Contactless Interactive Fall Detection and Sleep Quality Estimation for Supporting Elderly with Incipient Dementia

Abstract: In recent years, the demographic change in conjunction with a lack of professional caregivers led to retirement homes reaching capacity. The Alzheimer Disease International stated that over 50 million people suffered from dementia in 2019 worldwide and twice the amount will presumably be affected in 2030. The field of Ambient Assisted Living (AAL) tackles this problem by facilitating technical system-aided everyday life. AUXILIA is such an AAL system and does not only support elderly people with dementia in an early phase, but also monitors their activities to provide behaviour analysis results for care attendants, relatives and physicians. Moreover, the system is capable of recognizing emergency situations like human falls. Furthermore, sleep quality estimation is employed to be able to draw conclusions about the current behaviour of an affected person. This article presents the current development state of AUXILIA.

Keywords: AAL, dementia, omnidirectional camera, behavioural analysis, fall detection, sleep quality estimation

1 Introduction

The German population is expected to rise until 2035 and reach just over 83 million people [1]. At the same time, the amount of people requiring professional care rises even faster and will go from 3.5 million today to over four million by 2035 and climb even further to 5.1 million by 2050 [2]. In an effort to cope with the resulting shortage of professional caregivers, technical eHealth solutions are becoming popular in application-driven research in the last couple of years. As a result, in the field of Ambient Assisted Living (AAL) behaviour analysis-based technical assistance systems are being developed to support different user groups, such as the elderly or professional caregivers.

2 Related Work

A detailed description of the state of the art of fall detection methods is out of scope and can be found in [7, 8]. Fall detection methods can be based on wearable sensors and ambient sensors, like microphones, vibration detectors and cameras. Wearable sensors interfere with the daily routine and require the user to cooperate, which cannot be expected from dementia patients. On the other hand, acoustic and vibration sensors suffer from interfering vibrations and audio signals originating from normal daily activities. Cameras, on the contrary, neither hinder the user in their routines nor suffer from significant interferences, when they are used in top view. Moreover, they allow tracking the user’s movement during the whole fall event and one sensor can monitor more than one person. Delibasis et al. [9] present a method for detecting falls using a top view omnidirectional camera. The fall detection method presented in this paper is based on their work. We introduce a temporal threshold to reduce false positives and embed this approach in an interactive system by combining image processing and speech recognition results.

The analysis and evaluation of a user’s sleep information can add value in terms of long-term analysis of their general behaviour. Conclusions can be drawn, if they appear particularly conspicuous or confused on any given day. Measuring sleep quality basically consists of recognizing the different phases of sleep, specifically the rapid-eye movement (REM) and non-REM phases. A common standard to measure these phases is polysomnography (PSG), a technique that requires a clinical setting with physical attachments [10]. However, in...
the context of dementia it is not an option to require a patient to wear additional sensors to obtain information about their sleep. A well known and in the past often used indicator to determine sleep quality is the movement during sleep using images [11, 12]. With the sleep quality estimation proposed in this paper, we introduce an approach to distinguish between light and deep sleep. Additionally, evaluations are performed to determine the necessary parameters of the algorithm.

3 Interactive Fall Detection

Fall detection is essential in AAL applications, especially if they are directed at elderly people. This section presents the interactive fall detection system implemented by AUXILIA.

3.1 Interactive Fall Detection

The proposed interactive fall detection system is based on a combination of audio and video, captured by an omnidirectional camera and a microphone, respectively. Both sensors are placed on the ceiling of the flat. Speakers complete the interactive system, allowing it to communicate with the user. Image and audio processes run constantly and independently of each other. The image process is responsible for detecting the person in each video frame and classifying their pose as standing or lying. At the same time, the audio process detects predefined keywords spoken by the user. The video and audio algorithms send their outputs to a third process, which initiates an interaction protocol when a fall is detected. This protocol consists of a pre-designed dialog based on questions and instructions, which is built upon the user’s answers as well as their movements. This interaction allows discarding false positives, thus avoiding unnecessary external interventions.

