Biomedical Signal Processing and Control 40 (2018) 102–110

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Research Paper

Use of kinematic and mel-cepstrum-related features for fall detection based on data from infrared depth sensors

Paweł Mazurek*, Jakub Wagner, Roman Z. Morawski

Institute of Radioelectronics and Multimedia Technology Faculty of Electronics and Information Technology, Warsaw University of Technology, Nowowiejska 15/19, 00-665 Warsaw, Poland

A R T I C L E I N F O

Article history:
Received 7 March 2017
Received in revised form 26 July 2017
Accepted 5 September 2017

Keywords:
Classification algorithms
Data acquisition
Data processing
Event detection
Infrared image sensors
Public healthcare
Sensor systems and applications

A B S T R A C T

A methodology for acquisition and preprocessing of measurement data from infrared depth sensors, when applied for fall detection, combined with several approaches to the classification of those data, is proposed. Data processing is initiated with extraction of the silhouette from the depth image and estimation of the coordinates of the center of that silhouette. Next, two groups of features to be applied for fall/non-fall classification are extracted: kinematic features (various statistics defined on the position, velocity and acceleration trajectories of the monitored person) and mel-cepstrum-related features (components of the mel-cepstrum obtained by means of an unconventional set of mel-filters). Finally, the utility of these features in fall detection is assessed using three classification algorithms – *viz.* support vector machine, artificial neural network, and naïve Bayes classifier – trained and tested on two datasets consisting of, respectively, 160 data sequences (representative of 80 falls and 80 other human behaviours) and 264 data sequences (representative of 132 falls and 132 other human behaviours). The application of the combination of the kinematic and mel-cepstrum-related features yields highly accurate classification results – all classifiers achieved, depending on the dataset, 98.6–100% and 93.9–97.7% sensitivity. Thus, infrared depth sensors can be promising tools for unobtrusive fall detection. They provide data which can be in various ways preprocessed to form a basis for reliable fall detection. Appropriate selection of the feature sets directly affects the reliability of unobtrusive monitoring systems, and – indirectly – the quality of life of the monitored persons.

1. Introduction

The European and North-American populations are aging quickly. The problem of organised care over elderly persons, especially those suffering dementia is, therefore, of increasing importance. Hence the demand for research on new technologies that could be employed in care services for such persons. Its primary objective is to examine the applicability of various sensor systems for non-invasive monitoring of the movements and vital bodily functions, such as heart beat or breathing rhythm, of elderly persons in their home environment. The capability of such systems to detect dangerous events, such as person's fall, is of key importance since falls among elderly persons are the main cause of their admission and long-term stay in hospitals: it is the sixth cause of death for persons over the age of 65, the second for those between 65 and 75, and the first for those over 75 [1]. There are three main categories of monitoring techniques already applied in care practice – wearable, vision-based, and environmental [1,2]:

- The wearable techniques are based mainly on sensors, such as accelerometers combined with gyroscopes, worn by a monitored person [3,4]; those devices can take various forms, *e.g.* belts, necklaces or bracelets. The signals, acquired by the sensors, are transmitted via radio to a computer and analysed. Unfortunately, wearable techniques are considered intrusive – measuring devices have to be constantly attached to the person's body causing possible discomfort.
- The vision-based techniques are designed using fixed cameras that continuously record the movement of a monitored person [5,6]; the acquired data are processed by means of algorithms of pattern recognition that trigger an alarm in case of danger. The main limitations of these techniques are: the time and cost of installation, the limited space of application (within the range of cameras) and privacy violation.
- The environmental techniques are based on the installation of sensors in the places to be monitored – *e.g.* pressure sensors on chairs, cameras, and radio-frequency tags embedded throughout

* Corresponding author.
E-mail addresses: p.mazurek@ire.pw.edu.pl (P. Mazurek), j.wagner@ire.pw.edu.pl (J. Wagner), r.morawski@ire.pw.edu.pl (R.Z. Morawski).

http://dx.doi.org/10.1016/j.bspc.2017.09.006
1746-8094/

the home of the elderly persons, as well as in their furniture and clothing [7,8]. The limitations of these techniques are: the time and cost of installation and the considerable invasiveness with respect to the home environment.

Since several years numerous attempts have been made to apply technology of infrared depth sensors for monitoring of elderly persons. Those attempts are mainly motivated by the conviction that this technology may be less intrusive than vision-based solutions, less cumbersome than the wearable solutions, and less invasive with respect to the home environment than the environmental solutions. Moreover, the applicability potential behind this monitoring technique is related to the low price of infrared depth sensors which are installed in the mass devices such as Microsoft Kinect [9]. A 2014 review paper, authored by David Webster and Ozkan Celik [10], provides an excellent analysis of 48 best papers, selected out of 461 papers on Kinect applications in healthcare, published in 2011–2013; 25 of them are devoted to elderly care, 23 – to stroke rehabilitation. Some of the applications described there use both video cameras and depth-imaging cameras, the others – only the latter ones. On top of that, recent years, 2014–2016, brought numerous publications where the Kinect depth-imaging cameras are used for patients monitoring; good examples are [11–57].

