

Document downloaded from:

<http://hdl.handle.net/10251/51457>

This paper must be cited as:

Bosch Jorge, M.; Sánchez Salmerón, A.J.; Valera Fernández, Á.; Ricolfe Viala, C. (2014). Fall detection based on the gravity vector using a wide-angle camera. *Expert Systems with Applications*. 41(17):7980-7986. doi:10.1016/j.eswa.2014.06.045.



The final publication is available at

<http://dx.doi.org/10.1016/j.eswa.2014.06.045>

Copyright Elsevier

Fall detection based on the gravity vector using a wide-angle camera.

Marc Bosch-Jorge^{a,*}, Antonio-José Sánchez-Salmerón^a, Ángel Valera-Fernández^a, Carlos Ricolfe-Viala^a

^a*Instituto de Automática e Informática Industrial, Universidad Politécnica de Valencia, Spain*

Abstract

Falls in elderly people are becoming an increasing healthcare problem, since life expectancy and the number of elderly people who live alone have increased over recent decades. If fall detection systems could be installed easily and economically in homes, telecare could be provided to alleviate this problem. In this paper we propose a low cost fall detection system based on a single wide-angle camera. Wide-angle cameras are used to reduce the number of cameras required for monitoring large areas. Using a calibrated video system, two new features based on the gravity vector are introduced for fall detection. These features are: angle between the gravity vector and the line from feet to head of the human and size of the upper body. Additionally, to differentiate between fall events and controlled lying down events the speed of changes in the features is also measured. Our experiments demonstrate that our system is 97% accurate for fall detection.

Keywords: fall detection, artificial vision, feature selection, feature extraction, new features based on gravity vector, monocular camera, wide-angle camera, calibration, low cost

1. Introduction

People in the First World are aging. Life expectancy at birth in the First World has increased to nearly 80 years during this century and the last.

*Principal Corresponding Author

Email address: mbosch@ai2.upv.es (Marc Bosch-Jorge)

Moreover, remaining life expectancy at old ages has also increased, meaning that elderly people live for longer. (Lutz et al., 2008) states the rate of population ageing in the current century will exceed that of the previous. This has serious consequences in society, because of the need to take care of these people and the increasing related health care costs (Bech et al., 2011). This has led to the development of care, monitoring, and ambience assisted technologies (Chaarouai et al., 2012).

Falls are one of the major issues in elderly health. A fall can have severe consequences, which can be worse if the person is not assisted in a short period of time, which can happen if they also lose consciousness or are unable to call for help.

During recent years several fall detection systems have been developed. (Mubashir et al., 2013) provides a recent survey on principles and approaches for fall detection. There are three major approaches which are based on wearable sensors, ambience sensing and vision based systems. Wearable sensors measure motion, such as accelerations (Mathie et al., 2004) and periods of inactivity (Sixsmith & Johnson, 2004), posture of the person (Kangas et al., 2009), or a combination of both (Luo & Hu, 2004). These are good at fall detection, but have some drawbacks: they can be uncomfortable, people may forget to wear them and they are prone to produce false positives. The second approach takes measures of the ambience of the room being monitored, such as sound (Zhuang et al., 2009) and vibrations (Alwan et al., 2006) in the room.

The third approach is based on computer vision systems. Computer vision systems are aimed to distinguish between different activities of the person being monitored, analyzing a video stream. With just one camera it is possible to monitor an area with independence on the number of people present. Action recognition is a popular research area (Poppe, 2010), but current methods are computationally very expensive. Action detection has usually better performance rates, as it only correlates observed sequences to labeled video sequences.

Fall detection through computer vision is difficult due to the multipart nature of the human body. The human body is composed of several parts which can be moved freely, thus making the process of identifying and locating people more difficult (Ferrari et al., 2008). This problem could be lessened by using human parts which are usually detectable, such as the head, waist or feet.

Several cameras are needed to monitor an entire house. This issue can be

partially solved with wide-angle cameras, which use wide field of view lenses and can thus monitor larger areas. This type of lens produces images that are highly-distorted, which must be corrected.

Some people may have privacy concerns because of the use of a computer vision systems. (Rajpoot & Jensen, 2014) provides a general model for video surveillance systems and identifies privacy requirements.

This paper presents a new approach to fall detection based on monocular wide-angle cameras by calibrating these cameras and correcting their distortions. In this scenario a new set of features is proposed which results in 97% accuracy for fall detection.

