Towards a Single Sensor Passive Solution for Automated Fall Detection

Michael Belshaw, Babak Taati, Jasper Snoek, and Alex Mihailidis

Abstract—Falling in the home is one of the major challenges to independent living among older adults. The associated costs, coupled with a rapidly growing elderly population, are placing a burden on healthcare systems worldwide that will swiftly become unbearable. To facilitate expeditious emergency care, we have developed an artificially intelligent camera-based system that automatically detects if a person within the field-of-view has fallen. The system addresses concerns raised in earlier work and the requirements of a widely deployable in-home solution. The presented prototype utilizes a consumer-grade camera modified with a wide-angle lens. Machine learning techniques applied to carefully engineered features allow the system to classify falls at high accuracy while maintaining invariance to lighting, environment and the presence of multiple moving objects. This paper describes the system, outlines the algorithms used and presents empirical validation of its effectiveness.

I. INTRODUCTION

With the aging population in Canada (and around the world), Canadians are beginning to experience the effects of an over-extended health care system. One approach to alleviate the healthcare pressures resulting from this demographic change, is to promote and support *aging in place*. Aging in place, is the concept of enabling and empowering older adults to remain independent and live in their own homes for as long as they wish. This not only lessens the strain on the healthcare system, but also improves the quality of life of the elderly.

Unfortunately, without proper considerations, aging in place may pose health and safety risks, such as those arising from falls within the home and subsequent complications. In fact, falls and fall-related injuries are a major cost factor to the healthcare system in Canada [1]. An older adult who has suffered a fall may lay on the ground for an extended period of time, even days, before receiving proper healthcare. This is a major risk factor, particularly for older adults living alone. Such delays often result in major, even life-threating, health complications and could incur significant cost. This paper focuses on the development of an intelligent vision-based in-home fall monitoring system for the purpose of providing prompt response to older adults in case of an emergency.

Personal Emergency Response Systems (PERS) provide victims of falls or other dangerous events with an outbound call for help. The standard PERS device is a wearable panic button which is carried by the person and must be manually triggered once an emergency situation arises. This device relies on the following criteria: the person is wearing the button, is conscious, and is capable of pushing the button *after*

the event. However, these criteria are not always met for fall victims. Our solution automatically monitors the events and actively detects falls without the need for victims to initiate a call for help, or to even remember to wear the device. Computer vision and video analysis techniques are used for automated fall detection. A complete working prototype has been designed, implemented, and tested.

II. PREVIOUS WORK

Several vision-based fall detection methods have been proposed in the past [2]–[6]. These techniques typically use a small (< 10) set of features (visual cues), usually operate in controlled environments (e.g. with controlled lighting), and are unable to function in the presence of multiple subjects within the scene. These techniques do not yet fully address the challenges of real life situations in which rooms have complex lighting (e.g. with external windows, multiple moving subjects and dynamic environments).

Our previous work demonstrated a single-camera ceiling-mounted PERS that detected falls in real home environments [7]. The system maintained an adaptive background model to extract a silhouette of the active region. This background model was derived from an single Gaussian and was able to detect foreground regions and their cast shadows. The silhouette and shadow region were then processed through a neural network to distinguish *fall* versus *no-fall* events based on shape and velocity features.

Two in-home trials were conducted in real living rooms, in which able-bodied residents were asked to occasionally simulate falls, in order to evaluate the system. Upon improving the system following the first trial, the system was able to detect 100% of the simulated falls in the field of view (FOV) of 48° by 61°. However, numerous false alarms did occur even with the addition of post processing (5.43 per day).

The real-world tests indicated that complex lighting situations (such as light from windows) and multiple occupants generated the majority of the false positives when operating at five frames per second (fps).

