Smart Home Care Network using Sensor Fusion and Distributed Vision-based Reasoning

Ali Maleki Tabar
Wireless Sensor Networks Lab
Dept. of Electrical Engineering
Stanford, CA 94305
maleki@stanford.edu

Arezou Keshavarz *
Wireless Sensor Networks Lab
Dept. of Electrical Engineering
Stanford, CA 94305
arezou@keshavarz.net

Hamid Aghajan
Wireless Sensor Networks Lab
Dept. of Electrical Engineering
Stanford, CA 94305
aghajan@stanford.edu

ABSTRACT

A wireless sensor network employing multiple sensing and event detection modalities and distributed processing is proposed for smart home monitoring applications. Image sensing and vision-based reasoning are employed to verify and further analyze events reported by other sensors. The system has been developed to address the growing application domain in caregiving to the elderly and persons in need of monitored living, who care to live independently while enjoying the assurance of timely access to caregivers when needed. An example of sensed events is the accidental fall of the person under care. A wireless badge node acts as a bridge between the user and the network. The badge node provides user-centric event sensing functions such as detecting falls, and also provides a voice communication channel between the user and the caregiving center when the system detects an alert and dials the center. The voice connection is carried over an IEEE 802.15.4 radio link between the user badge and another node in the network that acts as a modem. Using signal strength measurements, the network nodes keep track of the approximate location of the user in the monitoring environment. The network also includes wall-mounted image sensor nodes, which are triggered upon detection of a fall to analyze their field-of-view and provide the caregiving center with further information about the user's status. A description of the developed network and several examples of the vision-based reasoning algorithm are presented in the paper.

Categories and Subject Descriptors

I.4.8 [**Scene Analysis**]: Object recognition; C.3 [**Special-purpose and Application-based Systems**]: Real-time and embedded systems

*Arezou Keshavarz was in an internship program at the Wireless Sensor Networks Lab, Stanford University.

VSSN'06, October 27, 2006, Santa Barbara, California, USA.

General Terms

Algorithms

Keywords

Home monitoring, Elderly care, Posture detection, Distributed processing, Vision-based reasoning

1. INTRODUCTION

Wireless sensor networks have the potential to dramatically increase our ability to develop user-centric applications to monitor and prevent harmful events. The availability of inexpensive low-power sensors, radios, and embedded processors enables the deployment of distributed sensor networks to provide information to users in different environments and to offer them control over unwilling situations. Networked sensors can collaborate to process and make deductions from the acquired data and provide the user with access to continuous or selective observations of the environment. To be of practical use, in most situations these devices need to be small in form factor, low power, lightweight, and unobtrusive.

A rapidly growing applications domain in which sensor networks can have a significant impact on the timeliness and affordability of the service is assisted living and monitored care for the elderly and persons in need of such care. Traditional solutions to such needs often limit the independence of the person receiving care. Wireless sensor networks can offer a solution to users who care to live independently while enjoying the assurance of timely access to caregivers when needed. Approximately 30 percent of the people over 65 years of age and living in the community face an accidental fall each year, while the number is higher in institutions. Although less than one fall in 10 results in a fracture, a fifth of fall incidents require medical attention [8].

There are several applicable solutions for fall detection that are based on using accelerometers. In [16] a smart fall sensor is described as a device for being integrated into a garment. Also the Vivago system proposed by Särelä *et al.* [17] is a wrist-worn device with a manual alarm button and movement measurement.

Employment of wireless communication technologies to report accidental falls has been studied in [18], which describes a tele-monitoring system based on Short Message Service (SMS), in which accelerometer data from the device worn by the monitored user is transmitted hourly as an SMS message, directly from the portable unit to a remote server.

Similar systems based on using Personal Digital Assistants (PDAs) are reported in other articles.

The wireless sensor network developed in this work employs multiple sensing and event detection modalities, and uses visual information acquired by a set of image sensors as the basis for making interpretations about the status of persons under care. In a home-based monitoring application, the location, duration of stay at one position, sudden moves, and body gesture of the user can be used in combination with rule-based logic to evaluate various hypotheses about their well being. Multi-modal sensing allows for making various user-centric measurements in which the frequency and sequence of the measurements taken by the different elements in the network can be determined adaptively according to the nature of the expected events that occur around the user.

The rest of the paper is structured as follows. In section 2 a survey of related work on indoor monitoring and surveillance is included and motivations for our approach are discussed. Section 3 provides motivation for our approach and an overview of the proposed system. In section 4 the architecture of the system is presented and the different functional components of the network are described. The network includes several components such as accelerometer node, cameras, wireless communication modules, voice transmission module for the IEEE 802.15.4 radio, modem-based connection to monitoring center, and functions such as radio-based triangulation and position estimation. The proposed solution also includes a vision-based algorithm module that is employed upon a fall alert to validate the report and further analyze the status and posture of the user via image analysis. Section 5 includes a description of the proposed human detection and posture estimation algorithm. These algorithms have been developed to be lightweight so they can operate on the microprocessors of the boards equipped with image sensors. To conclude, section 6 includes a discussion of the proposed approach and further development directions.

2. RELATED WORK

Rapid advances in the technologies of image sensors and embedded processors enable the inclusion of vision-based nodes in various smart environment applications [1]. Image sensors provide rich sources of information both for human observation and for computer interpretation. By acquiring an information-rich data type from the environment, image sensors allow the network to extend its operation from measuring simple effects to knowing about the status of the user, analyzing the event, and possibly even responding to the event. The richness of data in image sensing also means a need for effective processing power to make deductions and high-level interpretations from the data. In-node image processing enabled by todays embedded processors helps mitigate the need to transfer raw images across the network, and hence reduces the constraints on bandwidth management of wireless networks. Employment of multiple image sensors allows for covering the different areas of the environment or having different viewing angles to the same scene. Overlapping views enable collaboration between the network nodes to combine attributes and derive detailed and more confident deductions about the events and status of the user.