The following sections of this paper are organized as follows. Section 2 gives an overview of the current state-of-the-art with regard to features used in vision based fall detection systems. Section 3 introduces the method used to correct image distortion in a wide-angle camera. Section 4 describes the proposed new features. Section 5 shows the experiments and section 6 shows the results of our approach. Finally, in section 7 we present our concluding remarks.

2. Preliminary

Current computer vision based fall detection systems use different features extracted from the person being monitored. The features can be 2D, when they are extracted directly from the images, or 3D, when a reconstruction of the real-world objects in the scene is performed prior to extraction of the features.

The first step in 2D feature extraction is to transform the image of the body. Bounding box, best-fit approximated ellipse and projection histograms are the three most common transformations. Several authors use one of these transformations or even a combination of them. (Liu et al., 2010) use a vertical projection histogram with a statistical scheme in order to identify and locate the person in the video and the ratio of the width and height of the bounding box and the absolute difference between these two values to detect falls. To differentiate between a fall event and a controlled lying down event, temporal evidence is measured from the experiments and verified by statistical hypothesis testing. (Foroughi et al., 2008) combine vertical and horizontal projection histograms with best fit ellipse of the body, along with the 2D speed of the head. (Nasution et al., 2009) propose a method for

posture identification based on projection histograms, but there is a singularity between the postures of bending and lying toward the camera. To solve this problem the angle between vertices of the current bounding box and the bounding box of the last standing position is calculated. Movement based events, such as running, jumping active and inactive events are detected using the speed of the bounding box. (Willems et al., 2009) fit an ellipse to locate people in the image and use a combination of the angle of the person, the aspect ratio of the bounding box and a vertical projection histogram as features. The approach presented in (Anderson et al., 2006) uses the aspect ratio of the bounding box and the covariance matrix of pixel distribution as features. (Töreyn et al., 2005) extract a wavelet of the moving pixels in the bounding box. They also take advantage of fact that video recording systems can record audio. Falls produce high amplitude sounds, so in their work they also extract a wavelet of recorded audio in order to distinguish falls from sitting or controlled lying events. Other works which use the aspect ratio are (Tao et al., 2005), (Miaou et al., 2006) and (Khan & Habib, 2009). (Miaou et al., 2006) make use of personal information to improve their results. (Khan & Habib, 2009) first detect large motion in a video sequence using Motion History Images, followed by a segmentation of the person using projection histograms. For fall detection they use a combination of the aspect ratio of the bounding box and the speed of change of width and height. (Olivieri et al., 2012) uses optical flow to detect falls and recognize other human activities.

3D can be extracted from calibrated video cameras, stereo-vision cameras and depth cameras. Features extracted from 3D data are the location of the body and its parts in real-world or certain dynamic parameters, such as speed, acceleration or orientation. (Rougier et al., 2006) use a monocular camera to track 3D head trajectory. The algorithm used is POSIT combined with a set of particle filters. A fall event is detected using the vertical and horizontal speed of the head relative to the world coordinate system. (Rougier et al., 2011) use a camera that provides a depth map of the scene to calculate the height of the centroid of the human relative to ground plane and body velocity. By using these two features their system is able to solve the problem of occlusion. (Auvinet et al., 2011) present a system where several cameras are used to perform stereovision. In their work they calculate a measure of the vertical distribution along the vertical axis. A fall event is detected when this distribution is abnormally near the ground for a certain length of time. (Anderson et al., 2009) use voxels to define a 3D representa-

tion of a person called "voxel person". Each voxel person is classified using fuzzy logic and assigned a linguistic representation (up-right, on-the-ground or in-between).

However, stereo-vision systems have practical disadvantages, since a sufficient number of overlapping views is needed to reconstruct 3D objects in the scene. Depth cameras are limited by their narrow-angle field-of-view. A simpler alternative is to process the frames captured from monocular cameras independently. Our method performs a simple 3D reconstruction using a homography between the image plane and the ground plane. The state-of-the-art of 2D features for fall detection is mainly based on bounding box properties and they are independently extracted from monocular and narrow-angle cameras, since these cameras have low radial distortions. Wide-angle cameras can monitor larger areas, but they have a significant radial distortion. (Nait-Charif & McKenna, 2004) also uses a wide-angle camera, but does not correct the distortion.

We introduce two new features based on the gravity vector for fall detection by using independent wide-angle video systems. One of the proposed features measures the angle between the projected gravity vector and the line from feet to head of the human. However, this feature presents a singularity when a monocular camera is used. This singularity is solved by adding a new feature based on the size of the upper body. Additionally, a time measuring feature is added to differentiate between fall events and controlled lying down events.

