

ULTRA-WIDEBAND RADAR AND VISION BASED HUMAN MOTION CLASSIFICATION FOR ASSISTED LIVING

Zhichong Zhou¹, Jun Zhang¹ and Yimin D. Zhang²

¹Department of Electrical and Computer Science Engineering, University of Denver

²Department of Electrical and Computer Engineering, Temple University

ABSTRACT

Fall detection for elderly is one of the most important areas in elderly healthcare. Both video and radar based detections are being developed for this purpose. This paper presents a new approach to classify different human motions through machine learning. In particular, our objective is to achieve high-accuracy fall detection through the exploitation of both video and radar data. Motion history image is applied to extract temporal features from video clips, and hidden Markov models are trained with the features extracted from both video and radar data to discern the types of motion. Experiment results indicate that the proposed approach provides improved performance in distinguishing falls from other motions such as sitting.

1. INTRODUCTION

Falls are the leading cause of injuries for elderly. Prompt fall detection saves lives and leads to effective treatment and cost reduction [1, 2]. As such, development of new techniques for prompt fall detection has attracted significant interests. Different sensing modalities, including inertial measurement unit-based wearable devices, video camera, and radar, have been proposed for this purpose [3–8]. Because wearable devices have shortcomings that they are intrusive, easily broken and must be worn or carried, we focus on non-invasive vision and radar modalities in this paper.

Video provides a non-invasive modality for motion classification and fall detection. In the field of human motion research, video-based classification is widely used with the advantage of direct perceiving and simplicity. Recognition of human motion within a video is considered a key problem in computer vision. Vision based approaches generally use videos or images to analyze motion features of a human body, and to distinguish features of fall activities from those of non-falls so as to achieve the function of fall detection.

Motion capture systems provide accurate three-dimensional (3-D) information of different human motions such as walking, running and crawling [9]. A Microsoft Kinect sensor is one such type of sensors with an attractive price [10]. However, in addition to privacy concerns, vision-based approaches have other limitations in the real life applications, e.g., the line-of-sight can be easily obstructed by walls and furniture. The Kinect sensor is also sensitive to external infrared sources which can significantly influence the captured depth of the video images. Furthermore, visual image data are sensitive to cluttered backgrounds. On the other hand, radar carries great potential to be one of the leading technologies due to its advantages of non-obstructive illumination, insensitivity to lighting conditions, privacy preservation and safety [8] [11]. The ultra-wideband (UWB) radars provide useful features about the spatial distribution of a target for human motion classification and fall detection [12]. In addition,

radar can obtain indirect but meaningful characteristics representing the moving trajectories [13].

In this paper, we examine and compare the above two different sensing modalities for fall detection. A Microsoft Kinect sensor is utilized to record human motions using RGB images, whereas a Time Domain PulsON UWB radar is employed in the experimental configuration to collect radar echo signals corresponding to the same human motions. The video and radar data are then used to examine and compare for their motion classification and fall detection performance. In the proposed approach, the motion history image (MHI) method is used to describe motion patterns, and the Hu Moments are exploited to scale the images into eight orders. In addition, principal component analysis (PCA) is used to reduce the dimension of radar data, and the k-means clustering algorithm is applied to extract useful data from both sensors. Finally, the hidden Markov model (HMM) is employed to build two motion models.

The rest of the paper is organized as follows. Section 2 describes the experiment scenario and procedures. In Section 3, we introduce vision and radar based models, and the proposed method is presented. Section 4 provides the experimental results with helpful discussions. At last, concluding remarks are presented in Section 5.

2. EXPERIMENT SETTING

In this work, a Microsoft Xbox 360 Kinect sensor is utilized as a camera to capture human motions and output RGB and depth images. A Time Domain PulsON P410 UWB radar, with a center frequency of 4.3 GHz and bandwidth of 2.2 GHz, is synchronized with and closely placed to the Kinect sensor. Figure 1 depicts the experimental environment. A stunt actor stands in front of the Kinect sensor and the radar with a horizontal distance of approximately 2 meters and performs falling and sitting facing the Kinect sensor and the radar.

