Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2008, Article ID 149304, 7 pages
doi:10.1155/2008/149304

Research Article

Falling Person Detection Using Multisensor Signal Processing

B. Ugur Toreyin, E. Birey Soyer, Ibrahim Onaran, and A. Enis Cetin

Department of Electrical and Electronics Engineering, Faculty of Engineering, Bilkent University, 06800 Bilkent, Ankara, Turkey

Correspondence should be addressed to B. Ugur Toreyin, ugur@ee.bilkent.edu.tr

Received 28 February 2007; Accepted 12 September 2007

Recommended by Eric Pauwels

Falls are of the most important problems for frail and elderly people living independently. Early detection of falls is vital to provide a safe and active lifestyle for elderly. Sound, passive infrared (PIR), and vibration sensors can be placed in a supportive home environment to provide information about daily activities of an elderly person. In this paper, signals produced by sound, PIR, and vibration sensors are simultaneously analyzed to detect falls. Hidden Markov models (HMM) are trained for regular and unusual activities of an elderly person and a pet for each sensor signal. Decisions of HMMs are fused together to reach a final decision.

1. INTRODUCTION

Detection of a falling person in an unsupervised area is a practical problem with applications in safety and security areas including supportive home environments. Intelligent homes will have the capability of monitoring activities of their occupants and automatically provide assistance to elderly people and young children using a multitude of sensors in the near future. Currently used worn sensors include passive infrared sensors, accelerometers, and pressure pads [1–5]. However, they may produce false alarms and elderly people simply forget wearing them very often. Computer vision-based systems may provide effective and complimentary solutions for fall detection [6]. Although visual systems are highly successful for detection of a fall, cameras must be placed in several parts of the house including bathrooms. Even if the video data is neither stored nor sent to an outside center for further processing, many people may find such practice disturbing.

A combination of passive infrared (PIR), sound, and vibration sensors provide an efficient solution for fall detection. In this paper, signals produced by these sensors are simultaneously analyzed to detect falling elderly people. Sound, PIR, and vibration sensors complement each other. For example, step sounds are hard to record, if there is a rug on the floor. However, low cost vibration sensors can be placed under a rug and they can capture vibrations due to a walking person or a pet. On the other hand, vibration sensors cannot be placed on hard floors. Instead, sound sensors can easily capture a fall on hard floors. PIR sensors easily detect the motion in a room but they cannot as reliably distinguish the motion of a pet from the owner as a sound sensor or a vibration sensor.

In this paper, signals produced by each sensor are processed separately in the wavelet domain. It is experimentally observed that the wavelet transform domain signal processing provides better results than the time-domain signal processing because wavelets capture sudden changes in the signal and ignore stationary parts of the signal. For our purposes, it is important to detect sudden changes rather than drifts or low frequency variations. Feature parameters are extracted from wavelet signals in fixed-length data windows and they are used in hidden Markov models (HMMs) which are trained according to possible human being and pet activities including falls.

In Section 2, analysis of the sound sensor signal is presented. The details of the PIR and vibration sensor data processing are described in Sections 3 and 4, respectively. In Section 5, experimental results are presented.

2. ANALYSIS OF THE SOUND SENSOR SIGNAL

In a typical intelligent supportive home environment, microphones can be placed in rooms and hallways. Audio signals captured by sound sensors can be used to detect a suddenly falling person. A typical nine seconds long stumble and fall

