# Human Movement Modeling and Activity Perception Based on Fiber-Optic Sensing System

Qingquan Sun, *Member, IEEE*, Fei Hu, *Member, IEEE*, and Qi Hao, *Member, IEEE*

***Abstract*—This paper presents a flexible fiber-optic sensor-based pressure sensing system for human activity analysis and situation perception in indoor environments. In this system, a binary sensing technology is applied to reduce the data workload, and a bipedal movement-based space encoding scheme is designed to capture people's geometric information. We also develop a nonrepetitive encoding scheme to eliminate the ambiguity caused by the two-foot structure of bipedal movements. Furthermore, we propose an invariant activity representation model based on trajectory segments and their statistical distributions. In addition, a mixture model is applied to represent scenarios. The number of subjects is finally determined by Bayesian information criterion. The Bayesian network and region of interests are employed to facilitate the perception of interactions and situations. The results are obtained using distribution divergence estimation, expectation–maximization, and Bayesian network inference methods. In the experiments, we simulated an office environment and tested walk, work, rest, and talk activities for both one and two person cases. The experiment results have demonstrated that the average individual activity recognition is higher than 90%, and the situation perception rate can achieve 80%.**

***Index Terms*—Fiber-optic sensor, human activity recognition, human movement modeling, situation perception, space encoding.**

## I. Introduction

HUMAN activity refers to a collection of actions and motions within a certain environment. Based on human geometric information (e.g., location, posture, gesture), human activity study aims to identify subjects' behavioral patterns and understand their situations (e.g., meeting, working, cooking). Human activity study is addressed in applications including healthcare, surveillance, house automation, and energy-efficient building [1], [2].

Conventional approaches to human activity recognition/analysis employ 1) video cameras [3]–[6] or 2) wearable sensors [7]–[9]. Video cameras acquire visible features of human activities. However, the challenges such as data association for multitarget, computation complexity in data processing, and data inconsistency under different environmental conditions, prevent their applications in real environments. Although wearable sensors may overcome those challenges, they have some shortcomings. For example, wearing sensors may change the behavioral habits of subjects under examination, and reduce the recognition accuracy. Additionally, wearing sensors is inconvenient. By contrast, nonintrusive sensors, such as thermal, pressure, and laser sensors have advantages including distributed measurements and computational complexity. More importantly, such sensors can work under various environmental conditions, such as poor illumination condition.

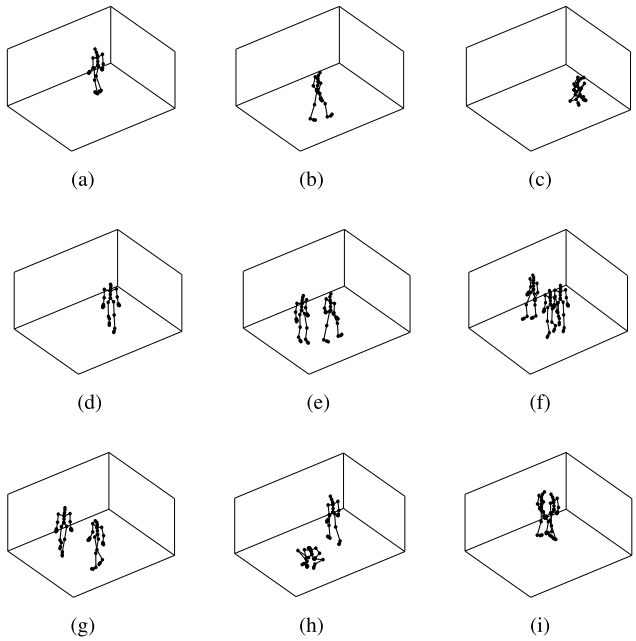



Fig. 1. Typical human activities: (a) walking while thinking, (b) walking (with intentions), (c) working in a fix position. General scenes: (d) one person, (e) two people, (f) three people. Specific situations: (g) two people walking, (h) one person walking and one person working, (i) two people talking.

Most human activity analysis research mainly emphasizes individual activity recognition but neglect contexts [10], [11]. For applications in a smart environment, situation perception is important for intelligent control and assisted living. Although the popular approaches such as hidden Markov models (HMM) [12], and robust data learning algorithms [13] have advantages in individual activity representation and classification, they are difficult to model and represent group activities. Furthermore, these methods usually involve high computational complexity and, hence, are unsuitable to real-time distributed sensor network applications.

In order to quickly and accurately analyze human activity, as well as perceive their situations in smart environment

Manuscript received April 18, 2013; revised August 19, 2013, November 29, 2013, April 17, 2014, and June 28, 2014; accepted August 5, 2014. Date of publication September 29, 2014; date of current version November 12, 2014. This work was supported in part by the U.S. National Science Foundation under the grant CNS-1059212. All results presented here do not necessarily reflect NSF's opinions. This paper was recommended by Associate Editor Z.-H. Mao.

Q. Sun is with the School of Computer Science and Engineering, California State University San Bernardino, San Bernardino, CA 92407 USA (e-mail: quanqian12345@gmail.com).

F. Hu and Q. Hao are with the Department of Electrical and Computer Engineering, University of Alabama, Tuscaloosa, AL 35487 USA (e-mail: fei@eng.ua.edu; qh@eng.ua.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/THMS.2014.2354046

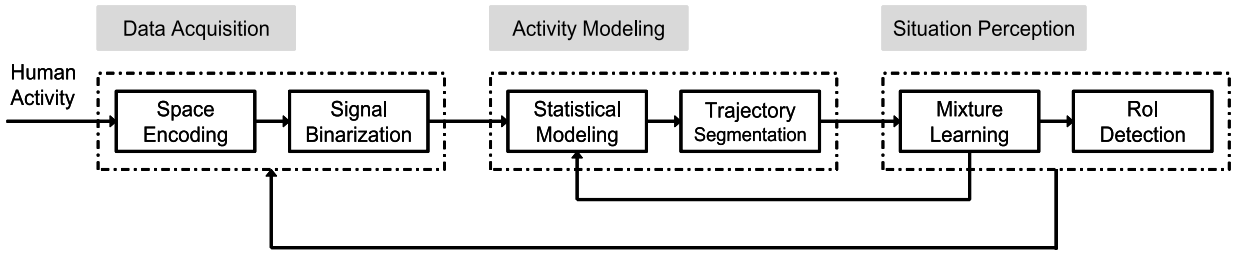



Fig. 2. System diagram for human activity modeling and situation perception.

applications, some technical challenges need to be overcome including

1) flexible, low-data-throughput sensing technologies to acquire human geometric information;
2) appropriate human activity representation models to reflect intrinsic features of human activities;
3) efficient methods to estimate situations with human activities and interactions.

Motivated by these challenges, we have built a flexible fiber-optic sensor-based human monitoring system. The system diagram is illustrated in Fig. 2. Such a system aims to estimate human activities, recognize scenes, and perceive situations (see Fig. 1) in an indoor environment. The contributions of our work include the following:

1) *A low data throughput, fast response sensing system for human activity analysis:* Our monitoring system is based on fiber-optic sensors and binary sensing technology. Fiber-optic sensors are not impacted by the environmental conditions, and do not affect users' daily activities. Such sensors can detect the footsteps and trajectories that reflect some features of human actions and activities, even certain psychological and physiological conditions [14]. Binary sensing reduces the high-dimensional data space to a 2-D space; therefore, the data throughput is reduced.
2) *A bipedal movement-based space encoding scheme:* Space encoding refers to the scheme that segments a space into different cells and uses codes to register these cells. Our space encoding scheme converts the two-foot structure to one code by using an "OR" operation. Additionally, a nonrepetitive encoding scheme is developed to remove the ambiguity generated by the "OR" operation and guarantee that each code is different from any other.
3) *An invariant activity representation model:* A successful activity representation should reflect the intrinsic features of activities. We build an invariant model based on the movement trajectories of people to represent their activities.
4) *A situation perception approach based on Bayesian network and region of interest:* For the daily activities in indoor environments, interactions occur with certain patterns. For example, when people have a conversation or meeting, usually they sit or stand close to each other. Our approach can discover the number of people and the interaction patterns. Thus, the situations can be perceived based on such information.

The rest of this paper is organized as follows: Section II reviews the related work. Section III presents the space encoding scheme, invariant activity representation model, and Bayesian network-based situation perception algorithm. Section IV describes the apparatus and methods for performance evaluation. Section V provides the result for proposed system and approaches. Section VI concludes this paper.

