Master's Thesis
석사 학위논문

Intelligent Video Surveillance: Behavioral Detection of Falls Based on Double-Layer Support Vector Machine

Hee Jung Yoon (윤 희 정)

Department of Information and Communication Engineering
정보통신융합전공

DGIST

2013

Master's Thesis
석사 학위논문

Intelligent Video Surveillance: Behavioral Detection of Falls Based on Double-Layer Support Vector Machine

Hee Jung Yoon (윤 희 정)

Department of Information and Communication Engineering
정보통신융합전공

DGIST

2013

Intelligent Video Surveillance: Behavioral Detection of Falls Based on Double-Layer Support Vector Machine

Advisor: Professor 손 상 혁
Advisor: Professor 박 태 준
Co-Advisor: Professor 문 상 준

by

Hee Jung Yoon
Department of Information and Communication Engineering
DGIST

A thesis submitted to the faculty of DGIST in partial fulfillment of the requirements for the degree of Master of Science in the Department of Information and Communication Engineering. The study was conducted in accordance with Code of Research Ethics[1].

01 . 04 . 2013

Approved by

Professor 손 상 혁 (Signature)
(Advisor)

Professor 박 태 준 (Signature)
(Advisor)

Professor 문 상 준 (Signature)
(Co-Advisor)

1) Declaration of Ethical Conduct in Research: I, as a graduate student of DGIST, hereby declare that I have not committed any acts that may damage the credibility of my research. These include, but are not limited to: falsification, thesis written by someone else, distortion of research findings or plagiarism. I affirm that my thesis contains honest conclusions based on my own careful research under the guidance of my thesis advisor.

Intelligent Video Surveillance: Behavioral Detection of Falls Based on Double-Layer Support Vector Machine

Hee Jung Yoon

Accepted in partial fulfillment of the requirements for the degree of Master of Science

01 . 04 . 2013

Head of Committee____________(Signature)
Prof. 손 상 혁
Committee Member____________(Signature)
Prof. 박 태 준
Committee Member____________(Signature)
Prof. 문 상 준

MS/IC 201123008

윤 희 정, Hee Jung Yoon, Intelligent Video Surveillance: Behavioral Detection of Falls Based on Double-Layer Support Vector Machine. Department of Information and Communication Engineering, 2012, 36 p.
Advisors: Professor Son, Sang Hyuk and Professor Park, Taejoon
Co-Advisor: Professor Moon, Sangjun

ABSTRACT

The elderly population is expected to more than triple by 2050 in the United States alone. This indicates the growing need of medical innovations that are expected to deliver convenient, user-friendly, and intelligent health care in the home. In particular, the use of computer vision and artificial intelligence offers a new promising solution to analyze human behavior and detect unusual events. We propose a novel method to detect unintentional falls, which are one of the greatest risks for seniors living alone. Our approach is based on a machine learning technique, in which we use a human skeletal joint data from a Red Green Blue Depth (RGBD) sensor, Kinect. To the best of our knowledge, this is the first fall detection research utilizing machine learning mechanism with joint information. After building a solid definition of various types of falls, the system is trained using Support Vector Machine (SVM) in two separate layers, which we call Double Layer SVM (DLSVM). Our evaluation results show that our proposed system can efficiently detect different types of falls facing various directions from the camera and is capable of accurately distinguishing an actual fall versus a fall-like behavior.

Keywords: activity detection, behavior recognition, fall detection, video surveillance, support vector machine

Contents

List of Figures

Ⅰ. INTRODUCTION

Today, those who are aged 65 and older exceed 35 million, representing 12.9% of the U.S. population [1]. As the Baby Boomer generation ages, the need for elderly care significantly increases, especially in emergency situations when they live alone or are left at home unsupervised. There are many risks that can cause harmful injuries and even cause deaths in these situations. Of many unintentional home injuries, falls are the most common cause of injury amongst the elderly population, resulting in hip fractures, severe head injuries, and joint dislocations. Falling is also the leading cause of accidental death in this age group. An estimated 10,000 elderly die each year from accidents related to fall. These incidents have a tremendous impact on medical cost. In U.S., more than 20 billion dollars are spent annually on seniors treating injuries caused by falls [2]. The Center for Disease Control and Prevention (CDC) estimates that by the year 2020, the annual direct and indirect cost of fall injuries is expected to reach 54.9 billion dollars [3]. Furthermore, studies show that 30 to 50 percent of elderly has a fear of falling down [4]. The fear itself increases the risk of falls, as well as resulting decreased activity isolation, and depression. In order to provide a convenient, cost effective method of detecting a fall for elderly individuals in emergency situations, we specifically focus on an efficient fall detection method using a surveillance system.

Previous studies for fall detection are limited to utilizing devices such as, help buttons or accelerometers [5]. However, it does not guarantee a fallen person to be notified to others when that person cannot reach for the button or forget to wear the device. There are also recent studies being done on external sensors, such as, floor vibration detectors [6]. These devices can be resolving solutions, but they require complex setup with floor dynamics, and research is still in their infancy. Additionally, another way to detect falls is using a video camera system with computer vision algorithms. For example, 2D cameras are used to measure the bounding box ratio of the person and detect the velocity of the fall [7]. However, the detection is not sufficient to determine an actual fall in comparison to a daily activity.

We present a novel approach to achieve reliable, real-time detection of unintentional falls by using an inexpensive Red Green Blue Depth (RGBD) sensor called Kinect. One of the major challenges to our proposed approach is specifying what constitutes many different types of fall of all directions from the camera versus fall-like behaviors (e.g. lying down, sleeping, etc.) and normal daily activities (e.g. walking, cleaning, etc.). Our main contribution in this paper is to overcome this challenge while keeping the accuracy rate high and false positive rate low.

Generating a catalog of falling behaviors is the first step of our approach. To address the challenge of behavioral detection, we specify the behavior with (x, y, z) skeletal coordinate data and provide a compact representation of the falling behavior. After building a solid definition of various types of falls, we train the system using a machine learning classifier called Support Vector Machines (SVM) in two separate classification layers: one for body shape detection and the other for motion sequence detection, which are collectively referred to as double-layer SVM (DLSVM). By combining our proposed DLSVM algorithm with Kinect, we implement a novel system called Intelligent Fall Detection System (INFADES). INFADES is designed to detect all types of falling behaviors and is capable of distinguishing an actual fall versus a fall-like behavior in an accurate manner. Moreover, our performance evaluation results demonstrate that its robustness is significantly increased to 97% accuracy. The methodology that we developed for fall detection system can be extended to a generalized behavioral video surveillance.

II. BACKGROUND – CURRENT FALL DETECTION APPROACHES

1. Accelerometer

The damage from unintentional fall is increased if the person cannot call for help in a fallen situation. One common approach to solve this issue is by using wearable device such as accelerometer [5] which is an electromechanical device that measures acceleration forces. As part of the process of measuring acceleration, the accelerometer yields information of vibration, inclination, and shock that contributes to the detection of falls.

As part of the process of measuring acceleration, the accelerometer yields information of vibration, inclination, and shock that contributes to the detection of falls. One of the more common examples of the modern accelerometer is known as Microelectromechanical System (MEMS) device [8]. This is a simple device that functions mainly with the cantilever beam and circuits that are designed for the purpose of detecting the presence of deflection senses. Gyroscope [9], which is a device for measuring body orientation based on the principles of angular momentum, are also used along with the accelerometer in order to detect falls at an earlier stage. Another accelerometer approach to detect falls is based on applying accelerometers onto Android smartphones [10, 11].

The major drawback of this technique is that the sensors are to be worn or carried everywhere, and require batteries which need to be recharged regularly for adequate functioning. If in any case such detectors are forgotten to be worn or recharged, which is very common for elderly individuals, no fall will be detected as the device is not triggered. DLSVM does not limit the use of (x, y, z) coordinate data from accelerometers as we know the extensive researches that are out there to support the motivation of unintentional falls. However, to address the concerns from this device, utilizing our system can significantly enhance the safety of the user in emergency situations.

2. Floor Vibration

In contrast to accelerometers, floor vibration-based fall detector [6] is another common research topic in detecting falls. The approach is based on the detection of fall vibration patterns that are significantly different from vibration generated by normal daily activity. The method using floor vibration sensors is inexpensive and preserves privacy; however, the performance is insufficient when the detection varies with different floor dynamics.