The Kinect sensor generates video recordings whereas the UWB radar receives the reflection of radar signal from the human body. The stunt actor's motions cause changes in characteristics of both video and radar signals, which contain adequate information of the motions. The expected range change of a fall is approximately between 2 and 3 meters, whereas that of sitting is between 0.5 and 1 meter.

3. THEORETIC BACKGROUND

3.1. Motion History Image

The MHI method has been applied as an effective tools to describe motion shapes and spatial distributions using motion sequences that imply the recency of human actions [14]. In order to describe the

motion in an image sequence, one can form an MHI of the target energy, and represent where the motion or a spatial pattern occurs. The advantages of MHI representations lie in that video images can be recoded in a single MHI frame. In this way, a small number of MHIs can span the time scale of human motions.

An MHI $H_\tau(x, y, t)$ is given by:

$$H_\tau(x, y, t) = \begin{cases} \tau, & \text{if } D(x, y, t) = 1, \\ \max(0, H_\tau(x, y, t-1) - \delta), & \text{otherwise,} \end{cases} \quad (1)$$

where x and y describe the two-dimensional (2-D) position, t is time, and $D(x, y, t)$ is an update function indicating that an object is present in the current video image. In addition, τ is the duration that decides the temporal extent of the movement, and δ is the decay.

Figure 2 give examples of falling MHI images with three different values of τ , i.e., $\tau = 5$, $\tau = 10$ and $\tau = 20$. Proper selection of the value of τ is important because a small value of τ may lead to loss of prior information of the motion in MHI, whereas a large value of τ makes the change of pixel values unapparent in the MHI templates.

3.2. Hu Moments

Moments are widely applied to characterize image patterns [15]. In order to extract features of the segmented MHIs obtained by temporal differencing between consecutive frames, we use the Hu moments [16], which are invariant to scale, translation and rotation. Eight statistic descriptors from the Hu moments are calculated for every MHI frame $H_\tau(x, y, k)$, where k is the index of the MHI frame.

Define a 2-D $(i+j)$ th order moment of the image function $f(x, y)$ as:

$$m_{ij} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^i y^j f(x, y) dx dy, \quad i, j = 0, 1, 2, \dots \quad (2)$$

If the image function is a sectional function, the moments of all orders exist and the moment sequence m_{ij} is determined by $f(x, y)$. In this case, the image function $f(x, y)$ is characterized by the moment sequence m_{ij} , $i, j = 0, 1, 2, \dots$. It is noted that the moments in (2) may vary when $f(x, y)$ changes by translating, rotating or scaling. The following central moments are used to obtain features that are invariant to image translation, rotation and scaling:

$$\mu_{ij} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^i (y - \bar{y})^j f(x, y) dx dy, \quad i, j = 0, 1, 2, \dots, \quad (3)$$

where $\bar{x} = m_{10}/m_{00}$ and $\bar{y} = m_{01}/m_{00}$. Eight Hu moments with $i, j = 0, 1, 2$, and 3 are used as the length of feature vectors from an MHI for classification.

3.3. Radar Signal Processing

Detection of human behavior with a radar relies on motion classification. Human motions cause changes in frequency, phase and time of arrival [17]. The radio frequency (RF) band of the UWB radar is 3.1 GHz to 5.3 GHz, which yields a range resolution of 6.8 cm. The k th pulse of the transmitted radar signal is denoted as $s_k(t)$ and its duration as t_r . The sampled signal vector consisting of L time samples is denoted as $\mathbf{s}_k = [s_{k,1} \dots s_{k,l} \dots s_{k,L}]^T$, where $s_{k,l} = s_k(lt_s + (k-1)t_r)$, $t_s = t_r/L$, t_s is the sampling interval, and $(\cdot)^T$ denotes vector or matrix transpose. Concatenating vectors \mathbf{s}_k , $k = 1, 2, \dots, K$, form an $L \times K$ received signal matrix

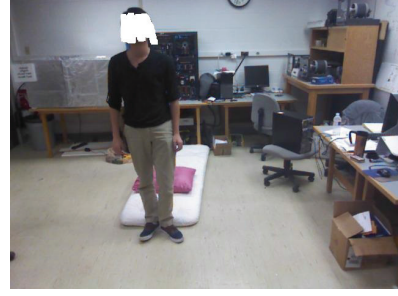


Fig. 1: Experimental scenario.