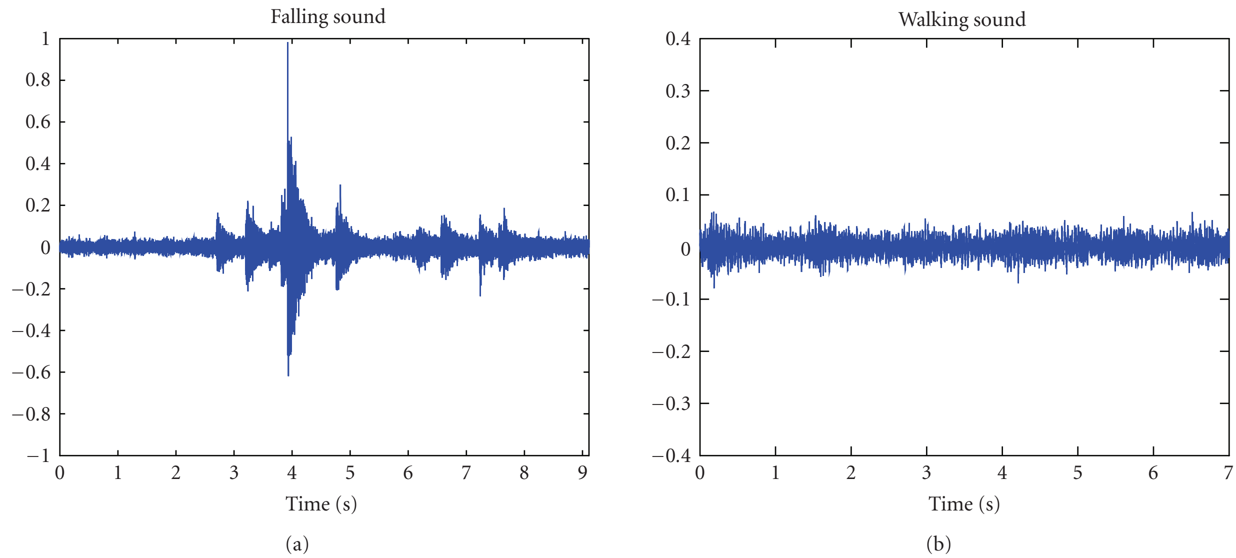

Figure 1: (a) Falling and (b) walking person sound recordings.

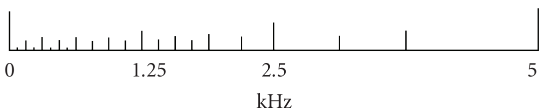

Figure 2: The subband frequency decomposition of the sound signal.

recording is shown in Figure 1(a), and step sounds are shown in Figure 1(b). In this case, the two sound waveforms are clearly different from each other. However, these waveforms may "look" similar as the distance from the sensor increases. For some other cases such as when TV set is on and loud, it may become even harder to distinguish a sound activity from the background noise. In addition, almost periodic nature of step sounds is hard to observe in the time domain signal but it becomes obvious after wavelet domain signal processing (compare Figures 1(b) and 3(b)). Another problem to be solved is that sound activity due to a person or a pet should be distinguished from the background noise.

Significant voice activity is detected using the Teager-energy-operator-based speech features originally developed by Jabloun and Cetin [7–9]. The sound data is divided into 1000-sample-long frames and the Teager-energy-based cepstral (TEOCEP) [7] feature parameters are obtained using wavelet domain signal analysis. The sound signal at each frame is divided into 21 nonuniformly divided subbands similar to the Bark scale (or mel-scale) giving more emphasis to low-frequency regions of the sound.

To calculate the TEOCEP feature parameters, a two-channel wavelet filter bank is used in a tree structure to divide the audio signal $s(n)$ according to the mel-scale as shown in Figure 2, and 21 wavelet domain subsignals $s_l(n), l = 1, \dots, L = 21$, are obtained [10–12]. The filter bank of a biorthogonal wavelet transform is used in the analysis [13]. The lowpass filter has the transfer function

$$H_l(z) = \frac{1}{2} + \frac{9}{32}\left(z^{-1} + z^{1}\right) - \frac{1}{32}\left(z^{-3} + z^{3}\right), \tag{1}$$

and the corresponding high-pass filter has the transfer function

$$H_h(z) = \frac{1}{2} - \frac{9}{32}\left(z^{-1} + z^{1}\right) + \frac{1}{32}\left(z^{-3} + z^{3}\right). \tag{2}$$

For every subsignal, the average Teager energy e_l is estimated as follows:

$$e_l = \frac{1}{N_l}\sum_{n=1}^{N_l} \left|\Psi[s_l(n)]\right|, \quad l = 1, \dots, L, \tag{3}$$

where N_l is the number of samples in the lth band, and the Teager energy operator (TEO) is defined as follows:

$$\Psi[s(n)] = s^2(n) - s(n+1)s(n-1). \tag{4}$$

The TEO-based cepstrum coefficients are obtained after log-compression and inverse DCT computation as follows:

$$TC(k) = \sum_{l=1}^{L} \log(e_l) \cos\left[\frac{k(l-0.5)\pi}{L}\right], \quad k = 1, \dots, N. \tag{5}$$

The first 12 $TC(k)$ coefficients are used in the feature vector.

The TEOCEP parameters are fed to the sound activity detector algorithm described in [6] to detect significant sound activity in a room.

