Emergency Fall Incidents Detection in Assisted Living Environments Utilizing Motion, Sound, and Visual Perceptual Components

Charalampos N. Doukas, *Student Member, IEEE*, and Ilias Maglogiannis, *Member, IEEE*

***Abstract*—This paper presents the implementation details of a patient status awareness enabling human activity interpretation and emergency detection in cases, where the personal health is threatened like elder falls or patient collapses. The proposed system utilizes video, audio, and motion data captured from the patient's body using appropriate body sensors and the surrounding environment, using overhead cameras and microphone arrays. Appropriate tracking techniques are applied to the visual perceptual component enabling the trajectory tracking of persons, while proper audio data processing and sound directionality analysis in conjunction to motion information and subject's visual location can verify fall and indicate an emergency event. The postfall visual and motion behavior of the subject, which indicates the severity of the fall (e.g., if the person remains unconscious or patient recovers) is performed through a semantic representation of the patient's status, context and rules-based evaluation, and advanced classification. A number of advanced classification techniques have been examined in the framework of this study and their corresponding performance in terms of accuracy and efficiency in detecting an emergency situation has been thoroughly assessed.**

***Index Terms*—Activity recognition, assisted living environments, body sensors, context awareness, event detection, human safety, patient telemonitoring.**

I. Introduction

TELEMONITORING is the physical status and health of humans or patients at home, and is an interesting solution compared to hospitalization in healthcare facility institutions since it offers a medical surveillance in a familiar atmosphere for the patient and can reduce the costs of medical treatment ([1]–[5]). Within the same context, the monitoring of human physiological data, in both normal and abnormal situations of activity, is vital for the purpose of emergency event detection, especially in the case of patients suffering from chronic diseases or elderly people living on their own. Special interest is paid in the detection of the severity of the case that can indicate injury level and assistance request type. Several techniques have been proposed for identifying such distress situations using either motion, audio, or video data from the monitored subject and the surrounding environment. The great challenge in such personal health systems is to provide less invasive monitoring technologies, increase mobility, and at the same time achieve high accuracy rates in patient status interpretation ([6]).

This paper introduces a solution to the problem of less invasive patient monitoring, describing the design and an initial implementation of a patient status awareness system that may be used for human or patient activity interpretation and emergency recognition in cases like elder falls and patient collapses. The proposed system utilizes motion information, audio, and video data, which are captured from both the patient area and the surrounding environment. Visual information and audio from the monitored site are acquired using overhead cameras and microphone arrays, respectively, while motion data and patient-generated audio sounds are collected through appropriate body sensors on the patient. Appropriate tracking techniques are applied to the visual perceptual component enabling the trajectory tracking of the subjects, and proper audio data processing and sound directionality analysis in conjunction to motion information and subject's visual location can verify fall and indicate an emergency event. Postfall visual and motion behavior of the subject indicates the severity of the fall (e.g., if patient remains unconscious or patient recovers and stands up). The severity analysis is performed through an ontological representation of the patient's context awareness, rules-based evaluation, and activity classification. A number of advanced classification techniques have been evaluated for this purpose and the performance of the classifiers has been assessed in terms of accuracy and efficiency. The innovation of the presented system against existing works (discussed in Section II) resides in four key elements:

1) the utilization of three separate information channels (motion, audio, and visual data) for patient status interpretation;
2) the information fusion and streaming capabilities of the latter data;
3) the ontology and rules-based evaluation for proper characterization of incidents; and
4) the context awareness concept, which is newly introduced in such systems.

The rest of the paper is organized as follows. Section II discusses the related work in the field of patient activity and fall detection. Section III describes the system architecture and the implementation of each module, while Section IV presents the incident semantic representation and rules-based-evaluation

Manuscript received April 19, 2010; revised September 9, 2010; accepted October 6, 2010. Date of publication November 9, 2010; date of current version March 4, 2011.

C. N. Doukas is with the Department of Information and Communication Systems Engineering, University of the Aegean, 81100 Mytilene, Greece, and also with the National Hellenic Research Institute, 11635 Athens, Greece, (e-mail: doukas@aegean.gr).

I. Maglogiannis is with the Department of Computer Science and Biomedical Informatics, University of Central Greece, Lamia, Greece.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITB.2010.2091140

approach proposed in this paper. Section V presents the experiments performed for the evaluation of the proposed platform along with the corresponding results, and finally, Section VI concludes the paper.

II. Related Work

Although the concept of patient activity recognition with focus on fall detection is relatively new, there already exists significant related research work, which may be retrieved from the literature ([5]–[14]). Information regarding the human movement and activity in assisted environments is frequently acquired through visual tracking of the subject's or patient's position. In [10], overhead tracking through cameras provides the movement trajectory of the patient and gives information about user activity on predetermined monitored areas. Unusual inactivity (e.g., continuous tracking of the patient on the floor) is interpreted as a fall. Similarly, in [13], omnicamera images are used in order to determine the horizontal placement of the patient's silhouettes on the floor (case of fall). Success rate for fall detection is declared at 81% for the latter work. A different approach for collecting patient activity information is the use of sensors that integrate devices like accelerometers, gyroscopes, and contact sensors. The latter approach depends less on issues like patient physiology (e.g., body type and height) and environmental information (e.g., topology of monitored site), and can be used for a variety of techniques enabling user activity recognition ([5], [8], [12]). Regarding fall detection, authors in [7], [11], and [14] use accelerometers, gyroscopes, and tilt sensors for movement tracking. Collected data from the accelerometers (i.e., usually rotation angle or acceleration in the *X, Y*, and *Z*-axes) are used in order to verify the placement of the patient and time occupation in rooms and detect abrupt movement that could be associated with fall. Detection is performed using predefined thresholds [5], [8], [9], [11] and association between current position, movement, and acceleration [7], [14]. In previous works ([15], [16]), we have presented a patient-fall-detection system based on such body sensors that utilized advanced classification techniques and Kalman filtering for producing more accurate results.

Sound processing has been also utilized for fall detection. Most of the related work focuses on collecting and analyzing sound data captured from the patient's close environment. In [17]–[19], authors present a sound analysis system enabling the detection of special sounds and their association with events related to specific activities or situations, where first aid is needed (e.g., falls, glass breaking, call for help, etc.). The sound event detection is based on feature extraction through discrete wavelet transformation (DWT), whereas classification to predefined events or vocal expressions is performed through a Gaussian mixture model (GMM) technique. In [20], Mel frequency cepstral coefficients (MFCCs) are used in order to detect a variety of sound signatures of both distressful and normal events. The examined sounds are categorized into classes according to their corresponding average magnitude levels that emerge from the application of Fourier transform on the sound signal. Cepstral coefficients are used as features fed into a GMM model for proper classification. Accuracy of proper classification achieves 91.58% according to the authors. The aforementioned methods are based on acquisition and processing of sound data that originates from user's monitored environment. In [16] and [21], we have proposed a different method for detecting patient distress situations utilizing sounds captured by microphones attached on body sensors and spectrogram analysis sound processing. This technique has provided satisfying accuracy in detecting body fall sounds and distress speech expressions, while it was proved more tolerant to background noise and sounds not originating from the patient.

The study presented in this paper integrates user movement information and sound using wireless sensors, visual tracking of the patient, and sound source localization using microphone arrays aiming at more accurate activity recognition systems. The proposed system is based on previous works by the authors in the context of movement characterization utilizing motion, sound and visual data individually ([15], [16], [21], [22]), and it is enhanced through semantic representation of the user's status and context awareness, while rules-based evaluation can provide an estimation of the severity of the incident (e.g., patient has recovered from fall, or patient is inactive, etc.). To our best knowledge, there is no relative work in the literature that combines both visual, sound, and motion sensor information, and uses semantics for improving human safety through incidents detection in assisted living environments.

III. Materials and Methods

This section provides important information regarding the materials used to collect patient data and the methods utilized in order to perform proper characterization of patient's status.

A. System Architecture

The presented system follows the architecture illustrated in Fig. 1. Camera devices and microphone arrays are installed at the patient's site. Special sensor nodes with networking capabilities are required for collecting and transmitting related activity data (i.e. accelerometer and sound data). These sensors can be attached on several locations on the subject's body. A monitoring node is required for collecting the aforementioned data and performing required processing in order to enable an estimation of the human status. Recorded video frames provide feed to the video tracker that tracks the movement of the patient's body and generates body shape features (i.e., coordinates of a bounding box containing the subject's body). Recorded sounds are utilized in order to detect emergency events like distress speech expressions or body fall sounds. Sound source localization provided by the microphone arrays can also be applied and facilitate the status awareness; background noise can be easily filtered through sound source redundancy. Additionally, in the cases, where the patient is the sound source, the localization of the latter in conjunction with visual trajectory information can provide more robust estimation of the actual incident and avoid false alarms generated by other sound sources.

