camera coordinates by means of the Eq. 2 transforming its normal vector (0, 0, 1, 0) from world homogeneous coordinates into camera ones as follows:

$$\hat{n} = \mathbf{M}^{-1} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \sin\beta\sin\theta \\ \cos\beta\sin\theta \\ -\cos\theta \\ 0 \end{pmatrix}. \tag{4}$$

It is clear from the Eq. 4 that the pan angle α is irrelevant in order to define the floor plane orientation in camera reference frame. At the same conclusion one can reach from Fig. 5b since the first rotation around the z-axis of α does not change the floor plane orientation with respect the camera reference frame. The previous consideration guarantees that the only useful camera calibration parameters are (H, θ, β). In order to detect the floor plane correctly (as it will be detailed below) it is useful to express the camera parameters (H, θ, β) in terms of floor plane coefficients. Given the estimated floor plane π_{F} in camera coordinates:

$$\pi_F\colon a_F x + b_F y + c_F z + d_F = 0 \tag{5}$$

and its normal vector ($\mathrm{a_F}$, $\mathrm{b_F}$, $\mathrm{c_F}$) (it is assumed that $\mathrm{a_F^2} + \mathrm{b_F^2} + \mathrm{c_F^2} = 1$, otherwise it can be normalized), from Eq. 4 the following relations arise:

$$\begin{cases} \sin\beta\sin\theta = a_F \\ \cos\beta\sin\theta = b_F \\ -\cos\theta = c_F \end{cases} \tag{6}$$

By using simple trigonometric considerations the camera parameters can be derived from Eqs. 5 and 6 as follows:

$$\begin{cases} \theta = \arccos(-c_F) \\ \beta = \arcsin\left(\dfrac{a_F}{\sqrt{1 - c_F^2}}\right) \\ H = d_F \end{cases} \tag{7}$$

when: $0 < \theta < \pi$, $-\frac{\pi}{2} \leq \beta \leq \frac{\pi}{2}$, $a_F^2 + b_F^2 + c_F^2 = 1$. Given the Eq. 5 of the estimated floor plane π_{F} and the person's centroid $\vec{C} = \left(c_x, c_y, c_z\right)$ in camera coordinates, the distance of $\vec{C}$ from the floor plane can be evaluated as follows:

$$h\left(\vec{C}\right) = \left|a_F c_x + b_F c_y + c_F c_z + d_F\right|. \tag{8}$$

The estimation of the external calibration parameters (θ, β, H) is accomplished during the installation of the device. Assuming that the camera is adjusted in order to look toward the floor (see A.1), the calibration plane is detected by a three-steps strategy: (1) detection; (2) filtering; (3) selection. The first step deals with the detection of enough large planes in the 3D point cloud, whereas in the second step detected planes are filtered out on the basis of some assumptions on camera orientation (defined by the given below Eq. 11). Finally, the third step selects the floor plane among all filtered planes. The planes detection algorithm searches iteratively the largest plane in the 3D point cloud removing points belonging to the detected plane at each iteration, as explained by the pseudo-code in Algorithm 1.

Algorithm 1 Self-calibration **Algorithm**: Planes detection

```
$S = \{(x_k, y_k, z_k): k=1,\ldots,25344\}$   # 3D point cloud
$N_0 = |S|$                                   # cardinality of S
$L_S = \{\}$
$L_\Pi = \{\}$
i=1
repeat
  $S_i$ is the largest subset in S fitting $\Pi_i$ with Ransac
  $\Pi_i = (a_i, b_i, c_i, d_i)$   # parameters of the Ransac fitted plane
  $S \leftarrow S \setminus S_i$
  $L_S \leftarrow L_S \cup \{S_i\}$
  $L_\Pi \leftarrow L_\Pi \cup \{\Pi_i\}$
  $i \leftarrow i+1$
until $(|S|/N_0 < p)$
```

Hence at a given iteration, the algorithm works with a subset of the 3D points used in the previous iteration. The detection procedure finishes when the size of the subset is lower than a prefixed percentage *p* (greater than 30) of the starting points. Since measured distances are normally affected by noise, planes are detected by using a RANSAC-based approach [32] which is robust to outliers. Let the i-th iteration of the algorithm, the RANSAC plane detector provides four parameters (a_i, b_i, c_i, d_i) describing the implicit model of the i-th fitted plane π_i in camera coordinates:

$$a_i p_x + b_i p_y + c_i p_z + d_i = 0, \tag{9}$$

with $a_i^2 + b_i^2 + c_i^2 = 1$, for each point $P = (p_x, p_y, p_z)$ belonging to the detected plane π_i. For each detected plane π_i the camera tilt and roll angles (θ_i, β_i) are evaluated by using Eqs. 7 and 9:

$$\begin{cases} \theta_i = \arccos(-c_i) \\ \beta_i = \arcsin\left(\dfrac{a_i}{\sqrt{1 - c_i^2}}\right) \end{cases} \tag{10}$$

The (θ_i, β_i) angles are used in the second step to filter out planes not satisfying the following constraints:

$$\begin{cases} -20^\circ \le \beta_i \le 20^\circ \\ 23.75^\circ \le \theta_i \le 66.25^\circ \end{cases} \tag{11}$$

Since not only the floor plane satisfies Eq. 11 but even all coplanar planes, the floor plane is selected as the farthest plane from the camera. Therefore, in the third algorithmic step the floor plane π_F is selected such that the subscript index F is:

$$F = \underset{1 \le i \le m}{\arg\max} \left\{ |d_i| \colon a_i^2 + b_i^2 + c_i^2 = 1 \right\}. \tag{12}$$

The self-calibration procedure was validated by using a MEMS-based Inertial Measurement Unit (IMU) [33] and a Laser Measurement System (LMS) both attached to the 3D range camera in order to derive ground truth data. The IMU sensor provided drift-free 3D orientation with a static accuracy better than 0.5°, whereas the LMS measures distances with accuracy of ± 1.0 mm. The calibration procedure was evaluated in several typical household environments such as living room, kitchen, bedroom, corridor and bathroom, and varying the following parameters:

(P.1) the percentage of floor occupancy by using three groups of objects: (a) carpet-like surfaces with thickness no greater than 5 cm, (b) furniture with height greater than 50 cm (like chairs, beds, nightstands, etc.), and (c) little objects (like poufs) having height ranging from 10 to 30 cm;

(P.2) the camera height from the floor plane, ranging from 2.00 to 2.70 m;

(P.3) the camera orientation β and θ angles, with $-20^\circ \le \beta \le 20^\circ$ and $23.75^\circ \le \theta \le 66.25^\circ$.

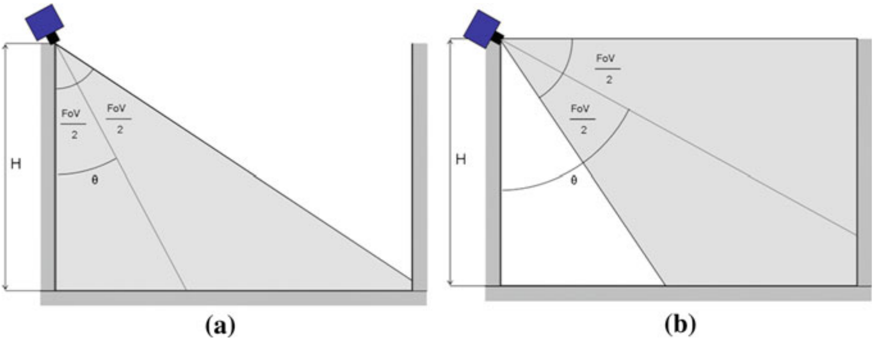


Fig. 6 The geometries involved into the definition of θ lower bound and upper bound are reported in **a**) and **b**) respectively. In order to avoid that camera captures same portion of wall the θ must be greater than FoV/2 that is 23.75°. On the other hand, given that in surveillance applications it is not useful camera captures more than 2.00 m on the opposite wall, the maximum useful value for θ is 90°-FoV/2 that is 66.25°

The camera height range was defined considering that normally the ceiling height is lower than 2.70 m and it is not recommended to install camera at a height lower than 2.00 m (to prevent both safety problem and saturation effects). The range for β seems to be reasonable, since in surveillance application usually one tries to accommodate camera without strong roll rotations. The maximum camera FoV is 47.5°, hence it was not useful to wall mount the camera with tilt angle lower than FoV/2 = 23.75° in order to prevent wall being captured by camera. The geometry involved into definition of θ lower bound is shown in Fig. 6a. Moreover, the maximum value of θ angle was defined considering that normally in surveillance applications it is not useful camera captures more than 2.00 m on the opposite wall and for this reason the maximum useful value for θ is 90°−FoV/2 = 66.25° as shown in Fig. 6b. Camera orientation and position values indicated by the previously mentioned P.2 and P.3 parameters allow to monitor virtually any floor portion inside a typical sized household room of about 4 × 4 m. The values of (H, θ, β) taken into account during the validation of the self-calibration procedure are summarized in Table 4 for a total amount of 324 camera configurations.

Since the measurement of people movements could be demoted by errors in camera calibration, the performance of self-calibration algorithm was evaluated. The calibration procedure was validated considering the relative errors E_H, E_θ and E_β defined as follows:

$$E_H = \frac{|H - \hat{H}|}{H},\ E_\theta = \frac{\left|\theta - \hat{\theta}\right|}{\theta},\ E_\beta = \frac{\left|\beta - \hat{\beta}\right|}{\beta}, \tag{13}$$

Table 4 Camera calibration parameters. Values used during tests

Parameter	Tested values
Height, H (m)	2.00, 2.14, 2.28, 2.42, 2.56, 2.70
Tilt angle, θ (deg)	23.75, 32.25, 40.75, 49.25, 57.75, 66.25
Roll angle, β (deg)	−20.00, −15.00, −10.00, −5.00, 0, 5.00, 10.00, 15.00, 20.00

where H, θ and β are the calibration parameters measured with camera attached IMU and LMS, while $\hat{E}$, $\hat{\theta}$ and $\hat{\beta}$ are the calibration parameters estimated by the self-calibration algorithm. Furthermore, the precision of the calibration procedure was evaluated for different percentage of available floor surface. The environments were arranged in order to obtain several percentage of floor occupancy considering both uncovered and covered floor with carpet surfaces in percentages ranging from 20 to 80 %.

