Abnormal human activities recognition: brief synthesis of vision based fall detection

Beddiar Djamila Romaissa
Research Laboratory on Computer Science’s Complex Systems, Larbi Ben M’hidi University
Oum El Bouaghi, Algeria
ad_beddiar@esi.dz

Nini Brahim
Research Laboratory on Computer Science’s Complex Systems, Larbi Ben M’hidi University
Oum El Bouaghi, Algeria

Denaib Ayyoub
a_denaib@esi.dz

Abstract—Various approaches of human activities recognition by vision in the context of elderly care have been proposed in the literature, particularly abnormal activities detection and analysis. This is due to the necessity to note that the observation of changes in the daily activities of individuals, enhanced and instrumental daily life activities are essential to detect medical conditions before they become critical. Taking care of the elderly has an obvious effect on their independence, their psychological status and their quality of life. Unfortunately, those proposals still ineffective and limited, due to the difference in lifestyle, habits and manners of performing a certain activity as well as the variety of activities from one field of application to another. Therefore, we present in this paper a brief synthesis and analysis of some existing works of fall detection.

Index Terms—human activities recognition; abnormal activities; synthesis; computer vision; elderly care; fall detection.

I. INTRODUCTION

Applications of video surveillance have attracted more and more researchers to propose new techniques of detection and analysis of human activities. Abnormal activities are studied in a particular way, given the information they provide for changes in the daily life of the elderly, their independence and the early detection of certain types of diseases before they become critical. Thus, their remote monitoring and assistance. Abnormal activities recognition starts to be a very interesting field of video surveillance because of the amount important information that is provided and which allow to afford an immediate intervention to preserve elderly lives. Falls, for example, are major health problem in elderly population that may cause loss of their independence, dehydration, hospitalization or moreover death because the person may stay immobile or unconscious for a long time.

By studying some of the state-of-the-art works of vision based abnormal activities recognition, we have been able to deduce some limitations that make the use of techniques and methods proposed for the detection of normal activities, inefficient. Abnormal activities recognition should be considered as a separate research axis and proposed methods should deal particularly with this kind of activities and their specific characteristics. Accordingly, this paper is a brief synthesis and analysis of some recent works of vision-based abnormal human activities recognition and a continuity of our work [1]. For this, the manuscript is organized as follows: In the second section, we define the abnormal activity as well as fall detection. We examine in the third section some approaches, proposed in the literature, of abnormal activities detection, in our case we are interested to fall detection. For each of the proposals, we discuss the motivation behind its suggestion, its stages and its validation mean. We summarize afterwards, each proposal in one diagram to finish with a conclusion where we recap our study.

II. ABNORMAL ACTIVITY AND FALL DETECTION

According to [2] an abnormal activity is defined as any out-of-ordinary and non-usual activity that may expose a person or group of people to danger in a particular context. An activity is considered abnormal when it is atypical and consists of undesirable acts. An abnormal activity is also an activity that occurs rarely and has not been expected in advance [3]. Furthermore, one can define fall as unintentional or sudden change of position of the body from an upright, sitting or lying to a lower inclining position [4]. It is an event that results in a person coming to rest on the ground or any lower level involuntarily [4]. A fall event can be decomposed into pre-fall phase, critical phase and post-fall phase and recovery. Moreover, fall detection consists in recognizing a fall among all daily life activities.

Fall detection systems are categorized according to the used device [5] into: non-computer vision based and computer vision based. The first category compromises wearable sensors, sensitive floor tiles, simple sensors and mobile devices. Wearable sensors should be worn all the day which may raise autonomy problem. The person may as well forget to wear them and they require batteries to be replaced or recharged regularly. Sensitive floor-based systems are more expensive and less robust because moving objects may trigger false alarms. So, non-computer vision-based methods are more
and more abandoned. The second category aims to detect inactivity, analyze changes of the moving human body and the 3D head motion. Video-based systems do not require to be worn, are easy to be mounted and they cover great surfaces which make them more suitable for elderly fall detection.