3. Video Based Fall Detection

Video surveillance offers a promising solution for automatic fall detection, as no body-worn devices are needed. Many researches have been done on detecting falls using image processing techniques. A commonly used method to detect falls is to analyze the ratio of the moving object's bounding box [12]. Due to problems of occluding objects, Rougier et al. [13] proposes to fit an ellipse on the foreground area, using a ceiling-placed, wide-angle camera. Furthermore, the works analyzed by Lee and Mihailidis [14] used a person silhouette and the 2D image velocity to detect falls with thresholds for normal zones of inactivity. Nait-Charif and McKenna [15] used ellipse representing the person and detected abnormal inactivates outside the normal inactivity zones such as chairs and sofas. Moreover, Rougier et al. [16], proposed a shape matching technique to track the person's silhouette along the video sequence. Despite the achievements that image processing has provided for visual fall detection, there are yet many challenges to overcome:

1) Video surveillance systems need to be robust to image processing difficulties, especially background modeling and object segmentation with issues of high video compression, shadows and reflection, and cluttered background. Problems occur when there is a poor job extracting all the relevant pixels required to track a person.

2) Visual fall detection is prone to high levels of false positive. Differentiating a falling behavior versus a normal daily activity is difficult to accomplish. As described by Foroughi [17], many current surveillance fall detection systems [18, 19, 20, 21] are unable to discriminate between an actual falling incident and an event when a person is merely lying down.

3) Tracking gait movements with the silhouette of the body shape is inefficient to capture the details of a motion. A definite feature of a body shape is required to monitor a movement and behavior of a falling motion.

III. SYSTEM OVERVIEW – INTELLIGENT FALL DETECTION SYSTEM (INFADES)

Utilizing video surveillance techniques supports the development of a ubiquitous environment with less human configuration. However, surveillance would require significant amount of knowledge or intelligence to process captured information. In contrast to using image processing methods to abstract human falling behavior, we use depth images that are captured from Kinect in order to work with 3D skeletal joint coordinate data. Kinect sensor is a low cost and easy-to-install solution to obtain depth images. Depth information is very effective for fall detection as it becomes possible to precisely track a person in the room. An important advantage is that, a person can be detected even in areas of low lighting.

Figure 1: Microsoft Kinect

1. Kinect

Kinect, shown in Figure 1, is a motion sensing input device released by Microsoft for controller free game play on the Microsoft Xbox [22]. Included in the device are both infrared (IR) and RGBD cameras. The depth image size has a maximum resolution of 640 by 480 and processes depth data with a frame rate of 30 fps. The RGBD camera has a maximum resolution of 1600 by 1200 (UXGA) to match the depth data with real images. OpenNI and NITE drivers, along with its libraries, are used for tracking individual skeleton joints with Kinect.

2. Architecture

3D joint data from Kinect is used for training and testing purposes on INFADES. We developed this standalone system by combining the OpenNI and LibSVM library, both an open source tool, in order to conveniently train and classify our data with DLSVM. Figure 2 shows the graphic user interface of INFADES. As shown on Figure 3, our algorithm makes use of the skeletal tracking API from OpenNI, and SVM training and classification API from LibSVM to construct DLSVM. The first layer classifies the body shape features and the second layer classifies the motion sequence of a fall; both of these layers are made more concrete in later sections.

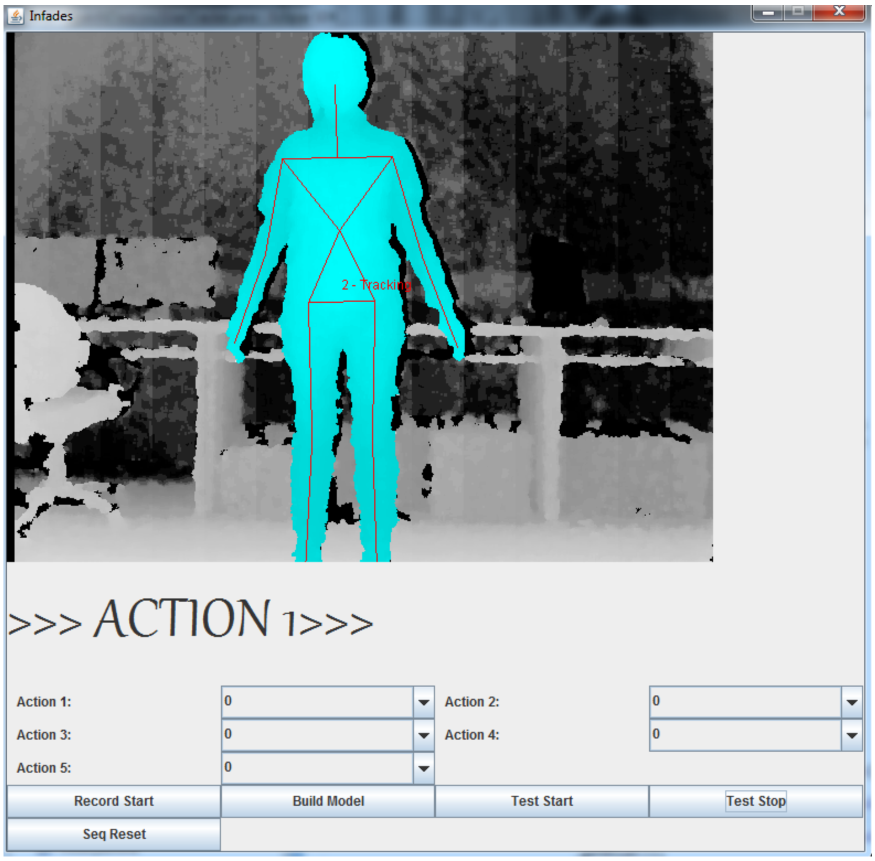


Figure 2: Graphic user interface of INFADES to train and test

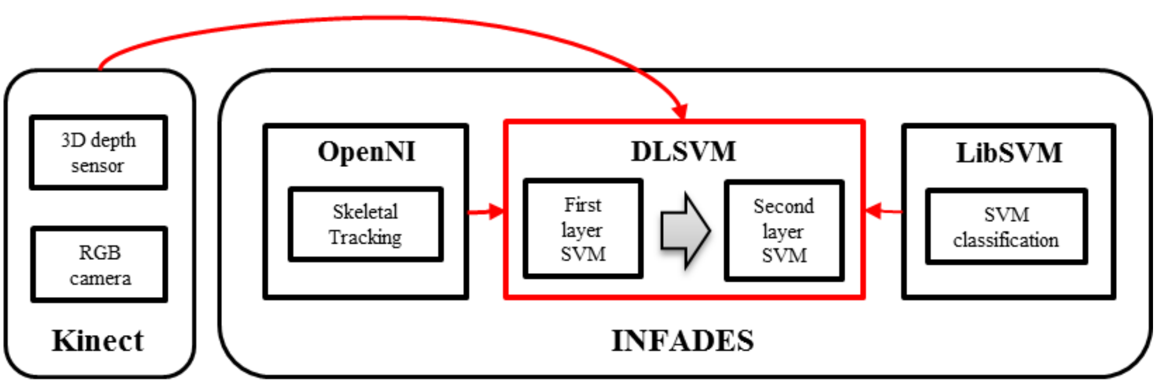


Figure 3: Architecture of system in deployment view of INFADES

IV. FALL BEHAVIOR

In this section, we describe the important factors that must be considered to define a falling behavior.

1. Types of Falls

It is crucial to first understand the most fundamental characteristics for a fall. However, doing so is challenging because there are many types of falls directed towards various different directions. To address this challenge, we gather a fixed collection of principal falls that capture most of the falling behaviors. We introduce four main types of falls according to downward direction: forward, backward, right, and left falls. We further divide these falls into four subtypes, each facing: front of camera, back of camera, right of camera, and left of camera. This yields 16 types of principal falls having similar characteristics of falling downwards. We use these principal falls in our training process so that the system can fully understand all different types of falls.

2. Skeletal Joints

When defining a fall, the 3D skeletal joints must be efficiently utilized so that the system can clearly understand the behavior. With this, we discover that all joints are not necessarily required when computing a fall. Using all the joints in the body will not only cause occlusion problems, but also build enormous data complexity because not everyone has similar body motions when falling down. For example, depending on how a person would fall, the wrist joint can be positioned near the body or swinging away from the body. Using joints that flexibly move in all different directions are difficult to compute with.