3.4 Background Modeling, People Segmentation and Tracking

All kind of vision-based recognition applications require that some pre-processing tasks are performed before that features are extracted. At this purpose a well-established framework is adopted including early vision algorithms for background modeling, foreground segmentation and people tracking. Since the adopted pre-processing framework is based on well-known concepts, only practical implementation details are given in this subsection omitting further theoretical details. Interested readers in pre-processing aspects can refer to [34] for further details. A Bayesian segmentation is used to detect the 3D elderly silhouette in depth images. In order to perform an automatic foreground extraction, an improved version [35] of the method proposed by Stauffer and Grimson [36] (Mixture of Gaussians—MoGs method) has been enhanced by considering depth information. In the traditional formulation of MoGs method, the probability that a pixel belongs to the foreground is modeled as a sum of K normal functions. For each pixel some of these Gaussians model the background and the others the foreground. In the suggested formulation, at each frame the model parameters are updated by using the Expectation Maximization (EM) algorithm and a pixel is considered to belong to the foreground if its depth does not belong to any of the Gaussians. The EM algorithm allows to update the Gaussian parameters according to a fixed learning rate that controls the adaptation speed. The well-known problem of this approach is the balancing between model convergence speed and stability. The MoGs scheme improves the convergence rate without compromising model stability: the background model is updated online and the global static retention factor of the traditional formulation is replaced with an adaptive learning rate calculated for each Gaussian at every frame. The segmentation involves a binary classification problem based on P(B|z), where z is the depth value at time t and B the background class. According to the background model process, let g(z; μ_k; σ_k) the probability that a particular depth belongs to the k-th Gaussian function G_k having weight ω_k ($P(G_k) = \omega_k$). With an explicit representation of the distribution P(z) as a mixture:

$$P(z) = \sum_{k=1}^{K} P(G_k)P(z|G_k) = \sum_{k=1}^{K} \omega_k \cdot g(z; \mu_k, \sigma_k) \qquad (14)$$

the posterior probability can be expressed in terms of the mixture components $P(G_k)$ and $P(z|G_k)$. Therefore, by using the Bayes rule the density $P(B|G_k)$ can be expressed as:

$$P(B|z) = \sum_{k=1}^{K} P(B|G_k)P(G_k|z) = \frac{\sum_{k=1}^{K} P(z|G_k)P(G_k)P(B|G_k)}{\sum_{k=1}^{K} P(z|G_k)P(G_k)} \qquad (15)$$

To estimate $P(B|G_k)$ a sigmoid function on ω/σ is trained using the logistic regression:

$$\hat{P}(B|G_k) = f\left(\frac{\omega_k}{\sigma_k}; a, b\right) = \frac{1}{1 + e^{-a\frac{\omega_k}{\sigma_k}+b}} \qquad (16)$$

with a = 96 and b = 3, evaluated by training. Once P(z) and $P(B|G_k)$ are estimated, foreground regions are those for which the relation $P(B|z) < 0.5$ is satisfied. The default threshold 0.5 worked quite well with a fitted sigmoid trained on representative data.

The whole segmentation process was implemented in C++ with the support of OpenCv library [37], guarantying real-time functioning. Once person's silhouette has been detected and its centroid (i.e., approximately near the center-of-mass) has been estimated, a tracking strategy allows to link people silhouettes in different time instants. A widely used approach for tracking is the Kalman filter [38] applied to each segmented object. This approach requires a high complexity management system to deal with the multiple hypotheses necessary to track objects. Due to the non-linear nature of human motion, a stochastic approach is used based on the ConDensation scheme (Conditional Density Propagation over time [39]) that is able to perform tracking with multiple hypotheses directly in range images (500 samples are used for people tracking). The 3D centroid is predicted frame-by-frame in range data, according to a state vector defined by merging position and velocity vectors of the centroid. The tracker is realized by thresholding the Euclidean distance between the predicted centroid position and its measured version in the adjacent time step. As discussed for the segmentation step, the ConDensation algorithm implementation in OpenCv library was used allowing to exploit the advantages of a low-level implementation.

3.5 The Fall Detection Strategy

A fall event is detected when the following events happen:

(1) the distance of the person's centre-of-mass (approximated with the silhouette's centroid) with respect the floor plane decreases below to 0.40 m within a time window of about 900 ms;

(2) the people silhouette movements remain negligible within a time window of about 4 s.

The centroid position vector over time $\vec{C}(t) = \left(c_x(t), c_y(t), c_z(t)\right)$ is estimated from range images by using the previously described algorithms for segmentation and tracking. The distance $h(t)$ of the centroid from the floor plane is equal to the z coordinate of the centroid in the world reference frame and hence it can be calculated by using the Eq. 8 as follows:

$$h(t) = \sin\beta\sin\theta \cdot c_x(t) + \cos\beta\sin\theta \cdot c_y(t) - \cos\theta \cdot c_z(t) + H \tag{17}$$

where (θ, β, H) are the previously described calibration. The proposed scheme for fall detection works when a whole human silhouette is detected and also when a partial occlusion occurs. The centroid height estimation was validated by using a test object with known height of 0.40 m and accommodated in nine different positions within a surface of 4 × 4 m as shown in Fig. 7. The height measurements were repeated for each camera position and orientation value reported in the previously discussed Table 4. For each measurement the following quantity was evaluated:

$$\Delta h = \left|\hat{h} - 0.4\right| \tag{18}$$

where $\hat{h}$ is the estimated height. The implemented fall detector is able to process range data real-time (up to 25 fps). Fall-detection performance was evaluated by

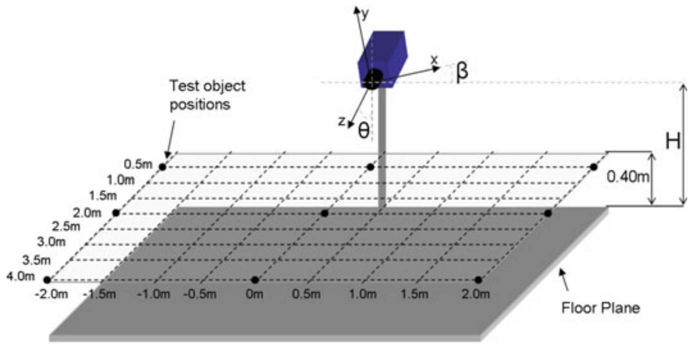


Fig. 7 Accuracy and precision of centroid height estimation by 3D camera was validated by using a test object with known height of 0.40 m and accommodated in 9 different positions within a surface of 4 × 4 m

using data collected during the simulation of falls in real-home scenarios such as living room, kitchen, bed room and bathroom. The performance of the overall system is quantified as suggested by Noury et al. [22] by using sensitivity and specificity measures, defined as follows:

$$\text{Sensitivity} = \text{True Positives} \,/\, (\text{True Positives} + \text{False Negatives}), \quad (19)$$

$$\text{Specificity} = \text{True Negatives} \,/\, (\text{True Negatives} + \text{False Positives}). \quad (20)$$

3.6 The Simulation Setup

The simulation of realistic fall events was performed with the involvement of 13 stuntmen. All participants were healthy male, from 30 to 40 years old and height between 1.55 and 1.95 m. A total amount of 460 actions were simulated of which 260 were falls in all directions (backward, forward and lateral) and with/without recovery post fall. The simulated falls were compliant with those categorized by Noury et al. [40] and they can be grouped into the following seven categories:

(F1) backward fall ending lying (FBRS),
(F2) backward fall ending lying in lateral position (FBRL),
(F3) backward fall with recovery (FBWR),
(F4) forward fall with forward arm protection (FFRA),
(F5) forward fall ending lying flat (FFRS),
(F6) forward fall with recovery (FFWR),
(F7) lateral fall (FL).

Each participant was involved in two sequences simulating ten falls (one for each type from F1 to F6 and four times the type F7) for each sequence. Since the falls in the lateral direction are associated with a high risk of hip fractures in elderly people [41], the simulation of this type of fall (F7) was mainly stressed. Moreover, in order to stress the reliability of the framework, the fall detector was validated in presence of occluding objects; each participant performed at least one half of falls (i.e. five falls in each sequence) occluded by a table, a chair or a sofa.

3.7 Experimental Results

3.7.1 Floor Detection

The precision of the self-calibration procedure was evaluated by repeating the procedure at increasing percentage of available floor both with and without carpet covering. Results without carpet are reported in Table 5.