3. System calibration

The calibration of the system consists of removing the distortion in the image and defining the transformation between the camera image plane and the ground plane of the room. The calibration process is solved in three steps. First lens distortion is calibrated to remove the distortion in the image. With undistorted images, the intrinsic and extrinsic camera parameters are computed. Finally, the homography H to define the transformation between the camera image plane and the ground plane of the room is computed. Calibration is done using several images of a chessboard template captured with the wide angle camera.

To calibrate the lens distortion, we use the calibration process proposed by (Ricolfe-Viala & Sanchez-Salmeron, 2011) for the rational function lens distortion model. According with (Ricolfe-Viala & Sánchez-Salmerón, 2010)

the rational function model performs better image correction if high distortion is present in the image. The rational function model was proposed by (Claus & Fitzgibbon, 2005) where distortion parameters are arranged in a 3x6 matrix and a six-vector of monomials in u and v as follows:

$$d(u, v) = \begin{bmatrix} A_{11} \cdot u^2 + A_{12} \cdot u \cdot v + A_{13} \cdot v^2 + A_{14} \cdot u + A_{15} \cdot v + A_{16} \\ A_{21} \cdot u^2 + A_{22} \cdot u \cdot v + A_{23} \cdot v^2 + A_{24} \cdot u + A_{25} \cdot v + A_{26} \\ A_{31} \cdot u^2 + A_{32} \cdot u \cdot v + A_{33} \cdot v^2 + A_{34} \cdot u + A_{35} \cdot v + A_{36} \end{bmatrix} \quad (1)$$

d represents the correction vector for pixel (u, v) in projective coordinates. The rational function lens distortion model can be computed with just one image of the calibration template.

Intrinsic and extrinsic camera parameters have been computed using several images of the calibration templates using the method proposed in (Ricolfe-Viala et al., 2012). This step is necessary to obtain the projection matrix P between the 3D world and the camera image plane and the homography H between the ground plane and the camera image plane.

The homography H is computed with just one image where the chessboard template was located on the floor using the method proposed in (Zhang, 1999). Since the plane of the chessboard template matches with the ground floor of the room, the homography H between the ground plane and the camera image plane is computed easily. Image is undistorted before using it to compute the homography H .

4. New features based on the gravity vector

A set of new features is defined for fall detection by using calibrated wide-angle cameras. Firstly, highly-distorted images are corrected and then features are independently extracted from each monocular camera. These new features relate the actual pose of the person with the theoretical pose of a person who is standing. The features used individually fail to detect falls in specific cases, but the combination of the features overcomes these failures.

To calculate the theoretical pose of a standing person we use the gravity vector (see fig. 1). The gravity vector, \mathbf{g}_v , is the projected normal vector to the real ground plane at a point \mathbf{f} . The point is located in the image and then its real-world position is calculated using the homography between the image plane and the ground plane. Then this real-world location is translated in

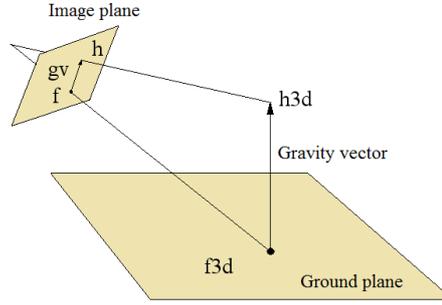


Figure 1: Gravity Vector

the direction of gravity and finally it is reprojected to the image using the projection matrix. This method is described in Algorithm 1.

Data: f , 2D feet location in the undistorted image plane;

H , homography between undistorted image plane and ground plane;

P , projection matrix to the undistorted image plane

Result: gv , projected gravity vector

$f3d \leftarrow H * f$;

$h3d \leftarrow f3d.z_translate(z_offset)$;

$h \leftarrow P * h3d$;

$gv \leftarrow h - f$;

Algorithm 1: Calculation of projected gravity vector

We use also the estimated height of the person. Once the feet and head are located in the image, an estimate of the height of the person is calculated by assuming the feet and the head are lie a normal line to the ground plane. The location of the feet in the real world is calculated with the homography between the image plane and the ground plane. Then, the real world line that projects in the head location in the image is calculated using the projection matrix. This line and the normal line to the ground plane that passes through the feet are arranged in a system of linear equations. Finally, this system is resolved using a SVD decomposition, which yields the closest point between the two lines. This method is described in Algorithm 2, and is also used to calculate the height of the waist.