III. CONTRIBUTIONS

This paper presents an improved version of the system and classifier of our earlier work [7] which accommodates some of the limitations observed during the in-home trials. These improvements include: the use of a wide-angle lens to increase the area of coverage; dealing with the resulting image distortion and lighting issues by using global features; enhancing robustness to abrupt lighting changes and lighting effects such as sunlight through a window; handling more
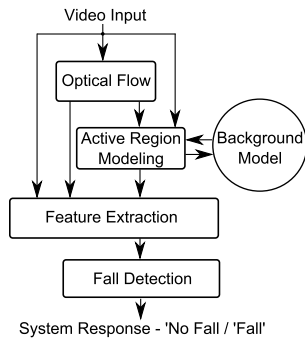
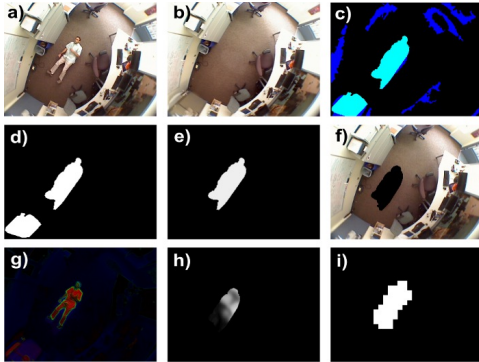Fig. 1.   Algorithmic Block Diagram of the System

Fig. 2.   Sample Fall Image and Its Processing: a) Input Image; b) Background Model; c) Foreground Blobs (light) and Shadow Blobs (dark); d) Foreground Blobs; e) Foreground Blob with Recent Activity; f) Foreground Masked Out; g) Background Subtraction Result; h) A Foreground Blob Region's Optical Flow Magnitude; i) Low Resolution Annotated Fall Blob Area

than one subject by tracking multiple active silhouette blob regions; enhancing background modeling to reduce false active silhouette regions associated with lighting changes by merging with image optical flow; collecting a large data set of simulated falls and normal events for training and testing purposes; and reducing the operating frame rate from five to two fps for lower data-rates, processing, and power requirements.

IV. SYSTEM DESIGN

Fig. 1 presents the block diagram of the algorithmic flow of the system. The algorithms are implemented in C++ and OpenCV on an Intel I7 processor.

An inexpensive Logitech QuickCam Pro 9000 captured video for training and testing at two fps at QVGA (320×240) pixel resolution. The camera was modified by removing the auto-focus lens and mount, and replacing them with a M12x0.5 lens mount and a board mount lens with focal length of 2.2mm.The wide angle lens increased the FOV to $92° \times 109°$.

A. Active Region Modeling

Active region silhouettes that were analyzed to determine the presence of falls were calculated with background subtraction and optical flow information. As described in our previous work, a background model was built by using a
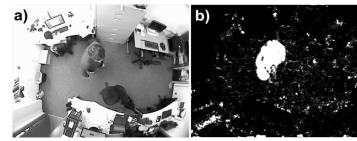

Fig. 3.   a) A Sample Video Frame; b) Optical Flow Magnitude

single Gaussian distribution centered at the estimated intensity of each pixel [7]. The model was then used to perform background subtraction to form silhouette images (active region *blobs*) and shadows (fig. 2(g),(d),(c)). The shadow regions were identified based on the isolation principles outlined in [8]. In contrast to our other work, in which only the largest active region blob was considered as the object of interest, the current system considers all active blob regions and their shadows. This multi-blob approach creates a difficulty with respect to the selection of the adaptation rates for updating the background model.

Many outlier blobs tend to occur from local lighting changes and objects being moved and should be adapted into the model. Previously, the largest foreground blob was always considered to be related to the object of interest and was thus adapted very slowly into the background model while the remaining blobs were adapted rapidly to account for lighting changes.