Techniques that include the use of cameras for reproting accidents are emerging. In [9] a system based on the Berkeley fall detector and a camera cell phone is proposed to provide live video feed from the user. In [3], [4] a system for retrieval and summarization of continuously archived multimedia data from a home-like ubiquitous environment is presented. Data are analyzed to index video and audio from a large number of sources. A video and audio handover scheme is implemented to retrieve continuous video and audio streams as the user moves in the environment. In [15] a posture determination using elliptical fitting and projection histogram is described. In addition to monitoring normal daily activities and detecting potentially adverse events such as falls, their UbiSense system is designed to capture signatures from the fall of patients by analyzing small changes in posture and way of walking. In [2], an indoor video surveillance for monitoring people in domestic environments is discussed. The solution consists of a moving object detection module, a tracking module designed to handle large occlusions, and a posture detector.

3. SYSTEM OVERVIEW

In this article we address the problem of elderly care by also considering the critical parameters in serving such applications such as cost, realizability, and power consumption. Accelerometer offers the most accurate and inexpensive way to detect a fall, so we chose to base our main fall detection system on it. Also by employing RSSI (Received Signal Strength Indicator) measurements between the user badge and a network of wireless nodes deployed in the environment, the approximate position of the user is calculated. The data is used both for movement tracking and for triggering the most suitable image sensors when a fall occurs. Movement tracking allows for creating an alert, for example, if for an extended period of time in the active parts of the day there is no user movement. Image processing is used to analyze the situation and determine the user's posture when alerts happen. This serves two purposes: One to determine the posture of the user after an accident, and two to reduce the number of false alarms caused by the rather sensitive accelerometer threshold settings typically used for ensuring fall detection. Reduction in the false alarm rate in turn allows the care providers to offer scalable service to a growing number of potential users. Our prototype design consists of three static nodes installed high on the wall or on the ceiling. This configuration also enables better estimation of the user's position by allowing direct paths with the user badge when RSSI is measured. Moreover, having three cameras makes it possible to either cover the entire space of the room or provide different perspectives to the user when needed.

The cameras are triggered by messages broadcasted by the user badge when an event is sensed upon analyzing the accelerometer signals. In-node image analysis and exchange of local decisions between the camera nodes allows the network to generate a report and raise an appropriate reporting flag. Having the image processing algorithms run on the local nodes also conserves the privacy of the user when there is no need to transmit visual data.

On the other hand, having multiple cameras in the solution also allows for some level of operation for the system if the user forgets to carry the badge. In such a case the main device used to detect a fall is not available. However, with the help of the cameras, which could be activated by the monitoring center or upon a timeout if they do not hear from the user badge for an extended period, some useful information about the situation of the user can be extracted.

As previously mentioned, total cost of the solution has been a major issue in the design. The accelerometers, microcontrollers, voice transmission circuitry, and the image sensor modules used in our prototype are all inexpensive devices with a bill of material for a small network in the range of 20-40 dollars. This is an attractive solution compared to other techniques based on using SMS devices or PDAs, as the proposed technique does not require a dedicated computer or addition of an expensive device to the communication devices carried by the user for the purpose of monitoring. However, it is possible to connect the proposed network to other devices such as computers, or employ other data transmission mechanisms to send the information to a data center. For example, the tracked position of the user or visual information collected by the cameras at alerts can be sent to the call center by employing a data modem scheme through an available computer. The voice transmission circuit allows the system to create a voice link between the user and the care center automatically or by user's demand and act as a 2-way phone.

4. HARDWARE ARCHITECTURE

Fig. 1 illustrates an example of how the prototype network operates. The network of sensors consists of a user badge node, several image sensor nodes, a node with a modem, and additional optional wireless nodes that help with localization of the user via signal strength measurements. Each node in the network includes an IEEE 802.15.4 radio and a local processing unit. The user wireless badge node is equipped with accelerometers and circuitry for voice transmission over the IEEE 802.15.4 radio. Periodic radio broadcasts by the user badge are received by the other nodes. By employing RSSI (Received Signal Strength Indicator) measurements of the user badge signals, the network determines the approximate position of the user. The position information is used for triggering the most suitable image sensors when a fall occurs. The fall alert is broadcasted by the user badge when the accelerometers on the badge record a significant change in their measured signal. Image processing is used to analyze the situation and determine the user's posture when alerts happen. Our prototype design consists of three static nodes installed high on the wall or on the ceiling. This configuration also enables better estimation of the user's position by allowing direct paths between the nodes when RSSI is measured. Moreover, having three camera nodes makes it possible to either cover the entire space of the room or provide different perspectives to the user when needed.

The voice transmission circuit placed on the user badge and one of the network nodes allows the system to create a voice link between the user and the care center automatically or by user's demand, and acts as a 2-way phone. The connection between the user badge and this node is established via an IEEE 802.15.4 radio link. The network node is equipped with a phone interface and uses the phone line to dial a stored number to make a call to the care center. The phone interface can also be used to upload the results of image analysis or an extracted scene obtained by the camera nodes to the monitoring center on demand. The operation of the network upon a triggering event is summarized in Figure 2.

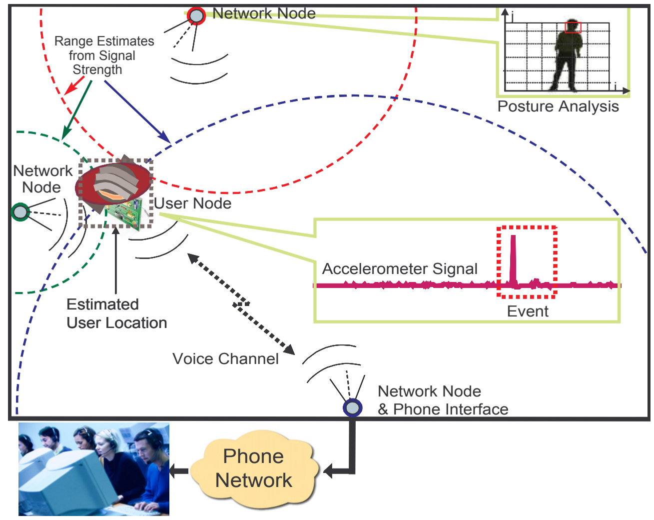


Figure 1: An overview of the developed smart care network. The user badge is equipped with accelerometers and a radio device. Camera nodes analyze views of the user upon triggering by the user badge.