## II. Related Work

There are mainly two kinds of sensor-based systems for human activity analysis. One is a wearable sensor-based system [15], and the other is nonintrusive sensor based one [16]. The most widely used wearable sensors for human activity analysis are accelerometers [17]. Through the measurement of the acceleration along different directions, human activities such as walking and running can be estimated [18], [19]. The biggest drawback of wearable sensors is that they cannot detect and recognize group activities, thus, are unable to determine context.

Nonintrusive sensors such as motion, laser, acoustic sensors, can perform long-distance sensing. Therefore, they are able to detect and analyze group activities [20], [21]. Some human activity analyses are based on such sensors [22]–[25]. Motion sensors are able to detect the motions of targets. They have been used to recognize gaits and gestures, and estimate human activities [24].

Acoustic sensors measure timing and energy response to calculate the distance. Although such sensors can perform well in outdoor environments, they cannot be used as the main sensing devices to detect human activities due to the multipath and clutter problems in indoor environments. Laser sensors have been utilized to track humans and model activities [26], [27]. However, these nonintrusive sensors occupy space and are difficult to deploy. Furthermore, such sensors are impacted by the environmental conditions (e.g., acoustic sensors have poor performance under noisy circumstances).

Pressure sensors can be easily deployed and not affected by environmental conditions. Three types of pressure sensors have been used for human detection and activity analysis: 1) piezoelectric [28], 2) polymer [29], and 3) fiber optic [30], [31]. Piezoelectric sensors are the most widely used pressure sensors used to measure human pressure with high resolution [32]. Polymer sensors can be used to develop lightweight pressure sensing systems. Fiber-optic sensor-based pressure sensing systems [33]

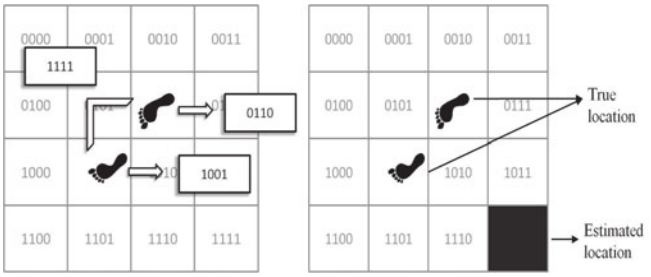



Fig. 3. (a) Theoretical binary space encoding scheme with four fiber-optic sensors. (b) Location estimation by using theoretical binary encoding scheme.

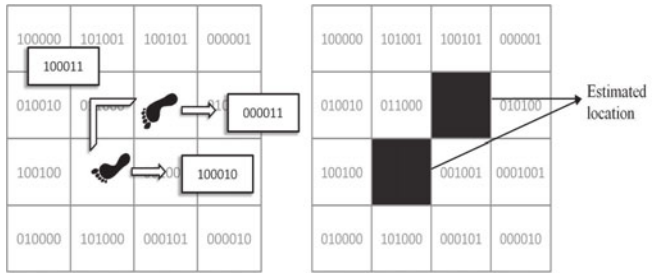



Fig. 4. (a) Proposed binary space encoding scheme with six fiber-optic sensors. (b) Location estimation by using proposed binary encoding scheme.

have the advantages of low hardware cost, good scalability, and convenient deployment.

The typical approaches for human activity analysis include single-layered approaches [34] and hierarchical approaches [35]. The signal-layered approaches mainly utilize spatial and temporal characteristics of human gestures to achieve activity recognition [36]. The popular methods for hierarchical human activity analysis are statistical approaches [37]. Such approaches use statistical state-based models to recognize gestures and activities. The problem for sensor-based monitoring systems is that such approaches are complicated and involve heavy computations. They are not appropriate for low-data throughput, low-resolution sensory data.

To study the group activity with pressure sensing systems, we need more efficient sensing schemes and statistical models. This paper is extended from our original work [38]. Here, we develop a new pressure sensing system based on low-cost fiber-optic sensors. To reduce data throughput, each sensor only detects the binary presence of weight pressure. When multiple binary sensor arrays are deployed, the whole sensing space is encoded by a set of binary codes. Accordingly, invariant statistical representation models based on trajectory segment and mixture model are developed for both individual and group activity analysis.

## III. Space Encoding-Based Human Activity Estimation and Situation Perception

### A. Binary Sensing Model

Sensing is a process of energy collection, and hence is an integral process. The conventional sensing process of human activity can be described by

$$\mathcal{O}(r_1,\, t) = h(t) * \int_{\Omega} \Omega(r,\, r_1)\psi_\theta(r,t)dr \tag{1}$$

where $\mathcal{O}$ is the sensory signal, $h(t)$ is the impulse function, $*$ is convolution, $\Omega$ is the geometric sampling function, $\theta$ represents the parameters of activity model $\psi$, $r$ and $r_1$ are coordinates of the subject and sensor, respectively [39].

The binary sensing process can be described by

$$\mathcal{B}(r_1,\, t) = \bigcup_{\Omega} I\left[\Omega(r,\, r_1) \cap \psi_\theta(r,t)\right] \tag{2}$$

where $\mathcal{B}$ is the binary sensory signal, $I$ is a logistic function whose output is "0" or "1." Equation (2) shows that the binary sensory signal only depends on the sampling function $\Omega$ and human subject activity $\psi$.

The general geometric sampling function $\Omega$ of the fiber-optic sensor could be described by a 2-D curve, which is $f(x,\, y) = 0$. In our proposed system, we have developed a space encoding scheme, which segments the space into encoded blocks. Therefore, the geometric sampling function in this case should be described by discrete blocks. As a result, it can be represented by $\Omega(x_i,\, y_i)$, where $(x_i,\, y_i)$ is the position of the $i$th block.

### B. Space Encoding for Bipedal Movements

A space encoding scheme converts space information into sampling geometry information. The optimal space encoding design depends on the sufficient intersections between sampling structure and human motions. An universal sampling structure for all human activities can lead to low sensing efficiency. For our fiber-optic sensing system, the floor space is segmented and encoded. In this paper, we encode a 4 m $\times$ 4 m area using 64 pieces of plastic mats and fiber-optic sensors. The size of each mat is about 0.5 m $\times$ 0.5 m. The fiber-optic sensors are deployed under these mats. By using different combinations of these fiber-optic sensors, each of these mats is encoded by a $n$-bit code ($n$ is the number of sensors). Through such an encoding method, human motions and activities are converted into a sequence of codes.

Here, we use an example to explain space encoding in detail. Suppose we want to encode a 16-block area. Ideally, each block can be encoded by one sensor. Thus, the 16-block area needs 16 sensors to encode. Obviously, such a scheme has poor scalability when encoding a larger area since in that case a large amount of sensors are needed. In comparison, we can choose a binary encoding scheme, namely, using multibit to encode one block. Then, theoretically, for such a 16-block area, only four sensors are needed to generate 16 binary combination codes ($2^4 = 16$). Such a scheme sharply reduces the amount of involved sensors.

However, the practical applications are more complicated than the theoretical cases. Because of the bipedal movement nature of humans, using such a theoretical binary encoding scheme will result in false alarms for localization. Therefore,

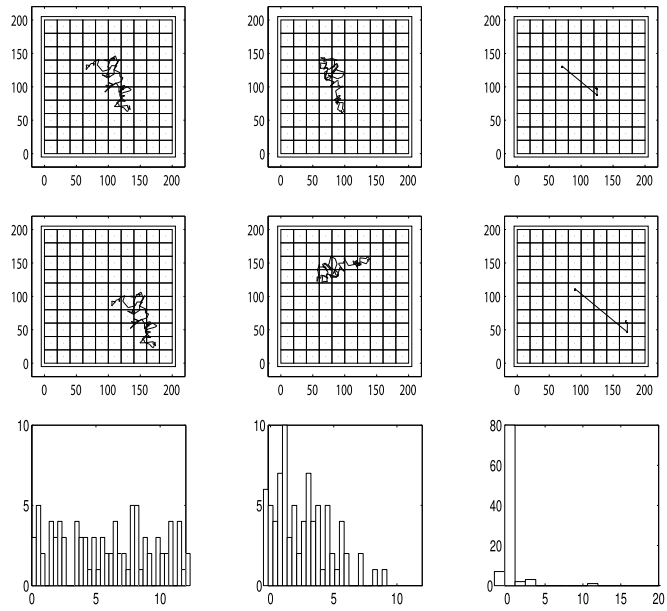


Fig. 5. Trajectories with uniform, Gaussian, and Levy distributions (first row). Translation, rotation, and scaling transformed trajectories (second row). Invariant representation of trajectory distributions (length histogram) (third row).

we propose a nonrepetitive encoding scheme which can avoid the false alarms.