The data are properly transformed in a suitable format for the classifier, and the classification phase begins. Based on a

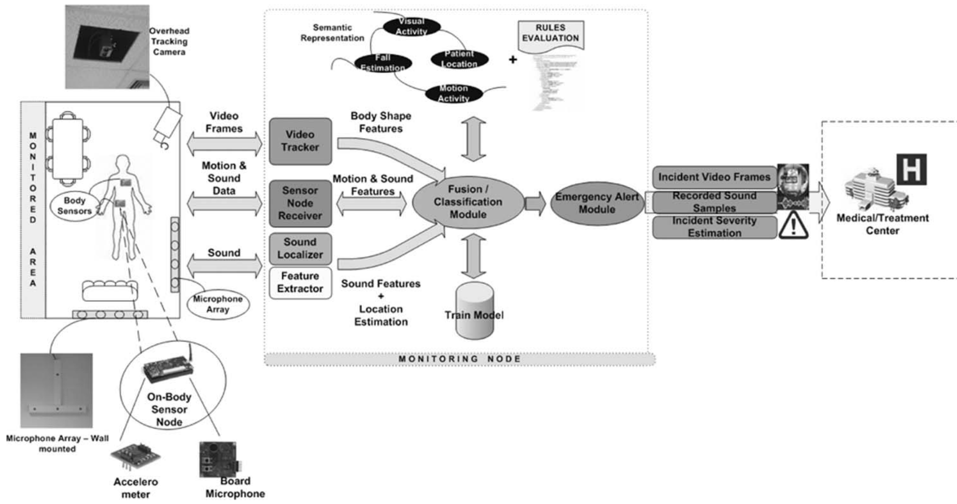


Fig. 1. Proposed system architecture illustrating basic modules: motion, sound, visual perceptual components, and respective equipment and the monitoring node.

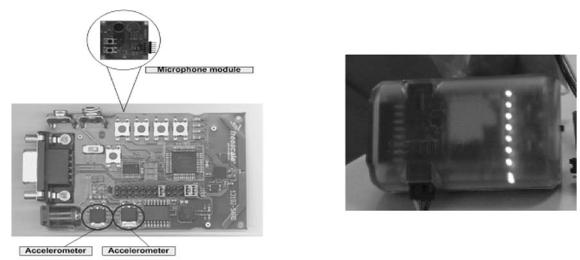
predefined classification model (i.e., train model), the patient status is detected (i.e., emergency status when an emergency event is detected, normal status otherwise).

Apart from the indication of an emergency incident (e.g., a patient fall), an estimation of the severity of the incident can be provided based on the patient's behavior after the fall as recorded visually; visual inactivity or soft activity combined with distress sounds originating from patient's location suggest that patient has not lost consciousness and is trying to recover from the fall. In case no visual or sound activity is recorded after fall estimation, higher severity of the incident might be estimated. In order to provide a more accurate estimation, a semantic model of the patient's status and context is built and proper rules evaluation follows.

Based on emergency event detection, the treatment personnel at a remote or local site can be alerted. In conjunction to the incident type and severity estimation, corresponding video frames and audio samples from the patient and the monitored environment can be transmitted facilitating the diagnosis process. Methods for analyzing visual, motion, and audio data that allow human body trajectory analysis, sound source localization, and incident detection are described in following Sections.

B. Motion and Sound Data Acquisition

Sensor data acquisition may be accomplished through wireless on-body (wearable) networks. On-body networks or wireless personal area networks (WPANs) are defined within the IEEE 802.15 Standard. The most prominent protocols for pervasive systems, such as the proposed system, are Bluetooth and ZigBee (IEEE 802.15.4 standard). The ZigBee has been developed as a low data rate solution with multimonth to multiyear battery life and very low complexity. It is intended to operate in an unlicensed international frequency band. The maximum data rates for each band are 250, 40, and 20 kbps, respectively. Two types of sensor nodes have been used in the implementation of the proposed system: a SARD ZigBee node [23] and a Sentilla Perk [24] sensor (see Fig. 2). Both of them contain



(a) (b)

Fig. 2. (a) SARD ZigBee node. The node acts as both receiver and transmitter. The RS232 interface provides connectivity with the monitoring device (e.g., a PDA) when the node is used as receiver. Two 3-D accelerometers and one microphone module are attached on the node. (b) Sentilla Perk node containing one 3-D accelerometer that can be attached on user and send motion data through the ZigBee wireless protocol. The plastic enclosure can protect the node from falls and makes it more suitable for carrying it on patient's body.

a 2.4-GHz wireless data transceiver RF reference design with printed circuit antenna, which provides the necessary hardware required for a complete wireless node using IEEE 802.15.4 (ZigBee) [25] packet structure. The first one includes an RS232 port for interface with a personal computer, whereas the second one uses a USB port interface instead. Both of them feature debug modules for in-circuit hardware debug, switches and LEDs for control and monitoring, and a low-power, low-voltage MCU (MicroController Unit) with 60KB of on-chip Flash, which allows the user flexibility in establishing wireless data networks. 3-D accelerometers for measuring acceleration at *X*, *Y*, and *Z*-axes have been attached on the nodes (SARD node contains two accelerometers, and Perk node one, respectively). A separate SensiNode [26] board has been also attached containing a microphone and additional sensors like illumination and temperature sensors. The Perk nodes are provided in a plastic robust small-sized enclosure (6 cm $\times$ 3 cm $\times$ 1.5 cm) making them more suitable for placing on patient's body and tolerating falls.

More than one sensor nodes can be placed on the patient's body. Preferable positions are close to user's chest and user's belt or lower at user's foot. The latter positions have proven based on conducted experiments to be more appropriate for distinguishing rapid acceleration on one of the three axes that is generated during a fall. Appropriate J2ME [27] and C code is developed and deployed on the nodes for reading the accelerometer values and transmitting them wirelessly to the monitoring unit. At the latter, a Java application built using the Sentilla IDE [24] receives the movement data and performs further processing, as described in the following sections. The *X, Y,* and *Z* acceleration values from both sensors are interlaced. In order to improve the accuracy of the latter decision, Kalman filtering [28], [29] has been applied on the sequence of the movement-type association of each acceleration data set according to the following algorithm. The measurement noise and acceleration noise factors have been set to 10 and 0.5, respectively, based on conducted experiments. The noise has been considered white and therefore a known covariance matrix has been used.

```
Start Kalman filter algorithm

Step 1:
Read acceleration value Xn from sensor
Step 2:
Calculate the noise covariance Nq and the Measurement
error covariance R based on the MeasurementNoise factor
and AccelerationNoise factor.

Nq = AccelerationNoise ^2 * [0.1^4/4 0.1^3/2; 0.1^3/2
0.1^2]
R = measnoise^2
For the previous 10 acceleration values Xi, i€[n-10,
n-1]:
                   Calculate the noise:
                   Noise = AccelerationNoise * Xi*
                   [(0.1^2/2); 0.1]
                   Calculate the measurement with the
                   estimated noise:
                   Meas = measnoise * Xi
                   z = Xi + MeasurementNoise;
                   Calculate the Innovation:
                   I = z - c * xhat;
                   Calculate the covariance of
Innovation
                   s = Xi * P * Xi' + R;
                   Calculate  the Gain matrix
                   K = a * P * Xi' * inv(s);
                   Calculate the estimate for the next
                   acceleration value:
                   Xest = a * Xest + K * I;
end
GoTo Step 1
end algorithm
```

Each filtered acceleration value (Xest, Yest, and Zest, respectively) is used as inputs to the classification process.

C. Human Body Visual Tracking

The goal of the body video tracker is to identify and track across time the frame regions containing human bodies. The tracker is built around a dynamic foreground segmentation algorithm [30] that utilizes adaptive background modeling. This is based on Stauffer's algorithm [25] to provide the foreground pixels. Stauffer's algorithm models the different colors every pixel can receive in a video sequence by GMMs. One GMM corresponds to every pixel at given coordinates across time. As a result, a map can be built, in which every pixel is represented by the weight of the Gaussian from its GMM that best describes its current color. This is the pixel persistence map (PPM): regions of the map with large values correspond to pixels that have colors that appear there for a long time, hence, they belong to the background. On the contrary, regions with small values correspond to pixels that have colors that appear there for a short time, hence, they are foreground. The deficiency of Stauffer's algorithm is related to the foreground objects that stop moving. In its original implementation, targets/objects that stop moving are learnt into the background. To avoid this in our system, the learning rates of the adaptation that increase the weights of Gaussians are not constant, neither across space, nor across time. Instead, they are spatiotemporally controlled by the states of Kalman filters [31] [(see Fig. 3(a)]. Every foreground area corresponds to a target being tracked by a Kalman filter. The foreground pixels are combined into body evidence blobs used for the measurement update stage of the Kalman filters. The states are used to obtain the position, size, and mobility of each target, the latter being a combination of translation and size change. This information is fed back to the adaptive background modeling module to adapt the learning rate in the vicinity of each target: frame regions that at a specific time have a slow-moving target, have smaller learning rates. The block diagram of the introduced body tracker is shown in–Fig. 3(a).