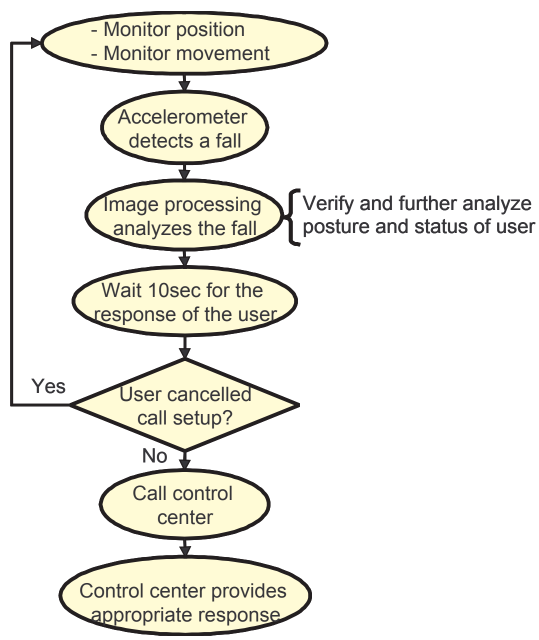
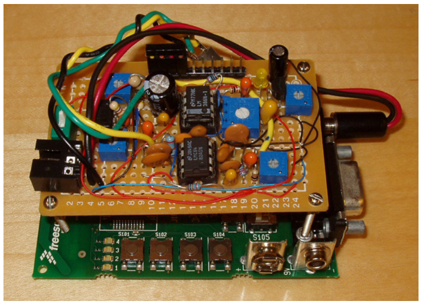
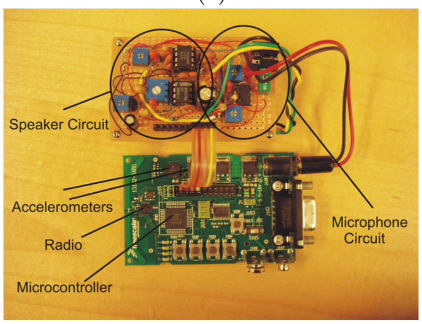


Figure 2: Flowchart of the overall fall detection and reporting operation.



(a)



(b)

Figure 3: (a) The user badge prototype consisting of a Freescale MC13192 sensor mote and a voice transmission circuit. (b) The main functional components of the user badge. The speaker circuit converts the micro-controller's PWM output to analog, and the microphone circuit amplifies the voice signal to be sampled by the micro-controller.

4.1 User Badge

The user badge is equipped with two accelerometers, which can measure acceleration of sudden movemements in the 3 main directions of x, y, and z. The z-axis accelerometer is a separate chip from the x-y accelerometer. The accelerometers are connected to the ADC pin of the micro-controller. Processing the signals reported by the accelerometers in the micro-controller allows for calling an alert and triggering image capture by the camera nodes. The x-y accelerometer measures the angle of the badge to the flat surface, making it possible to compare the tilt of the user badge with previous samples.

Various studies on the use of accelerometer signal for posture classification have been reported. The work reported in [19] demonstrates that two uniaxial accelerometers are sufficient to distinguish between sitting, standing, lying down, and dynamic movement. The work in [5] shows that using three accelerometers, it is possible to obtain complete information about the position, speed and acceleration of the person.

Figure 3 shows the implemented prototype of the user badge. The voice circuit is attached to the Freescale MC13192 SARD sensor application mote. The mote includes an MC 9S08GT60 micro-controller, which is a low power, low voltage MCU, an MC13192 2.4 GHz transceiver radio, and two accelerometers: MMA6261Q (x and y axis), and MMA1260D (z axis). Sampling rate of the accelerometers is chosen in our experiments to be 1000 per socond. The voice data is sampled at the rate of 8kHz by the MCU, and is converted to

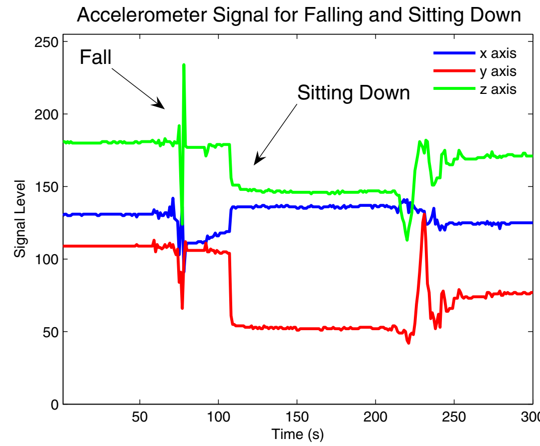


Figure 4: Accelerometer signals demonstrating differences between the events of falling and sitting down on a chair.

packets of size 110 bytes before being sent to the radio for transmission over an IEEE 802.15.4 channel.

In the implementation of the prototype in this work, we make two uses of the accelerometer signals. Simple peak thresholding is employed to trigger the camera nodes when a large amplitude is detected, while the accelerometer signals are also analyzed to add an additional reasoning mechanism to the decision making process. In figure 4 the measured values reported by the accelerometers are shown. The first sharp peak in the plots corresponds to the user falling down and the second sharp peak corresponds to sitting down on a chair. As these plots indicate, different movements registered by the accelerometers have different signatures, which can be used to distinguish between an accidental fall and a different action such as sitting down. In this case, the two postures can be distinguished by combining two measures: the sharpness of the registered peak, and the amplitude of the jump in the signal. Figure 5 shows a linear separation between the samples corresponding to the two postures when the two mentioned measures are employed to classify the reported posture in 20 experiments. The result of such signal analysis is combined with the vision-based reasoning module to improve the performance of the system as discussed in section 5.3.