### III. Abnormal activities detection

Several researchers have suggested new methods of recognizing abnormal human activities, especially based on normal human activities recognition. We are interested in this overview to fall detection. In the following, we summarize some of what have been presented in the literature. Researchers are increasingly interested to propose automatic monitoring systems in order to avoid the use of manual video surveillance systems that must ensure the follow-up of people over the long term and therefore manipulate too much data. Manual surveillance is not practical, inefficient and causes a difficulty of storage and display. In most surveillance systems, videos are obtained from multiple cameras, and are then compressed by the internal processor to be transmitted to the video server in the interest of storage or display.

The authors of [6] proposed a learning-based real-time system for detecting abnormal behaviors from surveillance videos. They used for that two approaches: an ACP approach to select the characteristics followed by an SVM classification and an optical flow approach. Their proposal differs from the state of the art as it allows detecting several abnormal behaviors using the same framework. It uses 20 primary components instead of detecting the parts of the body on contours. In addition, it does not use the position and velocity of the body parts for learning because their precise values can not always be obtained, and this generates false alarms. This system uses a background pattern or optical flow (in cases where the background is predominant) to detect the background from the foreground. A segmentation is then carried out using a projection histogram to divide the blobs containing several people according to the individuals. This segmentation is followed by body parts labeling, a distance or color-based follow-up of several people and finally a behavioral analysis was performed by the mean of the two approaches as mentioned above: Behavior analysis using learning approach and behavior analysis using optical flow approach. In the first case, the authors performed a preprocessing, they extracted contours from the centroid, calculated distances, used DFT to select 20 coefficients for each blob, extracted the most important features using supervised PCA, and finished with an SVM classification. The second approach is used in case of difficulty of obtaining the foreground image if an occlusion is observed. For that, they calculated the velocity of each pixel, carried out a filtering by 3x3 template followed by a thresholding to apply at the end the ordinary methods of optical flow for the classification. To test the performance of their system, they implemented experiments on a set of 625 examples. For each one, 4 frames are taken, while using 3 connected SVMs. Fig. 1 represents a diagram of their approach.

The authors of [7] suggested a new real-time solution for abnormal movement detection. It is dedicated to modern architectures of video surveillance where the limited computational force available is close to the quality of the camera for compression and communication. The algorithm is based on Macroblock motion vectors (global characteristics), which are generated in case of video compression, instead of pixels. This makes it possible to reduce the input data rate and thus facilitates the application in real time in limited calculation platforms. Their proposal differs from the others in the state of the art in that it is based on motion vectors; that it avoids segmentation and follow-up to ensure a semantic interpretation and that it lacks of assumptions about normal movements. The basic idea of this system is to extract the motion characteristics in order to estimate, by learning, the probability distribution of normal activities and to deduce consequently the abnormal activities which constitute improbable values in the processing phase. For this purpose, the authors followed two main phases (as presented in Fig. 2): the first phase involved the extraction of motion vectors from video, while the second phase consisted in extracting the motion characteristics from the previous vectors. In the first phase, a spatial and temporal compression of the video made it possible to obtain corresponding motion vectors is performed. These vectors are used as input of the second phase, an estimate of their probability distribution is made during learning, it was then compared to a statistical model and for each image, a feature vector is associated. Next, the authors selected the approximation histograms of the probability density functions of the motion vectors obtained to ensure the real-time quality of the calculations and finally decide on the activities. The proposed system is limited by several requirements, such as the reliability of regularly detecting events while reducing false alarms; The difficulty of efficiently characterizing the normal movements in order to discriminate between normal or abnormal movement and finally the difficulty of establishing a compromise between the quality of the camera and the limited calculation capacity available. In order to validate their proposal, the authors established an experimental that consists of following a pedestrian using a digital video camera Sony TRV 900. The system developed in C++ runs under Windows, Pentium 4 and the frames are acquired with a frequency of 75 frames Per second. Each video is 50 minutes of which 41 minutes are used for learning and 9 minutes for the test.