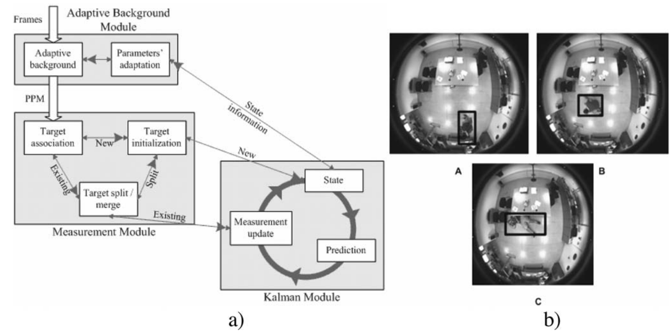


Fig. 3. (a) Block diagram of the body video tracker. Kalman filters spatiotemporally adapt the learning rates of the adaptive background algorithm, effectively avoiding learning of immobile foreground objects into the background. (b) Visualization of video tracking performance. The tracker detects the movement of the body and correlates it with the movement of a rectangular blob within the visual domain. Upper left *X, Y* coordinates and respective width and height of the blob are reported for each visual frame. Frame A corresponds to normal walking, Frame B to captured movement during fall, and Frame C illustrates the detection of body in horizontal position after fall.

Using the feedback configuration of the tracker, the learning of the slow-moving foreground objects into the background is slowed down long enough for the intended application, i.e., tracking people moving indoors and possibly falling down. The tracker results, as produced by the visual feed of an overhead camera, are illustrated in Fig. 3(b). Tracking through overhead cameras has been selected due to the fact that it provides a better visual representation of the monitored area and allows the tracker to gain a better estimation of the body shape when subject moves, falls, and lies still after fall. The presented tracker detects and tracks a rectangular blob around the detection of the moving body within the frames and reports the upper left corner coordinates and respective width and height of the blob. As indicated in Fig. 3(b), the size of the blob changes during the fall and after it.

D. Sound Processing and Event Detection

The detection of emergency events is also facilitated through appropriate sound processing of surrounding audio captured by the microphone arrays and patient sounds acquired by the body sensors. Microphone arrays are mostly utilized for sound source localization, whereas sounds captured from on-body microphones provide important features that can be properly classified and lead to event detection.

1) Sound Source Localization: An important aspect of the proposed system is the sound source localization that can lead to an estimation of the position of the individual in the event of an emergency. Localization can be performed using the estimation of the direction of arrival (DOA) of an acoustic source using time delay estimation (TDE). Typically, the problem is addressed using microphone arrays that collect data in frames so that the current estimate can be provided. The most popular approaches rely on defining the relative delay between a pair of microphones by means of comparing function that returns a peak at the correct DOA of the source. Common methods for TDE are the generalized cross-correlation [32] and blind source separation (BSS) [28]. The proposed system utilizes sound source localization using an implementation similar to BSS provided by [29].

Assuming the existence of two microphones, a single source would lead to the following discrete signal recorded at the *i*th microphone $i \in [1, 2]$

$$x_i(k) = s(k - \mathrm{T}_i) \tag{1}$$

where T_i denotes the time in samples that it takes for the source signal to reach the *i*th microphone. For the case of two microphones (see Fig. 6) and considering that $T_1 = 0$, the delay at i_2 is the relative $T = t_2$ between the two recorded signals. The corresponding DOA θ in degrees is defined with respect to the broadside of the array that is connected with *T* in the following way:

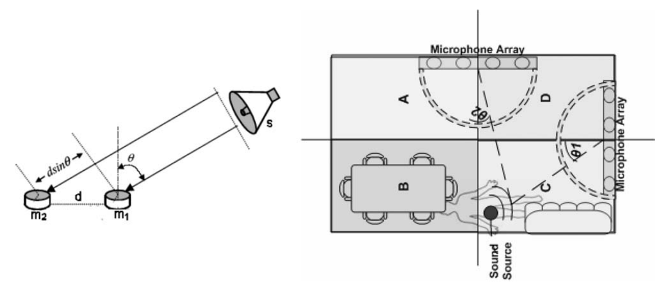
$$\theta = \arcsin\left[\frac{TC}{f_s d}\right] \tag{2}$$

where f_s is the sampling frequency of the recording system, and c the speed of sound. More details of the sound localization implementation can be found in [29].

The correlation of the sound source location and patient's body location is performed as follows.

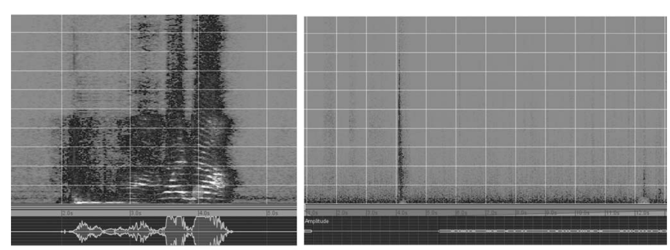
Consider two microphone arrays being attached on the walls of a monitored area, as show in Fig. 4(b). The arrays can cover a DOA $\Theta = 1800$. The area has been divided into four quadrants. When a sound is captured, each microphone array gives an estimation of the angle of arrival $\theta 1$ and $\theta 2$. Based on their values, the quadrant that contains the sound source can be easily determined. The presence of the patient within the latter can be then verified by the visual tracker as well. The deployment of a larger number of microphones per array and the introduction of arrays within the monitored array can increase the sound source localization accuracy by also allowing the utilization of more advanced techniques like angle tranquilization. The latter can be translated into the following algorithm:

```
Start Angle of Arrival Algorithm
T1,2 = timestamp of signal recorded at the microphones
for all microphones in the array:
          Calculate the average time delay T as:
          T = abs(T2-T1+T3-T2+T3-T1)
          Calculate the angle of arrival as:
          Θa = arcsin [(T * C)/ 22100] (C equals to the
          speed of light)
end
if(Θ1 > 90⁰ && Θ2 > 90)
          Quadrant = A;
else if (Θ1 > 90⁰ && Θ2 < 90)
          Quadrant = B;
else if (Θ1 < 90⁰ && Θ2 < 90)
          Quadrant = C;
else if (Θ1 < 90⁰ && Θ2 > 90)
          Quadrant = D;
end
End Algorithm
```



(a) (b)

Fig. 4. (a) DOA θ estimation using microphone arrays and TDE. (b) Estimation of patient location using sound source localization



(a) (b)

Fig. 5. Illustration of spectrogram analysis on (a) distress speech expression, and (b) sound generated by patient body fall.

If the quadrant-indicated area contains the subject's body, as indicated by the visual tracker, a binary feature with the value of 1 is used. Otherwise the feature has the value of zero.

2) Sound Feature Extraction: In this research paper, we have used spectrogram analysis, based on short-time-Fourier transform (STFT) for the detection of sounds features characterizing the fall of the human body or the vocal stress in speech expressions indicating distress events (see Fig. 5). Given a signal $x(t)$ and its Fourier transform $X(\tau, \omega)$, the STFT is defined as follows:

$$\mathrm{STFT}\{x()\} \equiv X(\tau, \omega) = \int_{-\infty}^{\infty} x(t)\omega(t - \tau)e^{-j\omega}\, dt. \tag{3}$$

The spectrogram is respectively given by the magnitude of the STFT of one function

$$\mathrm{spectrogram}(t, \omega) = |\mathrm{STFT}(t, \omega)|^2 . \tag{4}$$

In the most usual format of a spectrogram, the horizontal axis represents time, the vertical axis is frequency, and the intensity of each point in the image represents amplitude of a particular frequency at a particular time. Based on conducted experiments, the relative amplitude of a signal and the peak frequency at a given time can give a successful indication of a patient fall sound as captured by the microphone arrays; body falls generate low-frequency sounds with high amplitude. Using a threshold of >90% for relative signal amplitude and <200 Hz for peak frequency, the differentiation of a fall sound against other sounds is possible. More precisely, over a series of 20 sound samples containing both body fall sounds and background noise (e.g., radio, object falls, etc.), the detection of the fall was possible for 80% of them [21]. Different types of floor (i.e., wood, cement, and flooring tile) were also used. Neither the different floor types used, nor the various background noises seem to have influenced the performance of the system. The presented method has low computational complexity and can be easily integrated on sensor devices for real-time sound processing. The same analysis has been applied on vocal sounds in an effort to detect distress expressions.