Table 5 Mean and standard deviation of E_H, E_θ and E_β for percentage of available floor

Floor occupancy	E_H		E_θ		E_β		Δh (m)	
%	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.
20.00	0.4070	0.0302	0.1324	0.0311	0.0232	0.0072	0.3196	0.0456
30.00	0.0209	0.0027	0.0181	0.0035	0.0180	0.0029	0.0502	0.0034
40.00	0.0189	0.0028	0.0167	0.0038	0.0151	0.0030	0.0420	0.0030
50.00	0.0150	0.0028	0.0154	0.0039	0.0128	0.0028	0.0414	0.0028
60.00	0.0142	0.0033	0.0131	0.0041	0.0130	0.0029	0.0228	0.0024

Table 6 Mean and standard deviation of E_H, E_θ and E_β at different percentages of available floor and different percentages of carpet covering

Floor Occ (%)	Carpet Occ (%)	E_H		E_θ		E_β		Δh (m)	
		Mean	Std dev	Mean	Std dev	Mean	Std dev	Mean	Std dev
30.00	20.00	0.3484	0.0298	0.1159	0.0223	0.0219	0.0050	0.1256	0.0085
30.00	40.00	0.3477	0.0308	0.1167	0.0222	0.0232	0.0055	0.1244	0.0068
30.00	60.00	0.3490	0.0305	0.1155	0.0216	0.0213	0.0049	0.1080	0.0056
30.00	80.00	0.3480	0.0300	0.1172	0.0221	0.0227	0.0055	0.0858	0.0045
40.00	20.00	0.0208	0.0028	0.0187	0.0033	0.0176	0.0027	0.0570	0.0033
40.00	40.00	0.0216	0.0030	0.0176	0.0037	0.0175	0.0037	0.0534	0.0049
40.00	60.00	0.0199	0.0018	0.0185	0.0032	0.0182	0.0026	0.0596	0.0038
40.00	80.00	0.0209	0.0016	0.0177	0.0035	0.0172	0.0028	0.0542	0.0028
50.00	20.00	0.0187	0.0025	0.0167	0.0038	0.0147	0.0030	0.0522	0.0026
50.00	40.00	0.0194	0.0033	0.0161	0.0045	0.0154	0.0039	0.0560	0.0021
50.00	60.00	0.0187	0.0031	0.0178	0.0037	0.0153	0.0038	0.0380	0.0007
50.00	80.00	0.0193	0.0029	0.0177	0.0041	0.0140	0.0029	0.0218	0.0009

When the percentage of available floor was greater than 30 % the relative errors E_H, E_θ and E_β were less than 2 % with uncertainty less than 1.23 % (according to the 3σ rule in the theory of errors), whereas the measure inaccuracy of the centroid height was less than 5.0 cm and its uncertainty was less than 10.2 mm. The calibration procedure was evaluated also with carpet-like surfaces covering partially the floor plane. Mean and standard deviation of relative errors reported in Table 6 were estimated in correspondence of variable percentages of available floor surface (not occupied by furniture) and carpet-like surfaces. In the worst case in which many carpets were arranged in various positions and with long pile thickness (near to 5 cm) it was needed a percentage of available planar surface at floor level greater than 40 % in order to relative errors E_H, E_θ and E_β went below the 2 % with an uncertainty less than 1.35 %. In the same situation the measure inaccuracy of the centroid height was less than 6.0 cm with uncertainty less than 14.7 mm.

Some range images used by self-calibration algorithm during data collection in typical dwelling rooms are shown in Fig. 8. Column (a) reports intensity images

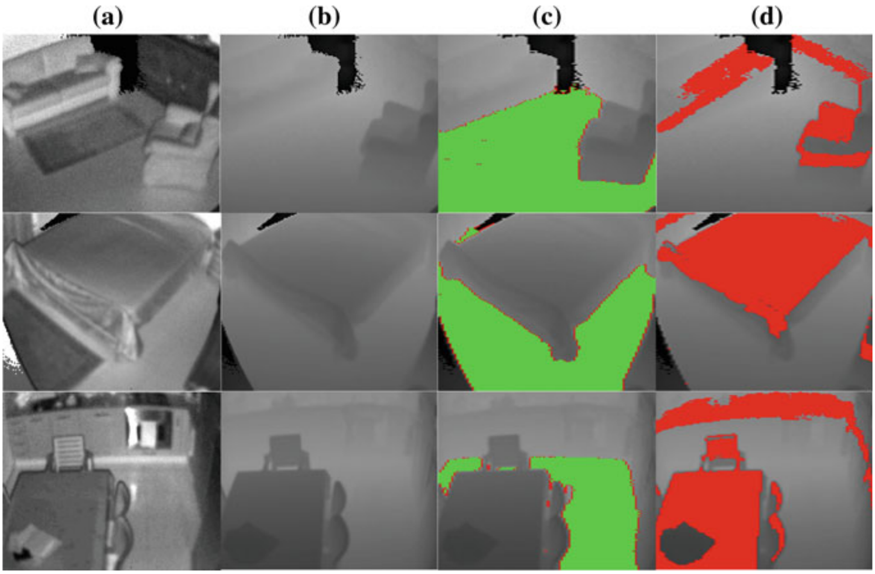


Fig. 8 Some range images used by self-calibration algorithm during data collection in typical dwelling rooms. Figure shows intensity images (column **a**), range images (column **b**), floor planes correctly detected in range images (column **c**) and rejected planes in range images (column **d**)

whereas column (b) reports corresponding range images. Floor planes correctly identified by self-calibration algorithm are shown in column (c), whereas rejected planes are shown in column (d). The first two rows in the figure are referred to rooms with a carpet as it is visible by corresponding intensity images. However, the carpet was not detected by 3D camera since its thickness of about 0.5 cm was lower than the maximum accuracy achievable at the distance of 4 m.

3.7.2 People Detection and Tracking

The position of the people centroid is estimated from the segmented silhouette in the range image and its evolution over time is pursued by using the tracking algorithm detailed in the previous section. Segmentation results are illustrated, as anticipated in the previous section, in Fig. 1 by reporting a critical situation in which passive vision is effected by camouflage effects. Since range camera is not sensitive to illumination or shadows, the elderly silhouette has been accurately segmented from range images (Fig. 1b). In order to emphasize the goodness of range images for segmentation, the same segmentation scheme has been applied to intensity images and the corresponding result is shown (Fig. 1d): the poor quality of the segmentation is due to both the inability of the system to model the background and the presence of camouflage effects (the garment presents chromatic information similar to the wall at the back). The presented results are obtained by using K = 3 Gaussian functions in the background modeling process with $\alpha = 0.005$ as learning coefficient. The time needed for segmentation is about of 15 ms whereas the classification step requires about 10 ms.

3.7.3 Fall Detection

The 3D centroid height profile over time allows to distinct falls from other activities by thresholding the centroid height and unmoving time interval as shown in Fig. 9. In Fig. 10 the typical trend of the centroid height during a fall is reported. During a fall, at least three phases can be distinguished [22]: the pre fall phase (indicated with I0 in Fig. 10), the critical phase (indicated with I1 in Fig. 10), the post fall phase (indicated with I2 in Fig. 10) and the recovery phase (indicated with I3 in Fig. 10). Simulated falls were detected by using three features: (1) the person's centroid height, (2) the critical phase duration, and (3) the post fall phase duration. The three thresholds identified during the analysis of recorded falls are reported in Table 7.

Fall-detection performance was evaluated by using the previously described dataset of simulated falls, with and without the presence of occluding objects such as tables, chairs, sofas, etc. Firstly, results without occlusions will be presented in the following. The first threshold TH1 alone was able to detect correctly all simulated falls achieving a sensitivity of 100 %, although it was not able to distinguish between a fall and a "fall with recovery" or between a fall and a "voluntary lying down on floor". A statistical visualization of results related to the threshold TH1 is shown in Fig. 11. The threshold TH1 alone correctly identified

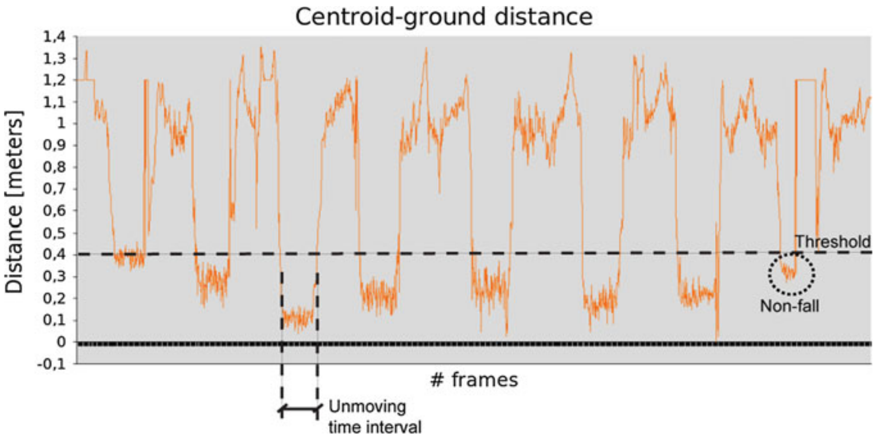


Fig. 9 Centroid height trend analyzed in order to detect falls. The first seven events are correctly classified as fall, whereas the last one is correctly classified as a non-fall since the unmoving duration is too short

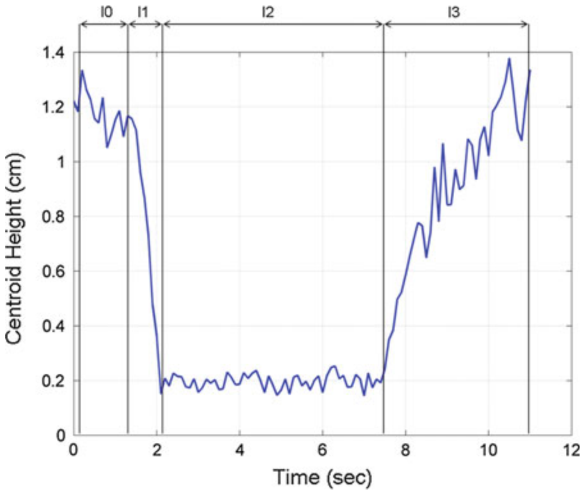


Fig. 10 The typical trend of the centroid height during a fall is shown. During a fall can be distinguished the following phases: the pre fall phase (I0), the critical phase (I1), the post fall phase (I2) and the recovery phase (I3)

63.5 % of ADLs as non-falls achieving a specificity of 63.5 %. By adding the second threshold TH2 a specificity of 79.4 % was obtained, since the threshold TH2 allowed to discriminate correctly a “voluntary lying down on floor” from an involuntary fall characterized by a shorter duration of the critical phase. The statistical visualization of TH2 discrimination capability is shown in Fig. 12. By using the TH1, TH2 and TH3 thresholds simultaneously a specificity of 100 % was achieved, since the threshold TH3 allowed to detect correctly falls with recovery as non-falls by considering the duration of the post fall phase shorter than 4 s in case of recovery. Conversely, in presence of occluding objects it was not possible to detect correctly all simulated falls. Although partial occluded falls happened behind a small object such as a chair were correctly handled, others seriously occluded falls such as those occurred behind a large table were prone to generate false negatives. Similarly, simulated falls with recovery gave rise to false positives due to the impossibility to detect occluded post fall movements. By using the three thresholds TH1, TH2 and TH3 defined in Table 7 a specificity of 97.3 % and a sensitivity of 80.0 % where obtained when falls were occluded by furniture. The previously discussed fall-detection performance is summarized in the following Table 8.

