Abstract—A toolbox for the automatic monitoring of elderly in a nursing home or in the natural home environment is proposed. Rather than monitoring vital signs or other biomedical parameters, the toolbox is focussed on the monitoring of activity patterns and changes therein. Activity information is derived from visual information using image processing algorithms. The visual information is acquired using 3D camera technology. Besides a traditional visual image, 3D cameras also provide highly accurate depth information. The 3D position of the subject is derived and serves as the primary information source for the different components in the toolbox.

I. INTRODUCTION

Over the last decades, life expectancy rates have increased dramatically. This has created a growing need for beds in nursing homes. This growing need causes governments to stimulate elderly to live longer in their natural home environment. As a result, research investigating monitoring possibilities for elderly is intensified, often focusing on falls, as falls are directly accountable for 40% of all elderly admissions to nursing homes or long-term care facilities [14]. Falls are the most common cause of injury for elderly people [13] as on average 33% of the seniors will fall within a year [12]. Besides several severe medical problems, falls also result in a decrease of life quality due to the often important fear for future falls [15]. This fear can become so great that people end up in a cycle: after a fall they become so afraid of falling that they limit their activities, this decreases fitness, mobility, balance and social interactions and increases the risk of subsequent falls [10], [1]. Although there exists a lot of work on the prevention of falls [18] and on the assessment of the risk on falls [17], falls will remain an important cause of the loss of independence of elderly.

Given the importance of falls among elderly, many technical aids have been developed in the prevention and detection of falls. Most often, a fall detector is a small portable device containing an accelerometer, some simple logic for thresholding the signal and a wireless communication component for alerting the base station. Detection results are typically very good, on condition that the device is worn correctly by the elderly. This, however, turns out to be problematic in many of the target populations, including people with dementia. Accelerometers are also employed for monitoring various kinds of activity patterns, both for elderly and other target groups (e.g. [6], [5]).

Fall detectors and other activity monitoring tools based on image processing techniques do not require an active cooperation of the subject being monitored, but are often less reliable. Previous work pursuing different strategies exists: One strategy consists of placing a wide-angle camera on the ceiling and detect falls in the 2D images [9], [8]. Others extracted posture information from a single 2D image and used that information for fall detection. We propose to use a 3D camera for performing visual fall detection. According to our knowledge, this approach is new.

In this paper, a framework for the monitoring of elderly is introduced, including fall detection capabilities. All components in the framework employ 3D pose information (i.e. 3D position of the subject in the room) of the monitored subject as their primary information source. At a coarse grain level, this information can be employed to monitor global activity cues (i.e. how often or how long is the subject active). At a finer level, it can be monitored how often the subject leaves the bed, walks around, sits in a sofa, etc. At the third level of granularity, one could monitor specific activities or events (fall detection, detection of cooking, bathing, ...). At each of those levels, the detected activity patterns convey information about the medical condition of the subject, as is reflected in the Activities of Daily Living scales [16], [3].

II. METHOD

A. 3D Camera

Besides a normal visual gray level image, a 3D camera also provides depth information. The camera is an active device, emitting modulated infrared light and is hence based on the time-of-flight (TOF) principle [11]. Compared to stereo vision, the accuracy of the depth information is higher. Moreover, the depth information is readily available, without heavy calculations. As there are no correspondences between the image pairs which have to be established, 3D cameras do not provide inaccurate depth information in regions with poor texture information. Currently, the image resolution is rather low (i.e. typically 176 by 144 pixels1), while the cost is very high. Once 3D cameras will become more widely used, the price is expected to drop significantly.

B. Pose recognition algorithm

The pose recognition algorithm consists of three steps. First, the human silhouette is extracted from the gray level image by subtracting the image from the background. The background image is calculated using a running average over many frames. The resulting image is thresholded, smoothed,
eroded and dilated in order to result in an image containing a few closed blobs (figure 1, lower image). The biggest blob in the image defines the human silhouette. In a second step, the center of the silhouette is calculated by fitting an ellipse to the blob (figure 1, right upper image). By thresholding on the width/height of the ellipse, blobs not related to human silhouettes can be discarded. Next, the 3D position of the silhouette’s center is calculated in a coordinate system defined by the room. This requires the spatial coordinates of the silhouette’s center to be transformed into room coordinates. This is done using a simple linear calibration method. The height above the ground of the silhouette is thresholded for classifying the pose as either standing, sitting ($z \leq 70\text{cm.}$) or lying ($z \leq 35\text{cm.}$).