The latter can be translated into the following algorithm for calculating the average sound peak frequency and average signal relative amplitude features:

```
Start Sound Feature Extraction Algorithm

Xs[] Five second segment of recorded data
for i=0 to all the number of recorded segments:
         Calculate STFT for the given Xs[i];
         Calculate spectrogram for STFT[i];
         for: all the signal segments:
                  Find maximum signal amplitude, Amax
                  Find coherent signal amplitude, Ac,
where Ac>0.9*Amx
                  for: all the Ac:
                           if Fpeak of Ac < 200 Hz
                           Use Fpeak and Ac for
classification
                  end
         end

end
End Algorithm
```

The following section provides more information on how motion, visual, and audio data can be collected in order to achieve the optimum motion analysis and fall detection.

IV. Experimental Protocol Details

In order to combine the aforementioned information channels and detect emergency fall incidents, an experimental protocol has been defined. The protocol describes issues like the movement types that can be analyzed, the suggested placement of the sensors for optimum results, and technical details like sampling rates and testbed setup. Three different combinations of movement types have been considered for assessment; 1) simple walk; 2) simple walk and fall; and 3) simple walk and run. The sensors have been placed on the subject's chest and belt (using special straps glued on sensors) in order to provide better estimation of the body movement and placement with respect to the ground. Each experiment containing one of the aforementioned movement types has an average duration of 120 s. Each individual performs at least two experiments including all three movement types. The sampling frequency (i.e., the rate sensors are collecting and transmitting data) is 20 Hz for the movement data and 22.1 KHz for audio data (default sampling rate of the microphone sensor). The frequency of falls in the second type of experiment is two or three falls per recording. The volunteers– subjects are directed to perform all movement types as realistically as possible, behaving like in real life (i.e., adding random stopping intervals in movement and changing acceleration at will). More specifically for the fall trials, the individuals are advised to initially walk within the experiment area (a flat room of 40 m^2 with obstacles like furniture) and then perform falls simulating events like stumbling on furniture or falling down because of loss of consciousness (e.g., in case of a heart attack). Each combined movement-type experiment (i.e., simple walk and run or simple walk and fall) can be considered to contain about 80% of its recording time of walking and the rest for running or falling). Postfall behaviors are also simulated by standing still (unconscious state), moving (trying to recover from fall), and getting up (recovering from fall). An overhead camera is capturing visual frames and two microphone arrays are capturing sounds (see–Fig. 1 for testbed setup). Recorded data are segmented into time segments of 5 s. Each segment is processed and the generated sound data and body motion features are utilized for creating classification models. Classification of all incoming data is performed every 5 s to maintain time granularity. All incoming data are time stamped and buffered until the classification process. Sound features consist of average peak frequency and average relative amplitude for the specific time segment, as calculated using spectrogram analysis. Respectively, body motion features are the standard deviation of the blob size generated by the visual tracker [(see Fig. 3(b)] and the average movement speed of the tracked body over captured frames. Finally, a binary feature (true/false) is used in order to indicate whether a detected sound has occurred within close proximity of the patient or not. Table I summarizes all the aforementioned features utilized for performing the experiments. The correlation of the motion and sound data with the patient body trajectory data can provide much more accurate results, as presented in the following section.

V. Incident Severity Estimation Through Semantic Representation and Rules-Based Evaluation

In order to semantically represent an emergency incident, as indicated by the motion, sound, and visual perceptual components, the ontology illustrated in Fig. 6 has been developed. An emergency incident can be characterized by its severity (e.g., high or low) based on fall estimation and more precisely if high or low visual and motion activity is identified after the fall, respectively. The patient movement ability level can also provide important information regarding the patient's ability to recover from falls, and finally the correlation of the sound source and the patient's location is also very important.

The ontological model has been developed within the Protégé [33] semantic framework using the ontology web language

TABLE I
DESCRIPTION OF UTILIZED MOTION, SOUND, AND VISUAL FEATURES

	Features	Short Description
Motion Features		
	X, Y, Z acceleration values	X, Y and Z acceleration values in [-1, 1] as obtained from on-body sensors.
Sound Features		
	Sound Proximity	Binary feature indicating whether a captured sound has been recorded in close proximity to the patient body. This information is generated by sound source localization and visual information (see Section 3.4.1)
	Average Peak Frequency	Numerical featured calculated using STFT transform on acquired sound signal (see Section 3.4.2)
	Average Signal Relative Amplitude	Numerical featured calculated using STFT transform on acquired sound signal (see Section 3.4.2)
Visual Features		
	Visual Blob size	The standard deviation of the blob (i.e. rectangular area containing subject's body) size as indicated by the visual tracker (see Section 3.3)
	Average Movement speed	The average movement speed of the tracked body over captured frames

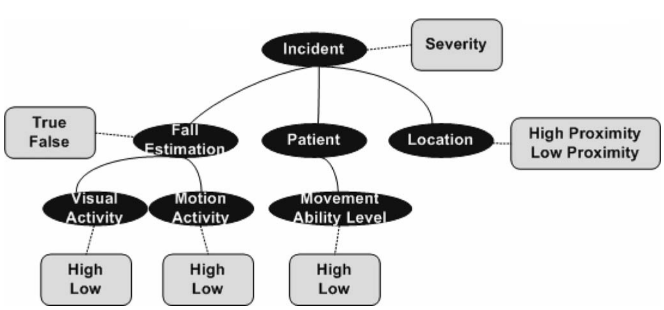


Fig. 6. Semantic representation of the ontology modeling an emergency incident based on fall estimation, patient, and location parameters

(OWL). The main advantages of the semantic representation of the incident in the context of the patient status can be summarized into the following.

1) Flexibility to modify and extend the contextual scheme by adding more classes. In case the parameters that define the context of the patient (e.g., status, environment, location, etc.) need to be modified, the ontological model can be altered without invoking modifications to the implementation modules or the architecture of the platform.
2) Using advanced semantic rule evaluation techniques, content adaptation decisions can be made according to a plethora of contextual parameters. The rules can be updated and extended without any need for system platform software modifications.
3) Additionally, ontologies are explicit because they define the concepts, properties, relationships, functions, axioms, and constraints that compose the contextual model. They are formal because they are machine readable and interpreted.

The creation of semantic rules required the description of the latter through abstract semantic languages like the semantic web rule language (SWRL) [34]. Within this context, the SWRL factory [33] mechanism and an integrated Jess rule engine [35] using the Protégé tool have been utilized. Jess provides both an interactive command line interface and a Java-based application programming interface (API) to its rule engine. This engine can be embedded in Java applications and provides a flexible two-way run-time communication between Jess rules and Java. The Jess system consists of a rule base, a fact base, and an execution engine.

Two indicative samples of SWRL rules follow that can be used within the presented framework in order to facilitate the decision on the emergency incident estimation:

```
Patient(?x)^Location(?y)^hasFallEstimation(?
x,?y, ?FallEstimation)^
hasVisualActivity(?x,?VisualActivity)^hasMot
ionActivity(?x,?MotionActivity)^
Location(?Proximity)^
swrlb:equals(?FallEstimation,?true)^swrlb:eq
uals(?Proximity,?High)
^swrlb:equals(?VisualActivity,?High)^swrlb:e
quals(?MotionActivity,?High)
->DefineIncidentSeverity(?Severity,"LOW")
```

```
Patient(?x)^Location(?y)^hasFallEstimation(?
x,?y, ?FallEstimation)^
hasVisualActivity(?x,?VisualActivity)^hasMot
ionActivity(?x,?MotionActivity)^
Location(?Proximity)^
swrlb:equals(?FallEstimation,?true)^swrlb:eq
uals(?Proximity,?High)
^swrlb:equals(?VisualActivity,?Low)^swrlb:eq
uals(?MotionActivity,?High)
->DefineIncidentSeverity(?Severity,"HIGH")
```

Both rules examine the correlation of patient's location to the sound source, the fall estimation, and postfall visual and motion activity. In both the cases, a fall has been detected and the body fall sound and/or other distress sounds have been captured in close proximity of the patient. In the first case, there is high visual and motion activity thus indicating that the patient has probably recovered from fall, whereas in the second case, low visual but high motion activity indicates that the patient is still on the floor trying to recover from fall. Thus, the first incident is characterized by low severity and the second by high severity, respectively. The first rule can also be modified to the following one, in order to avoid any false positives generated by the characterization of motion or sound data; in case an estimation of fall is generated but is followed by high visual and motion activity, and the movement speed of the body's visual trajectory is above a predefined threshold, then the subject has not fallen, but moves rather fast:

```
Patient(?x)^Location(?y)^hasFallEstimation(?
x,?y, ?FallEstimation)^
hasVisualActivity(?x,?VisualActivity)^hasMot
ionActivity(?x,?MotionActivity)^
Location(?Proximity)^
TrajectorySpeed(?Speed)^
swrlb:equals(?FallEstimation,?true)^swrlb:eq
uals(?Proximity,?High)
^swrlb:equals(?VisualActivity,?High)^swrlb:e
quals(?MotionActivity,?High)
^swrlb:equals(?speed,?High)
->DefineIncidentSeverity(?Severity,"VERY
LOW")
```

VI. Results and Experimental Evaluation

This section presents the results and experimental evaluation of the proposed system. The incorporated algorithms and tools for classifying the motion, audio, and visual perceptual components acquired by the methodology previously described are discussed in the following sections. The evaluation of the system involves the assessment of the system's accuracy in properly characterizing falls as well as the user-based evaluation in terms of acceptance and effectiveness and technical acceptability.