Therefore, the primary joints that are used for our fall detection are: head, right shoulder, left shoulder, and torso. The movements for these joints are fixed when yielding towards a downward direction and therefore can be used in order to define succinct description of the behavior. Moreover, since the primary joints are on the upper part of the body, it minimizes many issues with occluded objects. Figure 8d demonstrates how the primary joints move in an exemplified case of a fall with an occluded object covering the bottom half of the subject’s body. Additionally, four joints are used rather than utilizing a single joint to work with sturdier data that can help characterize and detect falls more efficiently. As shown in Figure 8, when one of the joints is not monitored, other joints can help support the detection.

3. Body Shape Features and Motion Sequence

A human behavior is defined as an observable activity or movement. In our case, a fall can be recognized by observing the detailed movement of how the person is falling over a short period of time. During the interval of a falling period, there are two important elements that are required to formulate the behavior: feature and motion sequence. Features are prominent characteristics that describe movement. Motion sequence, on the other hand, is a sequence of movement that describes a behavior. We believe that features reveal valuable characteristics that are used to generate motions sequences, which overall defines a behavior.

3.1 Body Shape Features

The features that we use are body shapes. A fall is not only characterized as a body yielding downward, but also as a body moving in a consecutive order due to the gravity pushing the body down in high velocity. For example, while slipping, the person loses balance of the body and abruptly falls down. We hence, categorize the body shapes that are created during the time the body hits the ground as features.

As shown on Figure 4, we divide the falling interval with five different body shapes, in which we describe with x, y, z coordinate data of the primary joints. For example, the first body shape is defined in a position when a person is standing or walking. In this case, the height of the primary joints would be at its highest position and would move parallel to the floor. The second body shape is in the position when a person is tilting towards the ground. Depending on the type of fall, the primary joints would yield around 45 degrees downward. The third body shape is a person in a leaning position. This body shape can be detected by the joints midway from the ground. The near floor position is the fourth body shape, which is characterized by a person just before a fall with the primary joints approximately 45 degrees up from the floor. The last body shape is in a fallen position, which is described by a person motionless on the ground. All joints would be near the floor. These five body shapes are used for the first layer of DLSVM later described in Section V.

3.2 Motion Sequence

Unlike human eyes, a system cannot detect a behavior merely with feature information. In other words, an actual fall cannot be detected simply with body shapes. The motion must be understood to recognize these behaviors. Hence, we construct a set of motion sequences utilizing flag mechanism of detected body shapes. Since the primary goal of our research is to detect the behavior of a fall, our system must first differentiate the falling behavior with other behaviors that are not associated with falls. We design our system to understand three specific behaviors: actual fall, non-fall, and normal behavior. Actual fall is when a person unintentionally falls down and may require assistant to get up or receive further care. The non-fall behavior is a fall-like behavior such as sitting, lying, or crouching to pick of an object. Normal daily activity includes behaviors such as walking, running, and doing chores. The sequences of these behaviors will be described in detail in the next section to coordinately explain the detection process for the second layer of DLSVM.

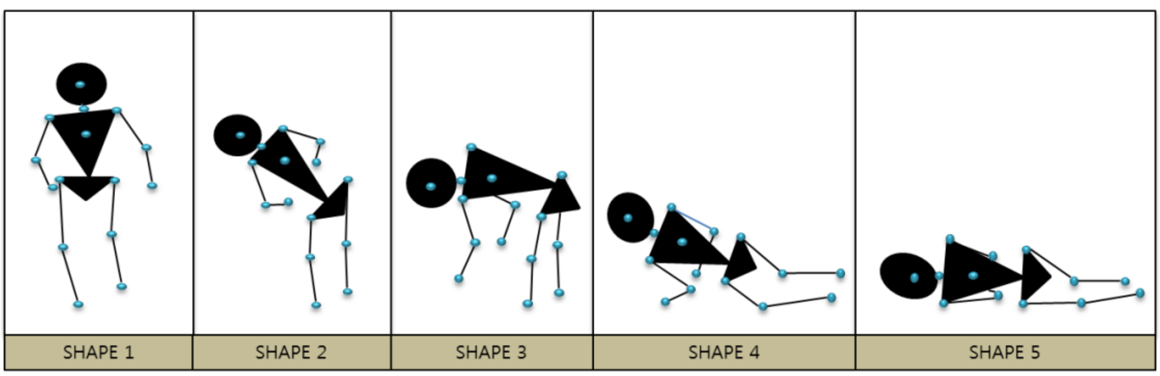


Figure 4: An example of the 5 different body shapes during the interval of a fall for the first layer of DLSVM

V. BEHAVIOR CLASSIFICATION BY MACHINE LEARNING ALGORITHM

1. Support Vector Machine

Support Vector Machine (SVM) is a promising statistical learning theory that has been effectively applied to various classification problems such as text categorization, bio-sequences analysis, object recognition, intrusion detection, and image retrieval. Based on its strong theoretical foundations by Vapnik [23] and

Joachims [24], the use of SVM reveals a large number of empirical successes among conventional classifiers such as naïve Bayes classifier, k-nearest neighbor classifier, and decision tree [25]. In particular, unlike SVM, classifier such as neural network is not suited to solve the classification problems with limited samples due to the proneness of over-fitting. Neural network uses empirical risk minimization, which minimizes the error on the training set. In contrast, SVM employs structural risk minimization, which minimizes the error on the training set with the lowest capacity. This difference overall equips SVM with greater ability to generalize [26].

We are given a labeled dataset (t(1), c(1)), …, (t(N), c(N)) to be used to train the SVM. The $t(i) \in Rn$ is an $n \times 1$ vector representing input data to the SVM. In our case, t(i) would symbolize the collection of (x, y, z) coordinates of the joints for our first SVM layer, and Boolean data for the second SVM layer. The c(i) is a class label representing the five classes of body shapes for the first SVM layer and 3 classes of behavior (i.e., fall, non-fall, and normal behavior) for the second SVM layer. R represents the real number and n shows the number of dimensions of the input to the SVM.

In general, SVM seeks to define a decision surface which provides the largest margin separating among the data classes while at the same time minimizing the number of errors. The resulting model is nonlinear, and the training is performed by the use of kernel functions. The kernel function k indicates a measure of similarity between a pattern t to be tested, and a pattern $\bar{t}$ from the stored training set. For example, given two pattern vectors t and $\bar{t}$ of dimension $n \times 1$, the kernel k can be represented as a canonical dot product:

$$k(t \cdot \bar{t}) = \sum_{j=1}^{n} t_j \cdot \bar{t}_j$$

where tj and $\bar{t}_j$ denote the jth element of vectors. Note that the dot product representation of kernels allows geometrical interpretation of the vectors in terms of skeletal joint data.

A key step in SVM is mapping of the input vectors t from their original input space Θ to the high-dimensional dot product space. There are two basic steps of SVM: (1) map the training data into a high-dimensional space via Φ, and (2) construct a hyperplane in the high-dimensional space that separates the multiple classes with maximum margin. Although there are many linear classifiers that can separate the classes, there is only one that maximizes the distance among the closest data points of each class and the hyperplane itself.

By solving a distance optimization problem given below, the solution to the linear hyperplane is obtained. The calculation results a classifier that works on previously-unseen examples, thus leading to good generalization. Although the separating hyperplane in the high-dimensional space is linear, it yields a nonlinear decision boundary in the original input space Θ. Moreover, it is important to note that the properties of the kernel function allow computation of the hyperplane without explicitly mapping the vectors in the high-dimensional space.

An input vector t can be represented as Φ(t) in the high-dimensional space. The computation of (Φ(t),Φ(ti)) in a high-dimensional space is reduced by using a positive definite kernel such that:

$$\left(\Phi(t), \Phi(t_i)\right) = k(t, t_i)$$

leading to binary decision functions of the form:

$$f(t) = \mathrm{sgn}\left(\sum_{i=1}^{N} c_i \alpha_i k(t, t_i) + b\right)$$

where the value b is a threshold of the decision boundary from the origin, and sgn is a signum function.

To apply our study using multiclass classification with m classes (m = 5 and 3 in the first and second SVM layers, respectively), the one-versus-all approach is used. The underlying basis of this approach is to reduce the multiclass problem to a set of binary problems, enabling the basic SVM to be utilized. Particularly for ti, there are m decision functions. The data ti then belongs to the class for which the above decision function has the largest value. Overall, classification for one-versus-all case is done by a winner-take-all strategy, in which the classifier with the highest output function assigns the class.

2. Double Layer Support Vector Machine (DLSVM) for Fall Detection

In order to achieve the detection of a fall as accurately as possible, INFADES utilizes SVM in two different layers, in which we call DLSVM. The first layer detects the shapes of the falling motions from individual video frames, while the second layer aggregates the results of the previous layer over a certain period of time to determine how similar the motions sequences of the shapes are to the fall behaviors.

