

Human Daily Activity Recognition Using Ceiling Mounted PIR Sensors

Xiaomu Luo^{1, a}, Tong Liu², Baihua Shen³, Jiaming Hong¹,
Qinqun Chen¹ and Hao Chen¹

¹School of Medical Information Engineering, Guangzhou University of Chinese Medicine,
Guangzhou, 510006, China

²Department of Electronic Science, Huizhou University, Huizhou, 516007, China

³School of Information Engineering, GuangDong University of Technology, Guangzhou, 510006,
China

^aemail: woodwood2000@163.com

Keywords: Pyroelectric Infrared (PIR) Sensors; Ambient Assisted Living (AAL); Wireless Sensor Networks (WSNs)

Abstract. Human daily activity recognition is the foundation of automatic Ambient Assisted Living (AAL) system. In the paper, we propose a sensing model which can capture the discriminative spatio-temporal feature of human motion in an efficient way. The object space is separated into distinct discrete sampling cells by reference structure, and the ceiling mounted Pyroelectric infrared (PIR) sensors are used to capture the time-varying signal induced by human motion. The GMM-HMM model is utilized to classify different human activities. We use a self-developed PIR sensor node mounted on the ceiling to conduct experiments in a real office environment. Promising experimental results confirm the validity of our model.

Introduction

Human activity recognition is an important research topic because it is the foundation of many applications, including automatic surveillance, human-computer interaction (HCI), ambient assisted living (AAL), and so on. The traditional vision-based action recognition paradigm has been widely used as the price of the cameras goes down and the techniques to process the video data have been well studied [1]. In general, computer vision-based processing includes background subtraction, motion segmentation, feature extraction [2]. Some representations of the feature, such as the geometric model of the human body, spatio-temporal patterns, appearance, contours and optical flow, have been demonstrated to be effective methods [3]. However, although numerous researchers have devoted to improve the performance of video-based action recognition, it is inevitably associated with some limitations. Firstly, they are susceptible to the change of light illumination and the variance of the background. Moreover, they are inflexible to be used in terms of sensing efficiency, as the high dimensional continuous data stream incurs intense computation burden. Hence, their scalable deployment is restricted, especially in the communication and computation constrained environment, i.e. wireless sensor networks (WSN).

To provide a supplement and alternative to traditional vision based approaches, some researchers utilized pyroelectric infrared (PIR) sensors for human activity recognition. PIR sensors belong to the thermal sensors, which only response to the incidence thermal flux. Hence, its output is only subject to the human motion within its Field of View (FOV), not affected by the illumination and background. In [4], Guan *et al.* represented human motion as a spatio-temporal energy sequence, and employed PIR sensor array with visibility mask to capture the energy variance. The nearest neighbor classifier was used to classify ten aerobic exercises performed by six persons. In [5], Liu *et al.* presents a distributed direction-sensitive infrared approach for fall detection. Three modules of PIR sensor were used to capture the synergistic motion patterns of head, upper-limb and lower-limb. When the falls happened, the discriminative feature of the human motion will be encoded. In [6], they further used pseudo-random-coded Fresnel lenses to acquire and compress the human motion, and hausdorff distance-based recognition method is employed to distinguish

different human actions.

Aforementioned studies have shown that the PIR sensors with the visibility modulation are competent to identify and classify human motions with high sensing efficiency. However, there are some limitation in practical application. Because their PIR sensors are side-view deployed, the performance of these systems relied on the assumption that the human motion occurs in a predefined region. For example, in [4], the authors assume that the distance between the PIR sensor array and the human body is 150 cm. This assumption is also held in [5][6]. Because the output amplitude of the PIR sensor is inversely proportional to the square of the distance between the human target and the sensors, the uncertainty of the distance in the real environment will definitely influence the performance of the system.

In this paper, we present a novel scheme for the indoor human activity recognition. Based on the optical flow analysis, we design the sensing model which can capture the distinctive feature of human motion. Two degrees of freedom of the reference structure segment the object space into distinct sampling cells. As the human performs different actions and crosses the boundaries of these sampling cells, the PIR sensors will capture the change of the thermal energy within each sampling cell and encode into low level data stream. Short Time Energy (STE) and the Gaussian Mixture Model and Hidden Markov Model (GMM-HMM) are used to model the data stream to achieve human activity classification. The ceiling mounted deployment of the PIR sensors will be flexible to cover large indoor areas. Experiments in real office environment are conducted to validate the proposed method.

Feature Analysis

To achieve human motion recognition, it is necessary to analyze the spatio-temporal feature of different activities. Based on the analysis, we design the corresponding sensing model.