4.2 User Position Tracking

The position of the mobile node is estimated using three static nodes installed in the room with known spatial coordinates. In the prototype system, the positions of the static nodes is entered into the GUI by clicking on a grid map indicating the operation area. Estimated position of the user is mainly used to trigger the most suitable camera nodes upon detection of an event. If the system is unable to determine the user's position for any reason, including in the case of unknown locations for the static nodes after deploymentg, all camera nodes are triggered by the event. Triangulation based on the RSSI received at three wall-mounted nodes yields an approximation of the user's position every time the user badge transmits a beacon signal. As [14] and several other papers have stated, with a small change in the position of the object, the RSSI value often changes considerably

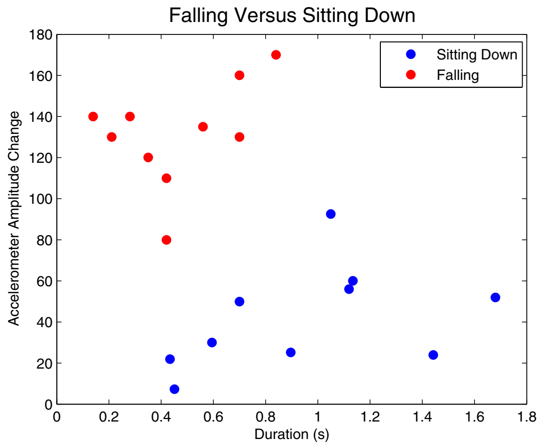


Figure 5: Results of 20 experiments indicating a linear separator can be used to classify falling and sitting down events based on the amplitude and the duration of the accelerometer signal.

with variations equivalent to 2-3 meters of distance change. To reduce this effect, a 1.5 second interval of samples is averaged; this interval is chosen experimentally. Figure 6 shows an example of user localization using RSSI.

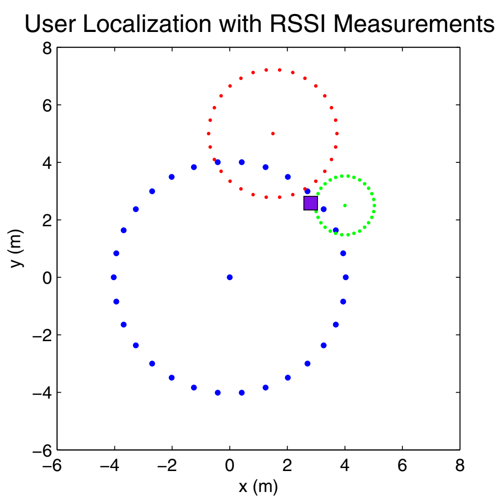


Figure 6: User's position is approximated by signal strength measurements from three nodes. The three circles represent range estimates obtained by each node. The dark square indicates the estimated user position.

To further improve the performance of the position estimation procedure, we also employ a weighting scheme based on the measured distance of the user from each static node. In our implementation, the weighing function used is inversely proportinal to the estimated distance of each node to the user. For instance, if the three static nodes estimate the user to be at ranges r_1, r_2, and r_3, respectively, the wights applied to their data in triangulation will be $\frac{R-r_1}{R}$, $\frac{R-r_2}{R}$, and $\frac{R-r_3}{R}$, where $R = \sum_{i=1}^{3} r_i$. With this weighting

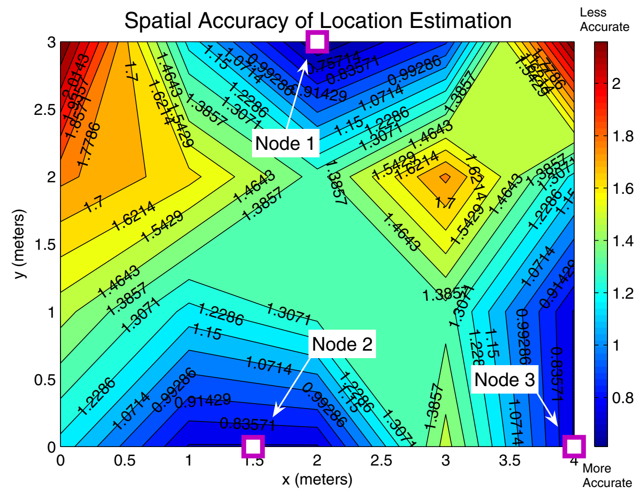


Figure 7: Accuracy of triangulation for different areas inside a 3×4 meter room. In this figure the room is divided into 20 blocks and the diffrence between the true and estimated positions is calculated.

function, the node which is closer to the mobile node has the greatest influence in the triangulation scheme, i.e. in defining the estimated position of the mobile node.

Figure 7 shows the accuracy of the different areas inside the room with this algorithm. It can be inferred from the figure that the accuracy is greater near the static nodes. The figure also shows that the points close to the static nodes have less variations.

Based on this accuracy analysis, we also adopt a rule to exclusively use one node's measurement if the node detects the user to be within a small range (e.g. 1-2 feet), in which case the disk indicating the range is assumed as the approximate position of the user. On the other hand, we use a rule to not use a node's measurement if the node's distance to the user is detected to be very large (e.g. 20 feet or more). These modifications provide robustness against the signal model issues and fluctuations of the RSSI at very close or very far ranges.

4.3 Camera Nodes

A network of three camera nodes is used to provide a full view of the room. Each camera may cover a part of the room and there might be some overlaps if the total area is small. Cameras are turned off in normal conditions to reduce unnecessary power consumption. At the time of the event, the camera nodes focusing on the position of the mobile node will be turned on and start capturing frames.

The cameras are used to investigate the event more carefully and to eliminate false alarms. For instance, we might not be interested in an alarm that is raised when the user is walking around the room in normal condition. Thus, the images captured from the cameras are processed to identify hazardous events that need to be reported.

Alternatively, the cameras could perform the position estimation task collaboratively based on simultaneous visual observations as reported in [13], which also provides an automated solution for the network node localization problem using an initial set of observations of the user.