In this paper, a study on fall detection by means of a system for unobtrusive monitoring, based on the infrared depth sensors, is presented; this study is a part of a broader research programme aimed at the investigation of combinations of infrared depth sensors with impulse-radar sensors when applied for unobtrusive patients monitoring. Those techniques have certain complementary advantages and disadvantages [58], and – when coupled – can increase the reliability of monitoring: impulse-radar sensors can be used for through-the-wall monitoring of movements, as well as detection of heart beat and breathing of a monitored person; on the other hand, infrared depth sensors can be employed for detecting person's falls. Hence the motivation to seek new approaches improving the effectiveness of the monitoring techniques; in particular – improving the capability of infrared depth sensors to detect falls.

Since fall detection is a complex classification problem, in order to reduce its dimensionality, it has been decomposed into two sub-problems: transformation of a sensor data sequence into a low-dimensional vector of features, characterising the monitored person's movement, and application of a standard classifier to that vector. In the considered case, a classifier is an operator (or an algorithm) attributing a label "fall" to a data sequence representative of an event considered to be a fall of the monitored person, or a label "non-fall" – otherwise.

The originality of the solution proposed here consists in a novel approach of the interpretation of data from the depth sensors, leading to the determination of two categories of features to be used for fall detection. The data are first transformed into the absolute 3D coordinates of the center of the silhouette of the monitored person; next, those coordinates are used as a basis for calculation of both groups of features, *i.e.* the kinematic features, such as maximum velocity or acceleration of a moving person, and mel-cepstrum-related features, *viz.* components of the mel-cepstrum obtained by means of a set of different filters than those applied in the known variants of the mel-cepstrum. The usability of the introduced features in fall detection is assessed using three different classifiers – *viz.* support vector machine, artificial neural network, and naïve Bayes classifier – trained and tested on a rich set of data sequences representative of both falls and fall-like events.

The paper is organised as follows. Section 2 is devoted to the methodology of data acquisition, preprocessing and transformation into space coordinates. In Section 3 the procedures for feature generation are introduced, and the applied classifiers are described. The results of the performance evaluation of the classifiers, followed by their discussion, are presented in Section 4. Finally, in Section 5 some conclusions are provided.

A preliminary version of the contents of this paper has been reported in three papers presented at the 2015 IEEE Conference on Intelligent Data Acquisition and Advanced Computing Systems, [41–43].

Table 1
Technical parameters of the depth sensor.

Parameter	Value
Data frame resolution	480×640 pixels
Data acquisition rate	30 frames per second
Angular field of view	45° vertically and 58° horizontally
Distance measurement resolution	*ca.* 1 mm for the distance of 0.5 m *ca.* 7 cm for the distance of 5 m
Measurement uncertainty	*ca.* ± 1 mm for the distance of 0.5 m *ca.* ± 4 cm for the distance of 5 m

2. Methodology of data acquisition and preprocessing

2.1. Data acquisition

2.1.1. IRMTv1 fall detection dataset

The IRMTv1 fall detection dataset contains raw measurement data acquired by means of two synchronised infrared depth sensors (labelled with S1 and S2 hereinafter) being parts of two Kinect apparatuses (model V1), whose most important technical parameters are shown in Table 1 (*cf.* references [9,59]). Configuration of the devices, relative to the observed area, is presented in Fig. 1. A monitored person has moved at the distance of *ca.* 1.5–4 m from each of them. The experiments have consisted in recording depth images – *i.e.* matrices of the integer numbers representative of the distance between the plane parallel to the sensor lens and points of an imaged scene – by means of both devices simultaneously. After the acquisition, the sequences were cropped to the length of 75 frames each; such length corresponds to the 2.5-s-long time window of analysis covering only the fall/non-fall motion of the monitored person, specific for each type of scenario.

A set of 20 fall scenarios and 20 scenarios corresponding to other actions of a monitored person, called non-falls, has been designed. Short descriptions of all scenarios are provided in Tables 2 and 3.

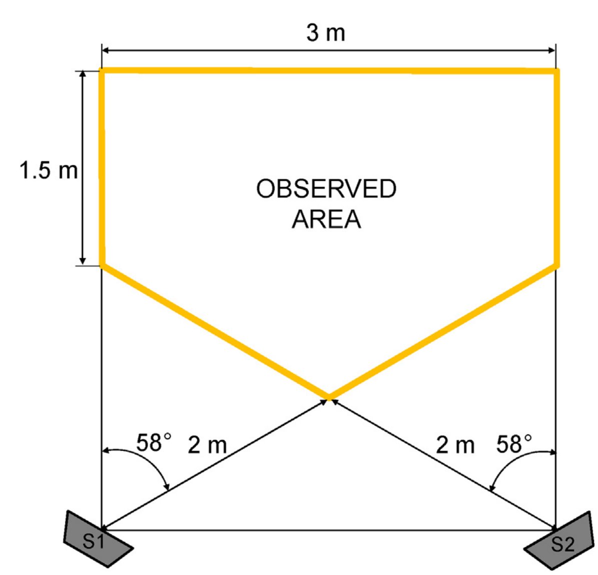

Fig. 1. Configuration of two depth sensors (S1 and S2) relative to the observed area.