The idea of the Internet of things is to allow several intelligent objects with different capacities to interact, communicate with each other and also cooperate with their neighbors in order to achieve common goals. The ability to provide services immediately, using this infrastructure has prompted researchers to use it for video surveillance systems. Particularly, to ensure assistance and supervision to elderly isolated individuals in society in order to enable them to live alone for as long as possible. This need created a serious social problem because of the considerable demographic change and the growing number of these elderly in developed countries.
(e.g. Japan, USA etc.). In addition, the lack of support from their families is, also a determining factor.

Since some services of the Internet of things can be provided immediately and that some diseases can be deduced by observing the activities of the elderly, the authors of [8] have implemented an Internet of things infrastructure for an abnormal human activities detection system for the purpose of supervision and assistance of the elderly. The proposed situation-aware system for the analysis and detection of abnormal activities of the elderly (an important field in the ubiquitous computer research paradigm allowing to detect the situation around the user, provide him with the necessary services which can even change the environment) was based on SVDD (support vector data description). It is composed of u_tile network sensors such as: pressure sensors and RFID antennas, making possible detecting the position and trace of elderly and a smart Plug developed by NTT to detect and control the status of the various devices of the house. This system, which used u_tile network sensors (locating human position compared to objects), included the extraction, analysis and classification of characteristics, the detection of the situation around the elderly, the creation of several spheres based on the different situations with several characteristics and finally the recognition by SVDD of the abnormal activities in the situations. Each situation was represented by: the zone, the status of the devices, actions of the elderly and its environment.

To validate their system, the authors first made a learning of the situations of a subject and then compared them with those detected by the system. For that, they used 5 subjects, and for each subject they recorded 40 minutes of which 5 minutes...
were to introduce the individual, 10 minutes to designate his usual activities and 24 minutes to designate his one-day activities. Fig. 3 demonstrates their proposal.

The growing need to fight terrorism in banks, airports, elevators and public places and the advent of low-cost sensors with ease of installation and use that can handle different types of data are motivations for researchers to propose systems for the recognition of human activities for video surveillance purposes. In particular, sensor networks have recently been used extensively in video surveillance because of their ease of installation, use in various environments, lower costs, efficiency and real-time quality. On the other hand, it is necessary to note that the observation of changes in the daily activities of individuals, enhanced and instrumental daily life activities are essential to detect emerging medical conditions before they become critical.

Therefore, a home surveillance system based on network sensors (Refer to Fig. 4), using RFID sensors is proposed in [9]. The basic idea of this system was to collect the daily life activities of the elderly, to construct an HMM model and to classify the activities into normal or abnormal using SVM. The proposed system is implemented by installing the RFID antennas at the entrance of the rooms and by attaching the passive tags to persons and objects, which made it possible to start recording after the tags have been detected by the antennas and to trigger an alarm for immediate intervention in case of necessity. The proposed system consisted of two main phases: the preprocessing phase and the detection phase. In the first phase and from a sequence input of normal activities, the authors applied the Baum Welch algorithm. The outputs of this algorithm and the test sequences were the inputs for modeling normal activities by applying the Viterbi algorithm, by modeling HMM sequences, applying the Forward algorithm and by processing the feature vectors. This model of normal activities is then used in the second phase by SVM to detect abnormal activities. To test the performance of their proposal, the authors used an RFID reminder system to collect the daily life activities data from two women and three men for 13 days, thus building a real data set. Next, they used a Matlab HMM toolbox to process data from the collected daily activities data.

The authors of [10] proposed an automatic method for detecting patterns of abnormal behavior. The method is based on Markov chains model (MCM) and is built on a mobile sensitive contactless monitoring platform MSSMP. It is composed of several steps (as shown in Fig. 5), the first of which consisted in transforming the triplets representing the human activities into singletons, K-means algorithm is subsequently used to analyze and classify the data collected from each room of the intelligent house. The next step was to estimate the probability distributions of each cluster to construct the transition probability matrix and finally to use a ratio to decide if an activity is abnormal or normal and trigger an alarm that can be canceled by the observed elderly in case of error of the system. On the one hand, the validation of this system is carried out by creating a real dataset where the activities of a 76-year-old man are captured during 128 days using infrared sensors. On the other hand, the proposal is compared with the Levenshtein distance algorithm.