Optical flow is the distribution of apparent velocities of movement of brightness patterns in an image [7]. The optical flow method can be employed to refine the analysis of the spatio-temporal feature as the human performs different activities. These images are taken from the video which the sampling rate is 25 frames/s . For each category of activity, we only select three frames and their corresponding optical flow vector images for visualization.

We divide the monitored region into four sub-regions; further, we aggregate the horizontal component of the flow velocity within these subregions separately, denote as horizontal motion energy (*HME*).

The *HME* reflects the horizontal component of the human motion that crosses the sub-region perpendicularly, and will be used as the cue to analyze the spatio-temporal feature of different activities. In [8], Wu found two velocity features of the fall: 1) the magnitude of both vertical and horizontal velocities of the trunk will increase dramatically during the falling phase, reaching up to 2-3 times that of any other controlled movement; 2) the increase of the vertical and horizontal velocities usually occur simultaneously, at about 300-400 ms before the end of the fall process, which are strongly dissimilar with the controlled human activities.

Based on the above analysis, the time-varying *HME* of each sub-region is the discriminative spatio-temporal feature which can be used to distinguish different activities. Hence, the feature-specific system should: 1) segment the monitored region into sub-regions and 2) use sensors to collect the energy variation of each sub-region. This is the inspiration of our sensing model design which will be elaborated in the succeeding sections.

Sensing Model.

To capture the discriminative spatio-temporal feature of different actions, the *HME* of each sub-region, the sensing model has to be designed correspondingly. Our model springs from the reference structure tomography (RST) paradigm, which uses multidimensional modulations to encode mappings between radiating objects and measurements [9].

The *object space* refers to the space where the thermal source moves. The *measurement space* refers to the space where the PIR sensors are placed. The *reference structure* which specifies the mapping from the object space to measurement space [9], is used to modulate the FOV of each PIR.

More specifically, with a Fresnel lens array around, each PIR sensor can detect sensitively the thermal fluctuation induced by human motion within its field of view (FOV). Each PIR sensor can be considered as a single pixel. The opaque mask covers part of the surface of the Fresnel lens array, acting as the geometric reference structure[10]. As a result, the PIR sensor could only sense part of its original FOV. Several PIR sensors with their own masks are multiplexing in one sensor node to modulate the visibility pattern of the object space. By this method, the three-dimensional (3D) object space is segmented into a lot of sampling cells, as contiguous points in each cell have the same unique visibility signature. The human target can be modeled by a vertical cylinder thermal source approximately. As the human body passes the boundaries of the sampling cells, PIR sensors will generate output with different characteristics corresponding to different human activities.

Assume there are M sensors in the measurement space, and the object space is divided into L discrete non-overlapping sampling cells. The measure vector of these M sensors at time instant t can be written in matrix form as

$$\mathbf{m}(t) = \mathbf{V}\mathbf{s}(t) \quad (1)$$

where $\mathbf{m}(t) = [m_j(t)] \in \mathbb{R}^{M \times 1}$ is the measurement vector, $\mathbf{V} = [v_{ji}] \in \{0, 1\}^{M \times L}$ is the measure matrix determined by the visibility modulation scheme, and $\mathbf{s}(t) = [s_i(t)] \in \mathbb{R}^{L \times 1}$ is the radiation state vector of the sampling cells.

In our sensing model design, each sampling cell corresponds to each sub-region of the monitored region. Thus, the spatial information is encoded into the measurement of the sensors by the visibility modulation. Further, the PIR sensors that are differential in nature are employed to transform the incident radiation into voltage measurements. Because the PIR sensors are ceiling mounted and sensitive to the human motion tangential to them, they provide the temporal information about the horizontal component of the human motion.

As a summary, through visibility modulation and employing PIR sensors, our sensing model could be used to capture the distinctive spation-temporal feature of different human activities.

Human Motion Classifier.

The short time energy (STE) has been proved to be effective in depicting the energy variation of the sine-like waveform[12], we employ it to represent the energy variation of the PIR signals. The STE of the n th frame of the i th PIR is defined as

$$p_i(n) = \sum_{m=0}^{K_n-1} |a_i(m) - avSTE_i(n)| \quad (1)$$

$$\text{with } avSTE_i(n) = \frac{1}{K_n} \sum_{m=0}^{K_n-1} a_i(m) \quad (2)$$

where $i \in \{1, \dots, L\}$ is the index of the PIR sensor, K_n is the total number of the sampling points in the n th frame, $avSTE_i(n)$ is the average energy of the sampling points, and $a_i(m)$ is the signal amplitude of the m th sampling point.

In our system, the sampling rate of the PIR signal is 25 Hz. The STE is calculated based on 2-second window with 1-second overlapping; it means that $K_n = 50$ in our implementation.

GMM-HMM Classifier.