Data: $[f, h]$, head and feet 2D location in the undistorted image plane;
 H , homography between undistorted image plane and ground plane;
 P , projection matrix to the undistorted image plane

Result: height, estimated height of h

$f3d \leftarrow H * f$;
 $A \leftarrow \begin{bmatrix} [1 & 0 & -c[0]] * P \\ [0 & 1 & -c[1]] * P \\ [1 & 0 & 0 & -f3d[0]] \\ [0 & 1 & 0 & -f3d[1]] \end{bmatrix}$;
 $v \leftarrow SVD(A)$;
 $h \leftarrow [v[2]/v[3]]$;

Algorithm 2: Calculation of height

4.1. Classification of standing

The first feature is α , the angle between the projected gravity vector and the line between the head and the feet of the person in the image. α varies between 0 and π (see fig. 2). α is close to zero if the person is standing but increases if the person is lying on the ground. α reaches its maximum value if the person is lying on the ground with their head and feet aligned with the optical axis of camera, with the head towards the camera. Algorithm 3 describes this method.

Data: $[h, f]$, head and feet 2D locations in the undistorted image plane

Result: α , angle between the gravity vector and the line from head to feet

$hf \leftarrow h - f$;
 $gv \leftarrow GravityVector(f)$;
 $pose_angle \leftarrow atan2(hf_y, hf_x)$;
 $gravity_angle \leftarrow atan2(gv_y, gv_x)$;
 $\alpha \leftarrow \pi - ||gravity_angle - pose_angle| - \pi|$;

Algorithm 3: Calculation of α

α has a singularity when the person is lying on the ground and their head and feet are aligned with the optical axis of the camera with the feet towards the camera. In this case, the angle remains close to zero and it is impossible to differentiate between the standing and lying position (fig 3).

To solve this singularity we add a feature based on the normalized size s of the upper body of the person in the image. If the person is standing, the size of their upper body in the image remains stable in the undistorted

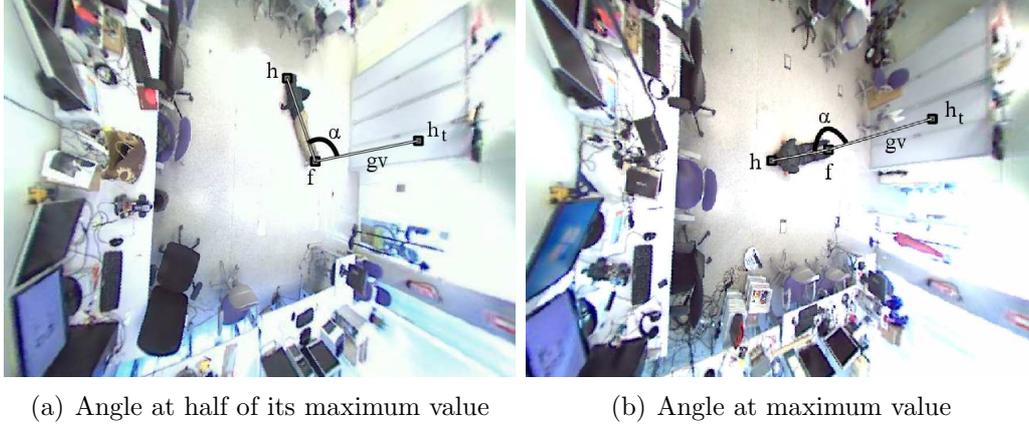


Figure 2: Different values of α when the person is lying down

image, but decreases when the person is lying on the floor due to the position of the camera on the ceiling. This decrease in size is particularly noticeable when the person is lying aligned with the optical axis of the camera, with the feet towards the camera (fig 3(b)). The actual size of the upper body, as , is normalized to the theoretical upper body size, ts , in the position of the person in the image, so this feature remains stable in the field of view of the camera. This method is described in Algorithm 4.

Data: $[h,w,f]$, head, waist and feet 2D locations in the undistorted image plane

Result: s , theoretical normalized upper body size in the image plane if the feet are located at f

```

 $f3d \leftarrow H * f$  ;
 $h3d \leftarrow f3d.z\_translate(head\_height)$  ;
 $w3d \leftarrow f3d.z\_translate(waist\_height)$  ;
 $h_t \leftarrow P * h3d$  ;
 $w_t \leftarrow P * w3d$  ;
 $ts \leftarrow ||h_t - w_t||^2$  ;
 $as \leftarrow ||h - w||^2$  ;
 $s \leftarrow as/ts$  ;

```

Algorithm 4: Calculation of s

This feature fails when the person is below the camera, because in this location the size of the upper body is highly variable. In this case, α remains the discriminating feature between classes.