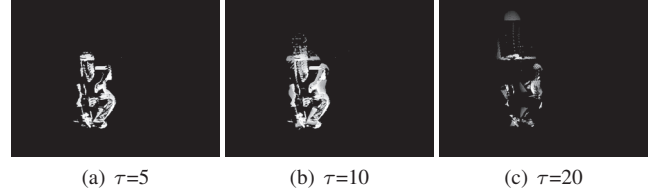


Fig. 2: Dependence on τ in the yielding MHI.

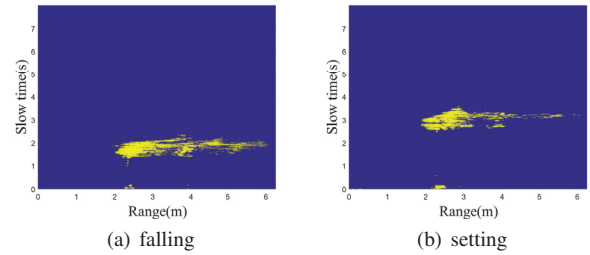


Fig. 3: Range-slow time plots of radar data corresponding to falling and sitting.

$\mathbf{S} = [\mathbf{s}_1 \dots \mathbf{s}_k \dots \mathbf{s}_K]$. A motion filter is first applied to obtain the target change detection among the data. The filtered radar signal is expressed as:

$$\mathbf{r}_k = [s_{k,2} - s_{k,1}, \dots, s_{k,L} - s_{k,L-1}]^T. \quad (4)$$

As such, we rebuild the filtered radar signal matrix as $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_K]$ with $L-1$ rows and K columns. Figure 3 shows the filtered result for falling and sitting.

It is still difficult to clearly classify falling versus sitting from the filtered results as those depicted in Figure 3. In order to enhance the contrast between these motions, a threshold is set to discard the values below it. The radar image is then converted to a 2-D logical matrix with its (k, l) th element expressed as

$$R_{k,l} = \begin{cases} 0, & R_{k,l} \leq Th_k, \\ 1, & \text{otherwise,} \end{cases} \quad (5)$$

for $k = 1, 2, \dots, K$ and $l = 1, 2, \dots, L-1$, where $Th_k = \sqrt{\frac{1}{L-1} \sum_{l=1}^{L-1} R_{k,l}^2}$ is the quadratic mean of each row, and $R_{k,l}$ is the (k, l) th element of matrix \mathbf{R} . This process eliminates the influence of low reflective body scatterers which may contaminate the received signals, and forms a new radar signal matrix $\hat{\mathbf{R}}$.

Signal processing is carried out based on this filtered matrix aiming at detecting moving targets and identifying the specific range between the targets and the radar. To reduce the dimension of the radar data matrix while preserving the motion characteristics, the PCA is used for dimension reduction. PCA performs an orthogonal transformation to convert radar signal $\tilde{\mathbf{R}}$ to a new coordinate system that consists of linearly uncorrelated variables. The new variables are referred to as the principal components. By choosing the first M principal components, we can reduce data dimension from $(L - 1) \times K$ of $\tilde{\mathbf{R}}$ to $M \times K$ of $\mathbf{\bar{R}}$ while most of the information in $\tilde{\mathbf{R}}$ is preserved. The singular-value decomposition of matrix $\tilde{\mathbf{R}}$ is given by

$$\tilde{\mathbf{R}} = \mathbf{A}\mathbf{\Lambda}\mathbf{B}, \quad (6)$$

where \mathbf{A} is the $M \times M$ left-singular matrix, $\mathbf{\Lambda}$ is an $M \times K$ rectangular diagonal matrix containing the singular values, and \mathbf{B} is the $K \times K$ right-singular matrix. The number of the principal components, M , is determined according to their total cumulative eigenvalues as the smallest integer that satisfies the following criterion,

$$\frac{\sum_{k=1}^M \lambda_k}{\sum_{k=1}^K \lambda_k} \geq 85\%, \quad (7)$$

where λ_k is the k th eigenvalue of covariance matrix $\tilde{\mathbf{R}}\tilde{\mathbf{R}}^H$, and $(\cdot)^H$ denotes conjugate transpose. In this paper, we select the value of M to be 60 to satisfy the above criterion.

