Distributed Vision-Based Reasoning for Smart Home Care

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

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

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

Abstract

In this paper we develop a distributed vision-based smart home care system aimed for monitoring elderly persons remotely. The cameras are triggered by a broadcast from the user badge when accelerometers on the badge detect a signal indicating an accidental fall. Tracking the approximate position of the user by the network allows for triggering cameras with the best views. Distributed scene analysis modules analyze the user's posture and head location; these information are merged through a collaborative reasoning module, which makes a final decision about the type of the report that needs to be prepared. The developed prototype also allows for a voice channel to be established between the user badge and a call center over the phone line. A description of the developed network and several examples of the vision-based reasoning algorithm are presented in the paper.

Keywords

Distributed image sensors, Remote patient monitoring, Vision-based reasoning, Smart elderly care

1 Introduction

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.

A growing application domain in which sensor networks can have a significant impact on the timeliness and affordability of the service for users 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.

Techniques that include the use of cameras for reporting accidents are emerging. In [7], 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 [11] 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.

The wireless sensor network developed in this work employs collaborative vision-based reasoning by camera nodes as the basis for making interpretations about the status of person under care. The vision-based analysis module is triggered when the user's wireless badge broadcasts a packet upon sensing a significant signal by one of its accelerometers. Image analysis allows for validation of the report and provides further analysis of the posture and status of the user.

2 System Overview

Figure 1 illustrates an example of how the prototype network operates. The system's configuration is based on the smart home care network we developed for elderly monitoring [12]. By employing RSSI (Received Signal Strength Indicator) measurements between the user badge and a network of 3 (or more) wireless nodes deployed in the environment, 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, which can further be used to reduce the number of false alarms. Our prototype design consists of three static nodes with cameras, installed high on the wall or near the ceiling. 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 acts as a 2-way phone. The connection between the user badge and the network nodes is established via an IEEE 802.15.4 radio link. One of the network nodes is equipped with a phone interface and uses the phone line to

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

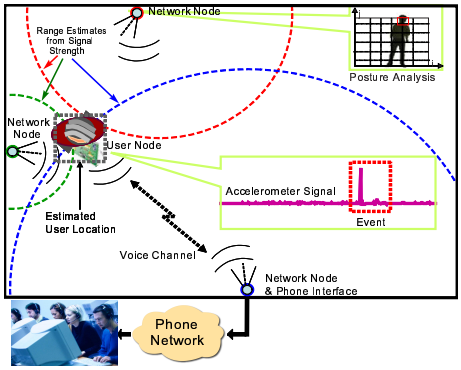


Figure 1. An overview of the developed smart care network. The user badge is equipped with accelerometers and a radio device. Camera nodes analyze views of the user upon triggering by the user badge.

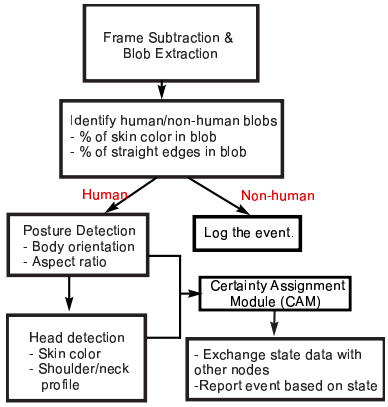


Figure 2. The above flow chart presents an overview of the distributed scene analysis module.

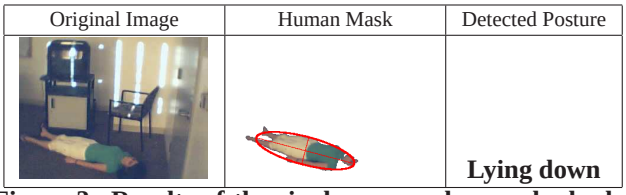


Figure 3. Results of the single-camera human body detection. An ellipse is fitted around the body and its major and minor axes show the orientation of the body mask.

make a call to the care center.

