33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Nighttime In-Home Action Monitoring For Eldercare

Zhongna Zhou, Erik Edward Stone, Marjorie Skubic, James Keller and Zhihai He

Abstract— **In this work, we develop a system to automatically monitoring actions of elderly people at home for safety enhancement and health monitoring. We use an Infrared camera embedded in a living environment to capture images. We study the characteristics of different clothing in Infrared images and develop an efficient silhouette extraction method for Infrared (IR) images using spatio-temporal filtering. We recognize human action using supervised learning methods. Our experimental results demonstrate that our proposed method is efficient.**

I. Introduction

Aging population shows a dramatic increasing during the past decade. The aging population explosion has become a social problem in both United States and many other countries. Independent lifestyles are highly desired by elderly people. However, independent lifestyles of older adults often come with high risks. To assist elderly persons' independent living, many smart home technologies have been developed to track and monitor activities of elderly persons at home with various sensors. Specifically, the video sensor is able to analyze human actions in visual scenes, which has significant advantages in providing rich contextual information of human activities. However, the application of such approach may usually be hindered by its dynamic usage environment in which the scenario varies over time [1]. The visual surveillance system, which is able to automatically detect anomalies at various situations and warn operators of dangerous activities, is extremely helpful to eldercare in providing real time video-based activity monitoring and functional assessment. Conventional surveillance systems consider using rule-based method to detect predefined dangerous activities [2, 4]. These approaches perform well in certain tasks such as fall detection [5], sleeping posture classification. Other methods consider action recognition and anomaly detection as supervised learning problems in which anomalies are treated as outlier to previous trained classifier of normal action patterns [3]. In this paper, we propose to develop a visual sensing system for nighttime action monitoring of elderly people at home(Fig.1). We capture images using Infrared cameras. We develop spatiotemporal filters for noise removal in IR images during very low-lighting conditions. We develop supervised learning methods to recognize their action patterns. The supervised learning algorithm allow the action model to be adapt to dynamic usage environments, thus achieve more robust and accurate results.

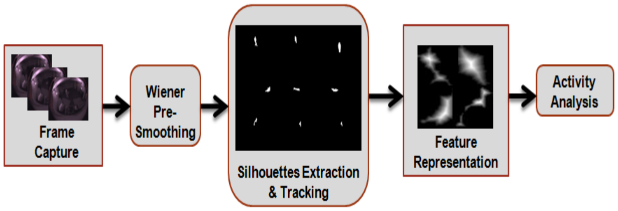

Fig. 1. Overview of our proposed system.

II. Data Collection and Silhouette Extraction

A. Data Collection

For algorithm development and performance evaluation, we have established a video database of night-time in-home action monitoring. Each action sample was captured using two fisheye Uni-brain cameras at a resolution of 640×480. As shown in Fig. 2, these two cameras have a large overlapped view, providing coverage of a bed, a couch, two chairs and two floor mats.

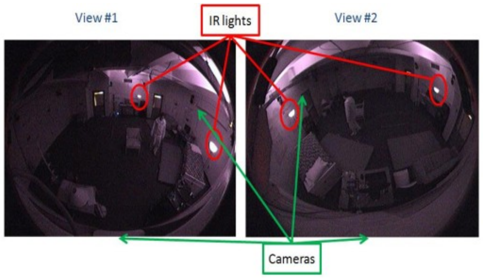

Fig. 2. Two sample images of our two camera views and their positions, red eclipses indicate the positions of the IR lamps, emitting IR lights which can't be seen by our eyes.

The cameras are embedded in two floor lamps, as shown in Fig. 3. The wavelength of the IR emitters we used is 850nm. In total, there are 216 individual infrared LEDs distributed in these two lamps and the total power draw is approximately 20 Watts. Each camera has a 180 degree horizontal field of view, and 131 degree vertical field of view. The fisheye lenses are used since they could gain more

Authors Z. Zhou, E. Stone, M. Skubic, J. Keller and Z. He are with the Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO. Corresponding author : Zhihai He (email: HeZhi@missouri.edu).