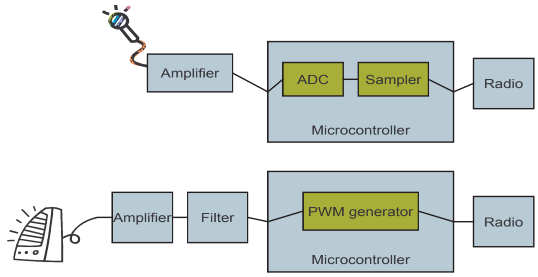


Figure 8: Voice transmission block diagram.

Using the camera nodes on occasions other than triggering by an accelerometer alert offers an attractive solution to cases when the user for example has forgotten to carry the user badge, in which case location tracking by cameras or on-demand visual monitoring by the care center are employed to monitor the status of the user. On the other hand, if other persons besides the user appear in the environment, tracking the user visually requires advanced vision-based analysis, while tracking by localizing the user badge via RSSI measurements is robust against such cases.

4.4 Voice Transmission Circuit

There are several ways to transmit voice over a wireless network. Approaches that we considered are:

1. Analog microphone and sampling by the micro-controller, which put the samples into packets to be transmitted over the wireless channel by the radio module. In the receiver, the micro-controller converts the samples into a Pulse Width Modulation (PWM) output and applies a Chebychev filter to convert PWM to analog signal, which is then amplified and is fed to the connected speaker.

2. Digital microphone with a codec that has Pulse Code Modulation (PCM) output. In the receiver, the PCM signal is converted to analog using the same codec chip, which is connected to a microphone.

3. Analog microphone connected to a delta modulation codec. In the receiver, the signal is converted to PWM, then a reverse codec is used to convert PWM to analog signal which is fed to the speaker.

The 2nd and 3rd methods have an advantage over the first approach in the voice quality as well as power consumption. For our prototype implementation we selected to use the first method, which provides adequate quality while requiring fewer circuit compoments. This solution would be attractive from a product design point of view as it uses the least expensive components among the mentioned choices. The block diagram of the voice transmission circuit is shown in figure 8. In this circuit the analog signal from the microphone is sampled at 8 KHz. Since there is an Analog Digital Converter (ADC) in the micro-controller unit, there is no need for external circuitry. The sampled voice is then sent via IEEE 802.15.4 radio communication to the receiving node.

4.5 Telephone Module

The telephone module is used to connect the home network to the call center when an event occurs. A major benefit of providing voice communication through the user badge is that the user does not need to go to the telephone, which is in fact impossible in some situations when an accident has happened. The phone interface module is able to dial a predefined number stored in the node's memory. The micro-controller on this node controls the actions of the phone interface module. First, the connection to the phone line is established by the off hook signal. Then the micro-controller puts the numbers on the interface module's port and the module dials the number. After the call is established a radio link between the node and the user badge is used to transmit voice between the user and the call center.

The phone interface module is activated in the case of an alarm or when the user presses the voice button on the user badge node. In alternative approaches such as the work reported in [18], the system is based on Short Message Service (SMS).

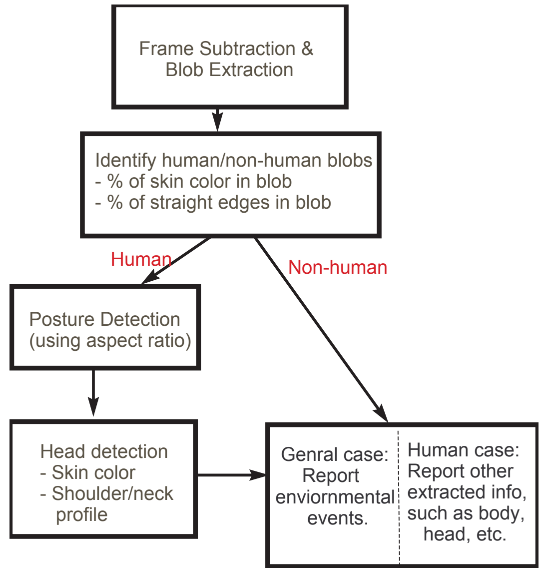


Figure 9: Flowchart of the vision-based posture analysis procedure.

Captured Image	Detected Non-Human Mask
	

Figure 10: The refrigerator door corresponds to a non-human object that has caused a change in the scene. The human body mask is correctly detected and analyzed separately as in figure 11.

5. VISION-BASED POSTURE ANALYSIS

Figure 9 shows the structure of the vision-based reasoning scheme for detection of the human body in the scene and posture analysis. The objects that appear in the scene due