When there is significant sound activity in the room, another feature parameter based on variance of wavelet coefficients and zero crossings is computed at each frame. Wavelet signals for each frame corresponding to the [2.5 kHz,

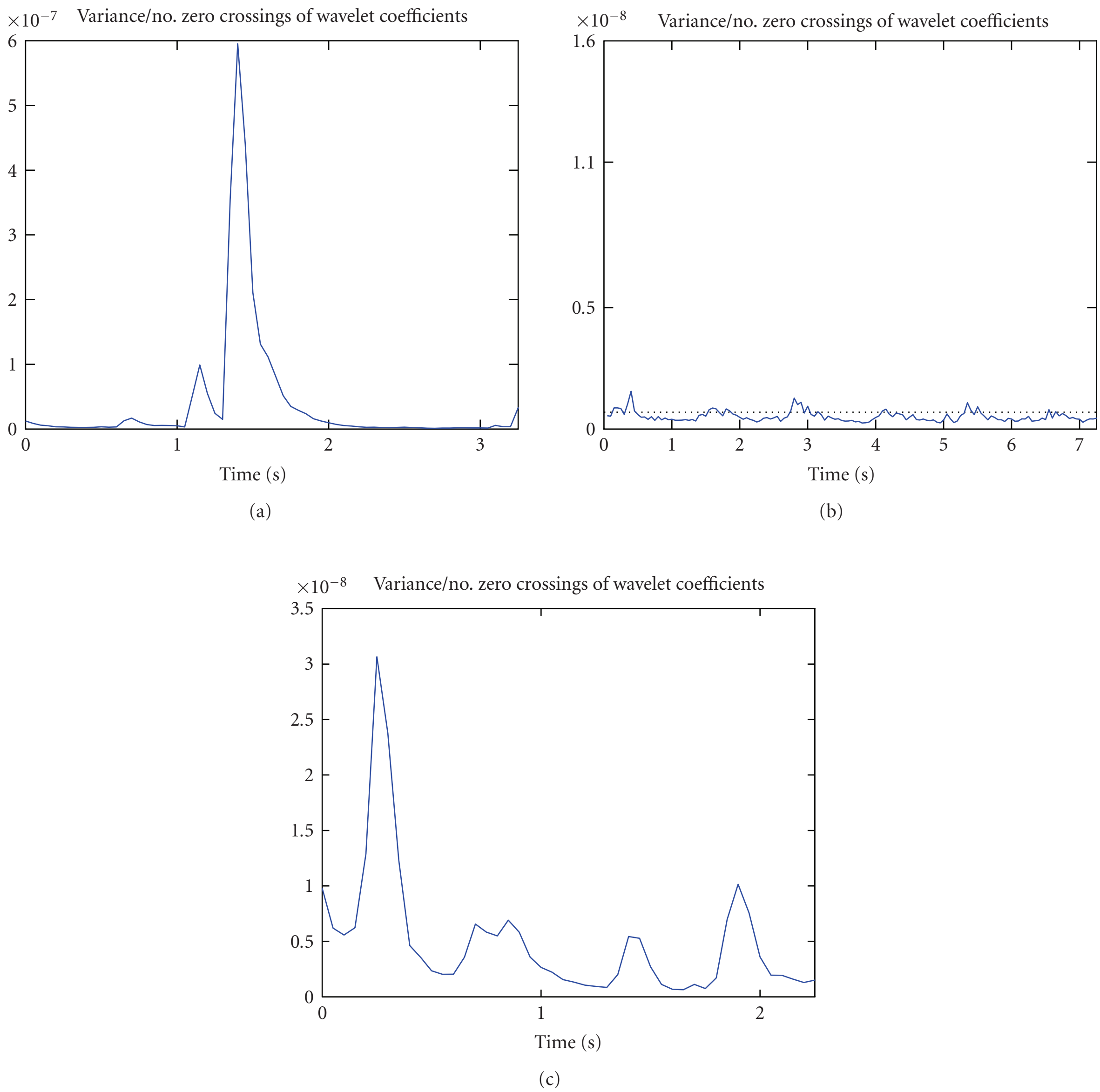

FIGURE 3: The ratio of variance of wavelet coefficients σ_i^2 over a number of zero crossings Z_i, $\kappa_i = \sigma_i^2/Z_i$: variations for (a) falling (1-2 seconds), (b) walking sounds, and (c) regular speech. Note that κ_i values for the walking case are an order of magnitude less than falling and regular speech cases. The threshold T is defined in κ-domain and marked with a line in (b).