Table 7 Fall-detection threshold values

Threshold	TH1	TH2	TH3
Measure	Centroid height	Critical phase duration	Post fall phase duration
Unit	Meters	Milliseconds	Seconds
Value	0.40	900	4

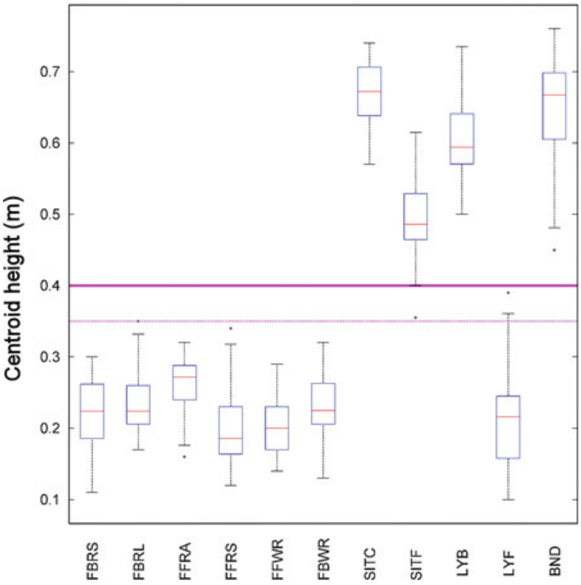


Fig. 11 Statistical visualization with boxplot of minimum centroid height value during falls and ADLs. The threshold TH1 alone correctly identified SITC, SITF, LYB and BND as non-falls, but it was unable to distinguish falls, falls with recovery (FFWR, FBWR) and voluntary "lying down on floor" (LYF)

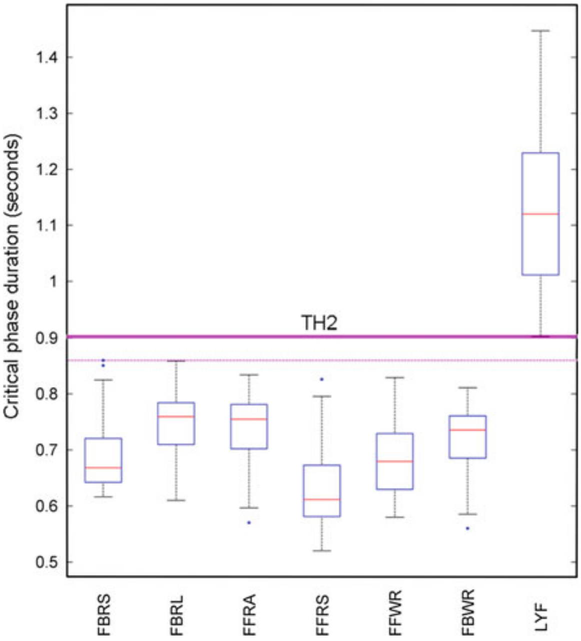


Fig. 12 Statistical visualization with boxplot of critical phase duration during falls. The threshold TH2 allowed to discriminate correctly a voluntary "lying down on floor" (LYF) from an involuntary fall characterized by a shorter duration of the critical phase

Table 8 Fall-detection performance

Thresholds	Sensitivity		Specificity	
	Without occlusions (%)	With partial occlusions	Without occlusions (%)	With partial occlusions
TH1	100	–	63.5	–
TH1, TH2	100	–	79.4	–
TH1, TH2, TH3	100	80.0 %	100	97.3 %

4 A TOF Camera-Based Framework for Posture Recognition

The human posture analysis is a highly active research area in computer vision, dealing with the ill-posed problem of inferring the pose and motion of a highly articulated and self-occluding non-rigid 3D object (a human body) from images. Traditionally, posture analysis algorithms are categorized on whether a body model is employed (either directly or indirectly) or not [42]. Model-based techniques use a priori information about human body shape in order to reconstruct the entire posture kinematics. Within this approach the body is usually represented with a stochastic region model or a stick figure model [4, 43]. Model-based approaches are quite expensive in term of computational resources and they are generally well suited for human motion capture in which the motion of significant segments of the human body must be tracked (i.e. head, arms and legs). On the other hand, model-free techniques estimate body posture directly on individual images without any preliminary information about the body shape and so allowing to overcome limitations of tracking features over long sequences [44, 45]. Within the model-free category two different approaches are investigated in literature: the probabilistic assemblies of parts and the example-based methods. In the first approach individual body parts are first detected and then assembled to infer the body posture [46]; whereas the second approach directly learns the mapping from image space to 3D body space [47]. Concerning the vision system, monocular and stereo/multiple view systems are the most investigated in literature of human posture recognition. Here similar considerations previously made for fall detection apply. Posture estimation from a monocular view is considerably more difficult than estimation from stereo/multiple cameras since a single view image can be strongly affected by perspective ambiguity making troublesome to correctly discriminate posture from unfavorable point-of-view [4, 5]. Furthermore, about one third of all DOFs are almost unobservable by using a monocular camera system [47]. Stereo and multiple view vision systems overcome perspective problems exploiting 3D geometric representation of human shape [48, 49]. However, stereo and multiple view vision are affected by many drawbacks discussed in the previous section concerning vision systems used in fall detection literature. Hence, the adoption of TOF vision is increasingly investigated also for human posture recognition [8–11, 15]. TOF cameras provide dense depth measurements at every point in the scene at high frame rates, allowing to disambiguate poses with similar appearance that can confuse monocular systems or overload stereo/multiple view systems due to the correspondence search between two or more views.

In this section, two different feature extraction approaches are presented, satisfying different requirements exhibited by AAL applications. In fact, gathered posture details and operational distance from the camera are usually inversely proportional: rehabilitation exercises can be performed at few meters from the camera (e.g., less than 3 m when the camera is upper pose on a television screen) and many postural details are required in order to check the correctness of exercise execution; while, conversely, critical events can occur at a greater distance from the camera but few postural detail are sufficient for critical event detection. This motivates the investigation of two feature extraction approaches having different discrimination capabilities in terms of gathered human postural information. The first is a topological approach in which the Morse theory is exploited in order to extract a Reeb graph-based skeleton representation of the human body [15]. A high level of detail can be achieved within a distance from camera up to 4 m. On the other hand, the second feature extraction approach advantages execution speed against details pursing a volumetric strategy based on the analysis of the 3D spatial distribution of the human body [50]. The discrimination capabilities of the two feature extraction approaches are evaluated by using a statistical learning methodology and compared on the basis of a common dataset of four basic human postures: standing, bent, sitting and lying down.

The pre-processing framework (background modeling, foreground segmentation, people tracking, and so on) is the same of those presented in the previous section devoted to detection of falls. Moreover, the same considerations apply for camera mounting setup and the self-calibration procedure. Instead, the TOF camera is the new MESA SR4000. Several improvements put the MESA SR4000 at the state-of-the-art in the field of TOF RIM over the previous MESA SR3000; for example the new camera is full noiseless (0 db), the power consumption has been reduced of about 50 % under normal operation, two highly accurate (manufacturer recommended) non-ambiguity ranges are now available (5.0 m at 30 MHz, and 10 m at 15 MHz) instead of one (7.5 m at 20 MHz), the camera focus is now adjustable in order to obtain accurate 3D data whereas the SR3000 does not have adjustable focus, only to cite a few of them. Further details on cameras comparison can be found in the comparative sheet provided by the manufacturer [31].

4.1 The Experimental Setup

Four main postures, standing, sitting, bent and lying down were simulated during basic Activities of Daily Living (ADLs) involving the interaction with common objects (i.e., tables, sofas, chairs, room/furniture doors, kitchen units, etc.) in order to evaluate the reliability of extracted features in typical home environments. The four main postures are simulated during the following five basic ADLs:

(A1) sitting down on a chair (height, 47 cm) and then stand up (SITC),
(A2) sitting down on floor and then stand up (SITF),
(A3) lying down on a bed (height, 52 cm) and then stand up (LYB),
(A4) lying down on floor and then stand up (LYF),
(A5) bent down to catch something on the floor (BND).

Postures simulation involved only ten subjects among the 13 previously said, ranging in age from 35 to 40 years and height from 1.72 to 1.95. Each subject performed four times every action from A1 to A5 for a total amount of 200 simulated tasks chosen from those actions that mainly might be confused with falls. Other actions in addition to those listed from A1 to A5 were performed such as walking around and dropping objects on floor.