1) Optical Flow Decay: In order to adapt the background model for multiple active region blobs, optical flow computations [9] were incorporated into the background adaptation approach. A similar method for enhancing background modeling with optical flow has previouly been investigated [10]. Specifically, in this work we used optical flow magnitude to control background adaptation rates. This optical flow method rewards objects that are moving periodically with slow adaptation rates and objects that are non-moving or slowly moving with fast adaptation rates. (See fig. 2 for example input and model images processed to form blob regions and shadows.) The magnitude of dense optical flow vectors were used to quantify the motion at each blob within the scene. Fig. 2(e),(h) shows the optical flow magnitude of a foreground blob region. Fig. 3 depicts an example video frame for which the optical flow magnitude highlights a moving person as well as some sparse random noise across the image.

The underlying assumption here is the motion cues can rid the system of the stationary object blobs and also blobs arising from lighting changes. Moved furniture generates large foreground regions which remain stationary for extended periods. Similarly, lighting changes typically do not result in significant optical flow magnitude. Very slow moving blobs are therefore adapted more quickly into the background. Large blob regions with significant motion, on the other hand, are more likely to correspond to humans and are adapted at much slower rates.

B. Features

There are three types of features chosen for the fall detector: Silhouette Features, Lighting Features, and Flow

Features.

1) Silhouette Features: Silhouette blob features similar to those in our previous work were chosen for the fall detector [7]. These features include the Hu moments [11] as well as various blob status features that describe shape and dynamics. These features are encoded not only for the scene's foreground blobs, but also for their corresponding shadow blob regions. In total there are 32 silhouette features.

2) Lighting Features: The lighting features consist of various global difference image pixel statistics which are sensitive to lighting changes. They are applied on the full image as well as an inverse blob masked difference image. (See fig. 2(f) for an example input image with a blob mask applied over the fallen person.) The statistics are calculated on a difference image resulting from previous frame subtraction and also background model subtraction – with and without the blob mask applied to the images. To enhance the amount of information obtained by the features, they are calculated in both the RGB color space and also in a lighting intensity invariant color space $c_1c_2c_3$ [12]. Empirical analysis has shown that these features are a strong indicator of lighting changes. In total there are 24 lighting features.

3) Flow Features: Optical flow [9] features were generated using optical flow magnitude image statistics in the same manner as lighting features. The statistics are calculated for both the entire optical flow magnitude image and also the blob region's optical flow image. (See fig. 2(h) for an example of blob optical flow.) In total there are six optical flow features.

V. FALL DETECTION

To classify each frame as a *fall* or *no-fall*, based on the silhouette and lighting features, machine learning techniques were used. A training set that is per-frame annotated with *fall* or *no-fall* information was created. In frames with fallen persons, their shapes were annotated in 20% resolution. (See fig. 2(i) for a low resolution annotation example.) The annotated training data was later processed by the machine learning algorithm to learn from the annotated examples. Once the training was complete the resulting model was tested with the new test input frames to evaluate the system.

A. Data Collection and Annotation

Training and testing data were collected from three office room settings. Two rooms were only used for training data while the other room was used for testing data. This training and testing room separation was performed so that the empirical evaluation was less room-dependent. All three rooms have one large window occupying about 50 percent of a wall. These windows allowed for dynamic lighting changes due to sunlight and cloud movement over the course of a day. During the data collection period, videos were recorded in each room at two fps. Over the course of three weeks, able-bodied participants were asked to perform several simulated fall postures on the floor in all three rooms.

For analysis, the data set was reduced to 195 fall event sequences to increase the speed of the training processes:

162 sequences, from two rooms, were used for training; the remaining 33 sequences, from the third room, were used for testing. Each sequence was generated by selecting approximately 7 minutes of frames around the simulated falls.

The video sequences were manually annotated with one of five states assigned to each frame: *before-fall*, *before-fall-transition*, *fall*, *after-fall-transition*, and *after-fall*. The transition states are areas of ambiguous labeling when a person is in the process of falling and is not yet on the floor. Further annotation (a low resolution mask at 20% of the native image resolution) of the fall silhouettes was performed for all *fall* state frames (fig 2(i)). During the training process, in cases where multiple active blobs were tracked, the silhouette mask was used to identify the blob that corresponded to the fall[1].