Through distributed scene analysis, each camera node processes the scene independently to identify whether a human body exists in the image. The posture and head position of the human body is then estimated and a certainty level is reached by each camera based on the consistency of the obtained results. Through collaborative reasoning among the three cameras, a final decision is made about the state of the user. Since each single camera cannot be expected to always extract all the desired information from the scene, utilizing a multi-camera data fusion technique will provide complementary information when needed, resulting in a more accurate understanding of the scene. As a result, the system will be able to make a more reliable decision, create a more efficient report, and reduce the number of false alarms sent to the central monitoring center.

3 Distributed Scene Analysis

Figure 2 outlines the distributed scene analysis process, through which body and head detection is performed on images obtained from each of the three camera nodes. A mask of the objects that have appeared in the scene due to the event is extracted through a two-fold process: background subtraction and blob segmentation. The resulting mask corresponds to the main blobs in the image.

3.1 Human Body and Posture Detection

The objective is to find which one of the blobs, if any, corresponds to a human. As described in [12], 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, two properties have been used as opportunistic measures to improve the chances of correct identification of a human blob: percentage of straight edges in each detected blob, and percentage of skin color in each blob.

After the human blob is identified, an ellipse is fitted on the blob [Figure 3]. The lengths of the major and minor axes of the ellipse and their orientations are used as a measure to identify the posture of the person.

3.2 Head Detection

Head detection has been addressed extensively in prior research work. The work in [10] detects the head of the object in 3D using elliptical template matching and updates the detection results after each successive frame. The works reported in [9], [6], and [5] base their head detection algorithm on skin color and search for blobs with pixels whose color corresponds to that of the skin. The work in [8] also uses skin tone to determine the areas with skin color, but it also uses information about the facial features, such as eyes and lips, to confirm which one of the possible skin-toned regions corresponds to the face.

Most rigorous head detection algorithms such as template matching schemes are computationally expensive. On the other hand, success rate of simpler algorithms is much lower, limiting their use in practical systems. Here, we have combined two computationally-efficient and complementary schemes as opportunistic measures to increase the success rate while keeping the computation level low. The two head detection algorithms used include a scheme based on skin color content and a scheme based on the neck-shoulder profile of the detected body mask [Figures 2 and 3]. Each scheme returns a blank mask in the case of failure.

3.2.1 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 a blob extraction process is performed on the skin mask to identify regions that fall within the skin color range. From the human body detection step [Section 3.1], an approximate ellipse around the body is available. The probability that the head is located at one of the two ends of this box along its larger dimension is higher; therefore, the skin mask is multiplied by two gaussians centered near these

Figure 4. Two examples of head detection using skin color. Note that in the example on the left, the mask has ruled out the refrigerator door as non-human event and correctly identified the human body as the human object.

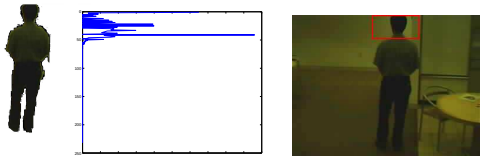


Figure 5. An example of head detection using the shoulder-neck profile. The left image shows the body mask. The middle graph shows the derivative in mask's width (windowed to emphasize the two ends of the body mask). The right image shows head detection result.

two ends in order to emphasize this property. 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. This eliminates 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 4], and is denoted as HMa. The skin color-based head detection only applies to cases in which the face is directed towards the camera.

3.2.2 Method 2: Head detection using shoulder-neck profile

The head detection algorithm based on the shoulder-neck profile detects the change in the width of the body mask in the area between the neck and shoulders. The head position is then approximated by a box encompassing the area in the body mask between top of the head and shoulders. In order to create the profile of the body width, the body mask is traversed along its larger dimension. The width of the mask in each cross section of the body mask is calculated. Furthermore, since the shoulder is most probably located at either of the two ends of the body mask, the width difference graph is multiplied by two gaussians centered near the two ends of the mask to reflect this emphasis. Similar to Section 3.2.1, 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 is identified by the peak index in the width difference graph. As shown in Figure 5, the head is approximated by a box covering the area between the shoulder and the top of the body mask. This mask is denoted as HMb. This scheme assumes that the front or back dimensions of the body are visible to some extent.