978-1-4244-4122-8/11/$26.00 ©2011 IEEE

IR lights than other lenses, thus could provide better picture quality with enhanced picture brightness. The IR emitters in two lamps are facing the ceiling so that the shadow is minimal according to our experiment. Fig. 2 shows hese two camera views, their positions, and the locations of the IR lamps.

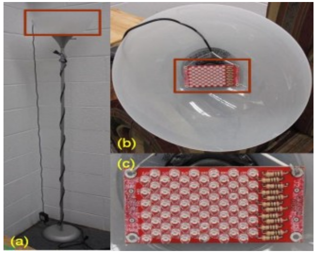

Fig. 3. The appearance of the IR lights deployed in the lamp. (a) The front view of IR lamp; (b) The top view of the IR lights with lamp;(c) The IR lights. The view of b, c are the red rectangular part showed in a, b respectively.

We recruited 6 volunteers to perform different actions in very low lighting conditions: walking, standing with hand motion, standing without hand motion, sitting down and stand up, sitting on a sofa, go to bed, getting up from bed, lying on bed, sleepless (with some body motion), sitting on bed, and lying down to bed. In addition, four abnormal actions were included: walking in the room and falling onto ground due to loss of balance, slipping off when trying to get up from a chair, falling when trying to get up from a bed, falling out of the bed when sleeping. In this paper, we focus on seven major actions: lying on bed, standing, walking, sitting, sitting on the bed, falling, falling from bed.

B. *The Spectral Characteristics of IR Video Clips*

In visible light conditions, the camera views are similar to what we see with eyes. But, in IR images, they are dramatically different. To develop an efficient image processing algorithm for IR images, we need to study the spectral characteristics of fibers and textiles of different human clothing in IR images. The IR spectral characteristics of different fibers and textiles are highly related to the fiber type, areal density, moisture regain and fabric construction [6]. Specifically, the spectral characteristics can be represented as:

$$\alpha_{\lambda} + \rho_{\lambda} + \tau_{\lambda} = 1 \tag{1}$$

in which α_{λ} is the spectral absorptivity, ρ_{λ} is the fabric reflectivity and τ_{λ} is the transmissivity. In our experiments, we observe that hydrophilic fabrics such as wool and cotton have higher spectral absorptivity and lower reflectivity than hydrophobic fabrics (like polyester). Particularly, polypropylene has the highest spectral reflectivity - when the fiber is dry, polypropylene fabric has the reflectivity as large as 60% - 70% in the near infrared region (NIR). The wool, on the other hand, has the lowest spectral reflectivity, which is close to zero throughout much of the spectrum. Moreover, when fibers are dry, spectral properties vary with fiber type. However, the IR characteristics of fabric may change with moisture regain, in which hydrophobic fibers are much more affected by moisture than hydrophilic fibers. Fig. 4 shows fiber samples in visible and IR lights, their fiber types are summarized in TABLE I.

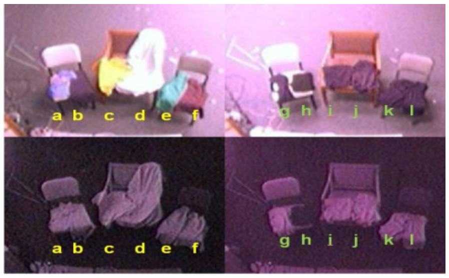

Fig. 4. The cloth color in visible and IR light condition; the images in the top row are collected in visible-light-condition; the images in the bottom row are collected in IR-light-condition; The fiber types are listed in TABLE I. The cloth h in IR light condition is dark because of its low areal density.