While achieving the training set model for the first layer of DLSVM, Figure 5 shows that five classes are set as the body shape of the falls described in Section IV C.1. It is important to note that the shapes are abstracted from 16 principal falls mentioned in Section IV A, which captures all types of falls. Since the body shapes differ for each type of falls, 16 features are represented for a single class. Figure 6 demonstrates that sixteen features are specified in each class, in which are interconnected with features from other classes, making 80 different ways of falling down. The input data of the first layer consist of x, y, z joint data of the primary joints described in Section IV B. We train a total of 1440 cases of a realistic unintentional fall, considering various distances from camera and height of the subject. Once the trained model is built, INFADES compares the real-time testing data with the trained model and outputs a binary sequence of detection or non-detection of the body shapes, which is the input data of the second layer of DLSVM.

The training set model of the second layer is achieved separately using this binary sequence. Multiple sequences are manually categorized into three different classes of behaviors described in Section IV C.2 (fall, non-fall, and normal behaviors) and output the detection the observed behavior. We found that all falls end on the ground. Therefore, when shape 5 is detected from the first layer, the system flags an initialization of a fall until the second layer of DLSVM confirms a behavior. Figure 7 describes the process of detection by using an activity diagram in Unified Modeling Language:

- When one or two shapes are detected including the fifth shape, the behavior will be determined as a normal daily activity.
- When three shapes are detected including the fifth shape, the behavior will be determined as a non-fall.
- When four or five shapes are detected including the fifth shape, the behavior will be determined as a fall only if the time from the second body shape to the last body shape is within a given time. In any falling movement, there is always acceleration due to gravity. Therefore, we set a time of 700ms to measure an actual fall. The behavior will otherwise be considered as a non-fall. Note that this interval is an adjustable parameter.

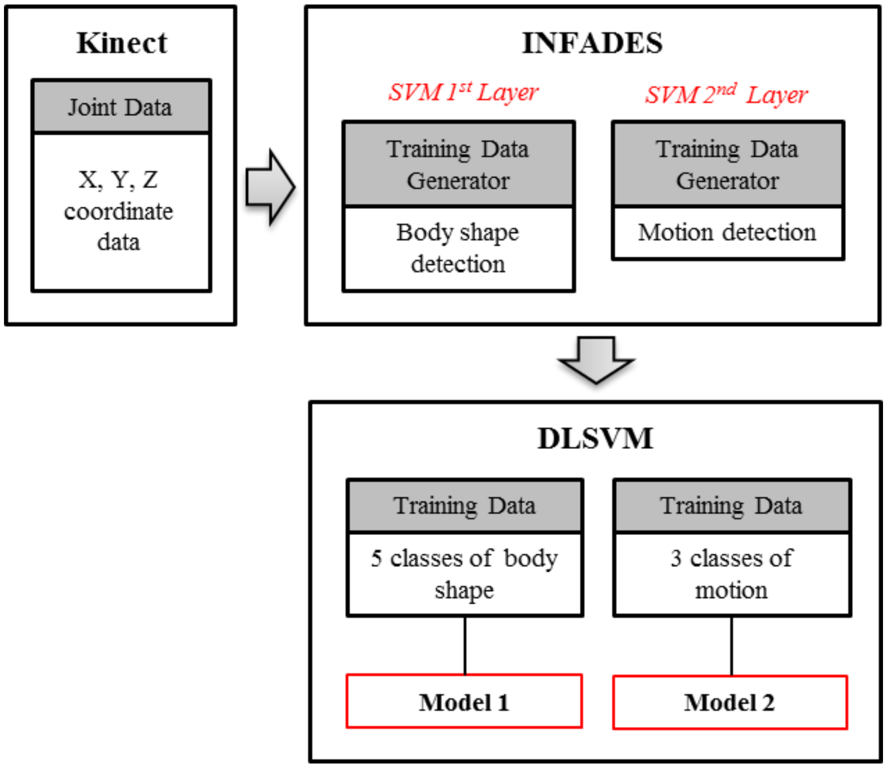


Figure 5. The training process: *x*, *y*, and *z*, coordinate data from Kinect is used by INFADES, more specifically in two layers of DLSVM to build “model 1” for the first layer and “model 2” for the second layer

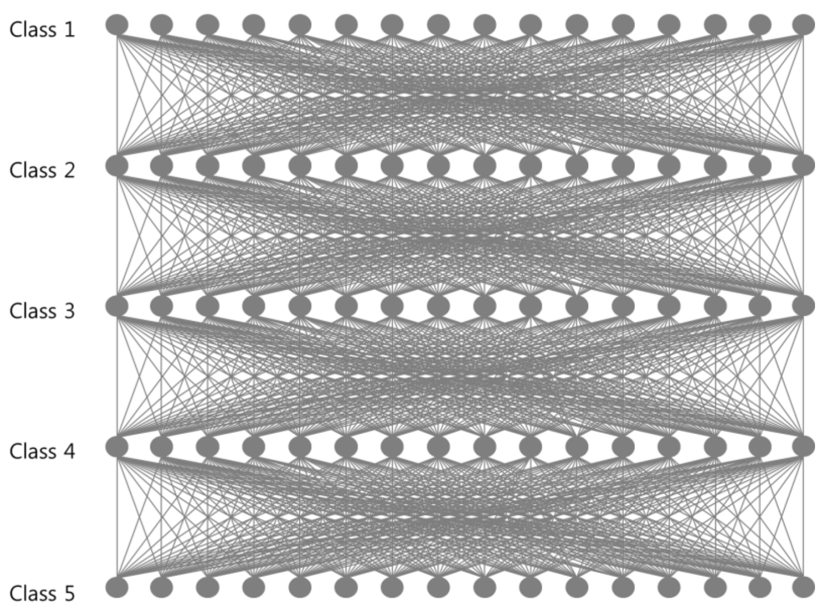


Figure 6. Features interactions: With 5 different classes (body shapes) and 16 different types of falls that are used, there is a total of 80 different ways of falling down