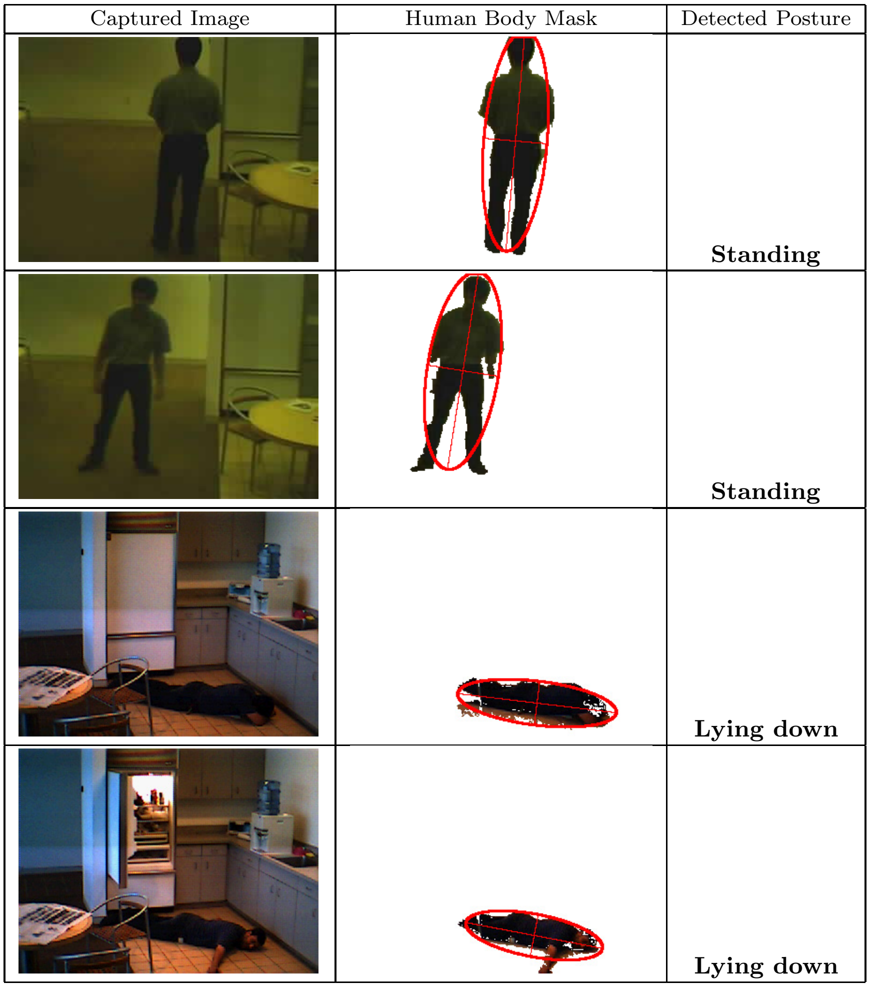


Figure 11: Examples of the human body mask and posture detection scheme.

to the event are extracted in a two-fold process: background subtraction and blob segmentation. A binary mask corresponding to object locations is constructed through background subtraction.The mask is further refined by applying small-region removal to eliminate the small noise effects in the mask. This is followed by the dual process applied to the inverted mask in order to eliminate the holes within the mask. The resulting mask corresponds to the main blobs in the image.

5.1 Human Body Detection

The objective is to find which one of the blobs, if any, corresponds to a human. Many properties of human body can be used to distinguish a human blob from a non-human blob. Since no individual metric can yield perfect results, these two properties have been used as opportunistic measures to improve the chances of correct identification of a human blob: percentage of straight edges available in each blob, and percentage of skin color available in each blob. Most man-made objects are composed of straight edges, whereas the human body has more round edges. Assuming that the event does not involve more than one human blob, this property is used to distinguish between human and non-human blobs. In order to do so, a straight-edge-detecting filter is applied to the mask, both in the horizontal and vertical directions. The ratio between the number of pixels corresponding to such edges and the total number of pixels within the blob is calculated for each blob.

If the percentage of straight edges is below a certain threshold (which was measured experimentally), the corresponding blob is identified as a human blob. If not, then the blob is identified as a non-human object [Figure 10]. The percentage of average skin color found within a blob also provides information about the nature of the blob. This ratio is obtained by counting the number of pixels that have color values within an average skin color range. Some examples of the human detection algorithm are demonstrated in figure 11. Note that the algorithm is not very sensitive to the specified thresholds; our aim here is to decrease the number of false alarms by eliminating the objects that demonstrate non-human characteristics.

5.2 Head Detection

Head detection has been addressed extensively in prior research work. The work of [12] detects the head of the object in 3D using elliptical template-matching and updates the detection results after each successive frame. The methods reported in [7] and [11] base their face detection algorithm

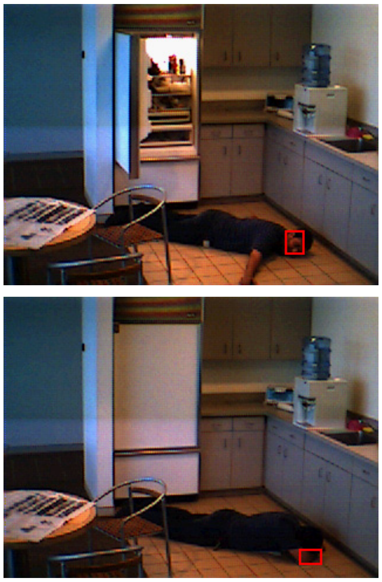


Figure 12: Examples of head detection using skin color.

on skin color and search for blobs with pixels whose color corresponds to that of the skin. A color histogram-based scheme is proposed in [6] for detection of human face. The work in [10] not only uses skin tone to determine the location of the skin colors, but also makes use of information about the facial features, such as eyes and lips, to confirm which one of the possible skin-toned regions corresponds to the face.

Using rigorous head detection algorithms such as template-matching are computationally expensive. On the other hand, the success rate of simpler algorithms is much lower than those requiring a higher computation level. Here, we have combined two computationally-efficient approaches as opportunistic measures to increase the success rate while keeping the computation level low. The two head detection algorithms proposed below are skin-color-based head detection and neck/shoulder-profile-based detection algorithms. The human body mask, obtained in section 5.1 is used as a starting point for determining the location of the head within the body mask.

5.2.1 Method 1: Head detection using skin color

An average range of skin color in the hue channel is chosen as an approximate measure for detecting the pixels that have skin color. A binary skin mask is constructed from these pixels, and the blob extraction process is performed on the skin mask to identify regions of pixels that fall within the skin color range. From the body detection step, an approximate box around the body is available. The probability that the head is located at the two ends of this box along its larger dimension is higher; therefore, the skin mask is multiplied by a gaussian at these two ends in order to emphasize this property on the skin mask. The gaussians are constructed to have parameters as a function of the body height; the mean of the gaussians has a distance of 5% of the body height from each end of the body; the variance of the gaussians is 10% of the body height. The objective is to eliminate any skin-colored regions that have been found in other parts of the body (such as hands, feet, etc.). The resulting head mask corresponds to skin-colored face regions [Figure 12].

The skin-color-based head detection only works for cases in which part of the face is present in the image. Furthermore, the picture quality and illumination issues can also affect the results of this algorithm. For instance, if the picture quality is too low and the face color is not visible enough, the algorithm fails to recognize the head. In the case of failure, a blank head mask is returned and the judgements of this algorithm are not taken into account.