A. Classification and Metaclassification Methods

Several advanced classification techniques have been utilized in order to build proper models for proper activity and status recognition. The selection of the specific algorithms was based on their utilization in related work for fall detection (see Section II). The features used for classification are summarized in Table I. The examined algorithms were: Bayes networks, naïve Bayes, naïve Bayes multinomial, support vector machines (SVMs), logistic regression, multilayer perceptron, nearest neighbor and K-nearest neighbor, neural networks, PART, NBTree, and SimpleCart. In addition, the following metaclassifiers have also been used:

AdaBoost [36]: Class for boosting a nominal class classifier using the Adaboost M1 method. Often dramatically improves performance, but sometimes overfits.

Classification via regression [37]*:* Class for doing classification using regression methods. Class is binarized and one regression model is built for each class value.

CVparameterSelection [38]*:* Class for performing parameter selection by cross validation for any classifier.

RandomSubSpace [39]*:* This method constructs a decision tree-based classifier that maintains highest accuracy on training data and improves on generalization accuracy as it grows in complexity. The classifier consists of multiple trees constructed systematically by pseudo randomly selecting subsets of components of the feature vector, i.e., trees constructed in randomly chosen subspaces.

NestedDichotomies [40]*:* A metaclassifier for handling multiclass datasets with two class classifiers by building a random class-balanced tree structure.

Dagging [41]*:* This metaclassifier creates a number of disjoint, stratified folds out of the data, and feeds each chunk of data to a copy of the supplied base classifier. Predictions are made via majority vote, since all generated base classifiers are put into the vote metaclassifier. This metaclassifier is useful for base classifiers that are quadratic or worse in time behavior, regarding the number of instances in the training data. Usually in this case, SVMs are used as base classifiers.

ThresholdSelector [42]*:* A metaclassifier that selecting a midpoint threshold on the probability output by a classifier. The midpoint threshold is set so that a given performance measure is optimized. Currently, this is the F-measure. Performance is measured either on the training data, a hold-out set, or using cross validation. In addition, the probabilities returned by the base learner can have their range expanded so that the output probabilities will reside between 0 and 1 (this is useful if the scheme normally produces probabilities in a very narrow range).

B. Evaluation of the Proposed System

The evaluation of the proposed system has been performed based on the experimental protocol and testbed setup, as described in Section IV. Two male volunteers of average height and weight at the ages of 28 and 35 wearing the sensors devices described in Section IV performed combinations of movement types. Twelve recordings in total have been utilized (each individual performing two experiments of three different motion combination types) that have provided 1440 s of recorded data (motion, sound, and visual data). The latter have been segmented into 5-s time frames (for sound processing) and annotated. The procedure involves the evaluation of classifiers, effectiveness of features, and information fusion, where the efficiency of the proposed classification model is calculated using a predefined procedure. The dominant method presented in literature, mainly used in situations, where the total amount of data is limited, in order to provide a sufficient amount of data for training and separately testing the developed model, is N-fold cross validation [43]. The most widely applied value for parameter N is 10, which is the value we selected for our experiments in order to verify each model's accuracy and performance: The 1440 s of recorded data were segmented into 5-s time frames resulted into 288 experimental time frames. Two hundred and sixty randomly selected time frames were used as a training data set, whereas the remaining 28 frames were used for testing. The latter process has been repeated ten times and the total error rate has been calculated from the average of each individual test error rate. The evaluation has been divided into two parts; initially, the characterization of motion using acceleration and sound data from the on-body sensors has been validated. Finally, the visual channel information has been added and the rules-based evaluation provides the essential fusion for complete fall incidents detection.

1) Motion Characterization Using Acceleration and Sound Data: Based on the number of sequential occurrence of a specific movement type, decision regarding a patient fall is taken. In order to improve the accuracy of the latter decision, Kalman filtering [28], [29] has been applied on the sequence of the movement-type association of each acceleration data set.– Fig. 7 represents the classification results and the significance of Kalman filtering from the conducted experiments (utilizing only motion and sound data) using the trained SVM model. Light

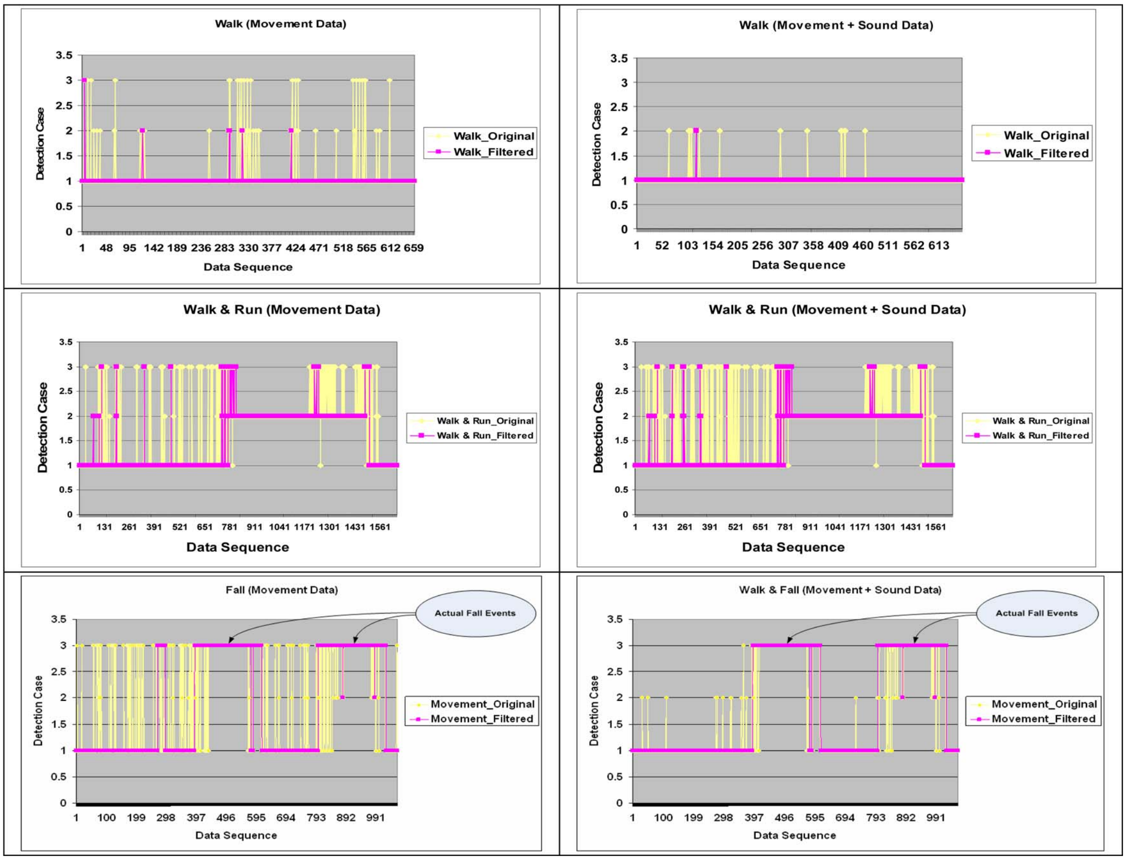


Fig. 7. Classification results from the conducted experiments using the trained SVM model. Light colored lines represent original results, whereas dark colored lines represent results after applying Kalman filtering. Diagrams on the left present accuracy results based on movement data, whereas diagrams on the right present accuracy results utilizing both motion and sound data.

colored lines represent original results, whereas dark colored lines represent results after applying Kalman filtering. The improvement in classification accuracy by utilizing both motion and sound features is also visible in Fig. 8 by the corresponding receiver operating characteristic (ROC) curves. Before applying Kalman filtering, the false positives for the case of falls were 60% of the total classified instances, and after applying Kalman filtering were minimized to 33%, respectively. For annotation purposes, the three movement types were associated with three integers: 1 for walk; 2 for run; and 3 for fall, respectively. Actual run and fall events are also annotated on the diagrams. For each experiment, two different diagrams were generated: one illustrating classification results based exclusively on acceleration data and one illustrating classification based on both acceleration and sound data. As it is indicated, Kalman filtering improves the overall detection by smoothing the sequential occurrences of run or fall events, respectively. In addition, the use of sound as additional classification feature has increased the accuracy of fall detections by minimizing the false ones in cases of simple walk and of walk and run. A threshold $t = 10$ has been selected for determining the occurrence of fall or run events from the total sequence of classified movement types (i.e., if sequential occurrence of fall movement types >10, then a fall is detected). Using the aforementioned classification and the latter threshold value, fall events were successfully detected in all cases, whereas run events were successfully detected at 96.72%.