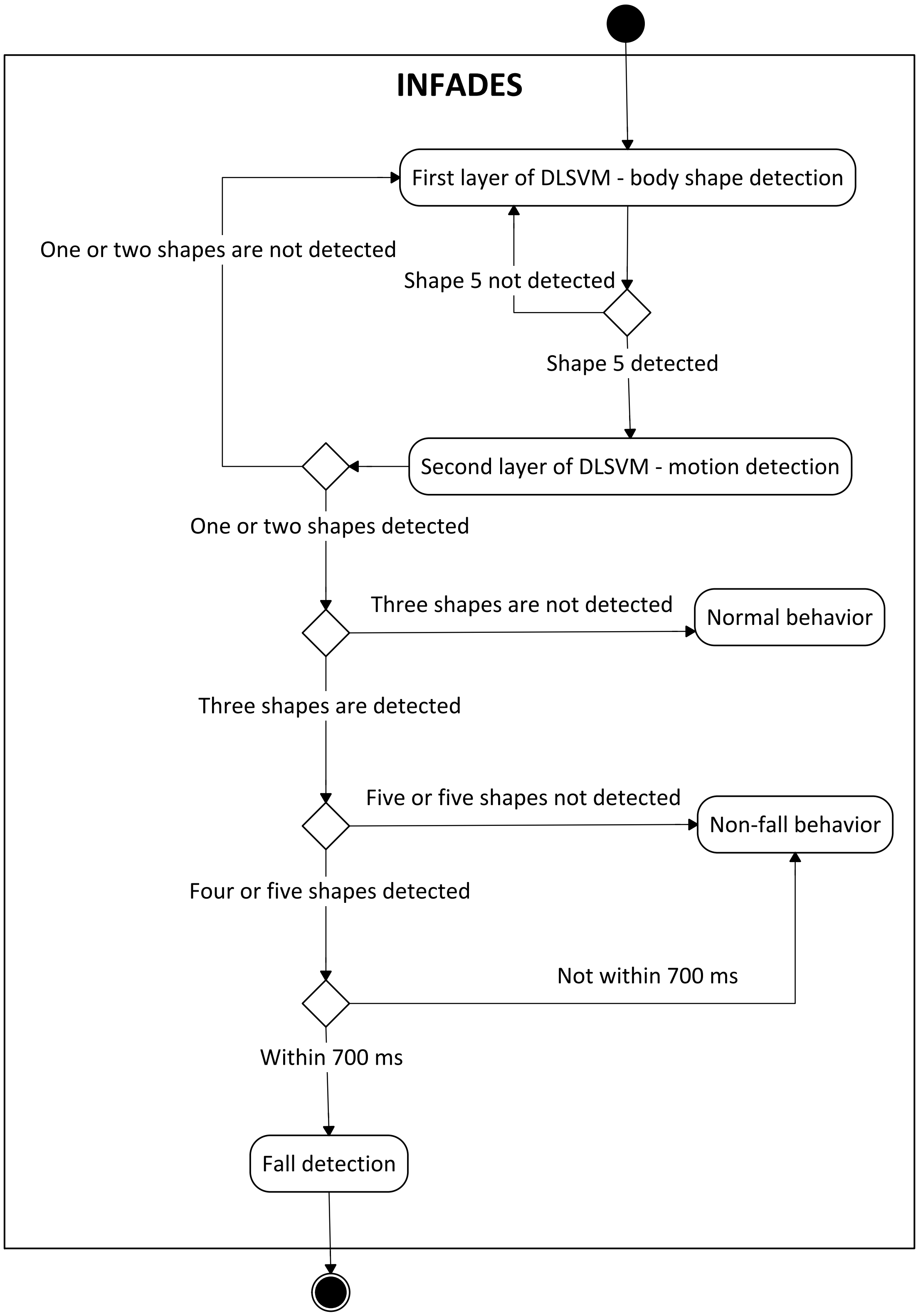


Figure 7. Activity Diagram of DLSVM: Process of classifying normal, non-fall, or actual fall behavior

VI. EVALUATION

1. Visualization of Joint Data

In order to evaluate the overall system performance, we start off by evaluating the pattern of the joints during exemplified cases of a fall, a non-fall, and a normal behavior. A fall with an occluded object is also shown to demonstrate the strength of our algorithm. Using only the y coordinate of the primary joints, we separately test the behaviors within a 10s interval. As shown in Figure 8, the patterns of the joints differ for all three behaviors. The constant, motionless pattern from Figure 8a shows a normal sitting behavior, while the gradual slope from Figure 8b reveals a lying down behavior. In contrast to these figures, the narrow slope and lack of movement after the slope from Figure 8c demonstrates an abrupt falling behavior. Figure 8d shows another example of a fall but with an occluded object covering the bottom half of the subject's body. Even though some joints are covered by an occluded object, a fall can still be detected with our system. This experiment demonstrates that we are capable of identifying different behaviors through the analysis of joint data patterns.

2. Accuracy of First Layer of DLSVM

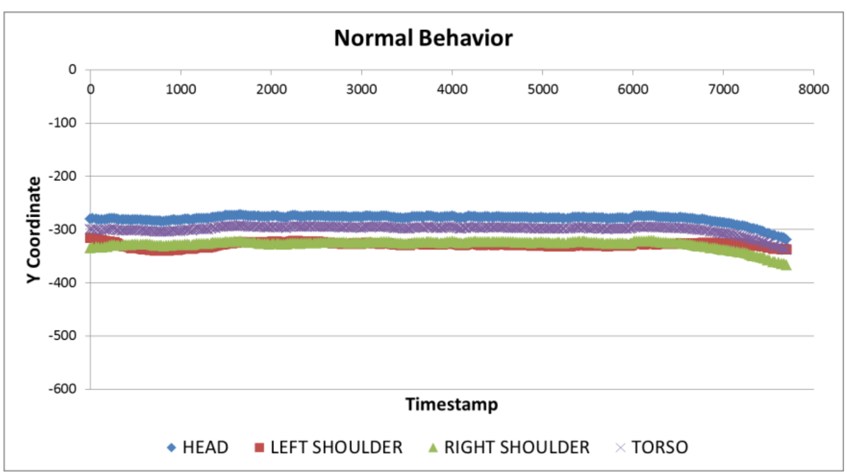
In formulating the sequence of the falling motion, the accuracy of detecting each body shape in the first layer of DLSVM is very crucial. Therefore, to quantify the accuracy, each shape is inspected individually using the parameter for probability estimates in the LibSVM library. In this experiment, 10 sets of 16 primary falls have been tested in an open setting. Falling was done on a 24 cm thick mat to allow realistic performance of falls. We collected a total of 160 data with an average sized male and a female in order to efficiently compute the accuracy of body shape detection.

An important point to consider is that when observing a single type of fall, gathering the data from all four directions is crucial since accuracy may vary from one side to another. In other words, by constantly changing directions of the camera to measure fallen data, accuracy may differ depending on how a person would fall. Therefore, for evaluation purposes, we place one Kinect on each side of the room to carefully extract data from every experimental falls.

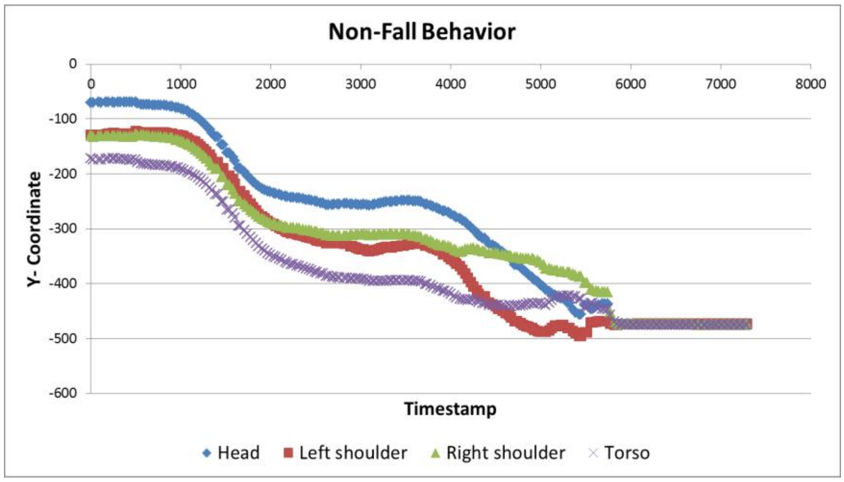
In demonstrating the results, we divide the experiment into four different sections by the direction of the camera. Figure 9a shows the accuracy comparison of body shapes from forward falls facing the camera, Figure 9b from back of the camera; Figure 9c from right of the camera, and Figure 9d from left of the camera. After computing the average accuracy value for each of the 10 testing sets, the results reveal that the body shapes are detected with high accuracy and proves that the first layer of DLSVM is efficient in triggering the second layer.

3. Performance Evaluation

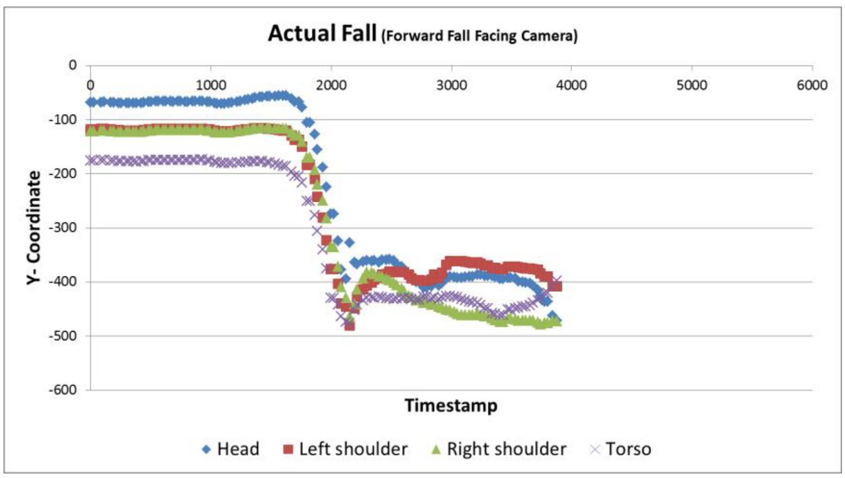
In order to evaluate the performance of our system, we separately tested two experiments: (1) fall experiment to evaluate the true positive and true negative values; (2) non-fall experiment to evaluate the false negative and false positive values. We worked with 250 random fall and 250 random non-fall cases with 5 different subjects (50 falls and 50 non-falls for each subject) of various height and weight in a realistic office setting. As in the previous evaluation discussed in Section VI B, both experiments were done on a 24 cm thick mat to allow realistic performance. As a result, the true positive value turned out to be 97.2%, the true negative value as 2.8%, the false negative value as 98%, and the false positive value as 2%.