B. Classifier Training

Fall detection was formulated as a per-frame binary classification task. That is, a binary classifier was trained to discriminate frames which were labeled as containing a *fall* from those which were a *no-fall*. Only the frames from the beginning of a fall sequence up to the *before-fall-transition*, and the first 12 *fall* frames were considered. If more fall frames were considered, slower adaptation rates would be necessary to prevent the silhouette from disappearing. However, a slow adaptation rate would result in the system not adapting quickly to moved furniture or lighting changes. Also, the transition frames were not included as they could not be reasonably labeled as *fall* or *no-fall* for training.

Three training approaches were taken to investigate which approach would better model the data. The data set was separated into two catagories, training and testing, based on rooms. This separation of rooms enabled us to better detect if the classifier works for new rooms. Classification performance was compared using three different machine learning classifiers, which vary in regularisation and the complexity of the decision boundary on the features: logistic regression, neural networks, and support vector machines [13].

VI. RESULTS

After training the classifier to detect the presence of a fall (per-frame), the classifier was used to classify the training and test data to determine system performance. The results in fig. 4(a) (train data) and 4(b) (test data) show the ROC curve for the classifier with various combinations of features used for training: L- indicates lighting features; S- indicates silhouette features; BG indicates image subtraction with the background model; Pre indicates image subtraction with the previous frame; Hu indicates the Hu moment features; and Status indicates the shape and dynamics features.

From these figures, it can be observed that the different features classify the training and testing data with various levels of success. The L-BG-Model, L-Pre-Model (Lighting

---

[1]The data set is planned to be freely available for research purposes.

(a) Train Data: Feature Results   (b) Test Data: Feature Results   (c) Test Data: LogReg, NN, and SVM Compared
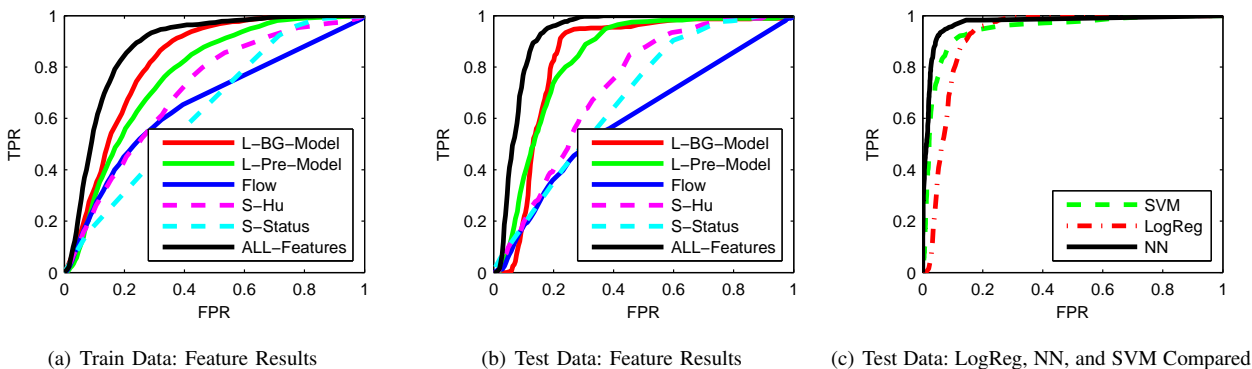
Fig. 4. Machine Learning ROC Curves

Features applied to both the previous frame difference and background model difference) appear to provide consistent, relatively high performance on both data sets. In contrast, the Flow (optical flow magnitude), with only six individual features, appears to have only minor success in classifying the fall data. The overall performance on the test data from the single room was slightly better than the results on the training data from the other two rooms. The improved performance was attributed to the patterns in the occupent traffic (more walking through, less sitting) and to having fewer frames with partially visible falls.

