

Recognizing Falls from Silhouettes

Derek Anderson, James M. Keller, Marjorie Skubic, Xi Chen, and Zhihai He

Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO 65203

Abstract—A major problem among the elderly involves falling. The recognition of falls from video first requires the segmentation of the individual from the background. To ensure privacy, segmentation should result in a silhouette that is a binary map indicating only the body position of the individual in an image. We have previously demonstrated a segmentation method based on color that can recognize the silhouette and detect and remove shadows. After the silhouettes are obtained, we extract features and train hidden Markov models to recognize future performances of these known activities. In this paper, we present preliminary results that demonstrate the usefulness of this approach for distinguishing between a few common activities, specifically with fall detection in mind.

Keywords—Eldercare, Silhouettes, Fall Recognition, Hidden Markov Models, and Video Privacy.

I. INTRODUCTION

MANY elders fall and are unable to get back up or call for assistance. As if the severity of the injury incurred during the fall was not enough, the problem is compounded by the realization that a long time could pass while they lie on the ground hurt until someone comes along and discovers them. In this paper, we introduce a method for recognizing falls from video sensors and discuss preliminary results based on a few common activities.

The video sensors discussed here are one component of a larger array of technologies being deployed in the TigerPlace community to help elders live longer and more independent lives [1][2]. A goal is to use the video sensors to collect information about daily activities of the elderly residents. From this, we can extract important information to perform automated functional assessment and detect abnormal events, such as people falling on the floor. We have discovered that many elders are reluctant to have video monitoring unless we can assure their privacy. To maintain the privacy of the residents, segmentation of the human from the video produces a silhouette, a binary map that only indicates the body position of the individual in the image. Features in the video that previously made the resident distinguishable are no longer present.

We have proposed a technique for extracting silhouettes based on statistically modeling a static background and then segmenting the human based on color information [3]. We extract a color feature for silhouette identification, and a brightness feature at each pixel for the detection and removal of shadows. This work was extended in [4] to address the difficulties of silhouette extraction in complex

and dynamic environments. High-level knowledge was fused with low-level feature-based classification to handle a time-varying background, and a decision process based on a fuzzy logic inference system was used to detach the moving objects from the human silhouette.

One central task in silhouette extraction is background modeling. Once a background model is established, those image regions with significantly different characteristics from the background are considered foreground objects. In [5], a Least Median of Squares method was used to construct a background model, and in [6], a differencing function was proposed to extract moving human silhouettes. Elgammal et al. [7] demonstrated a non-parametric kernel density estimation method to model and subtract the human from a background. The adaptive background updating handles small motion in the background scene, such as moving tree branches. A slightly different approach outlined in [8] is based on processing in the YUV color space, and is capable of detecting shadows and extracting the human body.

After the objects are segmented from their background, the problem becomes one of monitoring their activity. Brand has demonstrated monitoring office activity with hidden Markov models (HMM) [9]. They used entropic based HMMs to acquire models that are more interpretable, in terms of the parameters that the models estimate. Zhong presented an unsupervised method in [10] for detecting unusual activity using a combination of many simple features. Using a method similar to document-keyword analysis, he divided the video into equal length segments and classified the extracted features into prototypes, from which a prototype-segment co-occurrence matrix was calculated and used to determine abnormal activity. Zhong tested this procedure for a few different scenarios, one of which included typical activity in a nursing home.

In this paper we explain how to extract features from a silhouette and outline a procedure for training and performing activity recognition with HMMs. Preliminary results for falling, walking, and kneeling are used to demonstrate the effectiveness the technique.

II. SILHOUETTE BASED FALL RECOGNITION

As mentioned above, the first step in recognizing a fall involves the segmentation of the human from a known background. Video was captured at 3 fps on a Creative Labs WebCam Live. Figure 1 shows three frames and their corresponding silhouettes for the fall sequence we analyzed.