TABLE I
FIBER TYPES OF CLOTH IN FIG. 4

Label	Fiber Type
a - f	100% Cotton
g	55% Cotton + 25% Rayon +20% Polyester
h	63% Polyester + 33% Rayon +4% Spandex
i	100% Cotton
j	100% Polyester
k	80% Rayon + 20% Polyester
l	65% Polyester + 35% Rayon

C. *Silhouette Extraction*

Studies in field of object recognition in 2D video frames have demonstrated that silhouettes contain detailed information about human poses and shapes. Silhouette extraction is a background change detection technique whose accuracy depends on how well the background subtraction method performs. Usually, a further shadow removal is employed to ensure greater accuracy; and a binary morphological operation is used to fill up holes and remove noise from the extracted silhouettes. In our study, we are dealing with low resolution IR video clips. Compared to visible light images, the IR images are characterized by lower contrast and higher clutter noise (as shown in Fig. 5). As a result, silhouettes extracted from IR images are less accurate than those extracted from visible light images(as shown in Fig. 6). In this work, to reduce the clutter noise in IR images, we used smoothing filters. In our experiment, the median filter and Wiener filter are applied separately for performance comparisons. Fig. 7 shows two examples of IR images, the median and wiener filter filtered image results and their corresponding silhouettes. To evaluate the performance of pre-smooth filters, we manually labeled 80 frames as ground-truth. TABLE II summarizes the performance improvement of silhouette extraction using different smoothing filtering schemes. The detection rate is the number of correctly detected silhouette pixels over the total labeled silhouette pixels. It can be seen that with the similar detection accuracy, using filter achieves a significantly smaller error rate than the result not using filter; the performance of median and wiener filter are almost the same.

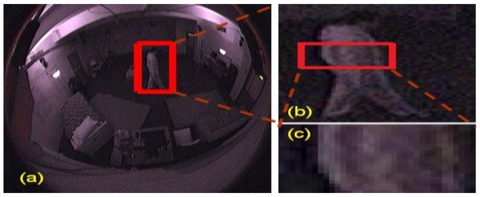

Fig. 5. One of the IR images; (a) The IR image; (b) The person in the IR image; (c) A specified part of the person.

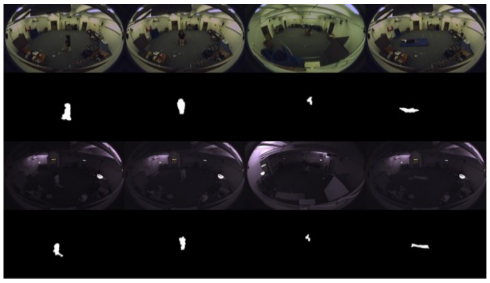

Fig. 6. Silhouette extraction results; the top 2 rows are visible-light- condition images and their corresponding silhouettes; the bottom 2 rows are IR-light images and their corresponding silhouettes; the action of each column from left to right are: walking, standing, sitting and laying on the ground respectively.

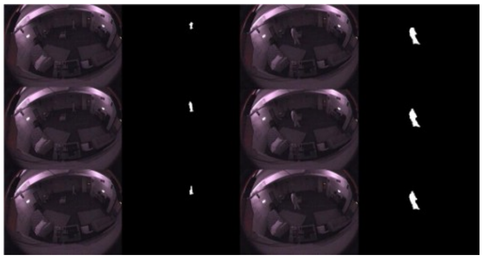

Fig. 7. Silhouette extraction results; the first rows are IR-light-condition images and their corresponding silhouettes; the bottom 2 rows respectively are median filtered and wiener filtered IR-light images and their corresponding silhouettes.