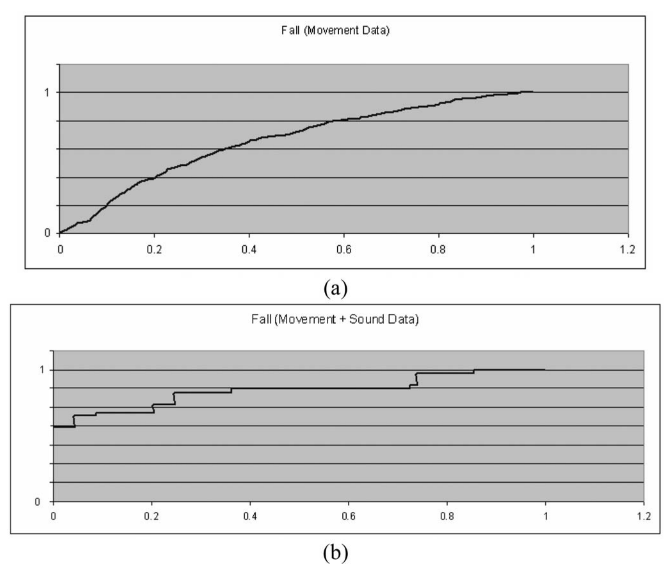


Fig. 8. ROC curves of SVM classifier performance in fall detection using: (a) motion data individually. Area under curve: 0.655, (b) motion and sound data. Area under curve: 0.885.

2) Evaluation of Motion, Sound, and Visual Information Channels: In addition, the classification results utilizing the algorithms described in Section VI, and the combination of motion, visual, and audio perceptual components and information fusion using semantic rules evaluation are presented in the Table II. A number of statistical analysis performance metrics have been used: the accuracy; the kappa statistic [44] (which measures the agreement between two raters who each classify N items into C mutually exclusive categories); and the root mean square error of the latter.

Additional evaluation metrics apart from accuracy are essential, since the latter is not sufficient itself for comparing the performance of different classifiers. For instance, lower kappa statistic (0.8923) and higher rms error (0.1337) in naïve Bayes than in Bayes NET (0.9392 and 0.1169, respectively) suggests that Bayes NET performs quite better than naïve Bayes, though the accuracy metric does not indicate such difference directly (96.49 versus 97.54, respectively). Bayes NET is also expected to perform better when including motion data, since naïve Bayes is based on the inherent assumption that features are conditionally independent and modeled by a normal distribution, which has been proved to be invalid when dealing with accelerometer data according to L. Bao *et al.* ([45], [46]).

As indicated by the evaluation results, the majority of the algorithms achieve high accuracy results. All fall incidents were successfully indicated in the case of SVM and AdaBoost meta-classifier, whereas related works in fall detection achieve up to 81% for visual information classification [13] and 91,58% for audio features classification [20]. The base classifier used with AdaBoost is the DecisionStump tree classifier, which performs regression based on mean-squared error. For SVM, the radial basis function kernel type has been used. The utilization of all the acquired perceptual components also improves the overall performance of the system, since fall determination accuracy is over 90% for the majority of classifiers with an average false positive range of 16.67% (in conjunction to 33% of false positive rate when using motion and sound data only). The utilization of rules-based evaluation (see Section IV) can potentially minimize false positives to zero.

TABLE II
EVALUATION RESULTS FOR THE DIFFERENT CLASSIFICATION ALGORITHMS USED ON MOTION, SOUND, AND VISUAL PERCEPTUAL COMPONENTS

Classification Algorithm	Algorithm Parameters	Correctly Classified Instances (%)
BayesNet	Simple Estimator, A: 0.5, search algorithm: hill climbing	97.54
NaiveBayes	No input	96.49
Logistic	Ridge value in log-likelihood: 1.0E-8	95.32
MultiLayerPerceptron	Learning rate: 0.3, learning time: 500 validation threshold: 20, num of epochs: 500	98.58
SVM	Kernel: RBF, C:1024, g : 0.125	100
IB1 (Nearest Neighbor)	No input	93.33
IBK (K Nearest)	KNN:1, Search Algorithm: LinearNNSearch [20], Euclidean distance	94.53
NNge	No Input	94.66
PART	Confidence Factor: 0.25,	99.29
NBTree	No input	100
SimpleCart	heuristic: true, numFold Pruning: 5	98.85
AdaBoost	Number of Iterations: 10, Weight Threshold: 100	100
ClassificationVia Regression	Classifier M5P, minNumInstances: 4	97.58
CVParameterSelection	Classifier: O-R [20]	88.17
RandomSubSpace	Classifier: REP-Tree [20]	99.87
NestedDichotomies	Classifier: J48 [20], confidence factor: 0.25	98.56
Dagging	Classifier: Support Vector, Kernel type: PolyKernel	86.8
ThresholdSelector	No Input	93.01

Classification Algorithm	Kappa statistic	Root mean squared error	Sensitivty/ Specifictiy (%)
BayesNet	0.9392	0.1169	97/93
NaiveBayes	0.8923	0.1337	96/88
Logistic	0.7389	0.2251	95/91
MultiLayerPerceptron	0.9478	0.0984	98/93
SVM	1	0.0181	100/100
IB1 (Nearest Neighbor)	0.5547	0.2734	93/88
IBK (K Nearest)	0.6547	0.2724	94/90
NNge	0.7094	0.231	94/92
PART	0.9674	0.0851	99/96
NBTree	1	0.0219	100/100
SimpleCart	0.9361	0.1193	98/95
AdaBoost	1	0.0211	100/100
ClassificationVia Regression	0.9347	0.1562	97/93
CVParameterSelection	0.2365	0.3343	88/80
RandomSubSpace	0.9044	0.1238	99/90
NestedDichotomies	0.9391	0.1222	98/96
Dagging	0.1895	0.3054	86/72
ThresholdSelector	0.6958	0.228	93/87

C. System User-Based Evaluation

The system in addition to its correctness on the detection of specific human activities was evaluated in terms of usability, user friendliness, and reliability. According to literature review, several factors can be utilized in order to evaluate patient-related applications. The most common way of measuring the aforementioned factors is the mean opinion score (MOS). In a subjective test, a number of people rate their quality of experience on a scale of 1 (bad) to 5 (excellent). The average of the scores is called an MOS. The resulting MOS depends on the range of experiences that were exposed to the group and to the type of experience being rated. Based on the evaluation performed by [47], four criteria of human factors evaluation are utilized: 1) technical acceptability; 2) operational effectiveness; 3) clinical appropriateness; and 4) equipment selection. Technical acceptability refers to issues like sensor communication, battery life, etc., operational effectiveness refers to system effectiveness (i.e., detection correctness), clinical appropriateness refers to usability as accepted by the treatment personnel, and equipment selection to issues like sensor wearability and convenience as judged by the patients.

In this context, a survey has been performed asking users and treatment personnel to evaluate the developed platform using

TABLE III
MEAN OPINION SCORE (MOS) FOR (A) TECHNICAL ACCEPTABILITY, (B) OPERATIONAL EFFECTIVENESS, (C) CLINICAL APPROPRIATENESS, AND (D) EQUIPMENT SELECTION AS INDICATED BY THE CONDUCTED SURVEY BETWEEN PATIENTS AND MEDICAL PERSONNEL THAT EVALUATED THE PRESENTED PLATFORM

	A	B	C	D
Patient	4.8	5.0	5.0	4.2
Personnel	4.2	4.6	5.0	4.8

the aforementioned method. A total number of ten individuals acting as patients (i.e., wearing the sensors) and ten experts in medical treatment have used the system and completed the survey. The corresponding results are presented in Table III. Both patients and personnel were asked to evaluate all four criteria, since all four of them can affect either the acceptability or the usability of the system. The performance of the system in fall detection has not been evaluated in the particular survey.