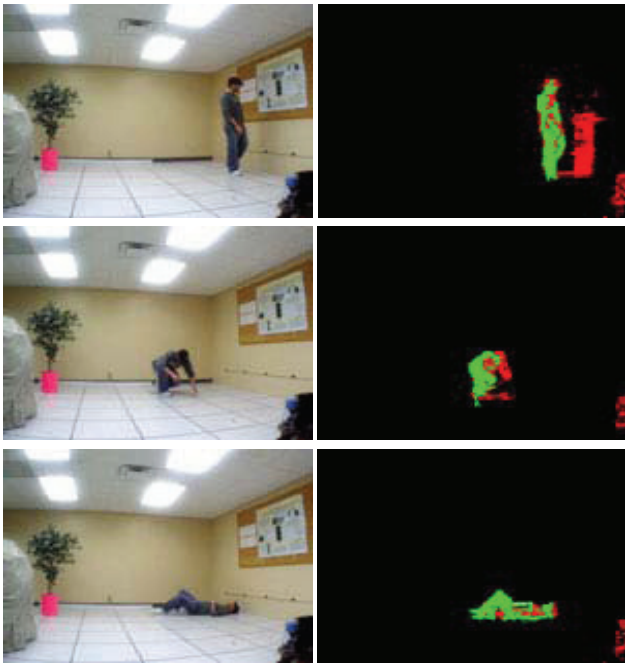


Fig. 1. Silhouette extraction results for a fall sequence. Each row shows the original image and its corresponding silhouette. Green areas indicate the silhouette and red areas are the shadow regions that are removed.

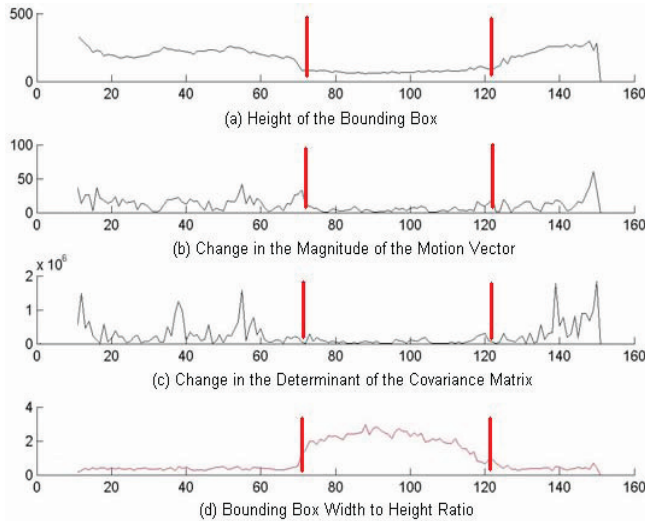


Fig. 2. Fall activity according to a few different silhouette features. The x-axis represents the frame number. Subject walked into the scene at frame 11, fell around frame 70, and got back up around frame 121. The vertical red bars show the frames where the user fell and got back up.

After the silhouettes are acquired, the next step involves extracting features and tracking their patterns over time. We continue to explore feature extraction but initially find that the width to height ratio of the silhouette bounding box and the off-diagonal term from the covariance matrix provide adequate features for fall detection. The bounding box width to height ratio indicates if the silhouette is larger in the vertical plane versus the horizontal plane, which for a person

standing, is typically the case. The off-diagonal covariance value is a measure of x-y variance. Figure 2 shows a few of the features that we have extracted, while figure 3 shows the features we are tracking.

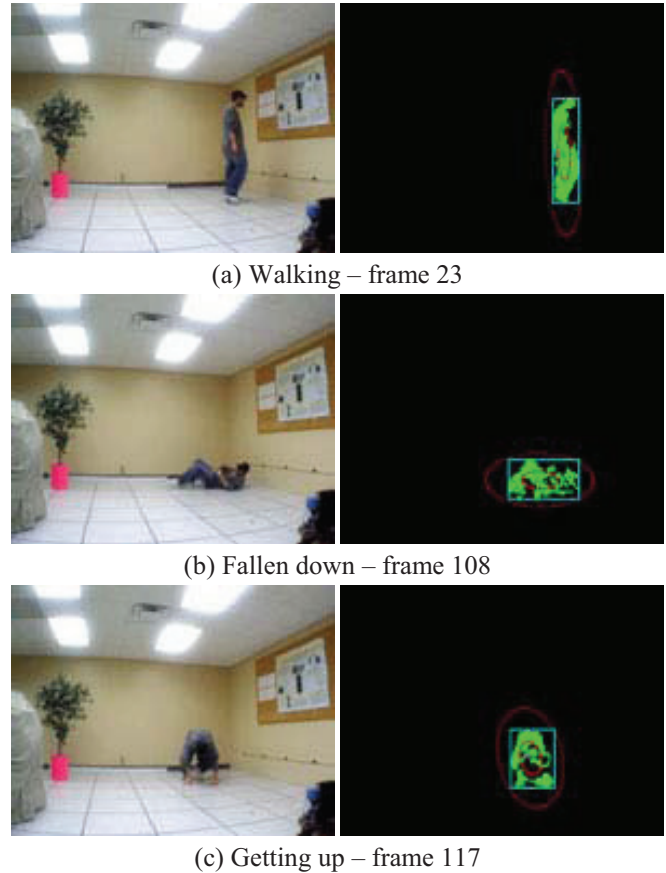


Fig. 3. Silhouette feature extraction results for the falling sequence. The bounding box is shown in light blue and the covariance matrix is shown in red at 1 and 3 standard deviations.