3.2.3 Head mask merging

Using two different schemes for head detection allows the algorithm to opportunistically employ one or both of their available results. The two masks obtained from these schemes are combined using an *OR* operation [Figure 6(a)]. If the two masks have an overlap (HMa $\cap$ HMb $\neq 0$), or if one of the head detection algorithms is unsuccessful (HMa=0 or HMb=0), then HMa$\cup$HMb is taken as the final head mask. The location of the head used in future steps is set to be the centroid of the final head mask. However, if the two head masks do not have an overlap (HMa$\cap$HMb=0) 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 locations will be reported to the monitoring center, if a report is deemed necessary [see Section 4].

3.3 Certainty Assignment Module (CAM)

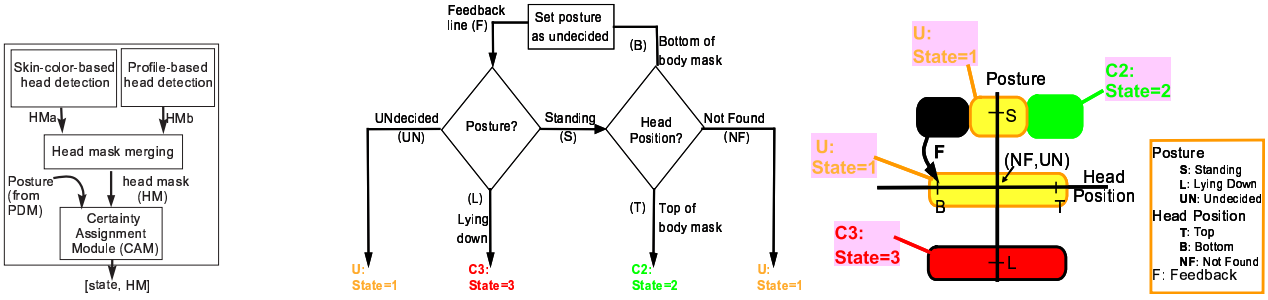
The results of posture detection and head detection modules are then passed to the certainty assignment module (CAM) to assign a state number to the result [Figure 6(a)]. This state number represents the certainty of the system about the user's condition. As shown in Figure 6(b) and further elaborated in Figure 6(c), each [*posture*, *head position*] combination gets mapped to a certain state number. Note that if an agreeing centroid for the head cannot be determined, the location of the head is unknown, which is mapped to the *NF* (Not Found) position in Figure 6(c).

The lying down posture most likely corresponds to a hazard situation; as a result, regardless of the location of the head, this condition needs to be reported to the monitoring center for further investigation. Thus, the highest state number (C3=3) is assigned to any condition which is identified as a lying posture, regardless of the location of the head. The standing up posture can only be verified if the location of the head is towards the top of the body, in which case C2=2 is assigned as the state. However, if the head location is towards the bottom of the body, the system identifies this as a conflict and runs the system through the feedback loop F, marked in Figures 6(b) and 6(c). This feedback loop allows the system to re-calculate the location of the head without taking into account information about the posture of the person. As a result, there will be no emphasis placed on either ends of the body [see Sections 3.2.1 and 3.2.2]. Finally, if either the posture or the location of the head is unknown (and the posture is not lying down), the assigned state is U=1.

3.4 Multi-Frame Event Analysis

The single-camera event analysis module can be used on multiple frames in order to obtain the trajectory of the person's movement, which could be analyzed for identifying hazardous situations. We captured a set of 30 frames from a fall event and passed the frames through the single-camera event analysis module. The trajectory of the centroid of the body mask was also obtained. The results of this experiment is shown in Figure 7.

Using this scheme, it is possible to make use of the subsequent frames to extract more information about the event. Furthermore, it is possible to use the information obtained from the successful rounds of image analysis to predict the



(a) (b) (c)

Figure 6. (a) A state representing the certainty about the user's condition is assigned by the CAM module. (b) The posture and head information is mapped to determine the state of the user. The x-axis maps the head location and the y-axis maps the posture. The assigned state reflects the criticality of the situation.