4.2 The Topological Approach

The topological features describe the human posture at a high level of detail; many body segments can be discriminated such as head, trunk, arms and legs, exploiting the full potential offered by range imaging. The intrinsic topology of a generic shape, such as a human body scan captured by a range camera, can be encoded in a graph as suggested by Werghi et al. [51]. A Reeb graph represents the hierarchical evolution of level-set curves on a manifold (that is a mathematical object more general than a classic surface) providing a powerful tool to understand intrinsic topology of any shape [52]. Moreover, defined a real-valued function on a manifold, the Reeb graph nodes represent the level-set curves of the function on the manifold. This function is called Morse function if it has no degenerated critical points on the manifold. Several Morse functions can be defined of which a few are depicted in Fig. 13. The directional height function is shown in Fig. 13a, b along horizontal and oblique directions, respectively. The radial distance function is shown in Fig. 13b and the geodesic distance function in Fig. 13d. For each function the respective level-set curves are highlighted with white colored lines or curves. In recognition applications the main aspect related to the Reeb graph concerns the invariance property under some transformations such as scale and rotation. Among all Morse functions reported in Fig. 13 only the radial distance and the geodesic distance are invariant under affine (translation, scale, rotation) transformations. In addition, the geodesic distance function is invariant under isometric transformations, i.e. those transformations that preserve the length of the path joining two generic points. The isometric invariance is very useful for posture recognition as shown in Fig. 13e in which the path joining the centroid C with the silhouette's left hand remains of the same length after a postural change. Furthermore, geodesic distance function allows to exploit the full potential offered by range imaging since it can be defined on a 3D mesh surface. Otherwise, by using

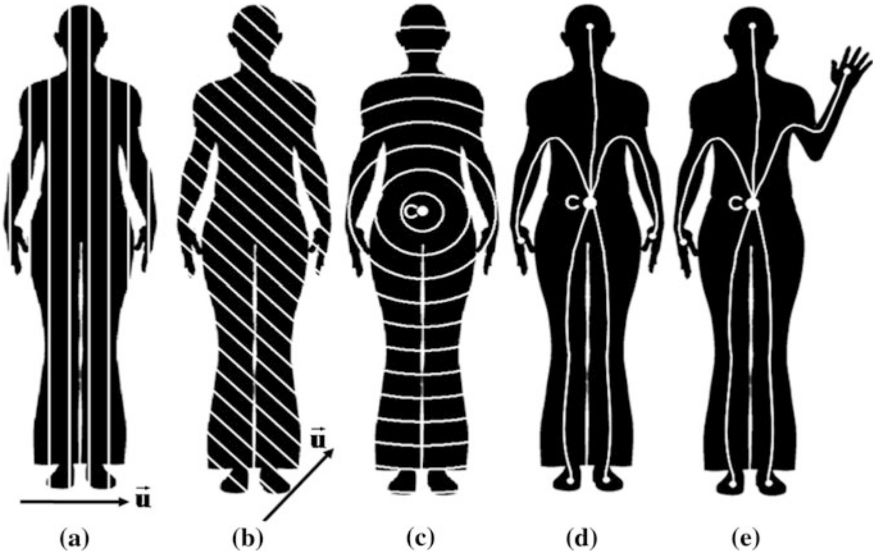


Fig. 13 Three Morse functions: Directional Height Function along the **a**) horizontal and **b**) oblique directions with level-sets depicted in white; **c**) Radial Distance Function with level-sets in white; **d**) Geodesic Distance Function with length paths in white and **e**) the same Geodesic Distance Function under a postural change

monocular passive vision the geodesic distance map is defined on a flatten foreground that can be affected by self-occlusions. This situation is depicted in Fig. 14, in which the geodesic distance map of Fig. 14c is defined on the flatten foreground of Fig. 14b obtained segmenting the original image in Fig. 14a: the mannequin's left hand results confused with the body trunk in the geodesic distance map. On the other hand, by using the range image shown in Fig. 14d a geodesic distance map can be computed without perspective ambiguity as shown in Fig. 14e.

Therefore, in this study the Reeb graph is extracted by using the geodesic distance function as described in the following. Given a range image, the Reeb function is defined as $f = G(x, y)$ with $(x, y) \in I$, where G is the geodesic distance map generated starting from the range image and I is the segmented image region. The geodesic distance map is generated in a two steps procedure: a connected mesh is computed from the range image and then geodesic distances are estimated by using the well-known Dijkstra algorithm on the connected mesh [53]. In Fig. 15c the geodesic map related to depth map of Fig. 15a is reported. Colors represent the distance of each surface point from the starting point (dark blue region): nearest points are blue, farthest ones are red. Whereas, Fig. 15b reports the connected mesh from which the geodesic map is computed.

Hence, starting from the geodesic-based Morse function (i.e., the geodesic map) the Reeb graph is extracted according to the methodology suggested by Werghi et al. [54]. Firstly, the co-domain of the real-valued Morse function f is subdivided in regular intervals as follows:

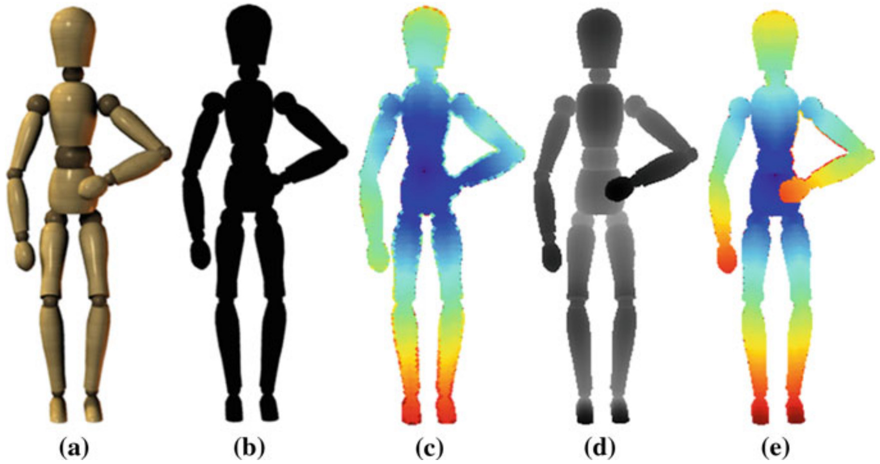


Fig. 14 Starting from the original image **a**), geodesic distance maps in **c**) and **e**) are computed from the flatten foreground **b**) and the range image **d**), respectively. Grayscale levels in the range image **d**) represent the distance of each point from camera: nearest points are dark, farthest are light. Colors in geodesic maps **c**) and **e**) represent the distance of each point from the starting point (dark blue region): nearest points are blue, farthest ones are red

$$z_0 = \min_{(x,y)\in I} f(x,y),\ z_N = \max_{(x,y)\in I} f(x,y),\ z_k = z_0 + k\frac{z_N - z_0}{N},\ \forall k \in \{0,\ldots,N\} \quad (21)$$

Then, $\Re_k$ support regions and S_k level-sets are defined, at the previously fixed intervals, as follows:

$$\Re_k = \{(x,y) \in I | z_k \le f(x,y) < z_{k+1}\},\ S_k = f(\Re_k),\ \forall k \in \{0,\ldots,N\}. \quad (22)$$

The Reeb graph is obtained by associating each level-set S_k to a graph node and linking together two graph nodes when the corresponding support regions are connected. More precisely, two support regions $\Re_k$ and $\Re_i$ are connected if the following condition is satisfied:

$$\exists(x_k,y_k) \in \Re_k, \exists(x_i,y_i) \in \Re_i \ \ni' \ \|P_k - P_i\| \le d, \quad (23)$$

$$P_i = \begin{bmatrix} X(x_i,y_i) \\ Y(x_i,y_i) \\ Z(x_i,y_i) \end{bmatrix},\ P_k = \begin{bmatrix} X(x_k,y_k) \\ Y(x_k,y_k) \\ Z(x_k,y_k) \end{bmatrix} \quad (24)$$

where $\|\cdot\|$ denotes the Euclidean distance between points P_i and P_k, whereas $X(\cdot,\cdot)$, $Y(\cdot,\cdot)$, $Z(\cdot,\cdot)$ are the world coordinates of each range image point indexed by $(x,y) \in I$ and d is a threshold defined according to the maximum distance between connected points and depends on the choice of N in Eq. 21. Figure 15d reports the Reeb graph related to the range image shown in Fig. 15a.

In order to define the feature vector, the Reeb graph is inspected looking for the graph nodes having the greater degree (i.e. the number of edges incident on a

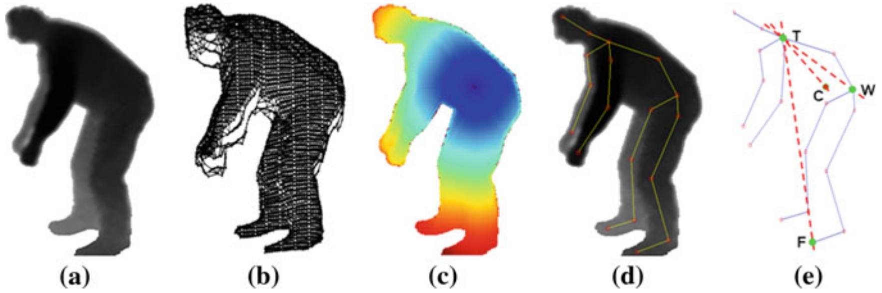


Fig. 15 Topological feature extraction approach. **a**) Original range image. **b**) Connected mesh computed starting from the 3D point cloud. **c**) Geodesic distance map. **d**) Reeb graph-based skeleton superimposed to the original range image. **e**) Line segments measured on skeleton during the feature extraction

node) named *T* and *W* in the Reeb graph shown in Fig. 15e, and the graph node having the minor height indicated as *E* in the same Fig. 15e. Hence, the topological feature vector is defined as follows:

$$\upsilon_T = \left(h_C, \angle\overline{TW}, \angle\overline{TC}, \angle\overline{TF}\right) \tag{25}$$

where h_{C} is the centroid's height with respect the floor plane and $\angle\overline{PQ}$ is the angle of 3D line segment $\overline{PQ}$ with respect the floor plane.