5.2.2 Method 2: Head detection using shoulder/neck profile

The shoulder/neck-profile-based head detection algorithm searches the body mask to see whether patterns similar to that of the neck-shoulder width change are observed in it. In fact, by recognizing the boundary between neck and shoulders, the head can be approximated by a box encompassing the area between top of the head and shoulders. In order to search for this pattern, the body mask is traversed along the larger dimension. The width of the mask in each row of the body mask is calculated. The shoulder is identified by a large change in the width of two subsequent rows. The subsequent width differences are plotted versus the number of body mask rows. Furthermore, the shoulder is most probably located at the two ends of the body mask; therefore the width difference graph is multiplied by a gaussian centered at the two ends of the box to reflect this emphasis [Figure 13]. Similar to section 5.2.1, the gaussians are constructed to have parameters as a function of the body height; the mean of the gaussians has a distance of 5% of the body height from each end of the body; the variance of the gaussians is 10% of the body height. The shoulder row is then localized by measuring the peak index in the width difference graph. As shown in figure 13, the head is approximated by a box covering the area between the shoulder and the top of the body mask.

This algorithm's performance is limited to images in which the frontal or back dimensions of the body are visible to some extent. In fact, if the person is facing the camera completely sideways, the algorithm fails to detect the head and returns a blank mask.

5.2.3 Head mask merging

Using two different schemes for head detection allows the algorithm to opportunistically employ one or both of their available results. Since the failure modes of the individual head detection schemes are caused by rather different factors, the two masks obtained from these schemes are combined using an OR operation. If the two masks have an overlap, or if one of the head detection algorithms is unsuccessful, then their union is taken as the final head mask. However, if the two head masks do not have an overlap and neither of the masks is blank, then there will be two potential distinct positions for the head in the image. As a result, the location of the head in the future steps will be set to unknown, but both head location candidates will be reported to the monitoring center, if a report is deemed necessary.

5.3 Performance and Future Work

Combining information obtained from different methods increases system reliability. For instance, in the head detection algorithm, the results of the skin-based detection

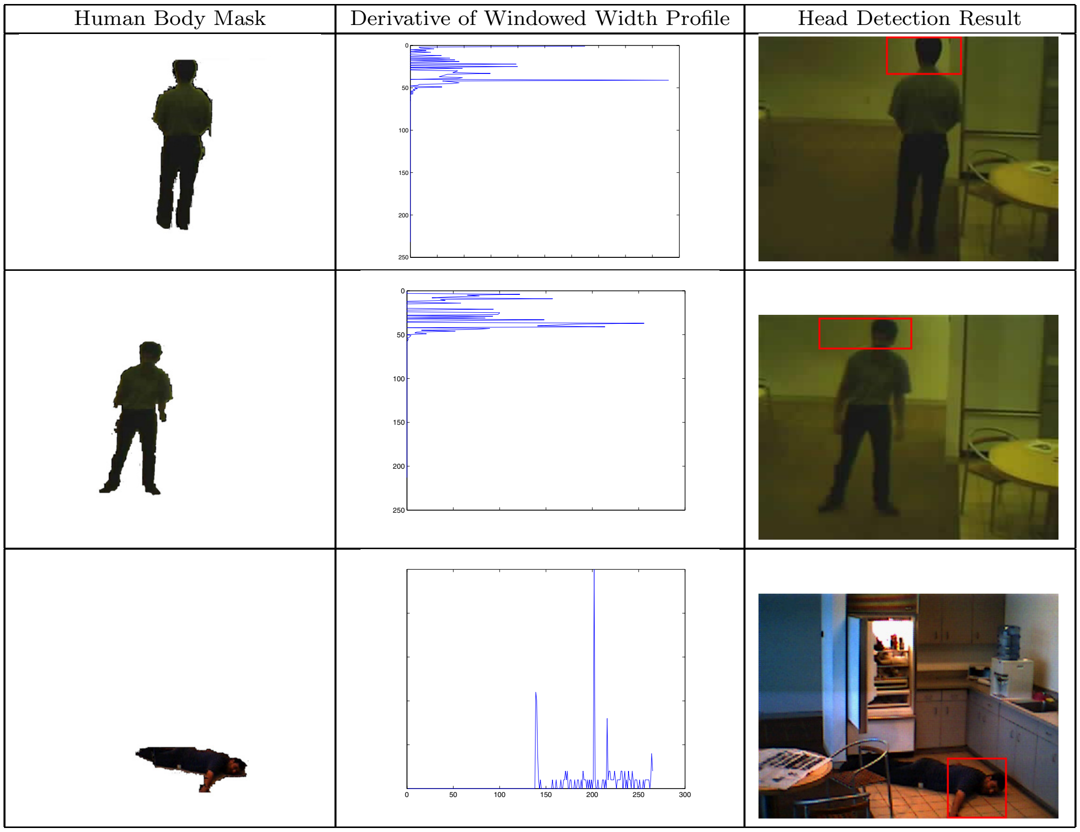


Figure 13: Examples of head detection using the shoulder-neck profile. The left column shows the detected body mask. The middle column shows the derivative of the mask width (windowed to emphasize the two ends of the body mask). The right column shows head detection result using this scheme.

Algorithm	Posture Detection	Head Detection
Accuracy	**94%**	**70%**

Table 1: Accuracy of the posture detection and head detection algorithms.

scheme and the profile-based detection scheme are combined to yield a final result. Although the performance of each of the head detection algorithms by themselves is limited, combining information from both of them improves the overall accuracy of the system. Using a total of 16 images, the accuracy of posture detection and head detection algorithms was measured. The results are shown in table 1.

Further data fusion among different camera views as well as the accelerometer data will also improve the system reliability [Figure 14]. By combining the results of different camera views, the information can be aggregated to provide a more reliable result and eliminate more false alarms. Furthermore, the signature of the accelerometer data can yield useful information about the falling process.

Handling problems such as structural changes in the environment and large illumination changes is the subject of further investigation. Such problems can be resolved if the background is refreshed at a suitable rate.