The design of the classifier is problem-specific. Gaussian Mixture Model and Hidden Markov Model (GMM-HMM) have been demonstrated as a powerful tool for modeling timevarying sequence data, such as speech [13] and video stream [14]. In our system, we employ GMM-HMM for the action classification.

The parameters of a GMM-HMM can be denoted compactly by N , M_G and $\lambda = (\mathbf{A}, \mathbf{B}, \mathbf{\Pi})$ for convenience. N is the number of hidden states and M_G is the number of Gaussian models. $\mathbf{A} = \{a_{ij}\}$ represents the hidden state transition probabilities matrix

$$a_{ij} = P(q_{t+1} = j | q_t = i), 1 \leq i, j \leq N$$

where q_t is the hidden state at time t , i and j are the labels of the hidden state. The transition probability distribution should satisfy the constrains of $a_{ij} \geq 0$ and $\sum_{j=1}^N a_{ij} = 1$. $\mathbf{B} = \{b_i(\mathbf{s}(n))\}$

denotes the probability density distribution of the observation vector

$$b_i(\mathbf{s}(n)) = P(\mathbf{s}(n)|q_t = i) = \sum_{m=1}^{M_G} c_{im} \mathcal{N}(\mu_{im}, \Sigma_{im})$$

where $\mathbf{s}(n) = [s_1(n), s_2(n), \dots, s_7(n)]^T$ is the observation vector, $s_i(n)$ is defined in Eqn. (8), c_{im} is the weighting coefficient for the m th mixture when the hidden state $q_t = i$. $\mathcal{N}(\mu_{im}, \Sigma_{im})$ is the Gaussian probability density function with mean vector μ_{im} and covariance matrix Σ_{im} . The weighting coefficients should satisfy the constraints $c_{im} \geq 0$ and $\sum_{m=1}^{M_G} c_{im} = 1$. It is the representation of how the Gaussian probability density functions are combined to represent the observed vector according to the hidden state. $\boldsymbol{\Pi} = \{\pi_i\}$ is the initial state probability vector

$$\pi_i = P\{q_0 = i\}$$

and satisfy the constraints $\pi_i \geq 0$ and $\sum_{i=1}^N \pi_i = 1$.

The parameters $\lambda = (\mathbf{A}, \mathbf{B}, \boldsymbol{\Pi})$ are learned from training data using Baum-Welch method [13]. This is done for each class separately. Based on the Bayesian rule, given the sequence $\mathbf{P} = [p_i(n)]$, the likelihood output $p(\lambda_i|\mathbf{P})$ is proportional to $p(\mathbf{P}|\lambda_i)$. That is, label the input sequence \mathbf{P} to the HMM with the highest likelihood,

$$i^* = \arg \max_i p(\mathbf{P}|\lambda_i)$$

where i corresponds to different activities.

Experimental Results.

In our experiment, the sensor node will be mounted on the ceiling $3m$ above the floor. Our sensor node mainly consists of Fresnel lens arrays, PIR detectors, Masks and CC2430 module. The CC2430 module combines the RF transceiver with an industry-standard enhanced 8051 MCU, 128 KB flash memory, 8KB RAM. After configure the CC2430, the data generated by the sensor node will be sent to the sink based on the 2.4G Hz IEEE 802.15.4 (ZigBee) protocol, and then the sink will transport the data to the PC host by RS232 serial port for data processing.

The experiments were carried out in an office environment. The monitored region covered by the sensor node was a cone with $3m$ radius. There were totally eight volunteers participated in our experiments, including three females and five males. The height of their ranges from $1.64m$ to $1.80m$, and the weight of them ranges from $50kg$ to $70kg$. Each volunteer emulated five kinds of activities, including fall, sitting down, standing up from a chair, walking and jogging. Every activity was emulated ten times by each volunteer at a self-select speed and strategy. Totally, we obtained 400 samples, 80 samples per each category.

The 10-fold cross-validations strategy is used to evaluate the performance of our system. The experiments were divided into two stages: the training and testing stage. In the training stage, 90% of the total samples were randomly selected to train the parameters $\lambda_i = (\mathbf{A}, \mathbf{B}, \boldsymbol{\Pi})$ for each GMM-HMM. The number of hidden states N and the number of Gaussian models M_G have to be specify manually selected before employ the Baum-Welch method (equivalently the EM method) [13]. We specify the N and M_G for different actions, as shown in Table I.