The most noticeable feature for the falling activity in figure 3 is the bounding box. When the subject is standing (figure 3.a) the width to height ratio is small. When the subject is on the ground (figure 3.b), the width to height ratio becomes much larger. As he is standing up (figure 3.c) the ratio appears similar. When the features are plotted for the fall sequence, the walking and falling frames appear as different clusters. These clusters are shown in figure 4.

These observations form a sequence over time and are used to train and perform classification with HMMs [11][12]. An HMM is a doubly stochastic process, i.e. there is an underlying stochastic process that is not observable (hidden) but can only be observed through another set of stochastic processes that produce the sequence of observed symbols [11]. Each activity category, such as falling, will have a separate HMM that is trained via the Baum-Welch procedure [11][12]. The most likely model is calculated for each observation sequence according to the forward-backward procedure with appropriate scaling [11][12]. For

the top model, the most likely state sequence is determined according to the Viterbi algorithm [11]. The number of consecutive times the model is in the same state can be combined with the frame rate to determine the time duration spent in each state. The probability density functions (pdf) for each state can be modeled as normal distributions. Due to the simple experimental design, we limited our state models to single Gaussian pdfs. Mixtures of Gaussians will be utilized as the complexity of the problem grows. Figure 5 shows the means and covariance matrices that the HMM learned for the fall sequence reported in this paper.

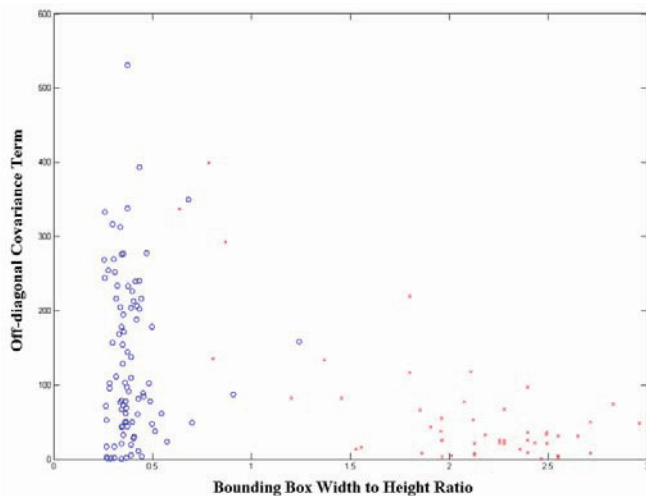


Fig. 4. Fall activity clusters. The walking frames are shown as blue circles while the falling frames are shown as red crosses.

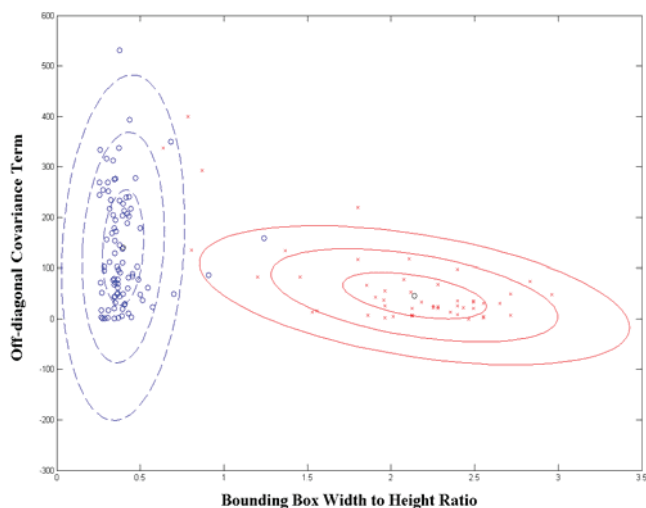


Fig. 5. Means and covariance matrices learned by the HMM for the fall activity. The blue curves are the Gaussian pdf learned by state 1, i.e. the walk state, shown at 1, 2, and 3 standard deviations. The red curves are the Gaussian pdf acquired for the second HMM state, i.e. the fall state, at 1, 2, and 3 standard deviations.