We propose to utilize a space encoding scheme to form the sampling structure to extract people's location and geometric information. Space encoding converts the 2-D geometric information to 1-D code information. Given the encoded space and deployed fiber-optic sensors, the response function of $i$th sensor is described as

$$\gamma_i = \begin{cases} 1, & \eta_i(x_i,\, y_i) < \rho \\ 0, & \eta_i(x_i,\, y_i) \geq \rho \end{cases} \tag{3}$$

here $(x_i,\, y_i)$ describes the mass center of the mat where the $i$th sensor is deployed, $\eta_i(x_i,\, y_i) < \rho$ means that a person is on the mat, and $\eta_i(x_j,\, y_j) \geq \rho$ means that the person is not on the mat. Combining the responses of all the sensors deployed under the same mat, the obtained $n$-bit binary stream forms a code. Then, this mat is encoded and represented by this code.

Space encoding aims to build a sampling structure with as few sensors as possible. If $n$ sensors are used to encode $m$ mats, usually they should meet the requirement $n \ll m$. By using theoretical binary encoding, only four sensors are needed to encode a 16-block area. However, because of the bipedal movement nature of humans, sometimes a person's two feet will trigger two adjacent codes simultaneously. Thus, the obtained sensory signal will be bitwise "OR" of these two adjacent codes. If codes $c_i$ and $c_j$ represent two adjacent mats, when both mats are triggered at the same time, then we get the combination code

$$c_{ij} = c_i \,||\, c_j. \tag{4}$$

Unfortunately, since 4 bits can only generate 16 codes, the combination of codes such as 0110 and 1001 is the same as the individual code 1111. Accordingly, the subject will be localized at the block with the code 1111. Obviously, an error arises (see Fig. 3). This error is a false alarm or ambiguity.

**Algorithm 1:** Trajectory Segmentation Algorithm

**Input**: code table: $C$, mat adjacent matrix: $A$, direction table: $D$
**Output**: segment length: $l$, speed: $v$

```
C_0 ← one code of C;
k ← 1;
n ← 1;
while k > 0 do
  C_k ← one code of C;
  check A;
  D_k ← C_{k-1} to C_k;
  while k > 1 do
    if D_k == D_{k-1} then
      continue;
    else
      l_n ← k − 1;
      v_n ← (k − 1)/T;
      k ← 0;
      n + +;
    end
  end
  k + +;
end
return l_n, v_n
```

In order to eliminate the ambiguity caused by the bipedal movement characteristics, we propose a nonrepetitive encoding scheme which can guarantee that each individual code and each combination of two adjacent codes are unique. The scheme can be represented as

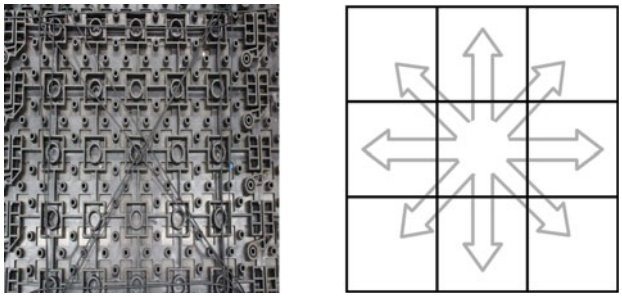
$$\begin{cases} c_i \neq c_j, & \forall\, i, j \in C \\ c_i \neq c_{ij}, & \forall\, i, j \in C \\ c_{ij} \neq c_{jk}, & \forall\, i, j, k \in C \end{cases} \tag{5}$$

where $i$ and $j$, $j$ and $k$ are adjacent blocks, and $C$ represents the code space. For a $4 \times 4$ block area, there are a total of 16 individual codes, and 42 combinations of two adjacent codes. Since $2^5 < 16 + 42 < 2^6$, with this scheme, six sensors are adequate to encode a 16-block area. An illustration of encoding 16 blocks area with six sensors can be found in Fig. 9(c).

Such a encoding scheme can avoid ambiguity. Fig. 4(a) shows a person at the center of the area; if codes 000011 and 100010 are triggered at the same time, the sensory signal will be coded as 100011. We can see that the sensory signal code is different from any individual code or combination of two adjacent codes. As shown in Fig. 4(b), we can estimate the person's location.

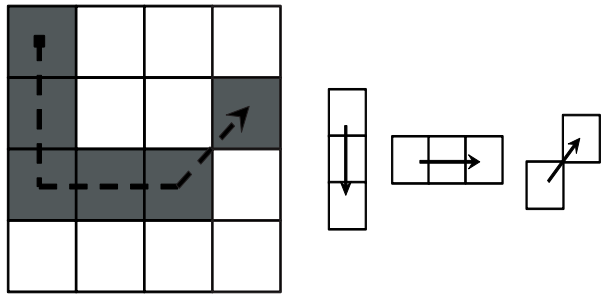
### C. Invariant Representation of Human Activity

The activity representation includes two aspects: 1) activity state and 2) statistical model. A successful activity representation should be invariant to a person's position, the room size,



(a) (b)

Fig. 6. (a) Encoded mat. (b) Mat-to-mat trajectory segment directions.



(a) (b)

Fig. 7. (a) Sample of human trajectory. (b) Segments of the human trajectory sample.

and sensors' location/orientation. Therefore, the position cannot be used to represent activities. In contrast, subject speed and trajectory segment are insensitive to geometric conditions and should be chosen as the states to represent activities.

Since our fiber-optic sensing system can detect the people's presence, each person's trajectory can be approximated by a set of linear segments. The lengths of these segments are also invariant to translation and rotation transforms. Its statistical distribution is invariant to the scaling transform. The person's position is not invariant to the translation and rotation transforms, but its statistical distribution is invariant to these transforms. Therefore, we define our activity representation model $\Phi$ as

$$\Phi_k = \{(L_k, V_k, t)\}, \quad t = 1, 2 \cdots T \tag{6}$$

here, $V$ represents the speed, $L$ is the trajectory segment, and $k$ represents the person number, and $t$ is the observation time.

The challenge here is to find the appropriate distributions to represent different activities. The general walking activity is random because of the effects of geometrical limitations and other people. Therefore, it can be modeled as a Gaussian distribution. For other complex activities, we need special models to represent them. In indoor environments, people usually work and rest. Such activity characteristics are similar to a Levy distribution, which is composed of many short walk lengths and few long walk lengths. Even for the same type of activity, it may present distinguishing patterns under different psychological conditions. For example, when people are walking while thinking, they may walk back-and-forth and repeat. Such an activity pattern can be appropriately represented by the uniform distribution. In the evaluation, we will further prove that the chosen distributions are consistent with the trajectory segment distributions of these activities. Fig. 5 illustrates the trajectories under uniform, Gaussian, and Levy distributions, and different rotation/translation/scaling transforms of these trajectories. It can be seen that invariant activity states and models can represent each set of similar activities.

### *D. Activity Pattern Estimation*

Based on the proposed invariant activity representation model, the activity pattern estimation issue becomes the trajectory segmentation problem. We propose a code-counting-based method to approximate the trajectory segment, that is, to calculate the number of codes along the same direction as the segment length. As shown in Fig. 6(b), we define eight mat-to-mat directions. The specific directions are represented by the arrows. Fig. 6(a) shows a sample of encoded mat.

For any mat $M_i$, represented by $c_i$, we record its adjacent mat matrix as $A_i$. When a person is present within the detection area, the starting code $c_i$ is recorded. With the movement of the person, once the next code $c_j$ is obtained, we calculate the movement direction, which is represented by $D_k$. Then, we repeat the operation to calculate the movement direction between two consecutive codes. If at the $n$th step, the movement direction is not $D_k$ anymore, the segment length and speed along $D_k$ can be approximated as

$$\begin{cases} l_n = k - 1 \\ v_n = (k-1)/T \end{cases} \tag{7}$$

where $T$ is the moving time along the direction $D_k$, $n$ is the number of segments observed, and $k$ is the index for measuring the length of each segment ($k$ will be reset whenever a new segment starts).