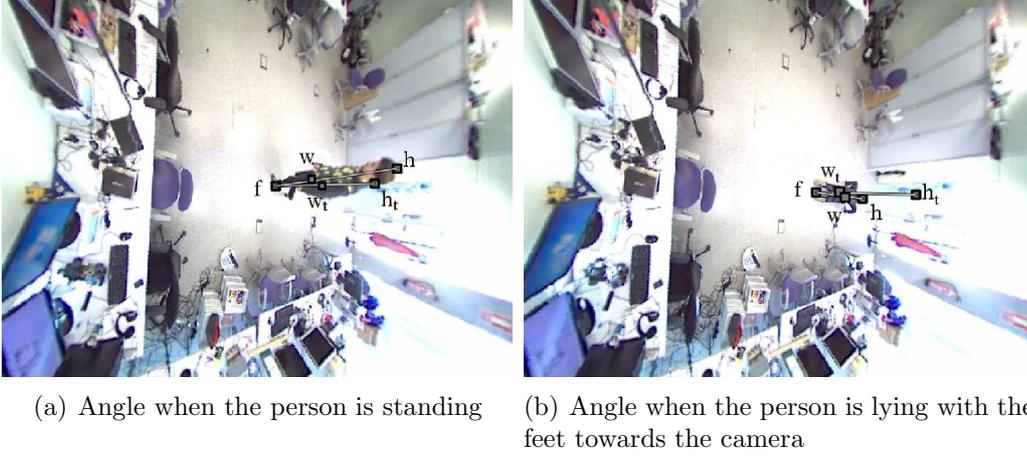


Figure 3: Singularity of α

4.2. Differentiation between a fall and lying down

Finally, falls are differentiated from controlled lying down actions using a time measuring feature. This feature, \mathbf{m} , is the number of frames where the ratio between the actual upper body size, as_i , and the mean of the actual upper body size, \overline{as} , in a sequence is above a certain threshold, th . Falls have a shorter duration than lying down, so this feature has higher values in the latter case. Algorithm 4 describes this method.

Data: $[h_i, w_i, f_i] i \in [1 \dots N]$ sequence of N head, waist and feet 2D locations in the undistorted image plane.

Result: m , number of frames where the normalized actual upper body size is above a certain threshold

```

for  $i \leftarrow 1$  to  $N$  do
     $as_i = ||h_i - w_i||^2$ 
end
 $m \leftarrow 0$ ;
for  $i \leftarrow 1$  to  $N$  do
    if  $as_i / \overline{as} > th$  then
         $m \leftarrow m + 1$ ;
    end
end

```

Algorithm 5: Calculation of m

5. Experiments

5.1. Dataset

We did not find any public dataset for our application, and so we recorded videos representing three different actions: walking, lying down and falling. We have a total of 335 videos, divided into 195 videos of walking, 40 lying down and 100 falling. The videos were recorded with an AXIS camera located on the ceiling of our laboratory. This camera has a resolution of 640x480 pixels and can record video at 15 fps. The actions were performed by four different people. Falls were performed forward, backwards and lateral. Some of the videos labeled as walking were recorded using an automated process, where a sequence was recorded when a person was detected in the scene. This process lasted 4 hours and yielded 98 videos. Each video consists of a sequence that is 80 frames long, the maximum time it takes to lie down, which was the longest action in our videos. As our work is focused on the features, each sequence was manually annotated, locating the feet, the waist and the head of the person, to avoid problems originated in the tracking algorithm. The dataset is available at (Bosch-Jorge et al., 2014).

5.2. Classifier

The data were classified using a SVM classifier. The implementation used is (Chang & Lin, 2011). We used a RBF kernel and a leave-one-out approach to separate data in testing and training sets. The parameters of the RBF kernel (C , γ) were optimized for each test using a cross-validation approach on the training set. For each test the data were scaled to $[-1, +1]$ range of values, as it is recommended by the authors of (Chang & Lin, 2011). Each test was classified with a binary classifier and a multi-class classifier. The multi-class approach used by (Chang & Lin, 2011) is one-vs-one with a voting strategy.