As indicated by the survey, the system has met great acceptability in the context of effectiveness (criteria B) and usability (criteria C). Technical acceptability has achieved some lower score due to the fact that users expected higher communication range between the sensors and the monitoring unit, and better battery life. Patients also noted that sensors could be more light and comfortable (criteria D). Future evolution of sensor technologies will address such issues improving communication, energy consumption, and wearability.

VII. DISCUSSION AND CONCLUSION

The costs of health care impose an enormous burden on the economy. The latest projections from the Centers for Medicare & Medicaid Services show that annual health-care expenditures are expected to reach $3.1 trillion by 2012, growing at an average annual rate of 7.3% during the forecast period or 17.7% of gross domestic product, up from 14.1% today [48]. Recent advances in communication and information technologies have impulsed the development of novel tools that enable remote management and monitoring of chronic disease, emergency conditions, and the delivery of health care on patient's site, saving time, travel, and other expenses. Health monitoring in home environments can be accomplished by establishing ambulatory monitors that utilize wearable sensors and devices that record physiological signals, sensors embedded in the home environment to collect behavioral and physiological data, or a combination of the latter. Studies for such noninvasive monitoring technologies have shown good acceptance rates by patients, presenting overall a positive impact on their perceived quality of life, as well as reducing hospitalization costs [49].

In this paper, an emergency fall incident detection platform has been presented that combines motion, visual, and audio information. It is a combined effort and elaboration of previous works of the authors ([15], [16], [21], [22]) assessing the latter perceptual components for motion characterization. Patient falls, especially in the case of elderly, are a great cause of injuries and happen both in home and hospital environments with great frequencies: a few thousands of incidents have been reported in the USA annually ([50], [51]). In the presented system, overhead cameras can track patient body movement, whereas microphone arrays record emergency sounds. Motion data and patient-generated audio sounds are collected through body sensors on the patient. Audio data processing and sound directionality analysis in conjunction to motion information and subject's visual location can verify fall and indicate an emergency event. Postfall visual and motion behavior of the subject indicates the severity of the fall based on semantic incident representation and rule-based evaluation. Proper rules among with information from all three channels can be used to minimize any false positives that can be generated by motion or audio characterization. Classification results along with user-based evaluation have shown promising results for the systems accuracy and acceptability in the context of incidents detection in assisted living environments. The system can also operate and provide estimations by individually utilizing the acquired data and contextual information. Even in cases, where visual information is not available, the previously recorded information (e.g., the motion trajectory of the patient going to the bathroom) in conjunction to context modeling through the ontology and rules evaluation (e.g., being in bath for several hours) could be used for estimating a distress situation (e.g., patient being unconscious in the bath).

All fall events have been simulated by volunteers, trying to be as much realistic as possible, under normal indoor lighting conditions, normal background noise, and relative distance to sensor receivers. Simulated falls can affect the overall system evaluation and performance; however, actual evaluation of the platform in a real environment by patients (e.g., the elderly) introduces the problem of collecting real falls and related incidents in a reasonable time frame. The evaluation results presented in the paper have been acquired using a relatively small number of subjects; however, initial results especially those concerning the characterization of falls against other movement types (i.e., walking and running) are very promising, especially when compared against results in related work (e.g., 81% fall detection in [13] using cameras, and 91.58% in [20] using sound information). The low false positive rates achieved (average false positive rate 16.67% when all perceptual components are utilized) are also very competitive against the values reported by related works in literature. The improvement in performance has been achieved when adding sound features to accelerometer data and when adding visual features to the latter as well. Therefore, resolving any potential redundant features issues has not been considered. Furthermore, the total number of features used for classification for all perceptual components is eight, and thus, no complexity or time training issues have been raised.

A limitation of the platform may be considered the equipment that needs to be installed within the monitored area (microphones and overhead cameras) and the sensors worn by patients. Despite the acceptability the system has met, as discussed in Section VI, the body sensor networks are still considered as invasive technology and require special treatment by users (e.g., proper body placement, battery replacement, etc.). Future evolution of sensor technologies will address such issues improving communication, energy consumption, and wearability, by consisting of more lightweight and less invasive sensors. Significant research effort is expected to be consumed in this field in the near future [52].

Extended clinical evaluations of fall-detection systems, like the one presented in this paper with more potential users and care experts, collecting real falls and related incidents in a reasonable time frame, and conforming to strict protocols for clinical evaluation, remain now to persuade industrial healthcare partners to invest in such technologies and produce corresponding commercial products. These studies of patient and elder outcomes require large numbers of participants and significant budgets, which are not always easy to find and escape the capabilities scope of scientific research.

The presented platform may also be extended to facilitate the monitoring of patient's behavior. Motion and sound data analysis can be utilized to recognize unexpected patterns in the behavior of patients diagnosed with cognitive impairments (e.g., dementia). Proper modification of the train model and semantic representation of the patient's context can help the assessment of the phenomenon progress and detect related incidents like amnesia attacks. The on-body wireless nodes can be enhanced with biosignal sensors (e.g., ECG, glucose, and temperature sensors) and provide a more complete assessment of the patient's status. Finally, methods like in [53] can be incorporated to prevent injury and improve human safety in cases of patient fall.

REFERENCES

[1] L. Lines and K. S. Hone, "Eliciting user requirements with older adults: Lessons from the design of an interactive domestic alarm system," *Universal Access Inf. Soc.*, vol. 3, pp. 141–148, Jun. 2004.

[2] R. Orpwood, C. Gibbs, T. Adlam, R. Faulkner, and D. Meegahawatte, "The design of smart homes for people with dementia: User-interface aspects," *Universal Access Inf. Soc.*, vol. 4, no. 2, pp. 156–164, Dec. 2005.

[3] V. G. Koutkias, I. Chouvarda, and N. Maglaveras, "A multiagent system enhancing home-care health services for chronic disease management," *IEEE Trans. Inf. Techol. Biomed.*, vol. 9, no. 4, pp. 528–537, Dec. 2005.

[4] N. Maglaveras, I. Chouvarda, V. G. Koutkias, G. Gogou, I. Lekka, D. Goulis, A. Avramidis, C. Karvounis, G. Louridas, and E. A. Balas, "The citizen health system (CHS): A modular medical contact center providing quality telemedicine services," *IEEE Trans. Inf. Techol. Biomed*, vol. 9, no. 3, pp. 353–362, Sep. 2005.

[5] N. Noury, T. Herve, V. Rialle, G. Virone, E. Mercier, G. Morey, A. Moro, and T. Porcheron, "Monitoring behavior in home using a smart fall sensor and position sensors," *Proc. 1st Annu. Int. Conf. Microtechnol. Med. Biol.*, pp. 607–610, Oct. 2000.

[6] I. Korhonen and N. Saranummi, "Personal health systems: Need, market place and challenges to their wide scale adoption," *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS)*, pp. 6152–6153, 22–26, Aug. 2007.

[7] N. Noury, "A smart sensor for the remote follow up of activity and fall detection of the elderly," in *Proc. 2nd Annu. Int. Conf. Microtechnol. Med. Biol.*, pp. 314–317, May 2002.

[8] M. Prado, J. Reina-Tosina, and L. Roa, "Distributed intelligent architecture for falling detection and physical activity analysis in the elderly," *(EMBS)*, pp. 1910–1911, Oct. 2002.

[9] K. Fukaya, "Fall detection sensor for fall protection airbag," *Proc. 41st SICE Annu. Conf.*, pp. 419–420, Aug. 2002.

[10] H. Nait-Charif and S. J. McKenna, "Activity summarisation and fall detection in a supportive home environment," *In Proc. 17th Int. Conf. Pattern Recog. (ICPR)*, pp. 323–236, Aug. 2004.

[11] J. Y. Hwang, J. M. Kang, Y. W. Jang, and H. C. Kim, "Development of novel algorithm and real-time monitoring ambulatory system using Bluetooth module for fall detection in the elderly," *Proc. 26th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 2204–2207, 2004.

[12] S. Wang, J. Yang, N. Chen, X. Chen, and Q. Zhang, "Human activity recognition with user-free accelerometers in the sensor networks," *Proc. Int. Conf. Neural Netw. Brain*, pp. 1212–1217, Oct. 2005.

[13] S.-G. Miaou, P.-H. Sung, and C.-Y. Huang, "A customized human fall detection system using omni-camera images and personal information," in *Proc. 1st Transdisciplinary Conf. Distrib. Diagnosis Home Healthcare*, pp. 39–42, 2006.

[14] F. R. Allen, E. Ambikairajah, N. H. Lovell, and B. G. Celler, "An adapted Gaussian mixture model approach to accelerometry-based movement classification using time-domain features," *Proc. 28th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 3600–3603, Aug. 2006.