TABLE II
THE FALSE DETECTION RATE WITH ACCURACY

	FP/(TP+FP)		
TP/GT	no filter	median	wiener filter
69%	25%	10%	10%
73%	29%	12%	12%

To further improve the silhouette extraction accuracy, we perform a contour tracking process on the smoothed images. Specifically, we assume that the movement of silhouette centroid in a very short time (e.g. 0.3 second) is linear, thus we use Kalman filter to achieve optimal silhouette tracking. Our experiment results show that silhouette tracking helps improve the silhouette extraction quality. When a high-quality silhouette (with sufficiently details and good accuracy) is available, the human pose can be readily identified by calculating the similarity between silhouettes and standard human pose template. In this work, we use the Medial Axis Distance Transform (MADT) [8] as a shape descriptor for human silhouettes. It assigns each internal pixel of a silhouette with a value reflecting its minimum distance to the boundary contour.

D. Human Action Classification

Silhouettes are extracted and normalized to a common size, represented by MADT feature vectors, then action recognition were performed using supervised learning methods. The Support Vector Machines (SVMs) are state-of-the-art large margin classifiers which have recently gained popularity within computer vision and pattern recognition. Therefore, we use the SVM classifier for action classification. Given a set of training data vectors {$\mathbf{x_1}$, .., $\mathbf{x_n}$} in R^d space and their labels y_i (say, y_i equals to -1 for negative samples or 1 for positive samples), the SVM classifier will search for hyper-plane that achieves best accuracy in separating the two data classes while maximizing the margin between two classes. All data vectors lying on one side of the hyper-plane are labeled as -1 while vectors on the other side are labeled as 1. If the training set is linearly separable, then a separating hyper-plane can be defined by the following inequalities:

$$y_i(\mathbf{w}\cdot\mathbf{x_i}+b)\geq 1, \forall i \in \{1,\ldots,N\} \qquad (2)$$

in which $\mathbf{x_i}$ stands for the data vector, y_i is the class label, $\mathbf{w}$ represents the normal vector that perpendicular to the hyper-plane, b is the offset. Finding the optimal hyper-plane is equivalent to minimizing $\|\mathbf{w}\|^2$ under constraints in (2). The performance of a SVM classifier could be affected by several other factors such as the choice of kernel, scaling, and feature selection. While studying the non-linear SVM model is worthwhile, however, the linear SVM model is more computational efficient. In this study, we use linear kernel for one-versus-all action classification.

III. Results

An example image sequences showing the performance of our silhouette extraction and tracking algorithm are in Fig. 8. Note that when the person moves away from the cameras, it becomes more difficult to extract the silhouettes. Our algorithm is able to effectively segment the human from the background and track the human very well. Fig. 9 plots a moving route of the silhouette centroid, the red line is the Kalman filtered silhouette tracking results and the blue line is the result without tracking. The results with tracking are more smooth than that without tracking. Fig. 10 shows the result of silhouettes with distance transform, in which the silhouettes are rescaled to a normalized size.

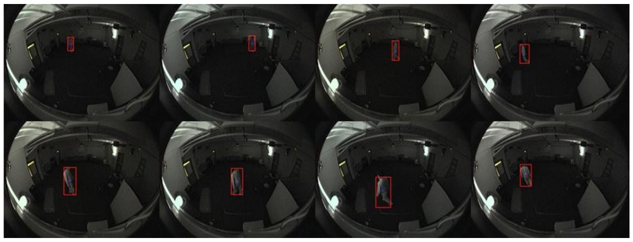

Fig. 8. Our proposed silhouette tracking results.

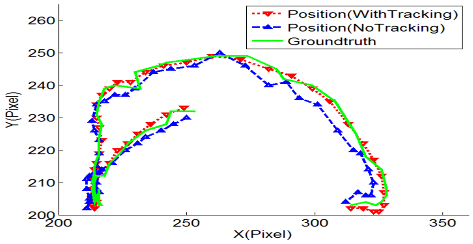

Fig. 9. The route of our proposed silhouette tracking results.