5.3. Parameters and feature extraction

The algorithms for features extraction presented in this work depend on the values of their parameters to work properly. α , s and m were calculated for the last frame of each sequence. *head_height* and *waist_height* were calculated for each sequence using the location of the head and waist in the first frame.

th has been determined empirically. We have analyzed the precision, recall, f1-score and accuracy for different values of th between 0 and 1, with

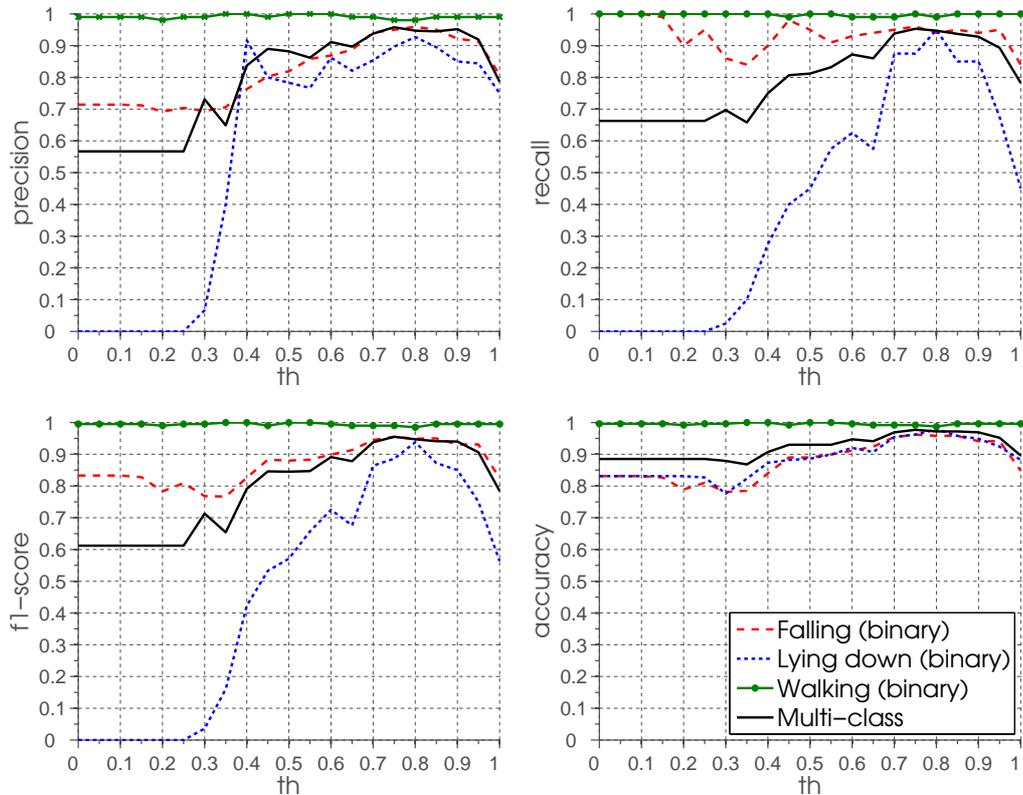


Figure 4: Evaluation of classification performance for varying th values.

a step of 0.05. The data were classified using the classifier of the previous section. The values are shown in figure 4. As can be seen in the figure, the best performance is achieved when th has a value of 0.8.

6. Results

The final results were obtained using a th value of 0.8.

The confusion matrices of the binary classification and multi-class classification are shown in tables 1 and 2, and the statistics are shown in table 3. The statistics are precision, recall, f1-score and accuracy for binary classification. We have used the macro-averaging measures and the average accuracy presented by (Sokolova & Lapalme, 2009) for multi-class classification. Precision, recall and f1-score have the same value in the multi-class classification because the confusion matrix is symmetric.

		Predicted class	
		Correct	Incorrect
Actual class	Falling	94	6
	Other	5	230
	Lying down	37	3
	Other	4	291
	Walking	195	0
	Other	1	139

Table 1: Confusion matrix for binary classification.

		Predicted class		
		Falling	Lying down	Walking
Actual class	Falling	94	4	2
	Lying down	10	35	0
	Walking	0	0	195

Table 2: Confusion matrix for multi-class classification.

	Class	Precision	Recall	F1-Score	Accuracy
Binary	Falling	0.949	0.940	0.945	0.967
	Lying down	0.902	0.925	0.914	0.979
	Walking	0.995	1.000	0.997	0.997
	Averaged	0.948	0.955	0.952	0.981
Multi-class	Macro-averaged	0.930	0.906	0.918	0.969

Table 3: Statistics for binary and multi-class classification.

7. Conclusions

In this paper we have proposed two new features for computer vision based fall detection. These features are angle between the projected gravity vector and the line from feet to head of the human and normalized size of the upper body. We perform a simple 3D reconstruction from a monocular camera to calculate these features. To perform the 3D reconstruction we need to calibrate the camera used by the system. Additionally, to reduce the number of cameras needed to monitor a large room we use a wide-angle camera. This type of camera uses a lens with high radial distortion, which must be corrected before the calibration process. As we use a single camera to monitor a large area, we have gained significant advantages in terms of system architecture without losing performance. Our system is more flexible, as cameras do not need to be calibrated one to one another, cameras do not need to overlap, cameras can drop in and out with little penalty, new features can be incorporated easily, and discriminative views can be exploited opportunistically. We believe our system is appropriate for widely-distributed camera systems thanks to its flexible architecture.