A system of automatic behavior analysis using IR motion sensors (u-sensor) has been proposed in [11], with the aim of assisting independent elderly wishing to live alone and improve their health. The proposed solution aims to ensure the privacy of the elderly, to provide continuous information about their activities and to warn their relatives and caregivers by means of an alarm triggered in case the person observed is in an abnormal situation. This system is composed of several steps that aim to monitor the daily life of the elderly. It estimated their normal activities and detected their abnormal activities automatically without disturbing them. The captured data is therefore preprocessed in order to extract the most
A video-based abnormal activities recognition system using R-transform and KDA whose goal is to take care of elderly has been proposed in [12]. This system concerned 6 types of predefined abnormal activities: forward fall, backward fall, chest pain, faint, vomit and headache. The system comprised five main steps: the captured data first passed through a preprocessing, where the background and the silhouette are segmented using the background subtraction, a thresholding is applied to the silhouette to facilitate the extraction of its characteristics (Refer to Fig. 7). R-transform is used at this level to extract directional features. The third step is used to calculate the feature vectors that are used by LDA or its non-linear extension KDA to improve the separation of activity classes. The next step consisted in generating code blocks and thus sequence symbols for each activity sequence with K-means and finally HMM is used for training and recognition. The proposed system is validated using a customized dataset containing 6 abnormal activities of 6 people (4 men and 2 women). For each activity, the authors took 10 sequences and 15 key silhouettes, which gave 150 silhouettes and therefore 900 for each person. Fold cross validation was established where the activities of all subjects except 1 on each iteration are captured for training and the remaining activities are used for the test; The procedure is repeated until everyone is involved.

Deformation of the body shape, the centroid, the perimeter and the principal axis of the silhouette are significative features for fall detection. This is due to the fact that changes may happen drastically and rapidly during falls than during normal activities. Based on this, The authors of [4] proposed to detect falls using motion history images MHI and the shape of the body. The silhouette of the human body is extracted, using background subtraction methods, from videos captured using a webcam. Some preprocessing operations are then carried out involving Gaussian/Median filter and morphological operations. Afterwards, the centroid of the detected silhouette is calculated and finally activity analysis is implemented to identify falls using a threshold. Fig. 8 shows the phases of the proposed method. To evaluate the performance of their approach, the authors use the Le2i fall detection dataset [13] which contains fall video sequences. Similarly, the authors of [5] proposed a video surveillance fall detection system based on silhouette deformation analysis during and after the fall. 2D and 3D information of the human body are used. The authors extracted the background, and the human body shape is determined using an ellipse. They used afterwards the motion history image for motion quantification to analyze the human shape variations and orientation. Hence, falls are recognized as lack of motion for a while (as presented in Fig. 9). To validate their approach, the authors used a real-world dataset composed of 24 daily life activities and 17 falls.

In addition, the authors of [14] proposed a fall detection system based on multiple shape features and motion. The silhouette is extracted using background subtraction methods. Then, morphological filter and component labelling CC are used to improve the silhouette. To extract features, the authors used Finite State machine. The extracted features (Vertical velocity of the head, area, height/width ration, ferret diameter, orientation, Procrustes distance) are fitted to 3 classifiers: SVM with RBF kernel, K-Nearest Neighbor and Backpropagation Neural Network to classify postures. Finally, the authors tested the inactivity to detect a fall posture. Fig. 10 represents a diagram of their approach.

A novel visionbased fall detection approach that analyzes an extracted human body using descriptive human postures
is proposed in [15]. The system is composed of three major steps: human body extraction, human posture description and fall event recognition. To extract the body, the authors extracted the background using a background model, applied a morphological procedure and surrounded the body with an ellipse. Then, postures are represented using multidirectional statistical analysis based on Normalized Directional Histogram NDH. Dynamic (change rate of NDH) and static (statistical moments of NDH) features are extracted and are fitted to
an Acyclic Graph SVM for classification (Refer to Fig. 11). To validate their approach, the authors used a public multiple camera fall dataset which is composed of 736 samples of falls and daily life activities recorded with 8 cameras in 24 different situations.