4.3 The Volumetric Approach

In order to discuss the volumetric-based feature extraction process, the 3D point cloud U computed by the range camera is defined as follows:

$$U = \left\{P_i = (X_i, Y_i, Z_i) \in R^3 | i = 1, \ldots, M\right\} \tag{26}$$

where X_{i}, Y_{i}, Z_{i} are the 3D world coordinates of the point P_{i}. The volumetric features exploit global information included into the 3D point cloud by considering two 3D cylindrical volumes V_{UP} and V_{DW} of radius R_i, as shown in Fig. 16, centered on the centroid C of the point cloud U and having world coordinates $C = (X_{\mathrm{C}}, Y_{\mathrm{C}}, h_{\mathrm{C}})$ in which $h_{\mathrm{C}} > 0$ is the centroid height with respect the floor plane. Given the following subdivision of the cylinder's ray in regular intervals, $\forall k \in \{0, \ldots, N\}$:

$$R_0 = \min_{i \in \{1,\ldots,M\}} \|(X_i, Y_i) - (X_C, Y_C)\|,\ R_N = \max_{i \in \{1,\ldots,M\}} \|(X_i, Y_i) - (X_C, Y_C)\|,\ R_k = R_0 + k\frac{R_N - R_0}{N} \tag{27}$$

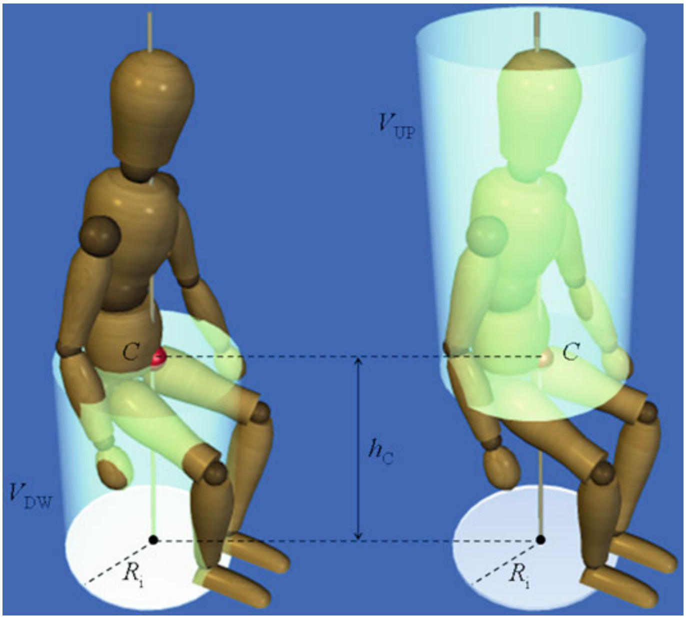


Fig. 16 3D cylindrical volumes used during the volumetric feature extraction

the total amount of points included in each cylinder is given by the following functions:

$$F_{UP}(k) = |\{i \in \{1,\ldots,M\} | \|(X_i,Y_i) - (X_C,Y_C)\| \leq R_k \wedge Z_i > h_C\}| \tag{28}$$

$$F_{DW}(k) = |\{i \in \{1,\ldots,M\} | \|(X_i,Y_i) - (X_C,Y_C)\| \leq R_k \wedge Z_i < h_C\}| \tag{29}$$

where |·| denotes the cardinality of a set. The volumetric feature vector can now be defined as follows:

$$\upsilon_V = \left(h_c,\, F_{UP}(N) - F_{DW}(N) \max_{1 \leq k < N} \Delta F_{UP}(k),\, \max_{1 \leq k < N} \Delta F_{DW}(k)\right) \tag{30}$$

where the operator Δ is the discrete derivative defined as follows:

$$\Delta F(k) \doteq F(k+1) - F(k) \tag{31}$$

In Fig. 17 the two functions defined by Eqs. 28 and 29 are plotted in correspondence of a 3D point cloud sampled for each main posture. The feature vector defined by Eq. 30 allows to keep very low the computational complexity of the feature extraction process although it is sufficient to discriminate reliably the four main postures since the spatial distribution of the 3D point cloud is dependent on the particular posture. Moreover, the computational simplicity is paid in terms of achievable level of detail in posture discrimination. Indeed, the feature vector υ_V is

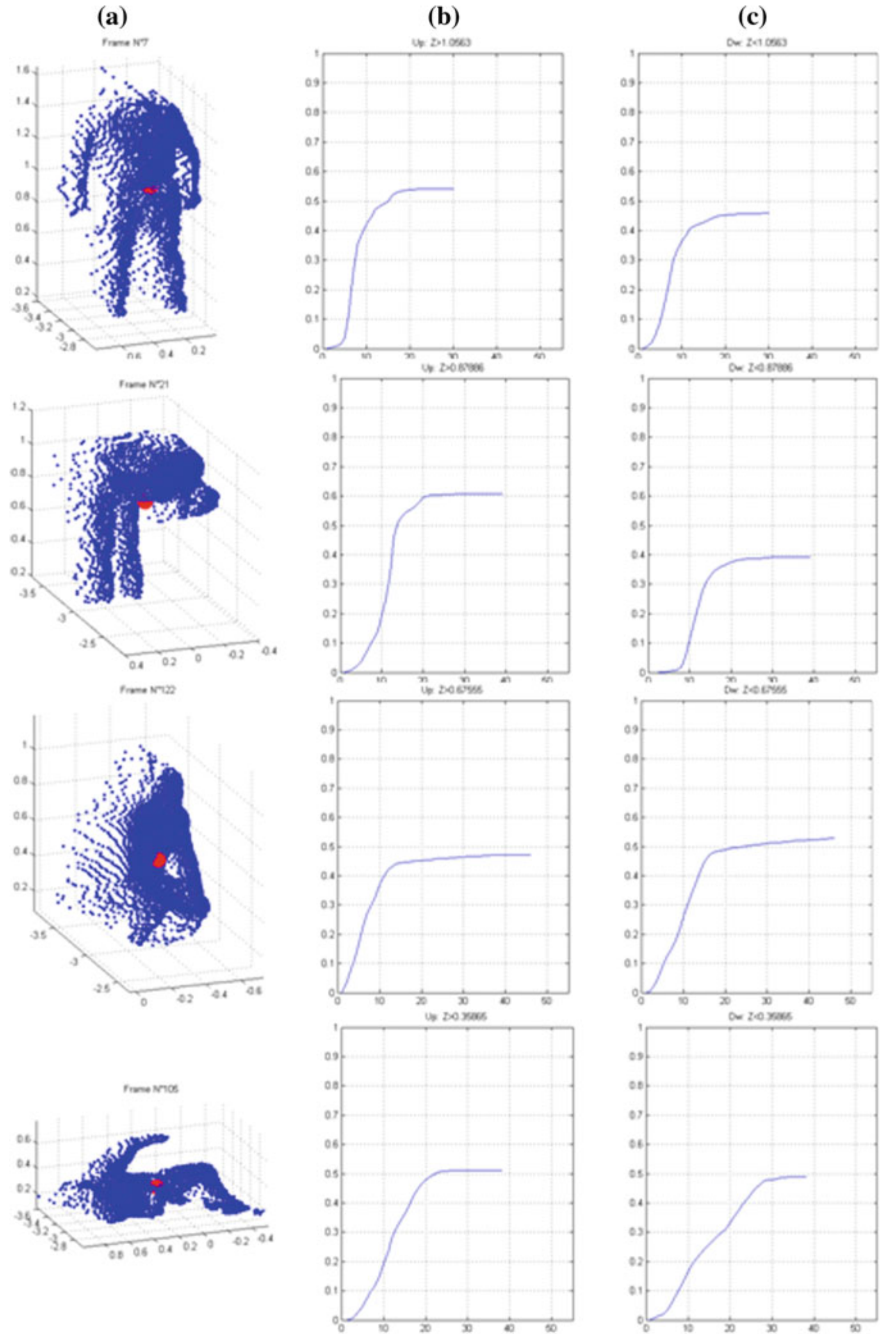


Fig. 17 Plotting of cylindrical volume cardinalities at the varying of ray values. **a**) 3D point clouds of the four main postures (from up to bottom): Standing, Bent, Sitting, Lying. **b**) Plots of the FUP function in correspondence of each posture. **c**) Plots of the FDW function in correspondence of each posture

unable to account for the position of body's segments like the υ_T feature vector does. However, the choice of the feature vector depends on the specific AAL application. If the positions of arms and legs are relevant (e.g., during the monitoring of rehabilitation exercises) then the topological approach is necessary; if it is needed to detect ADLs then the volumetric approach is sufficient (for fall-detection can be sufficient even the only h_C as it is discussed in the previous section).

4.4 Experimental Results

A good generalization ability during classification is definitely relevant since postures are not perfectly repeatable, the acquisition viewpoint varies in function of subject's position and some level of variation in range data is expected due to noise effects. This motivated the choice of a multi-class SVM classifier in conjunction with affine/isometric invariant features in order to discriminate the four main postures. Based on the principle of risk minimization, Support Vector Machines (SVMs) outperform other classifiers in terms of good performance in resolving non-linear and high dimensional problems with limited samples (high generalization ability). Moreover, SVMs try to find discriminative hyper-planes that maximize the margin between the classes overcoming, in a more natural way, the problem of over-fitting [54, 55]. The binary nature of SVM is adapted to the multi-class nature of the posture classification problem by using a one-against-one strategy. Since good results are documented in scientific literature related to posture recognition, a Radial Basis Function (RBF) kernel is used [56] and the associated parameters, namely regularization constant K and the kernel argument γ, are tuned according to a grid search procedure.