5.0 kHz] frequency band are obtained after a single stage wavelet filterbank. The variance, σ_i^2 of the 500-sample-long wavelet window and the number of zero crossings, Z_i, in each window i is computed.

A typical step sound is similar to a single syllable quasiperiodic speech signal. On the other hand, broken glass and similar sounds are not quasiperiodic in nature. As walking is quasiperiodic, the zero crossing value, Z_i, is small compared to noise like sounds. When a person stumbles and falls, Z_i decreases whereas the variance of the wavelet signal σ_i^2 increases compared to the background noise. Shouting and crying for help are voiced sounds and have more energy in higher frequencies. Therefore, Z_i decreases when a person shouts. So we define a feature parameter κ_i in each window i as follows:

$$\kappa_i = \frac{\sigma_i^2}{Z_i}, \tag{6}$$

where the index i indicates the window number. The parameter κ_i takes nonnegative values.

The sound signal due to regular speech has a varying σ_i^2-Z_i characteristic depending on the utterance. When vowels are uttered, σ_i^2 increases while Z_i decreases, which results in larger κ values compared to consonant utterances. Variation of κ values versus sample numbers for different cases are shown in Figure 3.

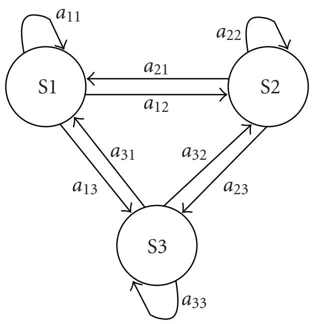

Figure 4: Three-state Markov model. Three Markov models are used to represent speech, walking, and fall sounds.

Activity classification based on sound information is carried out using HMMs. Three three-state Markov models are used to represent speech, walking, and fall sounds. In Markov models, S1 corresponds to the background noise or no activity. If sound activity detector (SAD) indicates that there is no significant activity, S1 is selected. If SAD detects sound activity in a sound frame, then either S2 or S3 is chosen as the current state according to the value of κ .

A nonnegative threshold value T that is small enough to reflect the periodicity in step sounds isintroduced in the κ-domain. In our implementation, we choose T as twice the standard deviation of κ values corresponding to no-activity portions of the input signal. If $|\kappa| < T$, we obtain S2; otherwise, S3 is attained as the current state. The classification performance of HMMs is based on the number of state transitions rather than specific κ values. Hence choice of T does not affect the values of the transition probabilities in different models as long as it reflects the almost periodic nature of step sounds.

In order to train HMMs, the state transition probabilities are estimated from 20 consecutive κ_i values corresponding to 20 consecutive 500-sample-long wavelet windows covering 125 milliseconds of audio data.

During the classification phase, a state history signal consisting of 20 κ_i values is estimated from the sound signal acquired from the audio sensor. This state sequence is fed to Markov models corresponding to walking, speech, and falling cases in running windows. The model yielding the highest probability is determined as the result of the analysis of the sound sensor signal.

The number of transitions between different states is large for a typical walking sound. Hence the probabilities of transitions between different states, a_{ij}'s, are higher than in-state transition probabilities, a_{ii}'s, for the walking model. On the other hand, feature parameter κ takes high values for a regular speech sound. Consequently, the value of a_{33} is higher than any other transition probabilities in the talking model. For the fall case, a relatively long no-activity/noise period is followed by a sudden increase and then a sudden decrease in κ values. This results in higher a_{11} value than any other transition probabilities. In addition to that, the number of transitions within, to and from S2, is notably fewer than those of S1 and S3. The state S2 in the Markov models provides hysteresis that prevents sudden transitions from S1 to S3 or vice versa, which is especially the case for walking.

3. PIR SENSOR DATA PROCESSING

Commercially available PIR sensors produce binary outputs; however, we capture a continuous amplitude analog signal indicating the strength of the received signal. The corresponding circuit is shown in Figure 5. The sampling rate is 300 Hz. A typical received signal is shown in Figure 6.