### III. EXPERIMENT AND RESULTS

An experiment validating the proposed method was performed. The 3D camera was positioned in the corner of a typical living room. A male subject of 27 years old was recorded by the monitoring system and its position was categorized as either standing, sitting or lying down. The footage contains recordings from different camera positions and was recorded over the course of a week, such that different light conditions and different clothing conditions were included.

In figure 2 a plot of the calculated height of the person above the ground is plotted over 1000 frames, while the subject is active in the living room. For clarity, both thresholds of 35 and 70 cm. are marked by a green and an orange line. The plot convincingly shows that lying down, sitting and standing can clearly be distinguished. We recorded ten sequences in which the subject was sitting, ten in which he was walking around and ten in which he was lying down on the ground. The “walking sequences” contained episodes in which the subject was bending to grasp something from a low table, ”sitting sequences” were recorded while sitting or lying back in different sofas and the ”lying sequences” comprised several different postures. The proposed algorithm was able to correctly categorize all thirty recorded sequences into the correct class, while the sequences were non-trivial.

### IV. DISCUSSION AND APPLICATIONS

Although the monitoring approach proposed here is targeted to the monitoring of elderly, the pose recognition method was evaluated by monitoring a younger subject. As it is generally believed that the age of the subject does not influence the reliability of static pose recognition, we preferred not to engage elderly in the validating as repeated transitions from lying down to standing would be very demanding to them. However, in the near future our activity analysis framework will be validated in the geriatric department of a hospital.

We believe that simply by investigating the trajectories defined by the 3D position of the monitored subject’s center, several interesting properties of the subject’s general condition can be derived. From coarse to grain, we currently monitor the total distance walked per day, the total time in bed, in the sofa and walking, the total amount of walking, sitting and lying episodes or transitions and the amount of falls. All this information is easily derived from the 3D posture.

For instance, a fall detector could already be implemented by simply ringing an alarm whenever the subject’s pose is categorized as “lying down on the ground”. Although this approach might be quite reliable in a nursing home setting, it would probably be prone to high amounts of false positives in a home environment. Typically, when you lie down on the ground in order to look for your wedding ring underneath some piece of furniture, the fall detector would ring the alarm. In order to overcome this, we have proposed an elegant learnable context model [7].

### V. ETHICAL ISSUES

Home monitoring projects employing camera technology as the one described in this document raise ethical issues. In general, we have to ask ourselves whether being continuously watched by a camera is the price we want to pay for medical
monitoring goals. Other important issues are the storage and access of the monitored data. The general architecture of the project we propose consists of two components: at the client side, there is the camera and possibly other sensors, together with a processing unit. At the server side, the monitored data is stored in the electronic patient record, summarized, visualized and so forth. The client and the server are connected through a regular Internet connection. As the image processing algorithms are running at the client side, no images are transmitted from the client to the server. Only the calculated activity information (e.g. the position of the patient) is transmitted. This entails that it is guaranteed that the captured 3D images are never stored, nor at the client or the server side and that they are never transmitted over the Internet. As a result, doctors or other personal at the hospital running this home monitoring service can never watch the images acquired by the camera. With respect to the activity information which is transmitted to the hospital, it is indeed true that patients reveal some aspects of their privacy. Concerning privacy, there is no difference between the information captured by our client and the information gathered by infrared sensors, accelerometer, bed pressure sensors and so forth [2], [4]. The activity information is transmitted through secure communication and is stored in the electronic health record and is hence only accessible by the medical staff responsible for the patient’s care. Whether patients want to reveal information about their daily behavior at home for the benefit of their health, is an individual choice. Patients can only be enrolled in such a home monitoring program after giving their full consent. They should be able to quit the program at all time and they should be able to temporarily prevent the monitoring tool from transmitting data.

VI. CONCLUSION

This paper presents a very simple but reliable 3D pose recognition algorithm applied on images obtained from a 3D camera. We described the reliability of the pose classification recognition algorithm applied on images obtained from a 3D sensor. This work was performed as part of the BRUCARE project, which is granted by the Brussels Capital Region. The authors also would like to thank Ken Pintelon for his contributions to this work.

REFERENCES