3.4. K-means Clustering

Clustering is an important procedure for data mining, which plays a momentous role for extracting meaningful information from data sources. The k-means clustering is a simple learning algorithm for clustering. The feature vectors from video and radar data corresponding to all the motion classes are partitioned into c clusters $S = S_1, S_2, \dots, S_c$ by the k-means clustering algorithm. Figure 4 shows the output radar based scatter plots of falling and sitting motions. These figures indicate that the features for these two types of motions are much better distinguished as compared with the radar imaging depicted in Figure 3.

3.5. Hidden Markov Model

HMMs, which use observable variables to learn the way of objectives, are known for their effectiveness in temporal pattern recognition. An HMM describes stochastic sequences as Markov chains where the states are related to a probability function. Consider an N -state HMM described as

$$\lambda = \{X, Y, \pi\}, \quad (8)$$

where X is the probability of transferring to another state q at next time $t + 1$ given the current state p at current time t , and Y is the probability of being observing symbol at state q . In the proposed approach, two HMM models are respectively designated as falling and sitting models. Testing sequence ϑ is classified in model $\lambda_{\hat{i}}$ with $\hat{i} = 1$ denoting falling and $\hat{i} = 2$ denoting sitting. That is,

$$\hat{i} = \arg \max P(\vartheta|\lambda_{\hat{i}}), \quad (9)$$

where $P(\vartheta|\lambda_{\hat{i}})$ is the likelihood probability, which implies that the HMM-based classification is based on the maximum probability.

The frame frequency of Kinect sensor is 30 frames/second and the duration for temporal extent in an MHI is 10 frames. As a result, 1/3 of the time duration is used to apply HMM.

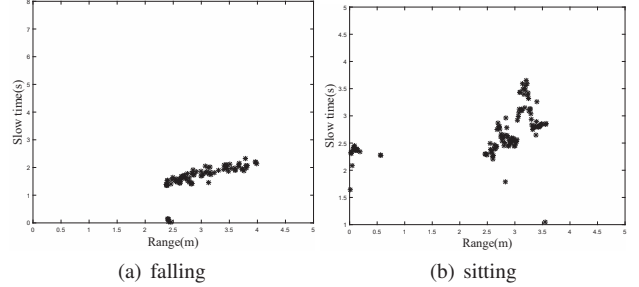


Fig. 4: Scatter plots of radar falling and sitting.

Table 1: Selection of training and testing data

Training		Testing	
96		60	
falling	sitting	falling	sitting
49	47	27	33

Table 2: Confusion matrix of video based approach

Activities	falling	sitting
falling	26	1
sitting	9	25

Table 3: Confusion matrix of radar based approach

Activities	falling	sitting
falling	27	0
sitting	4	29

4. EXPERIMENTAL RESULTS

Experiments as described in Section 2 are performed for data collection in order to verify the effectiveness of the proposed approaches. A Microsoft Kinect sensor is used to record the RGB video images, whereas a Time Domain UWB radar is used to collect radar reflections. Data collection using the Kinect sensor and the radar is performed simultaneously and synchronously. We produce the MHI with τ value equals to 10. In the series of experiments, the subjects fall towards the Kinect camera and the radar. There are seven subjects for fallings and six subjects for sittings. Specifically, 49 falling and 47 sitting are used as training data. These data are utilized to build the two types of motion models. In order to evaluate the accuracy of the built models, we used the other 27 falling and 33 sitting motions to be tested in each of the trained HMMs. The motion activities are listed in Table 1.