It is difficult to compare our method with others, because the features are of different nature. (Nait-Charif & McKenna, 2004) also uses a wide-angle camera, but does not correct the distortion of the camera, and does not provide any statistics for fall detection. Moreover, the use of different databases also difficulties the comparison between methods. However our method presents one of the highests accuracy rates of the state of the art. For instance, (Foroughi et al., 2008) and (Cucchiara et al., 2007) obtained 97.7% and 97.23% accuracy respectively. Our dataset is available at (Bosch-Jorge et al., 2014).

Our future work is focused on a person-variable time measuring feature, due to the fact that young people usually move faster than old people. A robust body-part location should be used to increase the reliability of the system. As an ultimate goal, we are aimed at person identification and action detection in a system composed of several cameras, located in different rooms, with tracking across the rooms. The system should be able to trigger different events depending on the action detected, on who has performed the action and on in which room the person is located.

Acknowledgement

This work was partially financed by Programa Estatal de Investigación, Desarrollo e Innovación Orientada a los Retos de la Sociedad (Dirección General de Investigación Científica y Técnica, Ministerio de Economía y Competitividad) under the project DPI2013-44227-R.

References

- Alwan, M., Rajendran, P. J., Kell, S., Mack, D., Dalal, S., Wolfe, M., & Felder, R. (2006). A smart and passive floor-vibration based fall detector for elderly. In *Information and Communication Technologies, 2006. ICTTA'06. 2nd* (pp. 1003–1007). IEEE volume 1.
- Anderson, D., Keller, J. M., Skubic, M., Chen, X., & He, Z. (2006). Recognizing falls from silhouettes. In *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE* (pp. 6388–6391). IEEE.
- Anderson, D., Luke, R. H., Keller, J. M., Skubic, M., Rantz, M., & Aud, M. (2009). Linguistic summarization of video for fall detection using voxel person and fuzzy logic. *Computer Vision and Image Understanding*, 113, 80–89.
- Auvinet, E., Multon, F., Saint-Arnaud, A., Rousseau, J., & Meunier, J. (2011). Fall detection with multiple cameras, an occlusion-resistant method based on 3-d silhouette vertical distribution. *IEEE Transactions on Information Technology in Biomedicine*, 15, 290–300.
- Bech, M., Christiansen, T., Khoman, E., Lauridsen, J., & Weale, M. (2011). Ageing and health care expenditure in eu-15. *The European Journal of Health Economics*, 12, 469–478.
- Bosch-Jorge, M., Sánchez-Salmerón, A.-J., Valera-Fernández, A., & Ricolfe-Viala, C. (2014). Dataset for fall detection. <https://mebiomec.ai2.upv.es/>.
- Chaaroui, A. A., Climent-Pérez, P., & Flórez-Revuelta, F. (2012). A review on vision techniques applied to human behaviour analysis for ambient-assisted living. *Expert Systems with Applications*, 39, 10873–10888.

- Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*, 1–27. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Claus, D., & Fitzgibbon, A. W. (2005). A rational function lens distortion model for general cameras. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* (pp. 213–219). IEEE volume 1.
- Cucchiara, R., Prati, A., & Vezzani, R. (2007). A multi-camera vision system for fall detection and alarm generation. *Expert Systems*, *24*, 334–345.
- Ferrari, V., Marin-Jimenez, M., & Zisserman, A. (2008). Progressive search space reduction for human pose estimation. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (pp. 1–8). IEEE.
- Foroughi, H., Aski, B., & Pourreza, H. (2008). Intelligent video surveillance for monitoring fall detection of elderly in home environments. In *Computer and Information Technology, 2008. ICCIT 2008. 11th International Conference on* (pp. 219–224). doi:10.1109/ICCITECHN.2008.4803020.
- Kangas, M., Vikman, I., Wiklander, J., Lindgren, P., Nyberg, L., & Jämsä, T. (2009). Sensitivity and specificity of fall detection in people aged 40 years and over. *Gait & posture*, *29*, 571–574.
- Khan, M., & Habib, H. (2009). Video analytic for fall detection from shape features and motion gradients. In *Proc. World Congress on Engineering and Computer Science* (pp. 1311–1316).
- Liu, C.-L., Lee, C.-H., & Lin, P.-M. (2010). A fall detection system using k-nearest neighbor classifier. *Expert Syst. Appl.*, *37*, 7174–7181.
- Luo, S., & Hu, Q. (2004). A dynamic motion pattern analysis approach to fall detection. In *Biomedical Circuits and Systems, 2004 IEEE International Workshop on* (pp. 1–5). IEEE.
- Lutz, W., Sanderson, W., & Scherbov, S. (2008). The coming acceleration of global population ageing. *Nature*, *451*, 716–719.
- Mathie, M. J., Coster, A. C., Lovell, N. H., & Celler, B. G. (2004). Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. *Physiological measurement*, *25*, R1.