Table I The specification of GMM-HMM.

| | Fall | Sit down | Stand up | Walk | Run |
|-------------------------|------|----------|----------|------|-----|
| # Gaussian Models M_G | 2 | 2 | 2 | 2 | 2 |
| # Hidden State N | 16 | 14 | 16 | 14 | 16 |

The confusion matrix of our experiment result is shown in Fig. 1. Although there are a few false classification, the performance of our system is encouraging. These false classification mainly lie in “walk” and “run”. The reason is that both “walk” and “run” are horizontal activities, which cross the entire monitored region covered by the sensor node. Some “fast walk” samples are similar to “slow run” samples, or vice versa, because the volunteers who participated in our experiments vary in height and weight. The high similarity of the motion performs by different person could contribute to this confusion.

| | Fall | Sit down | Stand up | Walk | Run |
|----------|------|----------|----------|------|------|
| Fall | 0.85 | 0.01 | 0.00 | 0.11 | 0.03 |
| Sit down | 0.03 | 0.97 | 0.00 | 0.00 | 0.00 |
| Stand up | 0.00 | 0.00 | 0.99 | 0.01 | 0.00 |
| Walk | 0.01 | 0.00 | 0.00 | 0.71 | 0.29 |
| Run | 0.01 | 0.00 | 0.00 | 0.21 | 0.79 |

Fig. 1. Confusion Matrix

Conclusion

In this paper, we present a compressive sensing model and its implementation for indoor human activity recognition. The sensing model is designed based on the optical flow analysis, and implemented by the reference structure. The discriminative spatio-temporal feature of the human motion is captured by the thermal differential PIR sensors. The GMM-HMM model is employed to classify different activities. The prominent advantage of the proposed approach is the low-dimensional data stream, robustness to the light illumination and do not need background subtraction. Encouraging experimental results show the efficacy of our sensing model. In the future, more sophisticated algorithms are planned to refine the activity classification results. Construction of large scale WSN, the fusion of different sensor nodes, will promote the performance of our system.

Acknowledgement

This work is supported by the National Natural Science Foundation of China (NSFC) under Grant No. 61301294 and No. 61401174, the Natural Science Foundation of Guangdong under Grant No. 2014A030310462, the Science and Technology Planning Project of Guangdong Province, China (Grant No.2011B031300015), and Youth Elite Project of GUCM.

References

- [1] P. Turaga, R. Chellappa, V. Subrahmanian, and O. Udrea, Machine recognition of human activities: A survey [J].*IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 11, pp. 1473–1488, 2008.
- [2] L. Shao, L. Ji, Y. Liu, and J. Zhang, Human action segmentation and recognition via motion and shape analysis[J]. *Pattern Recognition Letters*, vol. 33, no. 4, pp. 438–445, 2012.
- [3] D. Weinland, R. Ronfard, and E. Boyer, A survey of vision-based methods for action representation, segmentation and recognition[J]. *Computer Vision and Image Understanding*, vol. 115, no. 2, pp. 224–241, 2011.
- [4] Q. Guan, C. Li, X. Guo, and G. Wang, Compressive classification of human motion using pyroelectric infrared sensors[J]. *Pattern Recognition Letters*, vol. 49, pp. 231–237, 2014.
- [5] T. Liu, X. Guo, and G. Wang, Elderly-falling detection using distributed direction-sensitive pyroelectric infrared sensor arrays[J]. *Multidimensional Systems and Signal Processing*, 2011.
- [6] T. Liu and J. Liu, Design and implementation of a compressive infrared sampling for motion acquisition[J]. *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, pp. 1–15, 2014.
- [7] B. Horn and B. Schunck, Determining optical flow[J]. *Artificial intelligence*, vol. 17, no. 1-3, pp.

185–203, 1981.

- [8] G. Wu, Distinguishing fall activities from normal activities by velocity characteristics[J]. *Journal of Biomechanics*, vol. 33, no. 11, pp. 1497–1500, 2000.
- [9] D. Brady, N. Pitsianis, and X. Sun, Reference structure tomography[J]. *Journal of the Optical Society of America. A, Optics, image science, and vision*, vol. 21, no. 7, pp. 1140–1147, 2004.
- [10] P. Agarwal, D. Brady, and J. Matoušek, Segmenting object space by geometric reference structures[J]. *ACM Transactions on Sensor Networks*, vol. 2, no. 4, pp. 455–465, 2006.
- [11] M. Peng and Y. Xiao, A survey of reference structure for sensor systems[J]. *IEEE Communications Surveys Tutorials*, vol. PP, no. 99, pp.1 –14, 2011.
- [12] L. Lu, H. Zhang, and H. Jiang, Content analysis for audio classification and segmentation[J]. *IEEE transactions on speech and audio processing*, vol. 10, no. 7, pp. 504–516, 2002.
- [13] L. Rabiner, A tutorial on hidden markov models and selected applications in speech recognition[J]. *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [14] J. Yamato, J. Ohya, and K. Ishii, Recognizing human action in timesequential images using hidden markov model[C]. in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1992, pp. 379–385.