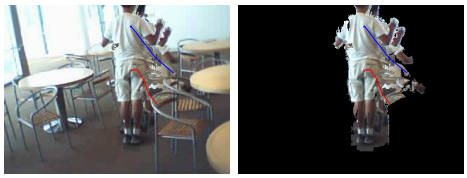


Figure 7. The body's centroid and the head's centroid are located in each frame using the single-camera event analysis module. The trajectory of the fall is shown on the superimposed frames.

location of the head in the frames where the module is not able to locate the head position.

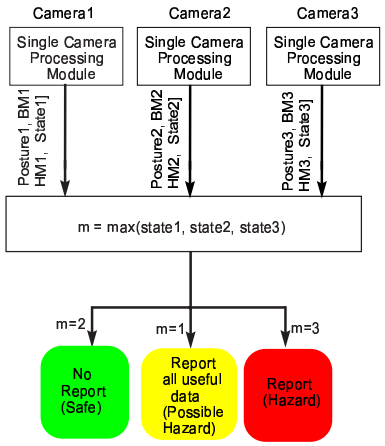


Figure 8. Through collaborative reasoning, the information obtained from the three cameras are combined to prepare a more efficient and more reliable report on the status of the user.

4 Collaborative Reasoning

Each of the distributed processing modules return information about the state and posture of the user as well as the obtained body mask (BM) and head mask (HM). Through collaborative reasoning, these information are combined to make a final decision about the type of report that needs to be prepared, as shown in Figure 8. The maximum state number received from the single-camera modules corresponds to the camera that is most certain about reporting the condition of the patient. All other states that are lower than the maximum value for the state are regarded less significant in creating the report. For instance, a maximum state of 3 implies that at least one of the cameras has identified the user in a lying down posture, which is interpreted as critical. All of the possible cases are handled as illustrated in Figure 8. Some examples of the state assignment process are shown in Figure 9. As shown there, the "RED" condition on the left is identified as a critical condition and is immediately reported to the monitoring center. The "GREEN" condition on the right is identified as a normal condition and no data is reported. Note that although the head is not in the field of view of camera 3 (thus resulting in a failure in that camera's head detection), the system is able to determine the status of the user correctly. This is due to the use of the feedback loop, which identifies and handles the conflicts within the head and posture detection processes.

Algorithm	Posture Detection	Head Detection	Overall Status
Accuracy	**94%**	**70%**	**90%**

Table 1. Accuracy of the posture detection algorithm, head detection algorithm, and the overall accuracy of the system.

5 Performance

As illustrated in Table 1, in our set of experiments, the single-camera posture detection function succeeded in 94% of the runs and the head detection was successful 70% of the time. These results were obtained by performing distributed head and posture processing on 16 images. Furthermore, 3 cameras were used to obtain images from 3 different angles. 10 frames from each camera were used to investigate the accuracy of the overall patient status assignment. It was observed that 90% of the times, the system was able to correctly identify the status of the person and produce a suitable report.