Given a trajectory sample as shown in Fig. 7(a), we can obtain the segmentation result shown in Fig. 7(b). As we can see, the trajectory sample is segmented to three parts along down direction, right direction, and right up direction, respectively. The trajectory segmentation algorithm is summarized in Algorithm 1.

With the sequences of the trajectory segments obtained from walking while thinking, walking with intentions, and working activities, the next task is to estimate distribution patterns of the trajectory segments. The probability density function of the trajectory segments can be modeled as

$$F(L) = F(l_1, \ldots, l_n | \theta) \tag{8}$$

here, $\theta$ depends on what kind of distribution pattern the $L$ yields.

In this paper, we investigate uniform, Gaussian, and Levy distributions for the segment patterns. The model is

$$G(l|\theta) = \begin{cases} \mathcal{U}(l \,|\, a, b), & \text{uniform pattern} \\ \mathcal{N}(l \,|\, \mu, \Sigma), & \text{Gaussian pattern} \\ \mathrm{L}(l \,|\, \mu, c), & \text{Levy pattern.} \end{cases} \tag{9}$$

Then, through maximum likelihood estimation the parameter $\theta$ can be estimated as

$$L(\theta; l_1, \ldots, l_n) = F(l_1, \ldots, l_n|\theta) = \prod_{n=1}^{N} F(l_n|\theta) \quad (10)$$

$$\theta^* = \underset{\theta}{\operatorname{argmax}} L(\theta;\, l_1, \ldots, l_n) \quad (11)$$

here, $\theta^*$ denotes the estimated value for the parameter $\theta$.

Kullback–Leibler (KL) divergence is a powerful method to measure the difference between two probability distributions. Hence, we utilize KL-divergence to measure the similarity between the estimated segment distribution and the standard distribution with $\theta^*$ as the parameter

$$D_{\mathrm{KL}}(F \,||\, G) = \sum_{n} \ln \left[ \frac{F(l_n \,|\, \theta^*)}{G(l_n \,|\, \theta^*)} \right] F(l_n|\theta^*). \quad (12)$$

The lowest divergence indicates the real distribution pattern of the trajectory segments.

### E. Mixture-Model-Based Scene Recognition

A situation refers to certain interactions among people or between people and the environment. It is a set of scenes in certain moments. A scene in this paper is defined to be determined by the number of people, which is the most useful precondition for the higher level interaction and situation perception.

Various situations may involve multiple people. However, the distributions of trajectory segment and speed contain little information about the number of people, as opposed to the distribution of their positions. Therefore, for the multiperson case, people's positions should be used to help scene recognition. In an indoor environment, especially in an office or a lab, different people act in different regions at most of the time. Hence, the action regions of multiple people forms a mixture.

In this paper, we propose to use a mixture model to represent the codes of multiple people and use expectation–maximization algorithm [40] to recognize the scenes. Mixture model approach aims to assign the observations into different clusters. Here, the clusters should reflect different regions where trajectories exist. In this model, the observation is the index of code $C_i$. The state variables to be estimated are the parameter of code distributions and the mixture coefficient.

In each E-step, we estimate the responsibility of component $k$ to the sample data $q(k)$, at iteration step $t+1$, given the measurement $C$ and the parameters of code distribution and component clusters $\beta$, at iterations step $t$,

$$q^{t+1}() = \underset{q()}{\operatorname{argmax}}\, \mathcal{G}(C, q^t(k), \beta^t) \quad (13)$$

$$\mathcal{G}[C, q(k), \beta] = E[\ln p(C, q(k)|\beta)]. \quad (14)$$

In each M-step, we optimize the parameters of code distribution and weight of clusters, given the updated responsibility of component $k$,

$$\beta^{t+1} = \underset{\beta}{\operatorname{argmax}}\, \mathcal{G}(C, q^{t+1}(k), \beta^t). \quad (15)$$

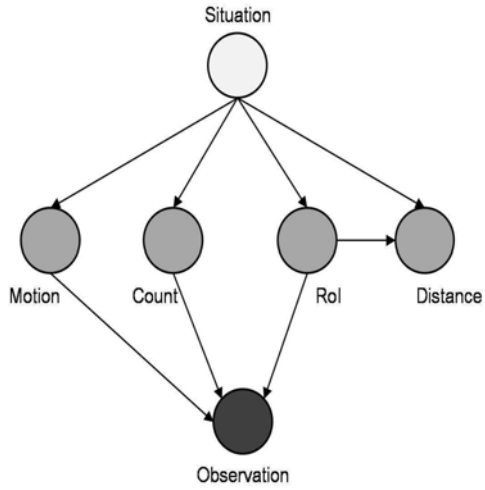



Fig. 8. Bayesian network-based situation reasoning model, showing the dependences between observations and hidden states.

The number of clusters $K$ can be determined by comparing the Bayesian information criterion (BIC) value for different numbers of the clusters. BIC can be represented as

$$BIC(K) = \simeq 2L(C, \theta) - dln(N) \quad (16)$$

where $L(x, \theta)$is the maximized mixture likelihood, $d$ is the number of free parameters, and $N$ is the number of observations.

The number of clusters $K$ is equivalent to the number of people. Thus, we can recognize the scenes through such a mixture model. Given the number of people, and the region of interest (RoI) information, we can perceive the specific situations.

### F. Bayesian Network Reasoning for Situation Perception

The situation can be perceived through Bayesian network reasoning (see Fig. 8). There are three levels in this graphical model. The shaded variable at the lowest level is the observation, the lighter shaded variables at the second level are some hidden features of a situation, and the unshaded variable at the highest level is the situation. The arcs correspond to causal dependencies between the variables. At the second level, the feature variables include motion, number of people, RoI, and the distance between RoIs. These features are extracted from the observations; therefore, the dependences among these variables can be represented as $p(M|O)$, $p(C|O)$, and $p(R|O)$. The distance information is closely dependent on RoIs; therefore, we denote their relationship by $p(D|R)$. With respect to the highest level situations, the dependences with the substantial features can be denoted by $p(S_i|M, C, R, D)$. Since we have obtained the information of motion and count in the previous parts, the challenge here is to find RoI.

RoI refers to the region where the interactions among people, or the interactions between people and the environment, could occur. In other words, a specific activity must happen in a certain area, and such area is the RoI. RoI can indicate the possible activities of people, thus, facilitating specific situation perception. RoI detection is a challenging problem in many sensing

systems. In our fiber-optic pressure sensing system, the space encoding scheme is able to localize people conveniently. Thus, it enables our system to achieve RoI detection.

The observations generated by our fiber-optic pressure sensing system are a stream of binary codes. Each code represents an encoded mat; namely, it indicates a person's location. Therefore, we can calculate the probability of codes in the data stream to obtain the probability of RoI. Let $c_i$ denote the count of $i$th code, then the appearance probability of code $i$ is

$$p_{T_j}(c_i) = \frac{c_i}{\sum_k c_k} \tag{17}$$

where $T_j$ is a subwindow of observation and $\sum_i c_i$ is the sum of all the codes in $T_j$ subwindow. Then, the RoI is determined by

$$\text{RoI} = Y_m \tag{18}$$

where $Y_m = \{i | p_{T_j}(c_i) > \zeta\}$, $\zeta$ is a threshold. Then, we get the normalized RoI

$$Y_n = \left\{ m | \frac{p_{T_j}(c_m)}{\sum p_{T_j}(c_i)} \right\}. \tag{19}$$

In order to extract people's spatial information in a certain situation, RoIs are segmented according to their location relationships. Given RoIs $Y_i$ and $Y_j$, the distance between the RoIs can be represented as

$$D = f_a \{Y_i, Y_j\} \tag{20}$$

where $f_a \{\cdot\}$ is the adjacency metric which calculates how many segment cells there are between two RoIs. Let $S_i$ denote the $i$th category situation. Given the conditional probabilities $p(D|R)$, $p(M|O)$, $p(C|O)$, and $p(R|O)$, we can estimate the situation $S_i$ based on the proposed Bayesian network by using

$$p(S_i|O) = \sum_{M,C,R,D} p(S_i|M,C,R,D)p(D|R)p(R|O) \times p(M|O)p(C|O) \tag{21}$$

where $p(S_i|M,C,R,D)$ is the conditional probability of situation, and it is determined by the category of situation and the relationships with the motion, count, RoI, and distance features.

