Expert Systems with Applications 41 (2014) 7980–7986

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

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

Marc Bosch-Jorge *,[1], Antonio-José Sánchez-Salmerón, Ángel Valera, Carlos Ricolfe-Viala

Instituto de Automtica e Informtica Industrial, Universidad Politécnica de Valencia, Spain

ARTICLE INFO

Article history:
Available online 11 July 2014

Keywords:
Fall detection
Artificial vision
Feature selection
Feature extraction
New features based on gravity vector
Monocular camera
Wide-angle camera
Calibration
Low cost

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.

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. Moreover, remaining life expectancy at old ages has also increased, meaning that elderly people live for longer. Lutz, Sanderson, and Scherbov (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, Christiansen, Khoman, Lauridsen, & Weale, 2011). This has lead to the development of care, monitoring, and ambience assisted technologies (Chaaraoui, Climent-Pérez, & Flórez-Revuelta, 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, Shao, and Seed (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 accelerations (Mathie, Coster, Lovell, & Celler, 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, Huang, Potamianos, & Hasegawa-Johnson, 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, Marin-Jimenez, & Zisserman, 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 and Jensen (2014) provides a general model for video surveillance systems and identifies privacy requirements.

* Corresponding author.
E-mail address: mbosch@ai2.upv.es (M. Bosch-Jorge).
[1] Principal corresponding author.

http://dx.doi.org/10.1016/j.eswa.2014.06.045
0957-4174/

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, Lee, and Lin (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, Aski, and Pourreza (2008) combine vertical and horizontal projection histograms with best fit ellipse of the body, along with the 2D speed of the head. Nasution, Zhang, and Emmanuel (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, Debard, Vanrumste, and Goedem (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, Keller, Skubic, Chen, and He (2006) uses the aspect ratio of the bounding box and the covariance matrix of pixel distribution as features. Töreyin, Dedeoglu, and Çetin (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, Turjo, Wong, Wang, and Tan (2005), Miaou, Sung, and Huang (2006) and Khan and Habib (2009). Miaou et al. (2006) make use of personal information to improve their results. Khan and 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, Gómez Conde, and Vila Sobrino (2012) uses optical flow to detect falls and recognize other human activities.

3D can be extracted from calibrated video cameras, stereovision 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, Meunier, St-Arnaud, and Rousseau (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, Auvinet, Rousseau, Mignotte, and Meunier (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, Multon, Saint-Arnaud, Rousseau, and Meunier (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 representation 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 distorsion. Nait-Charif and McKenna (2004) also uses a wide-angle camera, but does not correct the distorsion.

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 and Sanchez-Salmeron (2011) for the rational function lens distortion model. According with Ricolfe-Viala and 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 and Fitzgibbon (2005) where distortion parameters are arranged in a 3×6 matrix and a six-vector of monomials in u and v as follows:

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