III. EXPERIMENTS

Both nurses and elders identified falling as the most significant activity that the video sensor network should recognize. The following experiments were designed with this goal in mind. Below, we present an analysis of common activities such as walking, falling, and kneeling. These activities were selected in order to observe the differences between two different activities, falling and walking, and two more similar activities, falling and kneeling. Figure 6 shows a kneeling frame and the extracted silhouette features.

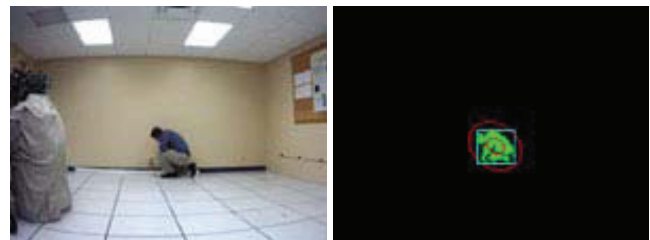


Fig. 6. Kneeling activity and its corresponding silhouette features: bounding box (light blue) and the covariance matrix (red) at 1 and 3 standard deviations.

The two sequences analyzed were 50 and 32 seconds long, respectively, comprising 150 and 96 video frames. Preliminary results from this initial set of features are very encouraging. We have thus far been able to train an HMM for each activity category and the trained models are able to correctly recognize the corresponding activities. The most likely state sequences for the falling and kneeling activities are shown in figures 7 and 8 respectively.

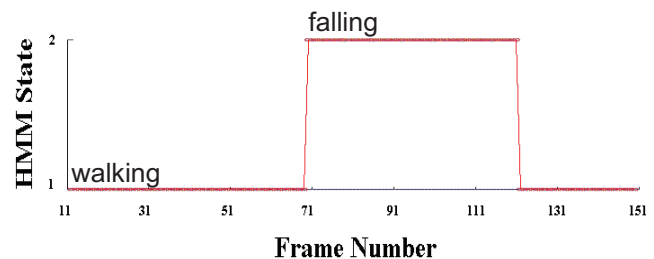


Fig. 7. Most likely state sequence for the fall activity according to the fall HMM. Subject fell around frame 70 and got up around frame 121.

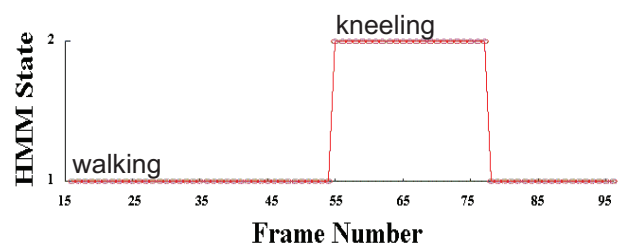


Fig. 8. Most likely state sequence for the kneeling activity according to the kneeling HMM. Subject kneeled around frame 54 and got up around 78.

The state sequences in figure 7 and 8 transition at the time points we expected and have no unexpected transitions. The walk state is learned by each HMM, i.e. in both the fall and the kneeling models. The four pdfs for the two HMMs are plotted in figure 9 to illustrate the similarities and differences between what was learned by the two models.

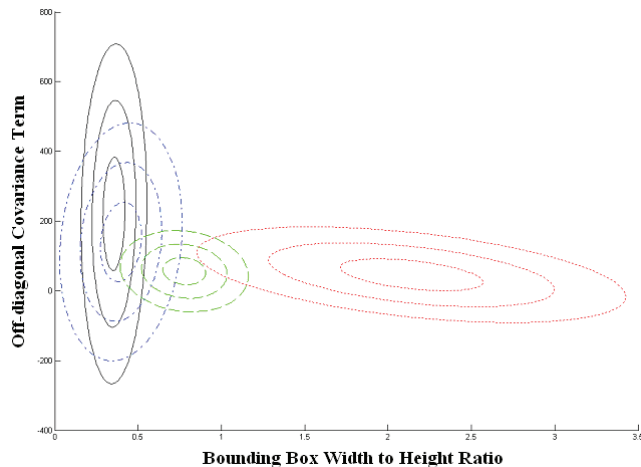


Fig. 9. Pdfs learned by the 2 HMMs (falling and kneeling). Black solid pdf is the walking state learned by the kneeling model, the blue semi-dashed pdf is the walking state learned by the fall model, the thick-dashed green pdf is the kneeling state learned by the kneeling model, and the highly dashed red pdf is the fall state learned by the fall model.