## IV. Evaluations

### A. Apparatus

A fiber-optic sensing system is a nonintrusive, flexible, and low-cost system. It can be used to measure strain, temperature, and pressure. We have built a fiber-optic pressure sensing system for human activity analysis and situation perception. The system is shown in Fig. 9(a). Fiber-optic sensors are deployed on the floor to measure human presence. The basic principle is using light propagation characteristics along an optical fiber. A fiber-optic sensing node is shown in Fig. 9(b). The emitter can generate and emit light signals. The light signals propagate through the optical fibers, and then the photo detector can receive them. When a person steps on an optical fiber cable, the light propagation is impacted, and then the intensity of light at detector side is reduced. Hence, we can sense the presence of humans through light intensity measurement.

Fig. 10 illustrates the experiment setup. Experiments are conducted in a 4 m $\times$ 4 m area, where 64 plastic mats are deployed to embed the fiber-optic sensors. The size of each mat is about 0.5 m $\times$ 0.5 m. Theoretically, the proposed ambiguity avoidance binary encoding scheme needs only six sensors to encode a 4 $\times$ 4, 16-block area. To avoid bends in the deployment of these fiber-optic sensors, in the evaluation, we use eight sensors to encode each 16-block area. The sampling frequency of our fiber-optic sensors is 4 Hz.

### B. Methods

To compare different encoding schemes, we investigate sensing efficiency and accuracy for each sensing scheme. Sensing efficiency here refers to the number of encoded blocks given the same amount of sensors. Sensing accuracy refers to the correct sensing rate. The data collected are sensory signals. Since we have encoded each block using a certain sensors, with the sensory data we can determine the positions of people. The experiments are conducted based on encoded areas as follows: 4, 9, 16, 25, 36, 49, 64, 81, and 100.

As for the other tasks, the collected data are also the sensory signals of fiber-optic sensors. The data to be processed are streams of binary codes. These streams of binary codes are shown in Fig. 11. They correspond to walking while thinking, walking with intention, and working activities; one-subject scene, two-subjects scene, and three-subjects scene; two subjects walking without interaction, one subject walking and the other subject working, and two subjects talking situations.

## V. Results

### A. Sensing Efficiency and Accuracy

Our proposed ambiguity avoidance space encoding scheme has lower demand for sensors than the ideal decimal encoding scheme and theoretical binary encoding scheme. Fig. 12(a) shows the sensing efficiency results of these encoding schemes. Encoding the same number of blocks, using binary encoding schemes, is more efficient and can save sensor resources. When encoding 16 blocks area, a decimal encoding scheme needs 16 sensors, theoretical binary encoding schemes requires four sensors, and our proposed encoding scheme needs six sensors. Although the proposed encoding scheme needs a few more sensors than the theoretical binary encoding scheme when encoding the same size of area, it can eliminate the false alarms caused by bipedal movement.

Fig. 12(b) shows the sensing accuracy results of binary encoding schemes based on an 8$\times$8 blocks area. Here, false alarm refers to the incorrect detection of the person's locations. Although both schemes have stable performance, the theoretical binary encoding scheme has a much higher false alarm rate than the proposed encoding scheme. The false alarm rate of the proposed encoding scheme is only about 0.04, which may be caused by the sensitivity of sensors and the loss of message packets. In comparison, the false alarm rate of the theoretical

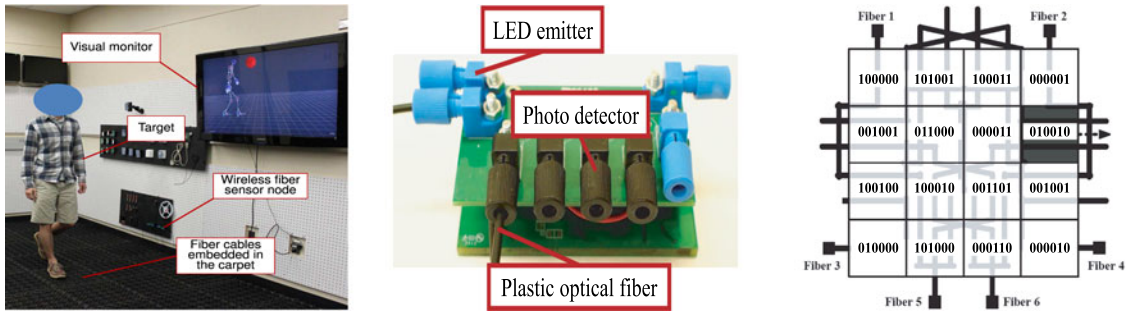



Fig. 9. (a) Fiber-optic pressure sensing system. (b) Fiber-optic sensor node. (c) Graphical illustration of space encoding using six fiber-optic pressure sensors.

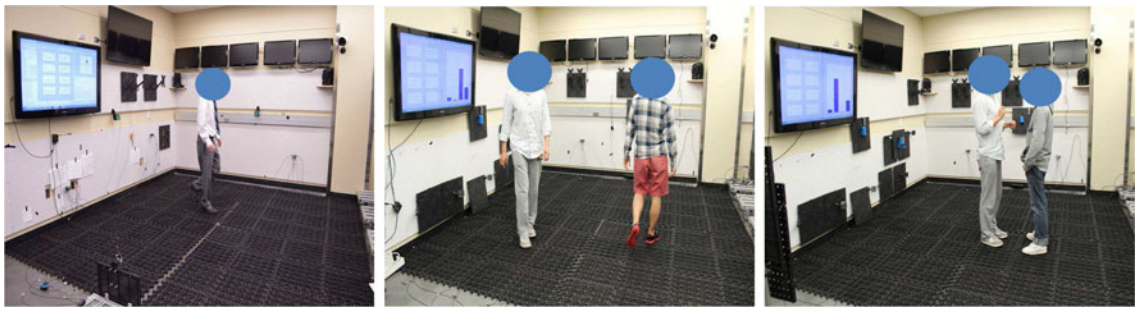


Fig. 10. System and experiment setup.

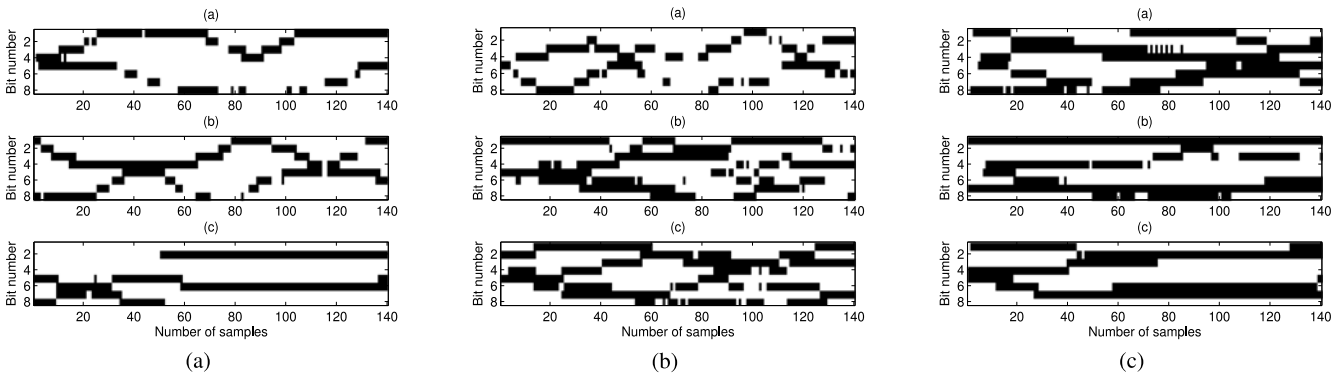



Fig. 11. Binary signals of human activities and situations. (a) Three kinds of investigated human activities: walking while thinking, walking (with intentions), and working. (b) General scenes determined by the number of human subjects: one-target scene, two-target scene, and three-target scene. (c) Specific situations in a certain moment: two targets walking, one target walking, and one target working, and two targets talking.

binary encoding scheme is around 0.13. The false alarm can be caused by sensors' sensitivity and packets' loss, but the major reason is the ambiguity generated by bipedal movement. Therefore, we can draw the conclusion that our proposed encoding scheme has a balanced performance between sensing efficiency and accuracy.