The implemented fall detection algorithm is based on the approach presented in [9]. This work uses the geometrical characteristics of a top view omnidirectional camera to decide whether a person is standing. In a top view omnidirectional image \( I(x,y) \), the major axis of a standing person’s silhouette is always radial with respect to the image centre. Therefore, if the silhouette’s major axis deviates from radial orientation for more than a given threshold, the silhouette is classified as fallen. This deviation can be determined by computing \( \cos \theta \), where \( \theta \) is the angle between the vector pointing from the centre of the image \( (x_0, y_0) \) to the silhouette centre \( (x_c, y_c) \), and the vector pointing from the silhouette centre in the direction of its major axis (see Figure 1).

As described in [9], the first step is to extract the user’s silhouette. To this end, background subtraction is performed on each frame \( I(x,y) \) by using a Mixture of Gaussian (MoG) algorithm [13]. The output of the MoG filter is processed by morphological operations in order to connect elements that can be part of the silhouette and to discard those which are not. The resulting binary image \( I'(x,y) \) is evaluated by calculating the eigenvalues \( \lambda_{1,2} \) of its structure matrix, where \( \lambda_1 > \lambda_2 \). If \( \lambda_1 \geq 1.5\lambda_2 \), the extracted silhouette is elongated and the direction of its major axis is given by the corresponding unit length eigenvector \( v_1 \). \( \cos \theta \) can be calculated as follows:

\[
\cos \theta = \frac{\left((x_0 - x_c, y_0 - y_c) \cdot v_1\right)}{\left\| (x_0 - x_c, y_0 - y_c) \right\| \left\| v_1 \right\|}
\]

Now \( \cos \theta \) is compared with a threshold \( L \) to classify the silhouette as fallen or standing. In our work, we incorporate an additional temporal threshold \( T \), which prevents the system from reacting to brief position changes or sporadic noise. For our application, the thresholds are fixed to \( L = 0.66 \) and \( T = 10 \) s. Therefore, if \( \cos \theta > 0.66 \), the silhouette is classified as standing. On the other hand, if \( \cos \theta \leq 0.66 \) for more than 10 seconds, the system communicates a fall detection.

Detection of a fall event activates a predefined protocol and a dialog is initialised through the speakers. A commercial speech recognition module (vicControl DSP 3.2) is used to detect keywords in the user’s answers, like “yes”, “no”, “doctor”, “help”, etc. In case the user asks for a doctor or does not answer at all, the system considers this an emergency. The emergency protocol is triggered as well if the user refuses any help but does not stand up after a given time. If the user refuses help
and stands up, a second protocol with different instructions verifies their mental fitness to rule out injuries. Failing this last test also leads to an emergency. The emergency state can easily be extended to an emergency services call with a commercial medical alert system.

### 3.2 Experimental Results

We tested the interactive fall detection in real time in a one-room flat. A total of 553 fall simulations were performed by three persons from several poses at various positions inside the home. We analysed a multitude of scenarios in order to test the interaction with the system, i.e. in every case different answers and reactions of the user were simulated. In 500 of these simulations the falls were detected successfully and the interaction with the user led to the right decision for the given situation: trigger an emergency or continue monitoring. This means a correct detection rate of 85%.

### 4 Sleep Quality Estimation

Mental health [14] and dementia [15] are related to sleep quality. However, the sleep quality evaluation is often only possible in a clinical environment using physical sensors like PSG. In this section we describe a contactless sleep quality estimation algorithm used in AUXILIA.