The strength of the received signal from a PIR sensor increases when there is motion due to a hot body within its viewing range. Therefore, it provides robustness against a possible confusion between typical voice activity and a fall analyzed by audio sensors only. Alarms produced by other sensors should be ignored when there is no motion in a room. On the other hand, the motion may be due to a pet or the owner. The PIR sensor data can be used to differentiate between the motion of a human being and an animal. Typically the PIR signal amplitudes for a person are higher than the amplitudes due to the motion of a pet as pets are smaller than human beings for a given distance as shown in Figure 7. However, a simple amplitude-based classification will not work because the IR signal amplitude decreases with distance. Another distinguishing factor is the speed of the motion. Pets move faster than human beings. This is reflected in the sensor output signal.

There is bias in the PIR sensor output signal, which changes according to the room temperature. Wavelet transform of the PIR signal removes this bias. Let $x[n]$ be a sampled version of the signal coming out of a PIR sensor. Wavelet coefficients obtained after a single stage subband decomposition, $w[k]$, corresponding to [75 Hz, 150 Hz] frequency band information of the original sensor output signal $x[n]$ are evaluated with the integer arithmetic high-pass filter, described in Section 2, corresponding to Lagrange wavelets [13] followed by decimation.

In this case, the wavelet transform coefficients $w[k]$'s are directly used as a feature parameter in an HMM-based classification. If the binary output of the PIR sensor indicates that there is no motion for the nth sample, then S1 is chosen as the current state. Similar to Section 2, we define a nonnegative threshold T_p in the wavelet domain. If there is a motion for the nth sample and the corresponding wavelet coefficient satisfies $|w[k]| < T_p$, we obtain state S2; otherwise, state S3 is attained as the current state.

Wavelet signal captures the high frequency information in the signal. Therefore, we expect that there will be more transitions occurring between states due to the motion of a pet.

For the training of the HMMs, similar to the audio signal processing step, the state transition probabilities for human being and pet models are estimated from 150 consecutive wavelet coefficients covering a time frame of one second.

During the classification phase, a state history signal consisting of 150 consecutive wavelet coefficients is computed from the received sensor signal. This state sequence is fed to the human being and pet models in running windows. The model yielding highest probability is determined as the result

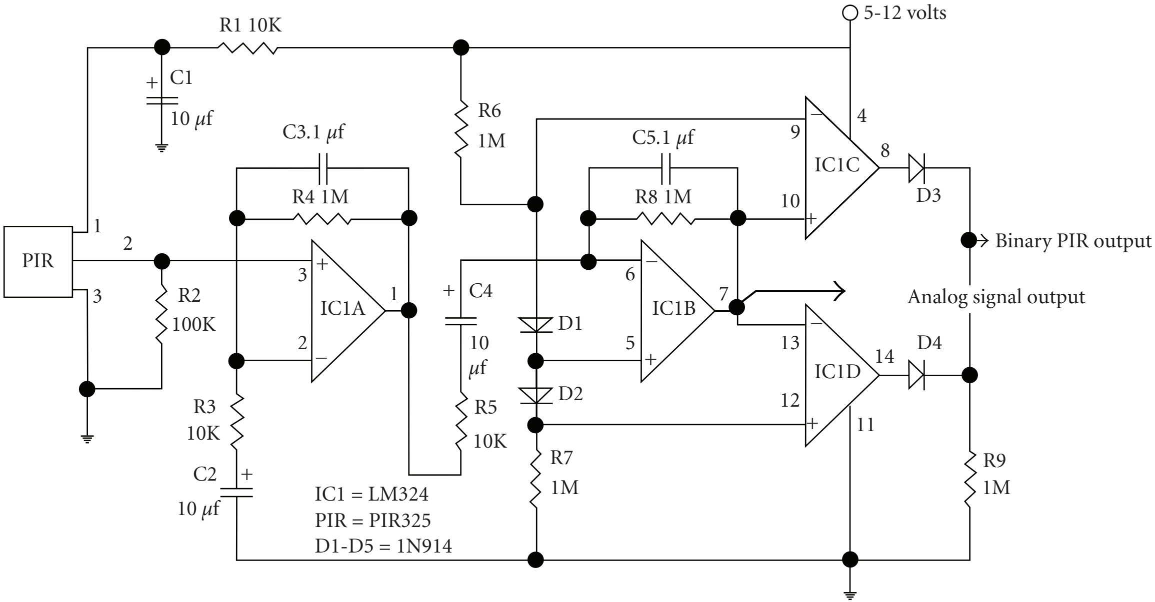

Figure 5: The circuit diagram for capturing an analog signal output from a PIR sensor.

of the analysis of PIR sensor data. The output of the sound-based decision system can be enhanced using the decision mechanism of the PIR sensor. For example, after a "fall" alarm is issued by the sound analysis system, there should not be any activity in the room or the only activity must be due to a pet. Also, when there is no activity in a room for a long time or only activity is due to a pet, a warning signal may be issued to the monitoring agency to check the elderly person.