In [16], a machine vision-based system was designed. As a preprocessing step, the median filter is applied to images and invalid pixels are removed, then the authors eliminated the background and extracted the contours. Later, the human body is surrounded by an ellipse and the authors suggested to extract the shape feature which gives the 2D position of the human body. SVM is used then to classify postures and the authors computed as well the distance from center of the ellipse to the floor to distinguish between falls and activities that are close to fall posture (as shown in Fig. 12). To evaluate the performance of their proposal, they used a dataset composed of 70 depth map videos [17] captured using a Kinect camera, 30 of them are different fall postures while the rest correspond to daily life activities. The videos were recorded in different rooms and different subjects have participated.

A Video-based fall detection system based on posture analysis and SVM was proposed in [18]. The authors extracted the silhouette from videos using background subtraction and applied some preprocessing operations to enhance the extracted human silhouette. As a second step, they extracted the shape and the movements features and used SVM to recognize postures (standing, sitting, bending and lying). Falls may occur before a lying posture, so they calculated the occurrence of the lying posture, if it persists among 12-25 frames, they concluded that it was a falling posture. To evaluate their approach, they used 3 datasets [19]–[21]. Fig. 13 represents stages of their approach.

A video-based fall detection system in indoor environments using Convolutional Neural Networks, CNN was proposed in [22]. The system is based on the fact that falls are specific changes of poses and tried to determine sequence of poses that may lead to a fall. After falling, lying is the most frequently pose so detecting falls may be performed by identifying the
lying posture. However, the system should distinguish between normal posture such as sleeping on the bed and lying on the ground after a fall. For that, the background was subtracted from the input videos in order to extract the human silhouette (as shown in Fig. 14). Then the extracted silhouette is fitted to the CNN that has the same architecture of the VGGNet in order to extract features and recognize activities. Finally, the speed of lying down is used as well to distinguish falls from other lying postures. To test the performance of their system, they created a dataset recording 5 activities from 8 different views (standing, sitting, lying, bending and crawling), performed by 5 subjects in 5 different indoor set-ups using the Kinect device.

Since 3D locations of major body joints may provide significative kinematic information, the authors of [23] suggested to extract body joints from depth cameras and used them as features for fall detection. They proposed a deep-learning-based approach using LSTM and transfer learning. Two networks are used in the proposed approach to implement transfer learning which allows extracting knowledge from one or more source tasks (the first network) and apply it to a target task (the second network). Skeleton sequences are fed to train a multiclass LSTM, and weights of the two first layers (relative to general features) are copied and transferred to the second network (2-class LSTM). After training the second network, results are refined to detect falls. To train the first network, the NTU RGB-D dataset is used. It contains 56880 videos of 60 actions performed by 40 subjects and to train the second network, the authors recorded a dataset of 11 actions performed by two people using Kinect device. Fig. 15 represents a diagram of their approach.

IV. CONCLUSION

In this paper, a brief synthesis and analysis of recent works on vision-based fall detection has been proposed. A lot of researchers are interested to fall detection because it is a major health problem in elderly population that may cause them death. However, proposed methods in the literature are still limited and inefficient. One can mention that fall detection databases contain simulated falls, because creating a real-world falls dataset is very time and resource consuming. These simulations have to be taken into consideration and should be adapted to work even for real world falls. Another limitation is the furniture of the room. A person can fall on a bed or a sofa and the system may not be able to recognize the posture as a fall because the person still at a height from the ground. We can conclude that a good fall detection system should ensure the safety and the privacy of the elderly as well as the quickness of intervention when a fall occurs. In this present paper, we have defined firstly the meaning of an abnormal activity and have analyzed, then some recent approaches of fall detection, where we explain their aims, their stages and the means used for their validation.

REFERENCES