#### 4.1 Methodology

The proposed system detects periods of restlessness by capturing images during the user’s sleep phases. Due to night-time light restrictions, additional infrared lighting is used to capture information without disturbing the user’s sleep. An overview of the sleep quality estimation process is shown in Figure 2. By shooting at night under infrared light, the images suffer from noise. Therefore each image $I$ is pre-processed with a Gaussian blur using a kernel size of $3 \times 3$. Employing an MoG algorithm, the previously adjusted image $I$ is used to model the background, resulting in a binary mask $M$. The algorithm needs two thresholds: $T_B$ defines the number of images used to model the background, while $T_I$ separates the foreground and background. Stronger movement leads to more foreground pixels in $M$. A set $B = \{M_0, \ldots, M_{T_{B}-1}\}$ of $T_B$ elements is collected, where $M = \text{mean}(M)$. Once $B$ reaches $T_B$ elements, its mean value $\bar{B}$ and the current timestamp $t$ are added to set $C = \{\bar{B}_0, \ldots, \bar{B}_{n-1}\}$ and $T = \{t_0, \ldots, t_{n-1}\}$ respectively. The sleep quality estimation process is stopped when $|T| = |C| = n$.

In order to classify time periods as light or deep sleep, a threshold proportional to the mean value of $C$ is defined as $T_S = p \cdot \bar{C}$ with $p \in \mathbb{R}$. A set of sleep states $S = \{S_0, \ldots, S_{n-1}\}$ is created, where $S_i = 1$ if $B_i > T_S$, otherwise $S_i = 2$. Due to the large number of elements in $S$ and $T$, post-processing reduces the size of both sets and merges them into $O = \{(t_i, S_i) \in (T, S) : S_i \neq S_{i-1}\}$ by only saving a new pair of values, when the sleep state $S_i$ changes.

#### 4.2 Experimental Results

For the evaluation of our sleep quality estimation, two datasets with sleep scenarios of different duration with were recorded by a frame rate of three frames per second (FPS). The test subject is 34 years old and has normal sleep. Scenario one, named Dataset 1, contains ~97000 images with a sleep duration of about nine hours, the second scenario, Dataset 2 contains ~120000 frames with a sleep duration of ten hours. During sleep periods a consumer sleep tracker (Nokia Withings) was used to collect ground truth data.

Since the sleep tracker and the proposed method produce different durations of sleep states, we cannot compare them directly. As a result, we employ dynamic time warping with Euclidean distance to compare the unsymmetric signals. During the experiments we use: $T_B = 500$, $T_I = 16$ and $p = 0.5$. We determine an appropriate value for $T_S$ from the set $\{0.5, 1, 3, 5, 10, 15, 20, 25, 30, 35, 40\}$, which describes the minutes to be buffered until a sleep state data point is created. The required number of images to be buffered is then calculated by 60 FPS $\cdot T_S$. Additionally, we evaluate Gaussian blur...
with kernel size of $3 \times 3$, histogram stretching for contrast adjustments using contrast limited adaptive histogram equalization [16], and both methods in combination as pre-processing methods.

As shown in Figure 3, the value with the smallest distance to ground truth states (lower is better) for both datasets is at $T_c = 5$ minutes and results in an average similarity of about 61%. It seems the results for $T_c < 3$ are too noisy compared to the ground truth, while for $T_c > 15$, too much information is lost due to averaging over the buffered images.

5 Conclusion

In this work we develop an interactive fall detection and a sleep quality estimation based on top view omnidirectional images from an image sensor. The detection of falls based on the combination of video and audio, where the two processes are running independently of each other. This approach leads to a detection rate of the system of 85%. The sleep quality estimation is conducted through background subtraction and subsequent detection and classification of the person’s movements. The evaluation of the approach is executed on two omnidirectional image sequences with infrared illumination, collected from one person, using a consumer sleep tracker for ground truth. The resulting accuracy is about 61%, which brings the algorithm close to body-based sensor technologies. Additional investigations are required to reach results comparable to polysomgraphically based systems. More investigation is needed in order to evaluate the presented approaches on users, who suffer from mild dementia.

Author Statement

Research funding: The implementation of the sleep quality estimation was funded by the European Regional Development Fund (ERDF) under funding code 100241945. The development of the interactive fall detection system was funded by the German Federal Ministry of Education and Research under grant number 03ZZ0431I. Conflict of interest: Authors state no conflict of interest. Ethical approval: The conducted research is not related to either human or animals use. Responsibility for the content of this paper lies with the authors.

References