### B. Human Activity Estimation

We have analyzed the distinguishing features of different activities or the same activity under different psychological and physiological conditions in Section III-C. Here, we use experiments to verify the analysis and discover the appropriate trajectory segment distributions. The experiment results of the three tested activity patterns (walking while thinking, walking with intention, working) are shown in Fig. 15. Although the trajectory segment distribution of walking while thinking is not perfectly consistent with an uniform distribution [see Fig. 15(a)], it still has the lowest KL-divergence with uniform distribution as shown in Fig. 12(c). In comparison, as shown in Fig. 15(b) and (c), the trajectory segment distributions of walking with intention and working activities reflect clearer statistical characteristics. Accordingly, they have the lowest KL-divergences with half-Gaussian and Levy distributions, respectively.

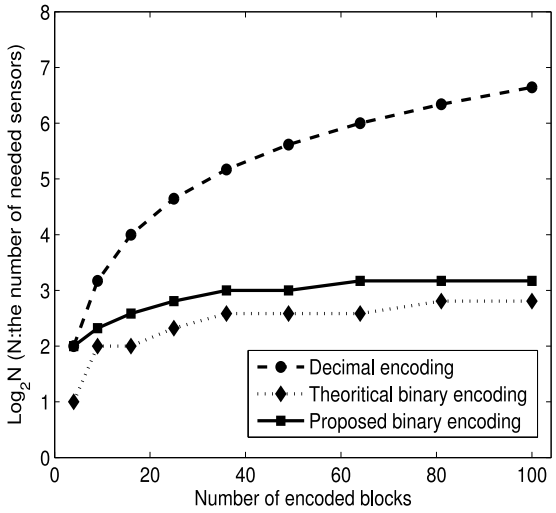



Fig. 12. Sensing efficiency comparison among three typical encoding schemes.

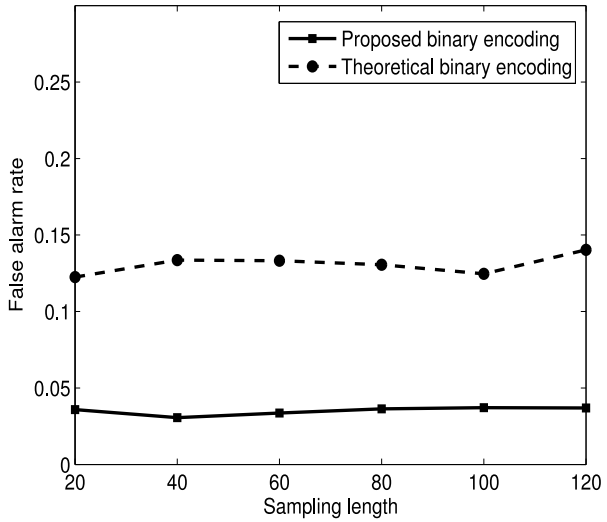



Fig. 13. Sensing accuracy comparison between two typical binary encoding schemes.

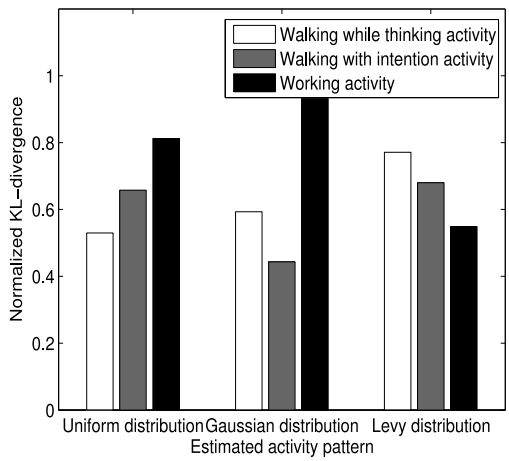



Fig. 14. Similarity estimation of different activity distribution patterns.

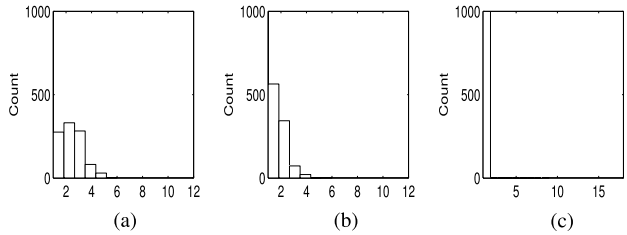



Fig. 15. (a) Trajectory segment distribution of walking while thinking activity. (b) Trajectory segment distribution of walking with intention activity. (c) Trajectory segment distribution of working activity.

Fig. 18(a) shows the recognition rate of these three activities. The recognition rates of walking while thinking and walking with intention activities are around 85%, while the recognition rate of working is close to 98% compared with the other two activities. Since the first two activities are both walking, they have higher similarity. However, working is different from walking; therefore, it does not have much ambiguity with walking.

### C. General Scene Recognition

For the general scenes, we aim to recognize the number of people in a scene. For multiple people scenes, the trajectories of different people have overlapped areas. The clustering accuracy decreases with the increment of the number of people. Fig. 16 shows the scene recognition result for the two person case. The histogram of code distribution is shown in Fig. 16(a). The distribution of code has a small overlap area. The clustering rate shown in Fig. 16(b) is 91%. Fig. 16(c) shows the convergence result, while the three person cases have a much lower clustering rate. More people lead to more complex situations. Hence, such scenes are harder to recognize.

The final scene recognition results are shown in Fig. 18(b). Based on the estimated statistical parameters, the scenes can be distinguished from each other, and can be identified with reasonable accuracy. The recognition rate of the one person scene can reach 87%, and that of the two person scene is close to 70%. The recognition rate of three people scene is lower as a large number of people can increase the repetition rate of the codes, and decrease the recognition rate.

### D. Specific Situation Perception

With respect to the specific situations, we investigate the two person case. The speed information is helpful to estimate individual activity. However, it does not contain spatial information. Hence, it cannot reflect the intrinsic characteristics of group activities and complex situations. We propose to use the RoI to enhance situation perception. The idea is inspired by the properties of human activity: when people interact with each other or interact with the environment, usually they will stop to do it, and when a subject stops, the code of the current region repeats. The high repetition rate of region codes indicates a certain interaction. This is the unique property of our proposed space encoding-based pressure sensing system.

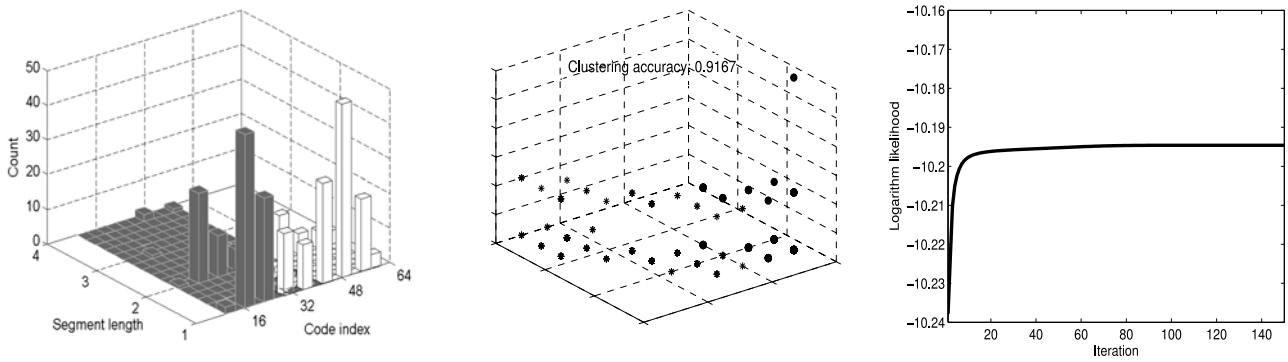



Fig. 16. (a) Histogram of code and segment length for the two person case. (b) Clustering result for the two person case. (c) Convergence of mixture learning.

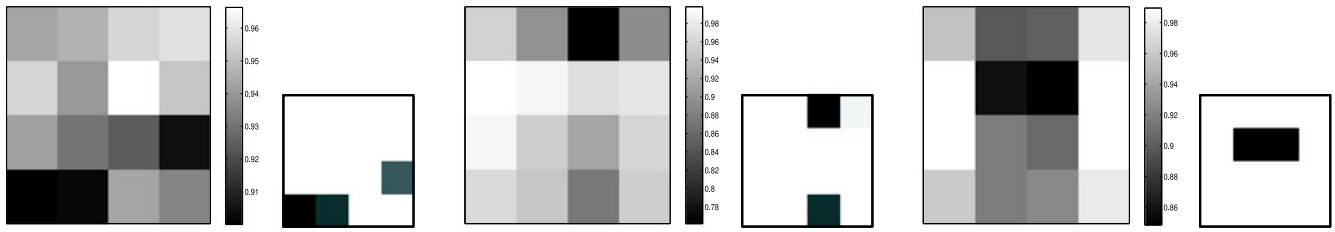



Fig. 17. Probability estimation for RoI.