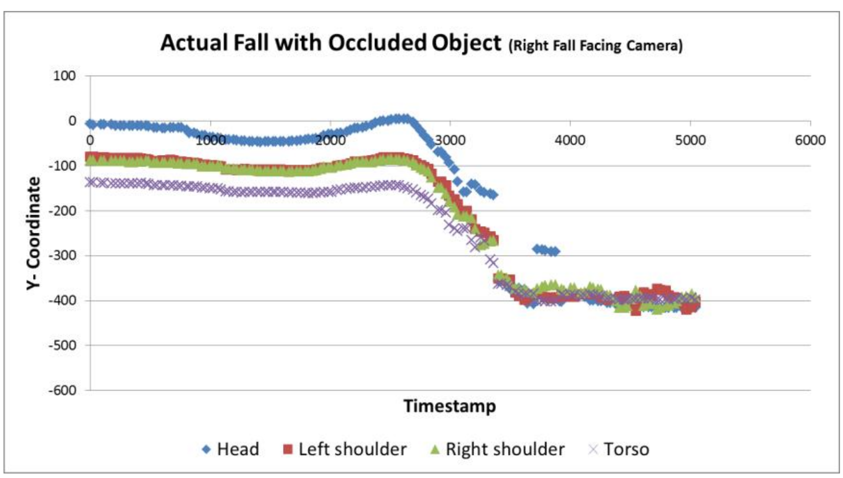
(a)



(b)

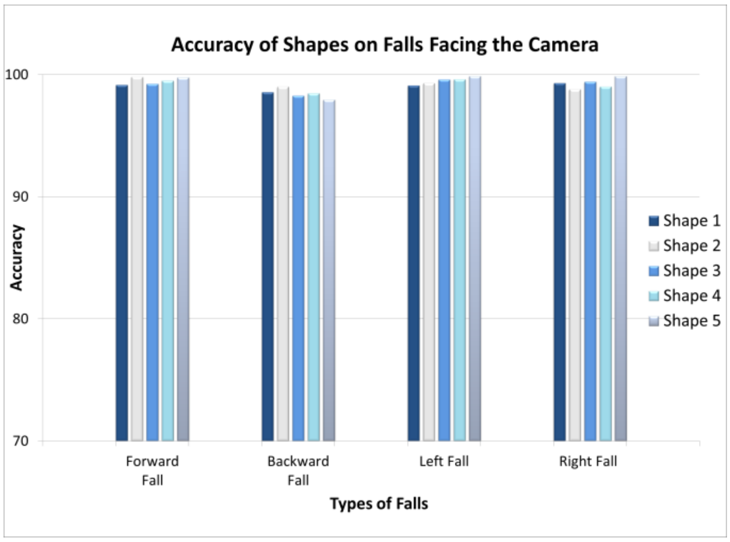


(c)



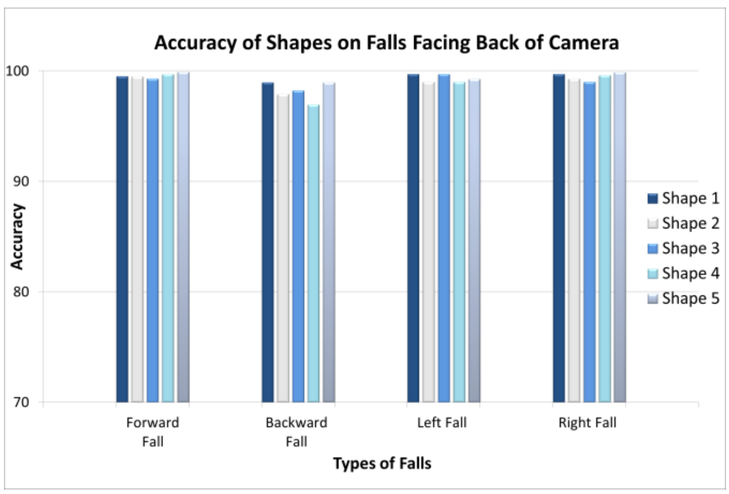
(d)

Figure 8. Joint data (head, left shoulder, right shoulder, and torso) visualization of different behaviors: (a) Normal sitting behavior (b) Non-fall going to sleep behavior (c) Actual fall (d) Actual fall with an occluded object



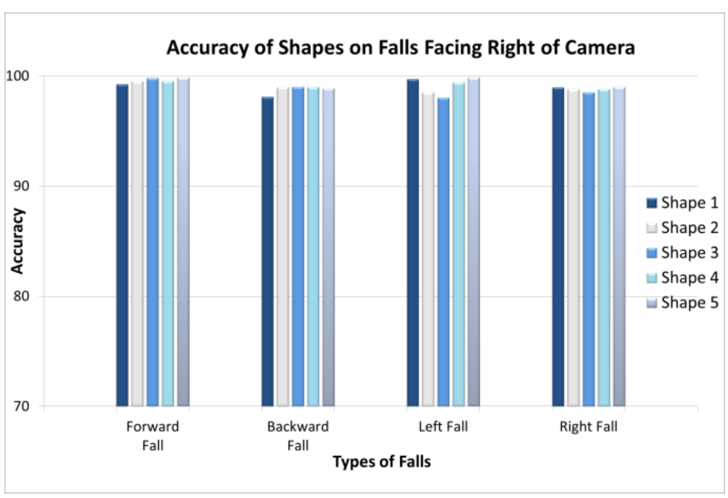
Accuracy of Shapes on Falls Facing the Camera
100
90
80
70
Accuracy
Forward Fall
Backward Fall
Left Fall
Right Fall
Types of Falls
Shape 1
Shape 2
Shape 3
Shape 4
Shape 5

(a)



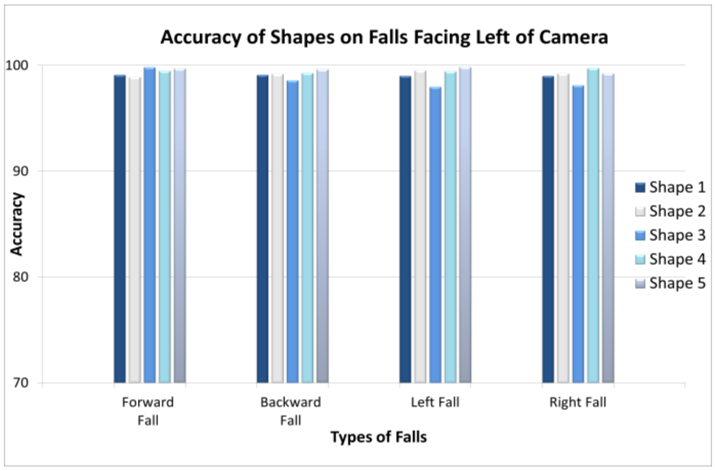
Accuracy of Shapes on Falls Facing Back of Camera
100
90
80
70
Accuracy
Forward Fall
Backward Fall
Left Fall
Right Fall
Types of Falls
Shape 1
Shape 2
Shape 3
Shape 4
Shape 5

(b)



Accuracy of Shapes on Falls Facing Right of Camera
100
90
80
70
Accuracy
Forward Fall
Backward Fall
Left Fall
Right Fall
Types of Falls
Shape 1
Shape 2
Shape 3
Shape 4
Shape 5

(c)



(d)

Figure 9. Accuracy comparison of body shapes during the first layer of DLSVM: (a) facing the camera (b) facing back of the camera (c) facing right of the camera (d) facing left of the camera

VII. RELATED WORKS

To the best of our knowledge, there has been no research utilizing joint data for fall detection. However, a few researches propose to use Kinect to detect falls. We evaluate and compare our system with three of recent studies.

Rougier et al. [27] proposed a Kinect-based fall detection system that uses human centroid height relative to the ground and body velocity to detect a fall. However, their evaluation does not prove strong results with large number of testing trials nor clearly indicate the subjects performing the experiment. Additionally, they do not mention about either the capability to detect different types of falls of various direction from the camera or information about differentiating actual fall versus a fall-like behavior.