Table 2 shows the result for video-based classification. 26 in 27 falling motions are correctly classified and 9 sitting motions are misclassified. Video-based recognition rate is 96.30% for falling detection and 75.76% for sitting detection. Table 3 shows the results obtained from the radar data that all the falling motions are classified correctly, and 29 in 33 sitting motions are successfully recognized. The radar-based recognition rate is 100% for falling and 87.88% for sitting. The results imply that the radar based approach give us better recognition especially for sitting motions.

In order to demonstrate the confidence of the classification models, cross validation has also been implemented for training and test-

Table 4: Cross validation results for video-based approach

		Fall	Sit	Fall Recognition Rate	Sit Recognition Rate
Group 1	Fall	9	1	90.00%	85.71%
	Sit	1	6		
Group 2	Fall	6	1	85.71%	87.50%
	Sit	1	7		
Group 3	Fall	6	0	100.00%	88.89%
	Sit	1	8		
Group 4	Fall	7	0	100.00%	87.50%
	Sit	1	7		
Group 5	Fall	5	0	100.00%	81.82%
	Sit	2	9		
Group 6	Fall	6	1	85.71%	75.00%
	Sit	2	6		
Group 7	Fall	4	1	80.00%	70.00%
	Sit	3	7		
Group 8	Fall	8	0	100.00%	87.50%
	Sit	1	7		
Group 9	Fall	10	2	83.33%	80.00%
	Sit	1	4		
Group 10	Fall	9	2	81.81%	100.00%
	Sit	0	6		
Average Accuracy				90.66%	84.39%

Table 5: Cross validation results for radar-based approach

		Fall	Sit	Fall Recognition Rate	Sit Recognition Rate
Group 1	Fall	10	0	100.00%	85.71%
	Sit	1	6		
Group 2	Fall	6	1	85.71%	100.00%
	Sit	0	8		
Group 3	Fall	6	0	100.00%	88.89%
	Sit	1	8		
Group 4	Fall	7	0	100.00%	75.00%
	Sit	2	6		
Group 5	Fall	5	0	100.00%	90.91%
	Sit	1	10		
Group 6	Fall	6	1	85.71%	87.50%
	Sit	1	7		
Group 7	Fall	5	0	100.00%	80.00%
	Sit	2	8		
Group 8	Fall	8	0	100.00%	100.00%
	Sit	0	8		
Group 9	Fall	10	2	83.33%	80.00%
	Sit	1	4		
Group 10	Fall	11	0	100.00%	100.00%
	Sit	0	6		
Average Accuracy				95.48%	88.80%

ing. All data have been randomly grouped into ten folds. We run 10 separate learning experiments in total to evaluate the recognition accuracy. For each experiment, we use 9 folds for training and the remaining one for testing.

Table 4 provides the confusion matrix and recognition rates for all 10 cross-validation experiments in video based approach. The classification accuracy is estimated as the average which are 90.66% and 84.39%, respectively. The average rates show that, compared with results in Table 2, the fall motion recognition accuracy drops from 96.30% to 90.66%, and the sit motion recognition rate increases to 84.39%.

Comparing with Table 3, which shows that the recognition rate for falling is as high as 100%, the average accuracy given by Table 5 shows a more convincing result at 95.48%. Meanwhile, the accuracy for classifying sit motion has improved from 87.88% to 88.80%.

5. CONCLUSION

In this paper, we have investigated the classification and recognition of human motions separately using video and UWB radar based sensing modalities. We proposed the MHI and Hu moment method to extract features of RGB images. For radar data, we applied motion filtering and PCA to reduce the data dimension and extract the features. The k-means clustering algorithm is utilized for vector quantization. Two HMMs for falling and sitting motions are trained for vision based and radar based data, respectively. From the classification results, we observed that the radar based method achieves higher classification performance with recognition rate of 95.48% in falling and 88.80% in sitting. In future work, fusion on these two sensing modalities will be investigated to further improve the recognition rate.