The best classification rates is experimentally found with the optimal parameters $(K;\gamma) = (1;32)$ for the topological approach and $(K;\gamma) = (1;64)$ for the volumetric one. A large dataset of 1200 samples, 300 for each posture, is collected in order to evaluate the classification performance. Postures are taken at various distances from the camera, ranging from 2.5 to 5 m. Confusion matrices are reported in Fig. 18 for both topological and volumetric features at distances of 2.5 and 5 m, whereas classification rates are reported in Fig. 19 for all intermediate distance values. As it is shown by reported results, topological features exhibit the best classification rate up to 3 m, whereas for distances greater than 3 m results are comparable with those of volumetric features. The training phase is done by using 200 samples, 50 samples for each posture, taken from only one viewpoint (the frontal view) in order to evaluate the generalization performance of the classification. Instead, the test phase is done by using the remaining 1000 samples taken from various viewpoints turning around the subject. The good generalization performance can be inferred by the reduced number of training samples adopted as support vectors, indeed about 40 % of the training set is used as support vectors. Results show the suggested features, both topological and volumetric, are suitable

(a)

	ST	BE	SI	LY
ST	1.000	0.002	0.000	0.000
BE	0.000	0.993	0.013	0.000
SI	0.000	0.005	0.983	0.006
LY	0.000	0.000	0.004	0.994

(b)

	ST	BE	SI	LY
ST	0.972	0.011	0.000	0.000
BE	0.019	0.967	0.026	0.000
SI	0.009	0.022	0.966	0.013
LY	0.000	0.000	0.009	0.987

(c)

	ST	BE	SI	LY
ST	0.988	0.009	0.000	0.000
BE	0.007	0.976	0.015	0.002
SI	0.005	0.014	0.982	0.009
LY	0.000	0.000	0.004	0.989

(d)

	ST	BE	SI	LY
ST	0.970	0.011	0.000	0.000
BE	0.018	0.972	0.025	0.003
SI	0.012	0.017	0.969	0.011
LY	0.000	0.000	0.006	0.986

Fig. 18 Confusion matrices for topological features **a**) at 2.5 meters and **c**) 5 meters. Confusion matri-ces for volumetric features **b**) at 2.5 meters and **d**) 5 meters

to exploit the full potential of range imaging most notably if used in conjunction with a classifier having good generalization capabilities (like SVM). Both topological and volumetric approaches (feature extraction and classification module) have been implemented on the embedded PC in c/c++ language achieving execution speed compliant with monitoring and surveillance purpose. When topological features were used the system worked at 5 fps with 87 % of execution time devoted to the feature extraction process. Instead, the system worked at 15 fps by using the volumetric features with an execution time of 60 % taken by the feature extraction process.

5 Discussion and Conclusion

The usage of TOF vision allows to solve some of the classic issues in background modeling and people segmentation, since depth information are not sensitive to illumination or shadows and can be used to detect more easily occlusions by exploiting the depth gap between people and occluding objects. Results shows the goodness of the proposed methods in real-time implementation and real AAL applications. The TOF camera experimented in this study is a state-of-the-art technology characterized by a very low noise and medium pixel resolution. Moreover, in order to keep this study as more general as possible, during data collection the TOF camera was set to a low integration time of 6 ms achieving so a noise level comparable with that of cheap cameras. The used camera is very compact and exploiting the proposed self-calibration algorithm it

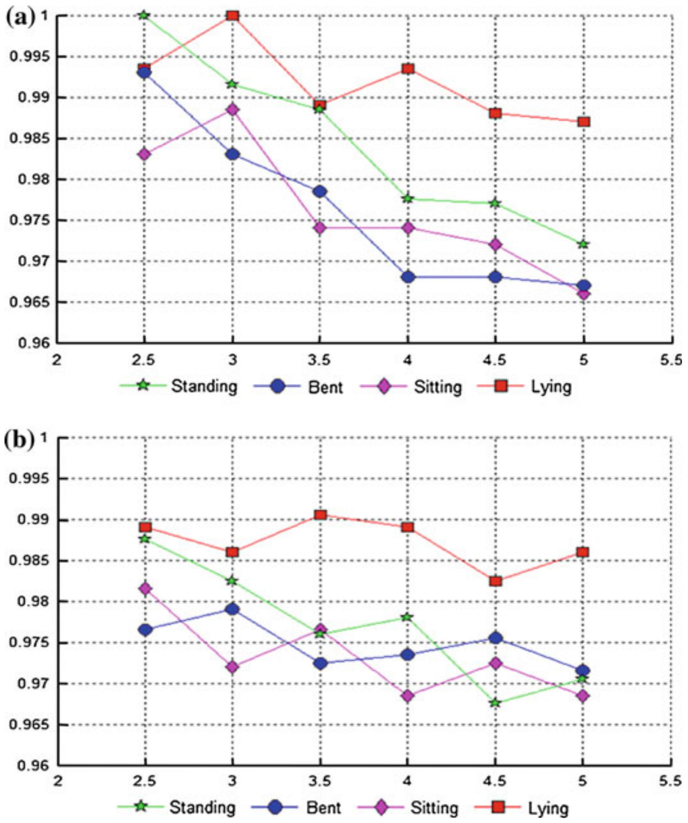


Fig. 19 Classification rates at varying of camera distance from 2.5 to 5 meters for both **a**) topological and **b**) volumetric features

can be installed simply without particular requirements or constraints. The suggested self-calibration procedure proved to be well suited for AAL application since it allowed to calibrate camera effectively without requirement of special calibration objects or user intervention but using only an automatic detection of floor plane that appeared to be always sufficiently visible (greater than 60 %) during falls recording in real dwelling rooms such as living room, kitchen, bed room, corridor and bathroom. The performance of the self-calibration algorithm was related to the amount of 3D points of the scene belonging to the floor plane. Better calibration estimations was obtained when at least the 30 % of the captured points belonged to the floor plane without floor covering and 40 % with carpet covering.

The presented fall-detector exploits the centroid height trend in order to detect fall events, thus the measurement precision is more important than accuracy. In other words, the systematic error does not affect fall-detection performance. The maximum estimated uncertainty in height measurement was less than 1.47 cm when camera was calibrated with floor plane variously covered by carpet and when at least 40 % of the captured points belonging to the floor plane. Moreover, between the TH1 threshold and the largest centroid peak value (35 cm) among all fall events was a difference of 5 cm (see dashed line in Fig. 11) that was definitely grater than the maximum uncertainty of 1.47 cm due to self-calibration procedure. The effect of the height measurement uncertainty on TH2 threshold was also evaluated. Taking into account all critical phase durations recorded during falls, the uncertainty in height measurement of 1.47 cm leaded to an uncertainty in the critical phase duration measurement less than 26 ms that was lower than the difference between the threshold TH2 and the maximum critical phase duration (860 ms) recorded during falls that was of 40 ms (see dashed line in Fig. 12). Since the third threshold TH3 was set to a very large time duration (at least of 4 s), the achieved precision did not play a critical role even in this case. Thus, the proposed threshold levels TH1, TH2 and TH3 provided a sufficient margin for successful detection of falls with respect to the height measurement precision. The three threshold in conjunction were able to detect all simulated falls without misdetection of ADLs as falls and vice versa, providing a 100 % of sensitivity and 100 % of specificity when occluding objects did not obstruct the camera's view. The tracking prediction mechanism allowed to estimate correctly the centroid height trend during simulated falls when some silhouette's portion was visible during critical and post fall phases (refer to Fig. 10). Fully occluded person movements during critical phase or during post fall phase gave rise to misdetections due to the impossibility to distinguish between a fall and a voluntary "lying down on floor and then stand up" (LYF) (critical phase occluded and post fall phase visible) or between a fall and a "fall with recovery" (FBWR or FFWR) (critical phase visible and post fall phase occluded). Instead, partial occlusions are correctly detected by evaluating the distance of the lower part of the segmented silhouette from the floor plane according to Eq. 11, allowing to adjust the position estimated by the particle filter. Thus, previously said misdetections demoted performance in presence of occluding objects leading to 97.3 % specificity and 80.0 % sensitivity. Segmentation and classification activities require about 25 ms per frame that seems reasonable since the minimum duration of the critical phase is about of 500 ms as indicate by Noury et al. [22]. Thus, a frame rate of 8 fps is fast enough for fall-detection purpose leaving available processing resources to be located for multiple 3D camera monitoring in order to deal with occlusions and limited FoV. It could be considered a limitation of presented studies that the trend of person's centroid height was determined from simulated falls in subjects with height greater than 1.55 m. However, the TH1 threshold not should be an issue for persons with height lower than 1.55 m, since they have a centroid height on average lower than taller persons. The TH2 threshold measured the critical phase duration. For persons with height lower than 1.55 m the critical phase duration could be shorter than that for person with higher height, thus TH2 should work correctly even for lower height. The TH3 threshold not should be an issue in any case since it works when the centroid height stays below the TH2 threshold. Results have been shown the feasibility to detect falls by using TOF camera highlighting related strength and weakness. The proposed fall-detector shows good performance also compared with other studies. In absence of occlusions performance is very similar to fall-detection system proposed by Bourke and Lyons [57] based on a bi-axial gyroscope sensor.

Other than for detection of falls, the capabilities of active vision have been demonstrated also for posture recognition in AAL contexts. Two feature extraction approaches, topological and volumetric, for the classification of four main postures (standing, bent, sitting and lying down) have been presented. The discrimination capabilities of the two feature extraction approaches are evaluated by using a machine learning approach and compared on the basis of a common dataset of simulated postures during simple ADLs. The different discrimination capabilities and execution speeds offered by the two approaches allow to satisfy different requirements exhibited by AAL applications. In fact, gathered posture details and operational distance from the camera are usually inversely proportional. For instance, rehabilitation exercises can be performed at few meters from the camera (e.g., less than 3 m) and many postural details are required in order to check the correctness of exercise execution, whereas critical events can occur at a greater distance from the camera (more than 3 m) but few postural detail are usually sufficient for detection of critical events. The topological features describe the human posture at a high level of detail exploiting the full potential offered by range imaging: many body segments can be discriminates such as head, trunk, arms and legs. As it is shown by reported results, topological features exhibit the best classification rate up to 3 m, whereas for distances greater than 3 m results are comparable with those of volumetric features. However, the high level of postural detail achieved with the topological features is paid in terms of computational workload (up to 5 fps). Volumetric features reflecting the spatial distribution of 3D point cloud provide a lower level of detail in posture discrimination, but they have the advantage to be less computationally expensive (up to 15 fps). The choice for one or the other depends on the specific AAL application. The results suggest high accuracy of topological features at distances up to 3 m, whereas beyond volumetric and topological approaches give similar classification performance (greater than 96.5 % in both cases).