The means and covariances learned by the two HMMs in figure 9 show an overlap in the distribution of the walking state (the semi-dashed pdf in blue and the solid black pdf). The falling state is the highly dashed red curves and it has good separation from the other distributions. The kneeling state is shown as thick-dashed green curves and it resides between the walking and falling states. This makes sense because the walking state has a low bounding box width to height ratio, the falling state has a higher ratio value, and the kneeling state has a ratio in between around a value of one. The model probabilities for each activity are slightly higher when both features are used, but the majority of information is captured by the bounding box width to height ratio.

IV. CONCLUSIONS AND FUTURE WORK

Recognizing falls for the elderly is an important task that can be approached by designing a video sensor network that is capable of segmenting a human from its background and tracking it over time. Once the human is segmented and any shadows are removed, features need to be extracted from the silhouette in order to use hidden Markov models (HMM) for temporal pattern recognition. HMMs are the workhorse of temporal pattern recognition and are a natural choice for this problem. The privacy of residents is ensured through the extraction of silhouettes, a binary map that indicates the body position only. Clusters form in the feature space for

the different activities and the resulting HMMs are capable of distinguishing between walking, kneeling, and falling. The most likely state sequence is an additional tool that can be combined with the frame rate to determining the amount of time the model spent in each state.

The work in this paper is preliminary and additional development and tests need to be run over a wider range of elderly activities. We also need to test on a broader range of subjects under a larger range of falling conditions. The width to height ratio of the bounding box has thus far proven to be a useful feature; however, additional features may be necessary or add additional benefit. Also, a single camera limits the viewing angle of the scene and, more importantly, the resident. It might work out that different features are necessary or multiple cameras in a single scene are required.

ACKNOWLEDGEMENTS

The authors were partially supported by NSF ITR grant IIS-0428420 and the U.S. Administration on Aging, under grant 90AM3013.

REFERENCES

- [1] M.J. Rantz, K.D. Marek, M.A. Aud, H.W. Tyrer, M. Skubic, G. Demiris, and A. Hussam, "A Technology and Nursing Collaboration to Help Older Adults Age in Place," *Nursing Outlook*, Jan-Feb 2005, vol. 53, no. 1, pp. 40-45.
- [2] M. Skubic, "Assessing Mobility and Cognitive Problems in Elders," AAAI 2005 Fall Symposium, Workshop on Caring Machines: AI in Eldercare, Arlington, VA, Nov., 2005.
- [3] X. Chen, Z. He, D. Anderson, J. Keller, and M. Skubic, "Adaptive Silhouette Extraction and Human Tracking in Complex and Dynamic Environments," *Proceedings, Int'l Conf. on Image Processing*, Atlanta, October 2006.
- [4] X. Chen, Z. He, J. Keller, D. Anderson, and M. Skubic, "Adaptive Silhouette Extraction in Dynamic Environments Using Fuzzy Logic," *Proc., WCCI*, Vancouver, BC, July, 2006.
- [5] L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette Analysis-Based Gait Recognition for Human Identification," *IEEE Trans. Pattern Analysis and Machine Intelligence*, pp. 1505-1518, December 2003.
- [6] Y. Kuno, T. Watanabe, Y. Shimosakoda, and S. Nakagawa, "Automated Detection of Human for Visual Surveillance System," *Proc. Int'l Conf. Pattern Recognition*, pp 865-869, 1996.
- [7] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric Model for Background Substitution," *6th European Conf. on Computer Vision*, Dublin, Ireland, June/July 2000.
- [8] O. Schreer, I. Feldmann, U. Golz, and P. Kauff, "Fast and Robust Shadow Detection in Videoconference Applications," *4th EURASIP-IEEE Region & Int'l Symposium on Video/Image Processing and Multimedia Communications*, pp 371-375, June 2002.
- [9] M. Brand, V. Kettner, "Discovery and Segmentation of Activities in Video," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22 (8), p. 844-851, August 2000.
- [10] H. Zhong, M. Visontai, J. Shi, "Detecting Unusual Activity in Video," *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, vol. 2, pp. 819-826, June 2004.
- [11] L. Rabiner., "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE*, 77 (2), p. 257-286, February 1989.
- [12] J. Bilmes., "A Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models," *Technical Report ICSI-TR-97-021*, Univ. of Berkley, 1988.