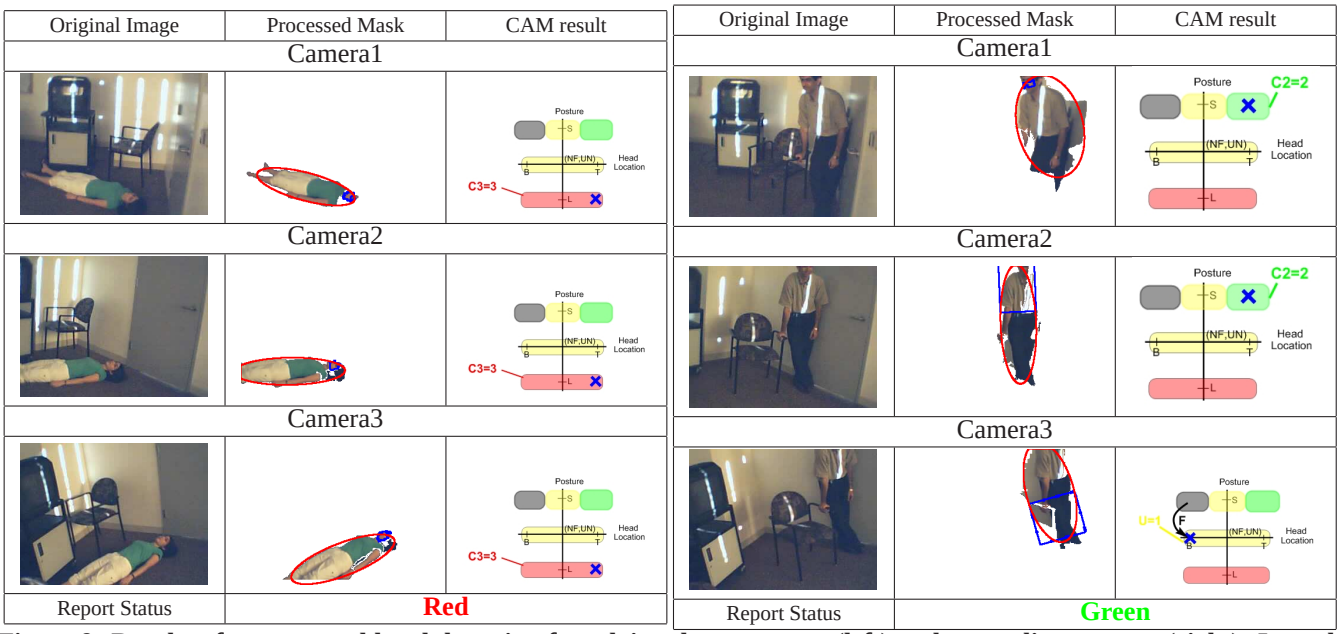


Figure 9. Results of posture and head detection for a lying down posture (left) and a standing posture (right). In each table, the first column shows the image captured by the 3 cameras; the second column shows the results of head and posture detection; the third column shows the certainty assignment plot through which the state of the user is identified.

Although the head detection component of the system might fail in certain camera views, the overall system is still able to prepare a reliable report by combining information from different camera views.

6 Conclusions

In this paper we construct a distributed vision-based reasoning smart care system for elderly persons. Distributed scene analysis modules make use of opportunistically available information in the image to assign a certainty state to the user's condition. Through collaborative reasoning, the information available from different views are fused together and the appropriate report is created and transmitted to the monitoring center for further investigation. By making use of different camera views, the number of false alarms is reduced, making the system more reliable and more efficient.

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, 2006.

[2] R. Cucchiara, A. Prati, and R. Vezzani. Posture classification in a multi-camera indoor environment. In *Proc. of ICIP*, 2005.

[3] G. C. de Silva, B. Oh, T. Yamasaki, and K. Aizawa. Experience retrieval in a ubiquitous home. In *Proc. of 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 *Proc. of 13th Annual ACM Int. Conf. on Multimedia, ACM Press*, 2005.

[5] J. Fang and G. Qiu. A colour histogram based approach to human face detection. In *Proc. of Int. Conf. on Visual Information Engineering: Ideas, Applications, Experience.* IEE, Visual Information Engineering Professional Network, July 2003.

[6] H. Fatemi, R. Kleihorst, H. Corporaal, and P. Jonker. Real time face recognition on a smart camera. In *Proceedings of Acivs*, Sept. 2003.

[7] 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 *Proc. of EMBEC*, 2005.

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

[9] 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 *Proc. of ProRISC*, Nov. 2003.

[10] B. Kwolek. Face tracking system based on color, stereovision and elliptical shape features. In *Proc. of IEEE AVSS Conf.* IEEE Computer Society, July 2003.

[11] 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 Proc. of 3rd Int. Conf. on Pervasive Computing*, May 2005.

[12] A. M. Tabar, A. Keshavarz, and H. Aghajan. Smart home care network using sensor fusion and distributed vision-based reasoning. In *Proc. of VSSN 2006*, October 2006.