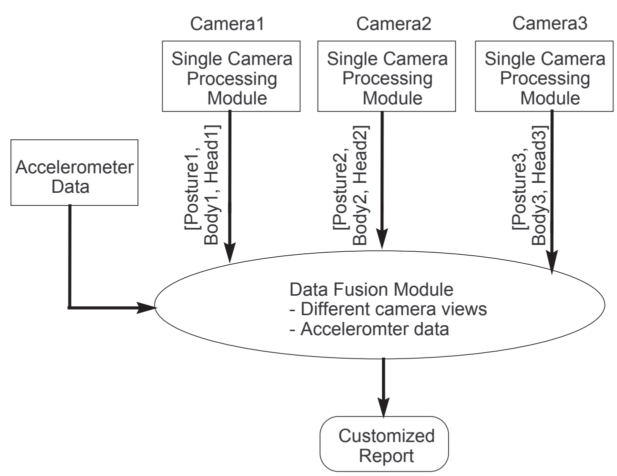


Figure 14: Information available from different camera views as well as the accelerometer data can be combined to increase the reliability of the system and reduce the number of false alarms.

6. CONCLUSIONS

This work addressed the application of elderly care in indoor environments under the constraints of a small form factor, low cost, low power, and being realizable with existing off-the-shelf devices. We characterized sensor-based fall detection techniques along with an RSSI-based triangulation method to detect accidents and track the approximate position of the user, and also combined a vision-based posture classification scheme to extract further information about the user when an alert occurs. Such additional information is often critical to access in very early stages of an accident for the follow-on emergency help.

The proposed approach is based on inexpensive hardware and is therefore affordable and scalable. Our experiments showed encouraging results in indoor environments such as hall and kitchen areas. While it is possible to add extra modules to send live video stream to the monitoring center, the basic prototype of the proposed approach readily includes a vision-based acquisition and processing system developed for node-based processing capabilities. Having the image processing algorithms run on the local nodes also conserves the privacy of the user when there is no need to transmit visual data.

7. REFERENCES

[1] M. Bramberger, A. Doblander, A. Maier, B. Rinner, and H. Schwabach. Distributed embedded smart cameras for surveillance applications. *IEEE Computer Magazine*, 39(2):68–75, Feb. 2006.

[2] R. Cucchiara, C. Grana, A. Prati, and R. Vezzani. Computer vision system for in-house video surveillance. *IEE Proceedings of Vision, Image and Signal Processing*, 152:2:242:249, APRIL 2005.

[3] G. C. de Silva, B. Oh, T. Yamasaki, and K. Aizawa. Experience retrieval in a ubiquitous home. In *Proceedings of the 2nd ACM workshop on Continuous archival and retrieval of personal experiences, ACM Press*, 2005.

[4] G. C. de Silva, T. Yamasaki, and K. Aizawa. Evaluation of video summarization for a large number of cameras in ubiquitous home. In *Proceedings of the 13th annual ACM international conference on Multimedia, ACM Press*, 2005.

[5] Y. Depeursinge. Device for monitoring the activity of a person and/or detecting a fall. US patent no.6 201 476 B1, March 2001.

[6] J. Fang and G. Qiu. A colour histogram based approach to human face detection. In *Prceedings of the Intrnational Conference on Visual Information Engineering: Ideas, Applications, Experience*. IEE, Visual Information Engineering Professional Network, July 2003.

[7] H. Fatemi, R. Kleihorst, H. Corporaal, and P. Jonker. Real time face recognition on a smart camera. In *Proceedings of Acivs 2003 (Advanced Concepts for Intelligent Vision Systems)*, September 2003.

[8] L. Gillespie, W. Gillespie, M. Robertson, S. Lamb, R. Cumming, and B. Rowe. Interventions for preventing falls in elderly people. Cochrane Database of Systematic Reviews 2003;(4):CD000340., 2003.

[9] T. R. Hansen, J. M. Eklund, J. Sprinkle, R. Bajcsy, and S. Sastry. Using smart sensors and a camera phone to detect and verify the fall of elderly persons. In *European Medicine, Biology and Engineering Conference*, November 2005.

[10] R.-L. Hsu, M. Abdel-Mottaleb, and A. Jain. Face detection in color images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:5:696:706, may 2002.

[11] R. Kleihorst, H. Broers, A. Abbo, H. Embrahimmalek, H. Fatemi, H. Corporaal, and P. Jonker. An SIMD-VLIW smart camera architecture for real-time face recognition. In *Proceedings of ProRISC 2003*, November 2003.

[12] B. Kwolek. Face tracking system based on color, stereovision and elliptical shape features. In *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS03)*. IEEE Computer Society, July 2003.

[13] H. Lee and H. Aghajan. Vision-enabled node localization in wireless sensor networks. In *COGnitive systems with Interactive Sensors (COGIS 2006)*, Paris, France, 2006.

[14] E. Leland, K. Bradford, and O. C. Jenkins. Robot localization and control. *Circuit Cellar*, Issue 188, March 2006.

[15] B. P. Lo, J. L. Wang, and G.-Z. Yang. From imaging networks to behavior profiling: Ubiquitous sensing for managed homecare of the elderly. In *Adjunct Proceedings Of the 3rd International Conference on Pervasive Computing*, May 2005.

[16] N. Noury. A smart sensor for the remote follow up of activity and fall detection of the elderly. In *Microtechnologies in Medicine and Biology 2nd Annual International IEEE-EMB Special Topic Conference*, May 2002.

[17] A. Särelä, I. Korhonen, J. Lotjonen, M. Sola, and M. Myllymaki. Ist vivago-an intelligent social and remote wellness monitoring system for the elderly. April 2003.

[18] C. Scanaill, B. Ahearne, and G. Lyons. Long-term telemonitoring of mobility trends of elderly people using SMS messaging. *IEEE Transactions on Information Technology in Biomedicine*, 10:2:412:413, April 2006.

[19] P. H. Veltink, H. B. J. Bussmann, W. de Vries, W. L. J. Martens, and R. C. V. Lummel. Detection of static and dynamic activities using uniaxial accelerometers. 4:4:375:385, December 1996.