- Miaou, S.-G., Sung, P.-H., & Huang, C.-Y. (2006). A customized human fall detection system using omni-camera images and personal information. In *Distributed Diagnosis and Home Healthcare, 2006. D2H2. 1st Transdisciplinary Conference on* (pp. 39–42). doi:10.1109/DDHH.2006.1624792.
- Mubashir, M., Shao, L., & Seed, L. (2013). A survey on fall detection: Principles and approaches. *Neurocomputing*, *100*, 144–152.
- Nait-Charif, H., & McKenna, S. J. (2004). Activity summarisation and fall detection in a supportive home environment. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on* (pp. 323–326). IEEE volume 4.
- Nasution, A., Zhang, P., & Emmanuel, S. (2009). Video surveillance for elderly monitoring and safety. In *TENCON 2009 - 2009 IEEE Region 10 Conference* (pp. 1–6). doi:10.1109/TENCON.2009.5395849.
- Olivieri, D. N., Gómez Conde, I., & Vila Sobrino, X. A. (2012). Eigenspace-based fall detection and activity recognition from motion templates and machine learning. *Expert Systems with Applications*, *39*, 5935–5945.
- Poppe, R. (2010). A survey on vision-based human action recognition. *Image and vision computing*, *28*, 976–990.
- Rajpoot, Q. M., & Jensen, C. D. (2014). Security and privacy in video surveillance: Requirements and challenges. In *ICT Systems Security and Privacy Protection* (pp. 169–184). Springer.
- Ricolfe-Viala, C., & Sánchez-Salmerón, A.-J. (2010). Robust metric calibration of non-linear camera lens distortion. *Pattern Recognition*, *43*, 1688–1699.
- Ricolfe-Viala, C., & Sanchez-Salmeron, A.-J. (2011). Camera calibration under optimal conditions. *Optics express*, *19*, 10769–10775.
- Ricolfe-Viala, C., Sanchez-Salmeron, A.-J., & Martinez-Berti, E. (2012). Accurate calibration with highly distorted images. *Applied optics*, *51*, 89–101.
- Rougier, C., Auvinet, E., Rousseau, J., Mignotte, M., & Meunier, J. (2011). Fall detection from depth map video sequences. In *ICOST* (pp. 121–128).

- Rougier, C., Meunier, J., St-Arnaud, A., & Rousseau, J. (2006). Monocular 3d head tracking to detect falls of elderly people. In *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE* (pp. 6384 –6387). doi:10.1109/IEMBS.2006.260829.
- Sixsmith, A., & Johnson, N. (2004). A smart sensor to detect the falls of the elderly. *Pervasive Computing, IEEE*, 3, 42–47.
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45, 427–437.
- Tao, J., Turjo, M., Wong, M.-F., Wang, M., & Tan, Y.-P. (2005). Fall incidents detection for intelligent video surveillance. In *Information, Communications and Signal Processing, 2005 Fifth International Conference on* (pp. 1590 –1594). doi:10.1109/ICICS.2005.1689327.
- Töreyn, B. U., Dedeoglu, Y., & Çetin, A. E. (2005). Hmm based falling person detection using both audio and video. In *ICCV-HCI* (pp. 211–220).
- Willems, J., Debard, G., Vanrumste, B., & Goedem, T. (2009). A video-based algorithm for elderly fall detection. In O. Dssel, & W. C. Schlegel (Eds.), *World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany* (pp. 312–315). Springer Berlin Heidelberg volume 25/5 of *IFMBE Proceedings*. URL: http://dx.doi.org/10.1007/978-3-642-03904-1_87.
- Zhang, Z. (1999). Flexible camera calibration by viewing a plane from unknown orientations. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on* (pp. 666–673). IEEE volume 1.
- Zhuang, X., Huang, J., Potamianos, G., & Hasegawa-Johnson, M. (2009). Acoustic fall detection using gaussian mixture models and gmm super-vectors. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on* (pp. 69–72). IEEE.