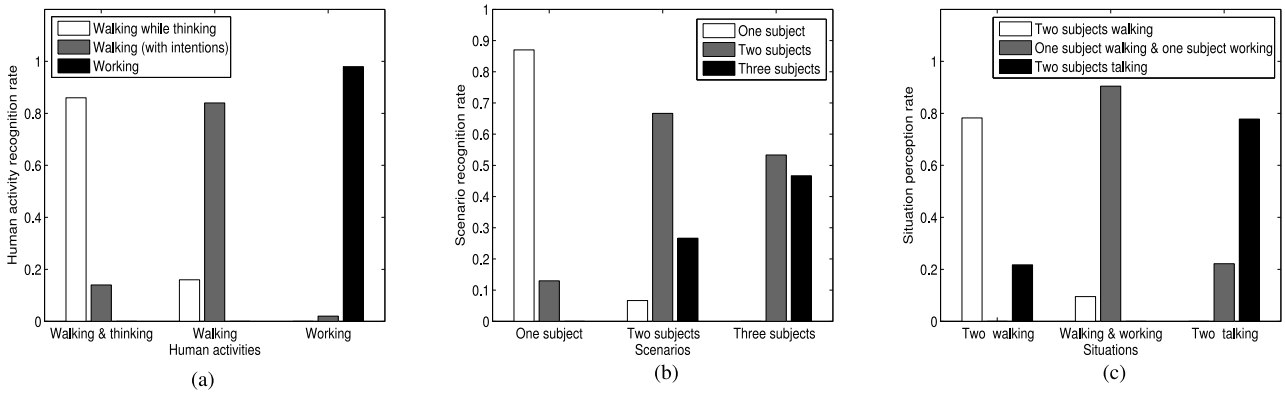



Fig. 18. (a) Performance of activity pattern estimation. (b) Performance of general scene recognition. (c) Performance of specific situation perception.

TABLE I
HUMAN ACTIVITY AND SITUATION PERCEPTION PERFORMANCE

| | One person (count) | | | Two persons (count) | | | | |
|---|---|---|---|---|---|---|---|---|
| Ground Truth | Walking | Working | Resting | Walking | Working | Walking & Working | Talking | Accuracy |
| Walking | **34**/40 | - | **2**/40 | **4**/40 | - | - | - | 0.85 |
| Working | - | **39**/40 | - | - | **1**/40 | - | - | 0.975 |
| Resting | **2**/40 | - | **35**/40 | - | - | - | **3**/40 | 0.875 |
| Walking | **4**/40 | - | - | **31**/40 | - | **5**/40 | - | 0.775 |
| Working | - | **1**/40 | - | - | **37**/40 | **2**/40 | - | 0.925 |
| Walking & Working | - | - | - | **2**/40 | **2**/40 | **36**/40 | - | 0.90 |
| Talking | - | - | **3**/40 | **3**/40 | - | **3**/40 | **31**/40 | 0.775 |

The results of RoI detection corresponding to the specific situations are shown in Fig. 17(a), (c), and (e). To make the result more readable, we use $(1 - p(c_i))$ to represent the probability of RoI. Therefore, the darker value represents the higher appearance probability of this region, and it is on the opposite side of the color bar. Such regions are RoIs since they indicate the possible human activities. However, they cannot indicate the human activities directly since the probabilities of RoIs are not normalized. In order to illustrate the differences of RoIs among a variety of situations, the normalized RoI probabilities are shown in Fig. 17(b), (d), and (f). In Fig. 17(b), there are two separated blocks, which indicate a two-person scene. Compared with Fig. 17(d) and (f), the blocks in Fig. 17(b) are lighter, meaning that the probabilities of these two blocks are smaller than that of Fig. 17(d) and (f). In the physical world, it means that these two people are walking. Fig. 17(c) shows that two RoIs are separated, and one block is much darker than the other one. We can infer that one person might stop and work there, thus generating a higher probability of that block. Based on the above analysis of the RoIs, we can understand the situation shown in Fig. 17(f). In this figure, two darker blocks are adjacent. One may conclude that two people might stop and talk there.

The RoIs and Bayesian network can provide more information related to the interactions and the dependencies. Therefore, we can successfully recognize some specific situations. The results of the specific situations recognition are shown in Fig. 18(c). The recognition rate of the two-subject walking situation can reach 77.5%, as well as the two-subject talking situation. For the situation that one person walks and the other works, we have obtained a higher recognition rate close to 90%. Table I provides the final results of human activity and situation perception. We can perceive walking, working, resting, and talking situations under different conditions using the proposed system and methods. We can see that for each situation, we can obtain a satisfactory perception accuracy and the average perception rate can reach 87%.

## VI. Conclusion

In this paper, we have presented a human activity analysis and situation perception system based on distributed fiber-optic pressure sensors. A set of space encoding and decoding techniques have been developed to acquire geometric information of bipedal movements of human subjects. Invariant human activity models are developed via statistical trajectory segment distributions. Bayesian network and RoIs are used to perceive interactions and situations among a group of human subjects.

The advantages of fiber-optic sensor systems in human activity modeling and situation perception include flexibility, robustness, and convenience. Such sensing systems make the fast, intelligent, and distributed monitoring practical in smart environment applications. This system is developed to achieve low-resolution and fast-response sensing, but not accurate high-resolution sensing. Therefore, it can be used as the primary sensing facility in low resolution required applications or the complementary facility to provide the context in high-resolution required applications.

Our proposed system has shown satisfactory results in activity pattern recognition and situation perception. Although the recognition rate of the three people scenario is not high, we can improve the accuracy in the following aspects:

1) Improve the deployment of fiber-optic cables with smooth angle variation. The principle of fiber-optic sensing system is using the light propagation along the optical fiber. We can deploy the fiber-optic cables in round shapes to reduce the light attenuation. Since the angles along a circle change smoothly, it is beneficial to light propagation. Thus, the signal intensity can be enhanced.
2) Improve the resolution by reducing the size of mat. Once the size of mat is reduced, the accuracy of trajectory segment estimation increases, and thus we can improve the recognition rate.
3) Deploy more sensors. The current prototype system only employs eight sensor arrays (with total 32 sensors) to encode 64 mats space. Using more sensors can enhance the information acquisition and enlarge the Hamming distance between the adjacent codes, thus improving the recognition accuracy.

## Acknowledgment

The authors would like to thank R. Ma and Dr. J. Lu for their help on experiments and data collections.

## References

[1] N. Kern, B. Schiele, and A. Schmidt, "Multi-sensor activity context detection for wearable computing," *Ambient Intell.*, pp. 220–232, 2003.
[2] V. Osmani, S. Balasubramaniam, and D. Botvich, "Human activity recognition in pervasive health-care: Supporting efficient remote collaboration," *J. Netw. Comput. Appl.*, vol. 31, no. 4, pp. 628–655, 2008.
[3] A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 3, pp. 257–267, Mar. 2001.
[4] C. Chen, K. Liu, and N. Kehtarnavaz, "Real-time human action recognition based on depth motion maps," *J. Real-Time Image Process.*, pp. 1–9, 2013.
[5] Q. Cai, Y. Yin, and H. Man, "DSPM: Dynamic structure preserving map for action recognition," in *Proc. Int. Conf. IEEE Multimedia Expo.*, 2013, pp. 1–6.
[6] J. Shen, P.-C. Su, S.-c. S. Cheung , and J. Zhao, "Virtual mirror rendering with stationary RGB-D cameras and stored 3-D background," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3433–3448, Sep. 2013.
[7] L. Wang, T. Gu, X. Tao, and J. Lu, "Sensor-based human activity recognition in a multi-user scenario," *Ambient Intell.*, pp. 78–87, 2009.
[8] C. Zhu and W. Sheng, "Wearable sensor-based hand gesture and daily activity recognition for robot-assisted living," *IEEE Trans. Syst., Man Cybern., Part A, Syst. Humans*, vol. 41, no. 3, pp. 569–573, May 2011.
[9] K. Liu, C. Chen, R. Jafari, and N. Kehtarnavaz, "Fusion of inertial and depth sensor data for robust hand gesture recognition," *IEEE Sens. J.*, vol. 14, no. 6, pp. 1898–1903, Jun. 2014.
[10] J. Ben-Arie, Z. Wang, P. Pandit, and S. Rajaram, "Human activity recognition using multidimensional indexing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 8, pp. 1091–1104, Aug. 2002.
[11] T. Gu, L. Wang, Z. Wu, X. Tao, and J. Lu, "A pattern mining approach to sensor-based human activity recognition," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 9, pp. 1359–1372, Sep. 2011.
[12] M. Brand and V. Kettnaker, "Discovery and segmentation of activities in video," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 844–851, Aug. 2000.
[13] N. Suzuki, K. Hirasawa, K. Tanaka, Y. Kobayashi, Y. Sato, and Y. Fujino, "Learning motion patterns and anomaly detection by human trajectory analysis," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, 2007, pp. 498–503.