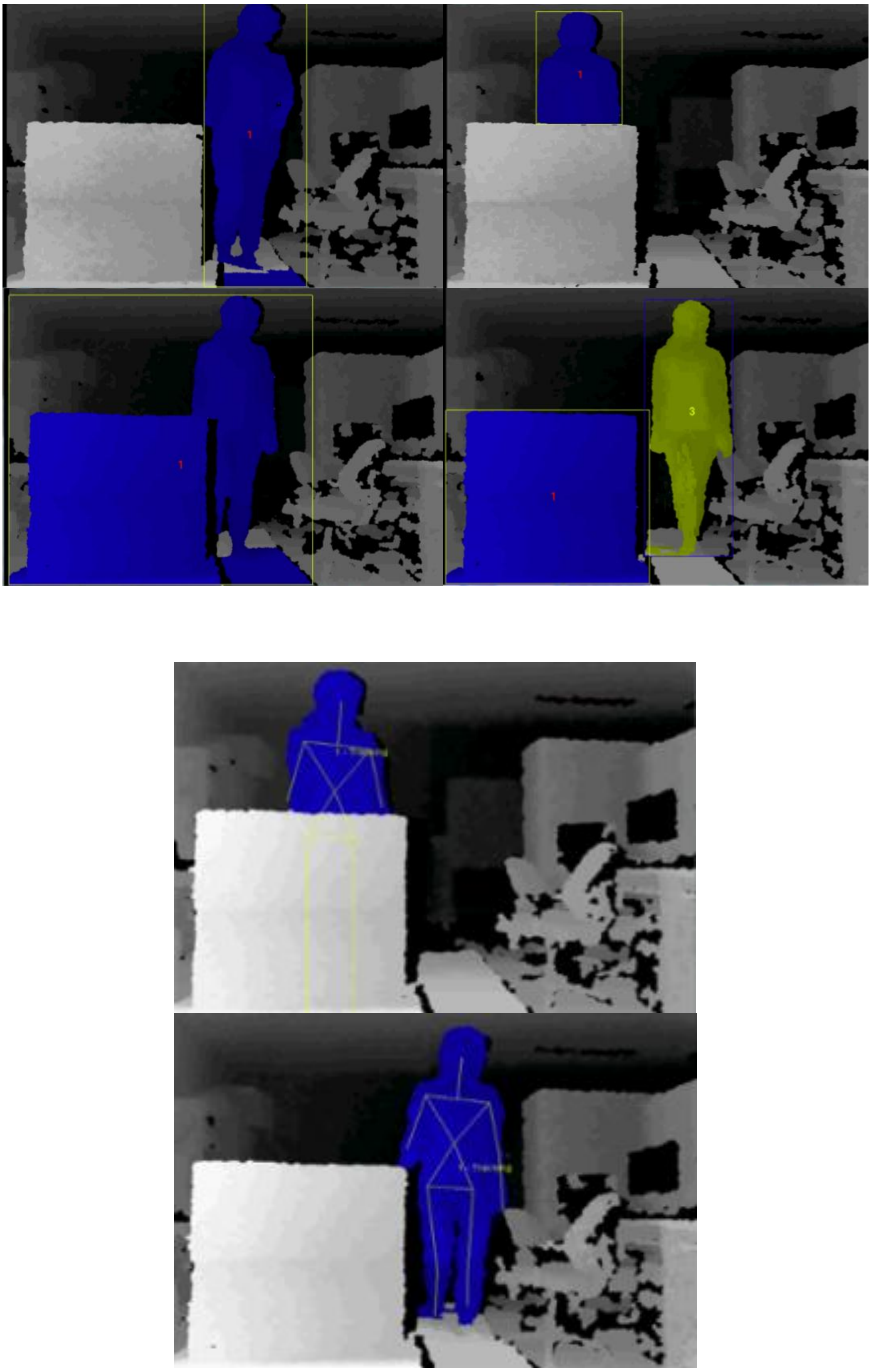
Additionally, the authors from [28] place Kinect 30cm above the floor and use image analysis in order to classify fall and other events such as feet in the front of the bed, fall, leaving the room, and activity in the room. However, their methodology detects the fall under the bed and does not describe a falling behavior by the entire body, which can produce high false-positive rate. Moreover, both of these studies do not mention about the capability to detect different types of falls of various direction from the camera nor information about differentiating actual fall versus a fall-like behavior.

A study by [29] measures the velocity based on the contraction or expansion of the width, the height, and the depth of the 3D bounding box. The velocity thresholds for the height and the width–depth composite vector of the bounding box, as well the duration of the fall are estimated by performing random search that optimizes the classification score in a training dataset. Their system can detect various types of falls as well as distinguish different activities. However, we implemented their algorithm and found that there were a few potential weaknesses.

First, the major drawback is that bounding boxes are sensitive when testing in a realistic environment. Very frequently, random bounding boxes were created on occluding boxed objects. The paper backs this issue by stating that the bounding box of the subject can still be detected even though another bounding box is created on an object. However, this is only true if the subject does not touch the object. Even with a slight touch, the system misunderstands the bounding box configuration and identifies the object part of, or as, the subject. Figure 10a demonstrates this example. This not only makes it inconvenient for the user to go through reconfiguration, but also increases potential risks in events of fall. In situations when the subject forgets to go through reconfiguration and happens to fall down, fall detection will not be triggered as the main bounding box is on the object. Our system, in contrast, utilizes skeleton to identify the subject's body. As shown in Figure 10b, detecting the skeleton of the subject significantly helps reduce occlusion and configuration problems since the system is initializing the human, not a box.

Additionally, we found that using simply the velocity of the fall was rather weak. Although using velocity may require less computation, it does not provide a behavior of a fall, and therefore makes the system too generalized. Moreover, velocities of the height, the width, and the depth may vary from person to person. New training would always be required to process the threshold for a particular subject. In contrast, we describe a human behavior by using shapes of the falling motion. Monitoring a falling motion helps the system to understand the behavior more clearly. For example, similarly to an observance from a human eye, the brain processes activity information to recognize a behavior. Likewise, we have our system understand the behavior of the motion rather than a being aware of velocity information.

Lastly, we evaluated the performance of their algorithm similarly to our evaluation of our system on Section VI C. We worked with 250 random fall and 250 random non-fall cases with the same 5 subjects (50 falls and 50 non-falls for each subject) in the same office setting. Despite the problems with the bounding box, the true positive accuracy rate turned out to be 95.6%, the true negative as 4.4%, the false negative as 89.6%, and the false positive as 10.4%.



(a)

(b)

Figure 10. Comparison of tracking tool used: (a) Bounding box from Primesense vs. (b) skeleton from OpenNI

VIII. CONCLUSIONS AND FUTURE WORK

Falls are a leading cause of injury-related morbidity and mortality. As the growing population of elderly increasingly motivates new healthcare systems at home, we applied fall behavior monitoring surveillance technique to a home health care environment. The key novelty of our approach is using Kinect to capture realistic skeletal data set of falls. In our implementation of INFADES, we linked the joint data and DLSVM algorithm together to define characteristics of a falling behavior. With a concrete definition of falls from DLSVM, INFADES was able to detect an actual fall and distinguish it from fall-like (e.g. lying down) behavior, while maintaining a high rate of accuracy. Our experimental results demonstrated that the combination of skeletal data and DLSVM significantly increases the robustness of fall detection.

There are many future works that are ahead of this study. First, I want to research and find the most optimize machine learning method to detect a behavior utilizing 3D joint data. In order to find the solution, I must experiment and evaluate other machine learning algorithms such as Liquid State Machine, Artificial Neural Networks, Decision Tree, and etc. Additionally, I want to expand this work so that the system can understand other behaviors, such as aggressive behaviors and criminal behaviors. The future mandates the need of an efficient way for utilizing surveillance CCTV camera without the need of a human behind the screen. Overall, through my future researches, I want to contribute to the development of a robust intelligent surveillance system.

References

[1] Administration on Aging. “Future Growth. A Profile of Older Americans: 2010”. 2010. http://www.aoa.gov/aoaroot/aging_statistics/Profile/2010/4.aspx.

[2] J. Davis, M. Robertson, M. Ashe, T. Liu-Ambrose, K. Khan, and C. Marra, “International Comparison of Cost of Falls in Older Adults Living in the Community: A Systematic Review,” Osteoporosis International, vol. 21, no. 8, 2010, pp. 1295-1306.

[3] “American Academy of Orthopedic Surgeons. OrthoInfo”. 2007. http://orthoinfo.aaos.org/topic.cfm?topic=A00101

[4] R. Rickover, “How the Alexander Technique can Help. Fear of Falling”. 2003. http://www.alexandertechnique.com/articles/falling/

[5] “Implementation of a Real-Time Human Movement Classifier Using a Triaxial Accelerometer for Ambulatory Monitoring

[6] M. Alwan, P.J. Rajendra, S. Kell, D. Mack, S. Dalal, and M. Wolfe, "A Smart and Passive Floor-Vibration Based Fall Detector for Elderly," Info. and Comm. Technologies, vol. 1, 2006, pp. 1003-1007.