Further tests were conducted with additional machine learning techniques to evaluate performance on the data set with all features considered. The techniques chosen were logistic regression, neural networks, and support vector machines. The results of the three techniques are depicted in fig. 4(c). A Multilayer Perceptron Neural Network achieved the best overall fall detection performance. Using validation data to select an operating point on the Neural Network ROC curve resulted in a true positive rate of 92% and false positive rate of 5% on the test set.

The test data set results from the neural network (33 fall sequences) had 115 false positive fall event frames. Visual inspection of these false positive frames indicated that *all* false positives occurred with persons in the field of view. Further fall classification was performed on 559 additional non-fall sequences in which no falls frames occurred. These additional sequences contained both empty and occupied rooms. The sequences were classified and the results showed that very few empty room false positive falls occurred. Thus, the improved classifier operating with a wide angle lens demonstrated substantial robustness to lighting change events.

VII. CONCLUSIONS AND FUTURE WORK

We have designed and implemented a vision-based personal emergency response system, which is able to handle multiple active regions, has a wide field of view that can capture a single large room, and is able to operate effectively under moderate lighting changes. A simulated fall data set was collected from three rooms with large external windows and this data set was partitioned into training and testing data for machine learning. Results indicate that the system

was able to handle multiple active regions in the scene while maintaining robustness to lighting changes.

For future work, we will: perform post-processing on the per-frame classifications; expand on the global lighting features; further decrease the fps; extend the data set; and perform in-home tests in large scale deployment.

REFERENCES

[1] SMARTRISK, "The economic burden of injury in canada," 2009.
[2] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Monocular 3d head tracking to detect falls of elderly people," in *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, 2006.
[3] D. Anderson, J. Keller, M. Skubic, X. Chen, and Z. He, "Recognizing falls from silhouettes," in *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, (New York City, USA), 2006.
[4] J. Spehr, M. Gvercin, S. Winkelbach, E. Steinhagen-Thiessen, , and F. Wahl, "Visual fall detection in home environments," in *International Conference of the International Society for Gerontechnology*, (Pisa, Italy), 2008.
[5] H. Nait-Charif and S. McKenna, "Activity summarization and fall detection in a supportive home environment," in *International Conference of Pattern Recognition*, (Cambridge, UK), 2004.
[6] T. Lee and A. Mihailidis, "An intelligent emergency response system: preliminary development and testing of automated fall detection," *Journal of Telemedicine and Telecare*, vol. 11, pp. 194–198, 2005.
[7] M. Belshaw, B. Taati, D. Giesbrecht, and A. Mihailidis, "Intelligent vision-based fall detection system: preliminary results from a real-world deployment," in *Rehabilitation Engineering and Assistive Technology Society of North America (RESNA)*, 2011.
[8] P. Rosin and T. Ellis, "Image difference threshold strategies and shadow detection," in *British Machine Vision Conference (BMVC)*, (Birmingham, UK), 1995.
[9] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," in *Proceedings of the 13th Scandinavian Conference on Image Analysis*, LNCS 2749, (Gothenburg, Sweden), pp. 363–370, June-July 2003.
[10] D. X. Zhou and H. Zhang, "Modified gmm background modeling and optical flow for detection of moving objects," in *IEEE International Conference on Systems, Man and Cybernetics*, vol. 3, pp. 2224–2229, 2005.
[11] M. Hu, "Visual pattern recognition by moment invariants," in *IRE Transactions on Information Theory*, vol. 8, pp. 179–187, 1962.
[12] T. Gevers and A. Smeulders, "Color-based object recognition," *Pattern Recognition*, vol. 32, pp. 453–464, March 1999.
[13] T. Joachims, "Making large-scale SVM learning practical," in *Advances in Kernel Methods - Support Vector Learning* (B. Schölkopf, C. Burges, and A. Smola, eds.), ch. 11, pp. 169–184, Cambridge, MA: MIT Press, 1999.