References

1. N. Foldi, L. Kaplan, J. Ly, O. Nikelshpur, M. Lucy, B. Jeffrey, ADL functions and relationship to cognitive status. Alzheimer's Dement. J. Alzheimer's Assoc. **7**(4), S244 (2011)
2. N. Shah, M. Kapuria, K. Newman, in *Embedded activity monitoring methods*. Activity Recognition in Pervasive Intelligent Environments, vol 4(13) (Atlantis Press, Amsterdam, 2011), pp. 291–311
3. B. Kröse, T. Oosterhout, T. Kasteren, *Activity Monitoring Systems in Health Care*, in Computer Analysis of Human Behavior, vol. 12 (Springer, London, 2011), pp. 325–346

4. A. Agarwal, B. Triggs, Recovering 3D human pose from monocular images. IEEE Trans. PAMI **28**(1), 44–58 (2006)
5. A. Fossati, M. Dimitrijevic, V. Lepetit, P. Fua, From canonical poses to 3D motion capture using a single camera. IEEE Trans. PAMI **32**(7), 1165–1181 (2010)
6. S.N. Lim, A. Mittal, L.S. Davis, N. Paragios, Fast illumination invariant background subtraction using two views: error analysis, sensor placement and applications. Proc. Comput. Vis. Pattern Recogn. **1**, 1071–1078 (2005)
7. R.I. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. (Cambridge University Press, Cambridge, 2004)
8. V. Ganapathi, C. Plagemann, S. Thrun, D. Koller, Real time motion capture using a single Time-Of-Flight camera, in *Proceedings of CVPR*, pp. 755–762 2010
9. W. Li, Z. Zhang, Z. Liu, Action recognition based on a bag of 3D points, in *Proceedings of CVPRW*, pp. 9–14 2010
10. C. Plagemann, V. Ganapathi, D. Koller, S. Thrun, Real-time identification and localization of body parts from depth images, in *Proceedings of ICRA*, pp. 3108–3113 2010
11. S. Oprisescu, C. Burlacu, V. Buzuloiu, Action recognition using time of flight cameras, in *Proceedings of COMM*, pp. 153–156 2010
12. C. Rougier, E. Auvinet, J. Rousseau, M. Mignotte, J. Meunier, Fall detection from depth map video sequences. Lect. Notes Comput. Sci. **6719**(2011), 121–128 (2011)
13. D. Falie, M. Ichim, Sleep monitoring and sleep apnea event detection using a 3D camera, in *Proceedings of 8th IEEE International Conference on Communications (COMM)*, pp. 177–180 2010
14. M. Grassi, A. Lombardi, G. Rescio, M. Ferri, P. Malcovati, A. Leone, G. Diraco, P. Siciliano, M. Malfatti, L. Gonzo, An integrated system for people fall-detection with data fusion capabilities based on 3D TOF camera and wireless accelerometer, in *Proceedings of IEEE Sensors*, pp. 1016–1019 2010
15. G. Diraco, A. Leone, P. Siciliano, Geodesic-based human posture analysis by using a single 3D TOF camera, in *Proceedings of IEEE International Symposium on Industrial Electronics (ISIE)*, pp. 1329–1334 2011
16. A. Leone, G. Diraco, P. Siciliano, Detecting falls with 3D range camera in ambient assisted living applications: a preliminary study. Med. Eng. Phys. J. **33**(6), 770–781 (2011)
17. www.xbox.com/Kinect
18. A. Kolb, E. Barth, R. Koch, TOF-sensors: new dimensions for realism and interactivity, in *Proceedings of Computer Vision and Pattern Recognition Workshops*, pp. 1–6 2008
19. S. Foix, G. Alenyà, C. Torras, Lock-in Time-Of-Flight (TOF) cameras: a survey. IEEE Sens. J. **11**(9), 1917–1926 (2011)
20. S.A. Guomundsson, H. Aanaes, R. Larsen, Environmental effects on measurement uncertainties of Time-Of-Flight cameras. Proc. IEEE ISSCS **1**, 1–4 (2007)
21. D. Lee, Effective Gaussian mixture learning for video background subtraction. IEEE Trans. Pattern Anal. Mach. Intell. **27**(5), 827–832 (2005)
22. N. Noury, P. Rumeau, A.K. Bourke, G. ÓLaighin, J.E. Lundy, A proposal for the classification and evaluation of fall detectors. J. IRBM **29**(6), 340–349 (2008)
23. M.C. Chung, K.J. McKee, C. Austin, H. Barkby, H. Brown, S. Cash, J. Ellingford, L. Hanger, T. Pais, Posttraumatic stress disorder in older people after a fall. Int. J. Geriatr. Psychiatry **24**(9), 955–964 (2009)
24. S. Sadigh, A. Reimers, R. Andersson, L. Laflamme, Falls and fall-related injuries among the elderly: a survey of residential-care facilities in a Swedish municipality. J. Commun. Health **29**, 129–140 (2004)
25. A. Shumway-Cook, M.A. Ciol, J. Hoffman, B.J. Dudgeon, K. Yorkston, L. Chan, Falls in the medicare population: incidence, associated factors, and impact on health care. J. Phys. Ther. **89**(4), 324–332 (2009)
26. S.R. Lord, C. Sherrington, H.B. Menz, *Falls in Older People. Risk Factors and Strategies for Prevention* (Cambridge University Press, Cambridge, 2007)

27. S. Elliott, J. Painter, S. Hudson, Living alone and fall risk factors in community-dwelling middle age and older adults. J. Commun. Health **34**, 301–310 (2009)
28. M. Shaou-Gang, S. Fu-Chiau, H. Chia-Yuan, A smart vision-based human fall detection system for telehealth applications, in *Proceedings of the 3rd Telehealth Conference*, pp. 7–12 2007
29. B. Jansen, R. Deklerck, Context aware inactivity recognition for visual fall detection, in *Proceedings of IEEE Pervasive Health Conference*, pp. 1–4 2006
30. R. Cucchiara, A. Prati, R. Vezzani, A multi-camera vision system for fall detection and alarm generation. Expert Syst. J. **24**(5), 334–345 (2007)
31. http://www.mesa-imaging.ch
32. M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6), 381–395 (1981)
33. http://www.xsens.com
34. C. Fabien, B. Deepayan, A. Charith, S.H. Mark, Video based technology for ambient assisted living: a review of the literature. Environments **3**, 253–269 (2011). IOS Press
35. D.S. Lee, Effective Gaussian mixture learning for video background subtraction. IEEE Trans. Pattern Anal. Mach. Intell. **27**(5), 827–832 (2005)
36. C. Stauffer, W. Grimson, Adaptive background mixture models for real-time tracking, in *Proceedings of IEEE Computer Vision and Pattern Recognition Conference*, pp. 246–252 1999
37. http://sourceforge.net/projects/opencvlibrary
38. R.E. Kalman, A new approach to linear filtering and prediction problems. Trans. ASME J. Basic Eng. **82**(Series D), 35–45 (1960)
39. M. Isard, A. Blake, CONDENSATION—conditional density propagation for visual tracking. Int. J. Comput. Vis. **29**(1), 5–28 (1998)
40. N. Noury, A. Fleury, P. Rumeau, A.K. Bourke, G.O. Laighin, V. Rialle, J.E. Lundy, Fall detection—principles and methods, in *Proceedings of the 29th IEEE EMBS*, pp. 1663–1666 2007
41. J. Parkkari, P. Kannus, M. Palvanen, A. Natri, J. Vainio, H. Aho et al., Majority of hip fractures occur as a result of a fall and impact on the greater trochanter of the femur: a prospective controlled hip fracture study with 206 consecutive patients. Calcif. Tissue Int. **65**, 183–187 (1999)
42. T.B. Moeslund, A. Hilton, V. Kruger, A survey of advances in vision-based human motion capture and analysis. J. CVIU **104**(2–3), 90–126 (2006)
43. J. Deutscher, I. Reid, Articulated body motion capture by stochastic search. Int. J. Comput. Vis. **61**(2), 185–205 (2005)
44. G. Mori, J. Malik, Recovering 3D human body configurations using shape contexts. IEEE Trans. PAMI **28**(7), 1052–1061 (2006)
45. R. Navaratnam, A.W. Fitzgibbon, R. Cipolla, Semi-supervised learning of joint density models for human pose estimation, in *Proceedings of BMVC*, vol. 2, pp. 679–688 2006
46. K. Mikolajczyk, D. Schmid, A. Zisserman, Human detection based on a probabilistic assembly of robust part detectors, in *Proceedings of European Conference on Computer Vision (ECCV)*, pp. 69–81 2004
47. C. Sminchisescu, A. Kanaujia, Z. Li, D. Metaxas, Discriminative density propagation for 3D human motion estimation. Proc. Comput. Vis. Pattern Recogn. **1**, 390–397 (2005)
48. L. Ren, G. Shakhnarovich, J.K. Hodgins, H. Pfister, P.A. Viola, Learning silhouette features for control of human motion. J. ACM TOG **24**(4), 1303–1331 (2005)
49. Q. Delamarre, O. Faugeras, 3D articulated models and multi view tracking with physical forces. J. CVIU **81**(3), 328–357 (2001)
50. A. Leone, G. Diraco, P. Siciliano, Topological and volumetric posture recognition with active vision sensor in AAL contexts, in *Proceedings of 4th IEEE International Workshop on Advances in Sensors and Interfaces (IWASI)*, pp. 110–114 2011

51. N. Werghi, Y. Xiao, J.P. Siebert, A functional-based segmentation of human body scans in arbitrary postures. J. T-SMCB **36**(1), 153–165 (2006)
52. G. Reeb, Sur les points singuliers d'une forme de Pfaff complétement intégrable ou d'une fonction numérique. C.R. Acad. Sci. Paris **222**, 847–849 (1946)
53. A. Verroust, F. Lazarus, Extracting skeletal curves from 3-D scattered data. Vis. Comput. **16**(1), 15–25 (2000)
54. C.J.C. Burges, A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Disc. **2**(2), 121–167 (1998)
55. I. Steinwart, A. Christmann, *Support vector machines* (Springer, New York, 2008)
56. H. Dongcheol, C. Wallraven, L. Seong-Whan, View invariant body pose estimation based on biased manifold learning, in *Proceedings of ICPR*, pp. 3866–3869 2010
57. A.K. Bourke, G.M. Lyons, A threshold-based fall-detection algorithm using a bi-axial gyroscope sensor. Med. Eng. Phys. J. **30**(1), 84–90 (2008)