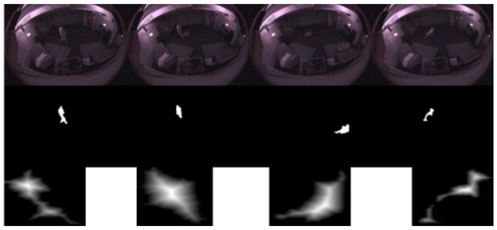

Fig. 10. Silhouette extraction results and their distance map; the images in each row from top to bottom are: original images, their corresponding silhouettes and distance map; The action of each column from left to right is: walking, standing, laying on the ground and sitting respectively.

TABLE III shows the binary classification results of seven types of action. Overall, the result is good. Since the proposed action classification method is based on the silhouettes' shape information (in the form of MADT shape descriptor in Section C), some errors may be introduced in due to the ambiguity between classes. For example, without additional context information about the action, it's difficult to tell some 'Bed-Falling' silhouette frames from 'Falling' silhouettes. In the future, in order to further improve the performance of silhouette extraction and get better classification results, the following aspects may be concerned: firstly, the interaction between human subject and other objects (such as bed, chair, etc.) may be taken into account. Secondly, in addition to the MADT shape descriptor, more features may be used to effectively represent the silhouettes.

IV. CONCLUSION

In this paper, we proposed a method for action monitoring under low light condition. We describe the development of the whole method which is able to efficiently extract and tracking human silhouettes. We then further represent the collected silhouette using distance transform features. And finally we applied SVM classification for action classify. Our experimental data demonstrate that the silhouettes are accurate and the proposed method is efficient.

REFERENCES

[1] K. Z. Haigh, L. M. Kiff, J. Myers and K. Krichbaum. The Independent LifeStyle AssistantTM (I.L.S.A.): Deployment Lessons Learned. The AAAI 2004 Workshop on Fielding Applications of AI, July 25, 2004, San Jose, CA. Pages 11-16.

[2] D. Anderson, R. Luke, J. Keller, M. Skubic, M. Rantz, and M. Aud. Linguistic summarization of video for fall detection using voxel person and fuzzy logic. CVIU, 113(1):80–89, 2009.

[3] F. Lv and R. Nevatia. Single view human action recognition using key pose matching and viterbi path searching. In Proc. CVPR, 2007.

[4] A. Nasution and S. Emmanuel. Intelligent video surveillance for monitoring elderly in home environments. In IEEE Workshop on Multimedia Signal Processing, 2007.

[5] N. Noury, A. Fleury, P. Rumeau, A. Bourke, G. Laighin, V. Rialle, and J. Lundy. Fall detection-principles and methods. In EMBC07.

[6] E.G. McFarland, W.W. Carr, D.S. Sarma and J.L. Dorrity. Effects of Moisture and Fiber Type on Infrared Absorption of Fabrics. Textile Research Journal 1999 69: 607

[7] Z. Zhou, W. Dai, J. Eggert, J.T. Giger, J. Keller, M. Rantz and Z. He, "A Real-time System for In-home Activity Monitoring of Elders," Proceedings, 31st Annual International Conference of the IEEE EMBC, Minneapolis, Minnesota, September 2-6, 2009, pp 6115-6118.

[8] H. Blum, "A Transformation for Extracting New Descriptors of Shape," Models for the Perception of Speech and Visual Form, Proc. Symp., pp. 362-380, 1967.

TABLE III
THE CONFUSION MATRIX FOR HUMAN ACTION RECOGNITION

	Bed Fall	Bed Lay	Bed Sit	Fall	Sit	Stand	Walk
Bed Fall	90.09%	0.47%	0.31%	9.12%	0%	0%	0%
Bed Lay	0.47%	92.77%	0%	6.60%	0%	0.16%	0%
Bed Sit	0.29%	0%	97.70%	1.87%	0%	0%	0.14%
Fall	8.58%	6.21%	1.92%	83.28%	0%	0%	0%
Sit	0%	0%	0%	0%	99.53%	0.16%	0.31%
Stand	0%	0.18%	0%	0%	0.18%	94.06%	5.57%
Walk	0%	0%	0.14%	0%	0.27%	4.27%	95.32%