4. VIBRATION SENSOR DATA PROCESSING

When there is a rug on the floor, it is very hard to capture any sound in a room. On the other hand, vibration sensors can be placed under the rug and vibration signals can be recorded. A typical output of a vibration sensor corresponding to a walking person is shown in Figure 8.

The peak in the signal is due to the pressure applied by a foot. In this study, a low-cost vibration sensor, ACH-01 manufactured by Measurement Specialties Inc., is used. It is observed that this sensor can capture the force applied by a foot or a falling person's body within an area of 25 cm^2. The rug used in our experiments has a thickness of 0.5 cm. Therefore, an array of sensors should be placed under a rug to cover the entire activity in a room.

When a person falls or sits on the floor, a multitude of sensors produces significant sensor outputs. In addition, the duration of sensor outputs is longer than a typical output due to a step, as shown in Figure 9. Moreover, vibration sensors can be placed under a mat or a couch to alarm for long-lasting inactivity.

A vibration signal due to a fall can be easily distinguished from a signal due to a step pressure by simply monitoring the duration of sensor outputs. In addition, several neighboring sensors produce output signals significantly larger in amplitude than background noise level at the same time during a fall.

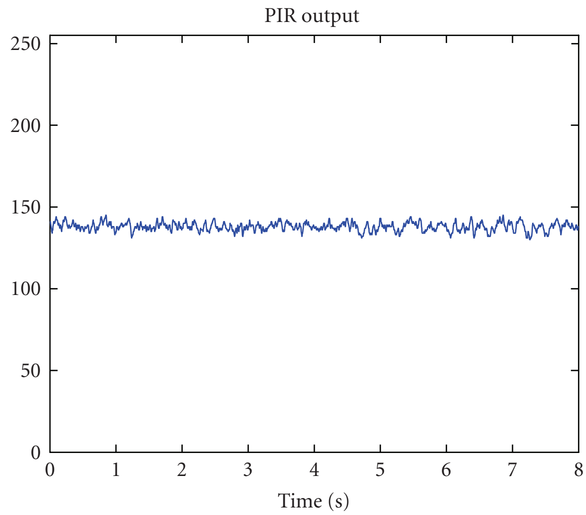

Figure 6: A typical PIR sensor output sampled at 300 Hz with 8 bit quantization when there is no activity in a room.

5. EXPERIMENTAL RESULTS

Models for sound and PIR sensor types are trained with four two-minute-long recordings of walking, falling, and speech signals of a single person and random activities of a pet. Falling detection results due to sound sensor outputs are compared with those which, when combined with the PIR sensor output, are presented in Table 1. Fusion of decisions from different sensors is realized by utilizing a logical "and" operation.

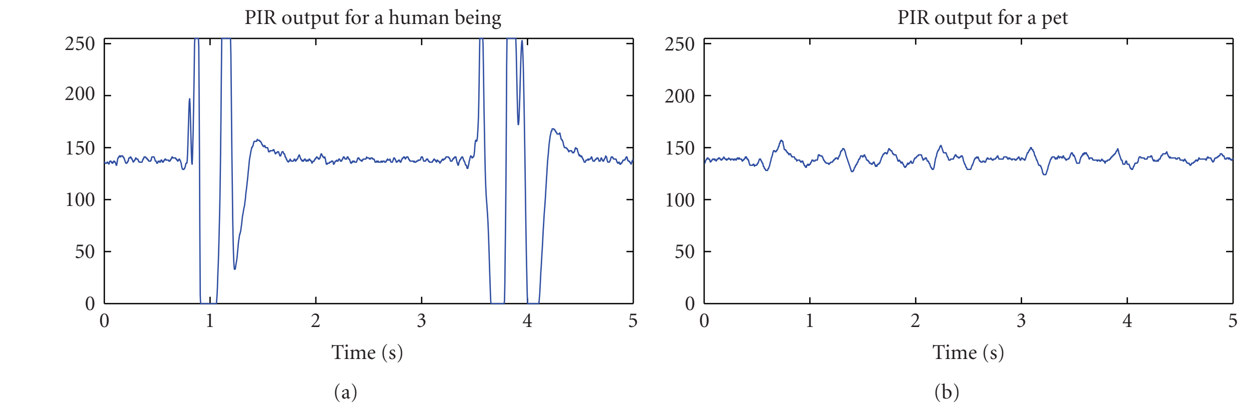

Figure 7: PIR sensor output signals recorded at a distance of 2 m for (a) a human being, and (b) a pet.