[15] C. Doukas, I. Maglogiannis, P. Tragkas, D. Liapis, and G. Yovanof, "Patient fall detection using support vector machines," presented at the Artif. Intell. Innovations 2007: From Theory to Applications, Int. Federation Inf. Process., 2007, vol. 247/2007, pp. 147–156. DOI: 10.1007/978-0-387-74161-1_16.

[16] C. Doukas and I. Maglogiannis, "Advanced patient or elder fall detection based on movement and sound data," in *Proc. 2nd Int. Conf. Pervasive Comput. Technol. Healthcare*, Tampere, Finland, Jan. 30–Feb. 1, 2008, pp. 103–107.

[17] D. Istrate, E. Castelli, M. Vacher, L. Besacier, and J.-F. Serignat, "Information extraction from sound for medical telemonitoring," *IEEE Trans. Inf. Theory Biomed.*, vol. 10, no. 2, pp. 264–274, Apr. 2006.

[18] M. Vacher, D. Istrate, and J. F. Serignat, "Sound detection and classification through transient models using wavelet coefficient trees," presented at the 12th EURASIP Eur. Signal Process. Conf., Vienna, Austria, Sep. 2004.

[19] D. Istrate, M. Vacher, and J. Francois Serignat, "Generic implementation of a distress sound extraction system for elder care," *Proc. 28th IEEE Eng. Med. Biol. Soc. (EMBS) Annu. Int. Conf.,* New York, pp. 3309–3312, Aug. 30–Sep. 3, 2006.

[20] N. C. Laydrus, E. Ambikairajah, and B. Celler, "Automated sound analysis system for home telemonitoring using shifted delta cepstral features," *Proc. 15th Int. Conf. Digital Signal Process.*, pp. 135–138, Jul. 2007.

[21] C. Doukas and I. Maglogiannis, "Human distress sound analysis and characterization using advanced classification techniques," presented at the 5th Hellenic Conf. Artif. Intell. (SETN), Syros, Greece, Oct. 2–4, 2008.

[22] C. Doukas, I. Maglogiannis, N. Katsarakis, and A. Pneumatikakis, "Enhanced human body fall detection utilizing advanced classification of video and motion perceptual components," presented at the 5th IFIP Conf. Artif. Intell. Appl. Innovations, Salonica, Greece, April 2009.

[23] (2010, Apr.). [Online]. Available: http://www.freescale.com.

[24] The Sentilla Perk Pervasive Computing Kit. (2010, Apr.). [Online]. Available: http://www.sentilla.com/perk.html.

[25] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 747–757, 2000.

[26] SensiNode. (2010, Apr.). [Online]. Available: http://www.sensinode.com.

[27] The JAVA ME Platform. (2010, Apr.). [Online]. Available: http://java.sun.com/javame/index.jspf.

[28] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1129–1159, 1995.

[29] F. Talantzis, A. G. Constantinides, and L.C. Polymenakos, "Estimation of direction of arrival using information theory," *IEEE Signal Proc. Lett.*, vol. 12, no. 8, pp. 561–564, Aug. 2005.

[30] A. Pnevmatikakis and L. Polymenakos, "Robust estimation of background for fixed cameras," presented at the Int. Conf. Comput. (CIC), Maxico City, Mexico, 2006.

[31] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME J. Basic Eng. (Series D)*, vol. 82, pp. 35–45, 1960.

[32] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, Aug. 1976.

[33] Protégé Ontology Editor and Knowledge Base Framework. (2010, Apr.). [Online]. Available: http://protege.stanford.edu/.

[34] The Semantic Web Rule Language Definition. (2010, Apr.). [Online]. Available: http://www.w3.org/Submission/SWRL/.

[35] The Jess Rule Engine. (2010, Apr.). [Online]. Available: http://www.jessrules.com/jess/index.shtml.

[36] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proc. 13th Int. Conf. Mach. Learning*, pp. 148–156, 1996.
[37] E. Frank, Y. Wang, S. Inglis, G. Holmes, and I. H. Witten, "Using model trees for classification," *J. Mach. Learning*, vol. 32, no. 1, pp. 63–76, 2011.
[38] R. Kohavi, "Wrappers for performance enhancement and oblivious decision graphs," Ph.D. dissertation, Dept. Comput. Sci., Stanford University, Palo Alto, CA, 1995.
[39] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach.*, vol. 20, no. 8, pp. 832–844, 1998.
[40] L. Dong, E. Frank, and S. Kramer, "Ensembles of balanced nested dichotomies for multi-class problems," *Proc. Int. Conf. Principles Practice Knowl. Discovery Databases (PKDD)*, pp. 84–95, 2005.
[41] K. M. Ting and I. H. Witten, "Stacking bagged and dagged models," *Proc. 14th Int. Conf. Mach. Learning,* 1997, San Francisco, pp. 367–375, 1997.
[42] I. H. Witten and E. Frank, *Data mining: Practical Machine Learning Tools and Techniques*, 2nd ed ed. San Francisco: Morgan Kaufmann, 2005.
[43] P. A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach*. London, U.K.: Prentice-Hall, 1982.
[44] J. Carletta, "Assessing agreement on classification tasks: The kappa statistic," *Comput. Linguistics*, vol. 22, no. 2, pp. 249–254, 1996.
[45] S. S. Intille, L. Bao, and L. Bao, (2003), *Physical activity recognition from acceleration data under seminaturalistic conditions*.
[46] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *Proc. 2nd Int. Conf. Pervasive Comput.*, 2004, pp. 1–17.
[47] D. P. Birkmire-Peters, L. J. Peters, and L. A. Whitaker, "A usability evaluation for telemedicine medical equipment," *Telemed. J.*, vol. 5, no. 2, pp. 209–212, July 1, 1999.
[48] Centers for Medicare and Medicaid Services, (2010, Apr.). [Online]. Available: http://cms.hhs.gov/.
[49] M. Alwan, S. Dalal, D. Mack, S. W. Kell, B. Turner, J. Leachtenauer, and R. Felder, "Impact of monitoring technology in assisted living: Outcome pilot," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 1, pp. 192–198, Jan. 2000.
[50] M. J. Krauss, S. L. Nguyen, W. C. Dunagan, S. Birge, E. Costantinou, S. Johnson, B. Caleca, and V. J. Fraser, "Circumstances of patient falls and injuries in 9 hospitals in a midwestern healthcare system," *Infect Control Hosp. Epidemiol.*, vol. 28, no. 5, pp. 544–550, May 2007.
[51] B. Jackson. (2009). Protecting the elderly from falls *National Safety Council*. [Online]. Available: http://www.nsc.org/resources/issues/articles/bj032902.aspx.
[52] P. F. Binkley, "Predicting the potential of wearable technology," *IEEE Eng. Med. Biol. Mag.*, vol. 22, pp. 23–27, May/Jun. 2003.
[53] T. Tamura, T. Yoshimura, and M. Sekine, "A study to demonstrate the use of an air bag device to prevent fall-related injuries," *Proc. of 8th IEEE Int. Conf. Bioinf. Bioeng. (BIBE)*, pp. 1–3, Oct. 2008.

Charalampos C. Doukas received the Diploma degree in information and communication systems engineering from the University of the Aegean, Mytilene, Greece, in 2005, where he is currently working toward the Ph.D. degree.

His main interests include video and image processing of medical data, medical ontologies and semantics, medical data classification, medical sensors, and data transmission over heterogeneous networks. He has published more than 40 papers in international scientific conferences, 13 journal papers and 5 book chapters. He is currently an external Researcher at the University of the Aegean University of Central Greece and at the National Hellenic Research Institute.

Dr. Doukas is member of the Technical Chamber in Greece, member of Hellenic Artificial Intelligence Society, member of IEEE EMBS Greek Society and peer reviewer in scientific journals like IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE and IEEE TRANSACTIONS IN MEDICAL IMAGING.

Ilias Maglogiannis (M'00) received the Diploma degree in electrical and computer engineering and the Ph.D. degree in biomedical engineering and medical informatics from the National Technical University of Athens (NTUA), Athens, Greece, in 1996 and 2000, respectively.

From 1996 to 2000, he was a Researcher in the Biomedical Engineering Laboratory, NTUA. In 2001, he was a Lecturer in the Department of Information and Communication Systems Engineering, University of the Aegean, Mytilene, Greece. Since 2008, he has been an Assistant Professor in the Department of Computer Science and Biomedical Informatics, University of Central Greece, Lamia, Greece. He has been a Principal Investigator in many European and National Research programs in biomedical engineering and informatics. His research interests include biomedical engineering, image processing, computer vision, and multimedia communications. He is the author or coauthor of over 100 publications in the aforementioned areas.

Dr. Maglogiannis has been on program committees of national and international conferences and is also a Reviewer for several scientific journals. He is a member of the International Society for Optical Engineering and the Hellenic Association of Biomedical Engineering.