6. REFERENCES

- [1] S. Sadigh, A. Reimers, R. Andersson, and L. Laflamme, "Falls and fall-related injuries among the elderly: A survey of residential-care facilities in a swedish municipality," *Journal of Community Health*, vol. 29, no. 2, pp. 129–140, 2004.
- [2] R. Igual, C. Medrano, and I. Plaza, "Challenges, issues and trends in fall detection systems," *Biomedical Engineering Online*, vol. 12, no. 66, pp. 1–24, 2013.
- [3] F. Bagala, C. Becker, A. Cappello, L. Chiari, K. Aminian, J. M. Hausdorff, W. Zijlstra, and J. Klenk, "Evaluation of accelerometer-based fall detection algorithms on real-world falls," *PLoS ONE*, vol. 7, no. 5, pp. 1–9, 2012.
- [4] L. Liu, M. Popescu, K. Ho, M. Skubic, and M. Rantz, "Doppler radar sensor positioning in a fall detection system," in *Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (San Diego, CA), pp. 256–259, 2012.
- [5] M. Wu, X. Dai, Y. D. Zhang, B. Davidson, M. G. Amin, and J. Zhang, "Fall detection based on sequential modeling of radar signal time-frequency features," in *Proceedings of IEEE International Conference on Healthcare Informatics (ICHI)*, (Philadelphia, PA), pp. 169–174, 2013.
- [6] Q. Wu, Y. D. Zhang, W. Tao, and M. G. Amin, "Radar-based fall detection based on Doppler time-frequency signatures for assisted living," *IET Radar, Sonar and Navigation*, vol. 9, no. 2, pp. 164–172, 2015.

- [7] T. R. Bennett, J. Wu, N. Kehtarnavaz, and R. Jafari, "Inertial measurement unit-based wearable computers for assisted living applications: A signal processing perspective," *IEEE Signal Processing Magazine*, vol. 33, pp. 28–35, March 2016.
- [8] M. G. Amin, Y. D. Zhang, F. Ahmad, and K. C. D. Ho, "Radar signal processing for elderly fall detection: The future for in-home monitoring," *IEEE Signal Processing Magazine*, vol. 33, no. 2, pp. 71–80, 2016.
- [9] X. Wang, M. Li, H. Ji, and Z. Gong, "A novel modeling approach to fall detection and experimental validation using motion capture system," in *Proceedings of IEEE International Conference on Robotics and Biomimetics (ROBIO)*, (Shenzhen, China), pp. 234–239, 2013.
- [10] Y. Li, K. Ho, and M. Popescu, "Efficient source separation algorithms for acoustic fall detection using a Microsoft Kinect," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 3, pp. 745–755, 2014.
- [11] S. Kianoush, S. Savazzi, F. Vicentini, V. Rampa, and M. Giusani, "Leveraging RF signals for human sensing: fall detection and localization in human-machine shared workspaces," in *Proceedings of IEEE International Conference on Industrial Informatics (INDIN)*, (Cambridge, U.K.), pp. 1456–1462, 2015.
- [12] A. Martone, K. Ranney, and R. Innocenti, "Automatic through the wall detection of moving targets using low-frequency ultra-wideband radar," in *Proceedings of IEEE Radar Conference*, (Arlington, VA), pp. 39–43, 2010.
- [13] J. Hao, X. Dai, A. Stroder, J. J. Zhang, B. Davidson, M. Mahoor, and N. McClure, "Prediction of a bed-exit motion: Multimodal sensing approach and incorporation of biomechanical knowledge," in *Proceedings of Asilomar Conference on Signals, Systems and Computers*, (Pacific Grove, CA), pp. 1747–1751, 2014.
- [14] M. A. R. Ahad, J. Tan, H. Kim, and S. Ishikawa, "Human activity analysis: concentrating on motion history image and its variants," in *Proceedings of ICCAS-SICE Joint Conference*, (Fukuoka, Japan), pp. 5401–5406, 2009.
- [15] Z. Huang and J. Leng, "Analysis of Hu's moment invariants on images scaling and rotation," in *Proceedings of International Conference on Computer Engineering and Technology*, (Chengdu, China), pp. 476–480, 2010.
- [16] M. K. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. IT-8, pp. 179–187, 1962.
- [17] A. Yarovoy, L. Ligthart, J. Matuzas, and B. Levitas, "UWB radar for human being detection," *IEEE Aerospace and Electronic Systems Magazine*, vol. 21, no. 3, pp. 10–14, 2006.