Table 1: Detection results and false alarms for 163-test recordings.

Audio signal content	No. of Recordings	No. of recordings in which a "fall" is detected		No. of recordings in which "false alarms" are issued	
		Using audio data only	Using both audio and PIR data	Using audio data only	Using both audio and PIR data
Walking + Speech	16	7	0	7	0
Speech	55	19	0	19	0
Walking	53	4	0	4	0
Falling	39	39	39	0	0

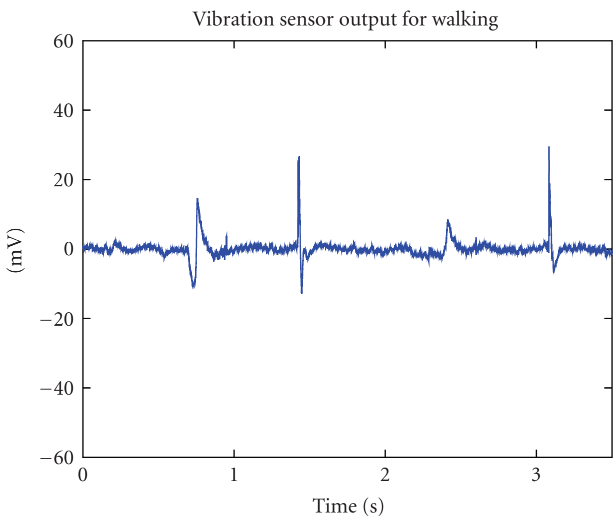

Figure 8: Vibration sensor output signal for a walking person.

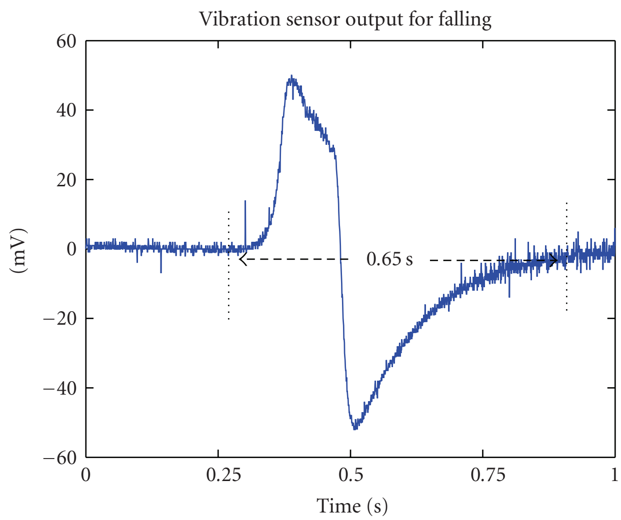

Figure 9: Vibration sensor output signal for a fall. The duration of a typical fall signal lasts more than 0.5 seconds.

A total of 163 recordings containing various activities are used for testing; 16 of the recordings contain both speech and step sounds, 55 contain speech without any motion, 53 contain step sounds, and 39 contain falling. When there is speech sound only or speech sound along with step sounds in the recordings, the system issues false alarms if only audio signal is used for the "fall" decision, as shown in the third and fifth columns of the table. It also issues alarms for recordings containing only walking sound. Last column of the table shows that false alarms are eliminated with the incorporation of the PIR sensor output signal in the decision.

This table does not include any experiments with a vibration sensor, but it is experimentally observed that the duration of a typical fall signal lasts more than 0.5 seconds. This is clearly larger than a step signal. Hence a vibration signal due to a fall and a signal due to step pressure are easily differentiable by just analyzing the duration of the sensor outputs.

6. CONCLUSION

In this paper, a method for detecting a fall inside an intelligent environment/building equipped with multitude of sound, vibration, and PIR sensors is proposed. Wavelet-based features are extracted from raw sensor outputs and are fed to a TEO-based sound activity detector. Similarly, PIR sensor outputs are also processed and sensor recordings containing various human and pet motions are used for training the HMMs corresponding to different activities including a fall. Vibration sensors are also used to detect human activity in rooms covered with rugs. Classification outputs from all sensors are fused together to reach a final decision.