[7] R. Caroline, C. St-Arnaud, and J. Rousseau, “Video Surveillance for Fall Detection”. Video Surveillance, InTech, 2011. 357-82.

[8] A. Bose, X. Hu, S. Kang, and T. Park, “Behavioral Detection of Malware on Mobile Handsets”. MobiSys 2008.

[9] A. Bourke and G. Lyons, “A Threshold-Based Fall-Detection Algorithm using a Bi-Axial Gyroscope Sensor”. Medical Engineering & Physics, 2008. Vol. 30, 84-90.

[10] J. Dui, X. Bai, Z. Shen, and D. Xuan, "PerFallD: A Pervasive Fall Detection System Using Mobile Phones". PERCOM Workshops, 2010.

[11] C. Tacconi, S. Mellone, and L. Chiari, "Smartphone-Based Applications for Investigating Falls and Mobility". ICST, 2011.

[12] B. Toreyin, Y. Dedeoglu, and A. Cetin, “HMM based falling person detection using both audio and video”. IEEE International Workshop Human-Computer Interaction, 2005, pp. 1–4.

[13] Rougier, C., Meunier, J., St-Arnaud, A. and J. Gousseau, “Fall detection from human shape and motion history using video surveillance”. 21st International Conference on Advanced Information Networking and Applications Workshops, 2007.

[14] T. Lee and A. Mihailidis, “An intelligent emergency response system: preliminary development and testing of automated fall detection,” Journal of Telemedicine and Telecare, vol. 11, pp. 194–198, 2005.

[15] H. Nait-Charif and S. McKenna, “Activity summarization and fall detection in a supportive home environment,” in Proc. 17th ICPR, vol. 4.2004, pp. 323–326.

[16] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Fall Detection from Human Shape and Motion History using Video Surveillance," Advanced Information Networking and Applications, vol. 2, 2007, pp. 875-880.

[17] H. Foroughi, A. Rezvanian, and A. Paziraee, “Robust Fall Detection using Human Shape and Multi-class Support Vector Machine,” 6th Indian Conference on Computer Vision, Graphics & Image Processing, 2008, pp. 413-420.

[18] S. Miaou, P. Sung, and C. Huang, “A Customized Human Fall Detection System Using Omni-Camera Images and Personal Information” Proceedings of Distributed Diagnosis and Home Healthcare Conference, 2006.

[19] R. Cucchiara, A.Pratti, and R.Vezani, “An Intelligent Surveillance System for Dangerous Situation Detection in home Environments” , in Intelligenza Artificiale, vol.1, n.1, pp. 11-15, 2004.

[20] C. Nait and S. McKenna, “Activity summarisation and fall detection in a supportive home environment”, Int. Conf. on Pattern Recognition, 2004.

[21] J. Tao, M. Turjo, M. Wong, and M. Wang, “Fall Incidents Detection for Intelligent Video Surveillance”, ICICS, 2005.

[22] G. Mastorakis and D. Makris, “Fall Detection System Using Kinect’s Infrared Sensor,” published online in Journal of Real-Time Image Processing, 2012.

[23] V. Vapnik, The Nature of Statistical Learning Theory. Springer, New York, 1995.

[24] T. Lee and A. Mihailidis, “An intelligent emergency response system: preliminary development and testing of automated fall detection,” Journal of Telemedicine and Telecare, vol. 11, pp. 194–198, 2005.

[25] T. Joachims, “Test Categorization with Support Vector Machines: Learning with Many Relevant Features,” 10th European Conference on Machine Learning, 1998, pp. 137-142.

[26] H. Shin and S. Cho, “Response Modeling with Support Vector Machines,” Expert Systems with Applications, 2006, pp. 746-760.

[27] C. Rougier, E. Anvienet, J. Rousseau, M. Mignotte, and J. Meunier, “Fall Detection from Depth Map Video Sequences,” 9th Int’l Conf. on Smart Homes and Health Telematics, 2011, pp. 121-128.

[28] C. Marzahl, P. Penndorf, I. Bruder, and M. Staemmler. “Unobtrusive Fall Detection Using 3D Images of a Gaming Console: Concept and First Results”. AAL-Kongress, 2012.

[29] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, and M. Finocchio, "Real-Time Human Pose Recognition in Parts from Single Depth Images," 24th IEEE Conf. on Computer Vision and Pattern Recognition, 2011, pp. 1297-1304.

요 약 문

지능형 비디오 감시: 3D 관절 데이터를 사용한 DLSVM 기반의 낙상사고 감지

독거노인의 수는 의학의 발전에 따라 증가되고 있으며, 미국에서만 독거노인의 인구수는 2050 년에 3 배로 증가될 예정이다. 독거노인의 증가에 따른 의료의 혁신이 병원에서뿐만 아니라 가정에서 필요하며. 한걸음 더 나아가, 편리하고 사용자와 친화적인, 그리고 지능형 의료 서비스가 필요하다. 컴퓨터 영상처리와 인공지능의 사용은 인간의 행동을 분석하고 이상 상황을 감지 할 수 있는 새로운 유망 솔루션을 제공할 수 있다. 이 연구를 통하여 혼자 사는 노인에 대한 가장 큰 위험 중 하나인 의도되지 않은 넘어짐, 낙상사고를 감지하는 방법을 제안하며, 구축하였다. 접근 방식은 마이크로소프트사에서 개발한 Kinect RGBD (RGB-Distance) 센서를 사용하여 인간의 골격 및 관절 데이터를 사용하며, 기계학습 기술을 기반으로 시작하였다. 많은 논문을 분석해본 결과, 이 연구는 현재 제출된 연구 중, 기계학습과 사람의 관절 정보를 사용하여 낙상사고 검출을 최초로 도입한 연구이다. 다양한 낙상의 종류를 따라 견고한 정의를 통하여, 전체 적인 시스템은 이 연구를 통하여 개발된 두 개의 서로 다른 레이어 지원을 하는 DLSVM (Double Layered Support Vector Machine)을 사용하여 구축 되었다. 평가 결과는 제안된 시스템이 효율적으로 다양한 방향을 향해 나타난 낙상 유형을 감지하며, 사고 또는 의도된 넘어짐인지에 대한 분류 능력을 확인 하였다.

윤희정

Acknowledgement

First and foremost, I would like to thank God for the guidance and blessings that he has given me to accomplish this work.

I would also like to express my sincere gratitude to thank my reach supervisors, Professor Sang Hyuk Son, Professor Taejoon Park, Professor Sam Chung, and Professor Moon Sangjun for giving me the opportunity to do research and providing invaluable guidance throughout this research. I could not have imagined having better mentors for my Master's study.

To my boyfriend, Ho-Kyeong Ra, I wish to offer my deepest thanks. He was and is a solid supporter which I rely again and again. Because of him, my experience here at DGIST, and this work exceeded my own expectation.

I'm extremely grateful to my parents for their love, prayers, and sacrifice for educating me for my future. I'm very much thankful for them to letting me come even come to Korea. I know how much you both always want me next to your sides. To my little brother who always supports and cares for our family, thank you. I'm sorry I'm not there to help out. Thank you for taking charge while I'm not there.

Finally, my thanks go to all the people who have supported me to complete the research work directly or indirectly.

Curriculum Vitae

Name : Hee Jung Yoon
Birth Date : August 11, 1988

Education

2003 – 2006: Gov. John R. Rogers High School
2006 – 2010: University of Washington, Seattle
2011 – Present: Daegu Gyeongbuk Institute of Science and Technology

Work Experience

2000 – 2011 Ginger Teriyaki as a part-time manager
2008 - 2009 University of Washington Health Sciences Financial Services as a student assistant
2009 – 2011 University of Washington Medical Center Communications Dept. as a hospital operator

Professional Activities

Volunteer Pharmacy Assistant at St. Joseph Hospital in the Out Patient Pharmacy
Volunteer Student Assistant at Good Samaritan Hospital in the Health Information Administration Department
Volunteer Student Assistant at University of Washington Medical Center in the Executive Director's office
West Coast, US Liaison for Young Generation Forum 2009
Organizer for US - Korea Conference 2011
Organizer for Young Generation Technical & Leadership Conference 2011

Honors and Awards

Best Presentation Award at the ICE Workshop 2011
Best Poster Award at the US-Korea Conference 2012