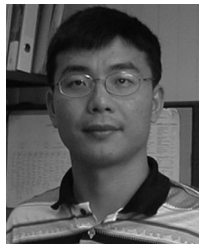
[14] R. J. Orr and G. D. Abowd, "The smart floor: A mechanism for natural user identification and tracking," in *Proc. Extended Abstracts Human Factors Comput. Syst.*, 2000, pp. 275–276.
[15] A. M. Sabatini, C. Martelloni, S. Scapellato, and F. Cavallo, "Assessment of walking features from foot inertial sensing," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 3, pp. 486–494, Mar. 2005.
[16] O. Urfaliglu, E. B. Soyer, B. Toreyin, and A. E. Cetin, "PIR-sensor based human motion event classification," in *Proc. IEEE 16th Signal Process., Commun. Appl. Conf.*, 2008, pp. 1–4.
[17] Z.-Y. He and L.-W. Jin, "Activity recognition from acceleration data using ar model representation and SVM," in *Proc. IEEE Int. Conf. Mach. Learning Cybern.*, 2008, vol. 4, pp. 2245–2250.
[18] Z. He and L. Jin, "Activity recognition from acceleration data based on discrete consine transform and SVM," in *Proc. SMC IEEE Int. Conf. Syst., Man Cybern.*, 2009, pp. 5041–5044.
[19] Y.-P. Chen, J.-Y. Yang, S.-N. Liou, G.-Y. Lee, and J.-S. Wang, "Online classifier construction algorithm for human activity detection using a tri-axial accelerometer," *Appl. Math. Comput.*, vol. 205, no. 2, pp. 849–860, 2008.
[20] Q. SUN, F. HU, and Q. HAO, "Mobile target scenario recognition via low-cost pyroelectric sensing system: Toward a context-enhanced accurate identification," *IEEE Trans. Syst., man Cybern. Syst.*, vol. 44, no. 3, pp. 375–384, Mar. 2014.
[21] F. Hu, Q. Sun, and Q. Hao, "Mobile targets region-of-interest via distributed pyroelectric sensor network: Towards a robust, real-time context reasoning," in *Proc. IEEE Sens.*, 2010, pp. 1832–1836.
[22] E. Tapia, S. Intille, and K. Larson, "Activity recognition in the home using simple and ubiquitous sensors," *Pervasive Comput.*, vol. 3001, pp. 158–175, 2004.
[23] L. Bao and S. Intille, "Activity recognition from user-annotated acceleration data," *Pervasive Comput.*, pp. 1–17, 2004.
[24] K. O. Arras, O. M. Mozos, and W. Burgard, "Using boosted features for the detection of people in 2D range data," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2007, pp. 3402–3407.
[25] J. Yang and Z. Fei, "ITGR: intermediate target based geographic routing," in *Proc. 19th Int. Conf. Comput. Commun. Netw.*, 2010, pp. 1–6.
[26] A. Panangadan, M. Matarić, and G. S. Sukhatme, "Tracking and modeling of human activity using laser rangefinders," *Int. J. Social Robot.*, vol. 2, no. 1, pp. 95–107, 2010.
[27] D. F. Glas, T. Miyashita, H. Ishiguro, and N. Hagita, "Laser-based tracking of human position and orientation using parametric shape modeling," *Adv. Robot.*, vol. 23, no. 4, pp. 405–428, 2009.
[28] M. Akiyama, Y. Morofuji, T. Kamohara, K. Nishikubo, M. Tsubai, O. Fukuda, and N. Ueno, "Flexible piezoelectric pressure sensors using oriented aluminum nitride thin films prepared on polyethylene terephthalate films," *J. Appl. Phys.*, vol. 100, no. 11, pp. 114 318–114 318, 2006.
[29] P. Srinivasan, D. Birchfield, G. Qian, and A. Kidané, "A pressure sensing floor for interactive media applications," in *Proc. ACM SIGCHI Int. Conf. Adv. Comput. Entertainment Technol.*, 2005, pp. 278–281.
[30] C. Kirkendall and A. Dandridge, "Overview of high performance fibre-optic sensing," *J. Phys. D, Appl. Phys.*, vol. 37, no. 18, p. R197, 2004.
[31] Q. Sun, *Indoor Scene and Human Activity Analysis with Wireless Binary Sensor Networks*. Ph.D. dissertation, The University of Alabama, Tuscaloosa, AL, USA. 2013.
[32] N. Griffith and M. Fernström, "Litefoot: A floor space for recording dance and controlling media," in *Proc. Int. Comput. Music Conf.*, 1998, pp. 475–481.
[33] Y. Zheng, N. Pitsianis, and D. Brady, "Nonadaptive group testing based fiber sensor deployment for multiperson tracking," *IEEE Sens. J.*, vol. 6, no. 2, pp. 490–494, Apr. 2006.
[34] Y. Ke, R. Sukthankar, and M. Hebert, "Spatio-temporal shape and flow correlation for action recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2007, pp. 1–8.
[35] A. Gupta, P. Srinivasan, J. Shi, and L. S. Davis, "Understanding videos, constructing plots learning a visually grounded storyline model from annotated videos," in *Proc. CVPR IEEE Conf. Comput. Vis. Pattern Recogn.*, 2009, pp. 2012–2019.
[36] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," in *Proc. 10th IEEE Int. Conf. Comput. Vis.*, 2005, vol. 2, pp. 1395–1402.
[37] N. Oliver, E. Horvitz, and A. Garg, "Layered representations for human activity recognition," in *Proc. IEEE Int. Conf. Multimodal Interfaces.*, 2002, pp. 3–8.
[38] Q. Sun, R. Ma, Q. Hao, and F. Hu, "Space encoding based human activity modeling and situation perception," in *Proc. IEEE Int. Multi-Disciplinary Conf. Cognitive Methods Situation Awareness Decision Support.*, 2013, pp. 183–186.
[39] Q. Hao, "An integral and differential geometric approach to behavioral information acquisition and integration via binary sensor networks," in *Proc. IEEE Sens.*, 2012, pp. 1–4.
[40] J. Luo, A. Brodsky, and Y. Li, "An EM-based ensemble learning algorithm on piecewise surface regression problem," *Int. J. Appl. Math. Statist.â¢*, vol. 28, no. 4, pp. 59–74, 2012.

**Qingquan Sun** (M'10) received the Ph.D. degree in electrical and computer engineering from The University of Alabama, Tuscaloosa, AL, USA in 2013.

He is currently an Assistant Professor in the School of Computer Science and Engineering, California State University San Bernardino, San Bernardino, USA. He has published around 20 journal/conference papers and book chapters. His research has been supported by U.S. National Science Foundation, U.S. Air Force Research Laboratory, and other sources. His research interests include intelligent sensing, distributed computing, wireless networking, signal processing, and machine learning in cyber physical systems.

**Fei Hu** (M'02) received the Ph.D. degree in electrical and computer engineering from Clarkson University, Potsdam, NY, in 2002.

He is currently an Associate Professor with the Department of Electrical and Computer Engineering, the University of Alabama (main campus), Tuscaloosa, AL, USA. He has published mor than 160 journal/conference papers and books. His research has been supported by U.S. National Science Foundation, U.S. Air Force Research Laboratory, Cisco, Sprint, and other sources. His research interests inlcude security, signals, sensors.



**Qi Hao** (M'07) received the Ph.D. degree from Duke University, Durham, NC, USA, in electrical engineering in 2006.

He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, The University of Alabama, Tuscaloosa, AL, USA. His research interests include smart sensor, intelligent wireless sensor network, and distributed information processing.