The proposed multiple sensor system may be used as a substitute for camera-based monitoring systems and complimentary solution for wearable systems. It can be used in cooperation with a wearable sensor and a push-button type call system. The proposed system can be further improved to handle false alarm sources like barking dogs, slamming doors, vacuum cleaning, and so forth. This can be achieved by training models similar to ones defined in Section 2. Another possible false alarm scenario is when a person intentionally sits on the floor and wiggles. If there is a false alarm, then he or she can simply cancel it using his/her wearable call device. It may also be used to increase the robustness of camera-based systems in an intelligent building.

ACKNOWLEDGMENTS

This work is supported in part by the Scientific and Technical Research Council of Turkey, TUBITAK Grant nos. EEEAG-105E065 and SANTEZ-105E121, and the European Commission with Grant no. FP6-507752 MUSCLE NoE project. Authors are grateful to Ergul family and their pet Sutlac for helping in recording PIR data.

REFERENCES

[1] N. M. Barnes, N. H. Edwards, D. A. D. Rose, and P. Garner, "Lifestyle monitoring: technology for supported independence," *Computing & Control Engineering Journal*, vol. 9, no. 4, pp. 169–174, 1998.

[2] S. Bonner, "Assisted interactive dwelling house: edinvar housing association smart technology demonstrator and evaluation site," in *Improving the Quality of Life for the European Citizen, Proceedings of the 3rd TIDE Congress*, pp. 396–400, Helsinki, Finland, June 1998.

[3] S. J. McKenna, F. Marquis-Faulkes, P. Gregor, and A. F. Newell, "Scenario-based drama as a tool for investigating user requirements with application to home monitoring for elderly-people," in *Proceedings of the 10th International Conference on Human-Computer Interaction (HCI '03)*, pp. 512–516, Crete, Greece, June 2003.

[4] H. Nait-Charif and S. J. McKenna, "Activity summarisation and fall detection in a supportive home environment," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04)*, vol. 4, pp. 323–326, Cambridge, UK, August 2004.

[5] W. P. Goforth, "Multi-event notification system for monitoring critical pressure points on persons with diminished sensation of the feet," US Patent No. 4647918, March 1985.

[6] B. U. Toreyin, Y. Dedeoğlu, and A. E. Cetin, "HMM based falling person detection using both audio and video," in *Proceedings of the International Workshop on Computer Vision in Human-Computer Interaction (ICCV-HCI '05)*, vol. 3766 of *Lecture Notes in Computer Science*, pp. 211–220, Springer, Beijing, China, October 2005.

[7] F. Jabloun and A. E. Cetin, "The teager energy based feature parameters for robust speech recognition in car noise," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '99)*, vol. 1, pp. 273–276, Phoenix, Ariz, USA, March 1999.

[8] D. Dimitriadis, P. Maragos, and A. Potamianos, "Robust AM-FM features for speech recognition," *IEEE Signal Processing Letters*, vol. 12, no. 9, pp. 621–624, 2005.

[9] S.-H. Chen and J.-F. Wang, "A wavelet-based voice activity detection algorithm in noisy environments," in *Proceedings of the 9th International Conference on Electronics, Circuits and Systems (ICECS '02)*, vol. 3, pp. 995–998, Dubrovnik, Yugoslavia, September 2002.

[10] E. Erzin, A. E. Cetin, and Y. Yardimci, "Subband analysis for robust speech recognition in the presence of car noise," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '95)*, vol. 1, pp. 417–420, Detroit, Mich, USA, May 1995.

[11] R. Sarikaya, B. L. Pellom, and J. H Hansen, "Wavelet packet transform features with application to speaker identification," in *Proceedings of the 3rd IEEE Nordic Signal Processing Symposium (NORSIG '98)*, pp. 81–84, Vigsø, Denmark, June 1998.

[12] R. Sarikaya and J. N. Gowdy, "Subband based classification of speech under stress," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '98)*, vol. 1, pp. 569–572, Seattler, Wash, USA, May 1998.

[13] C. W. Kim, R. Ansari, and A. E. Cetin, "A class of linear-phase regular biorthogonal wavelets," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '92)*, vol. 4, pp. 673–676, San Francisco, Calif, USA, March 1992.