HUMAN ACTIVITY RECOGNITION FROM BASIC ACTIONS USING FINITE STATE MACHINE

Nattapon Noorit[1], and Nikom Suvonvorn[2]
[1,2]Department of Computer Engineering, Faculty of Engineering, Prince of Songkla University, Hatyai, Songkhla, 90112
[1]nattapon.nr@gmail.com, [2]kom@coe.psu.ac.th

ABSTRACT.
High-level human activity recognition is an important method for the automatic event detection and recognition application, such as, surveillance system and patient monitoring system. In this paper, we propose a human activity recognition method based on FSM model. The basic actions with their properties for each person in the interested area are extracted and calculated. The action stream with related features (movement, referenced location) is recognized using the predefined FSM recognizer modeling based on rational activity. Our experimental result shows a good recognition accuracy (86.96% in average).

Keywords : human activity recognition, finite state machine, FSM recognizer, rational activity definition

1 INTRODUCTION

High-level human activity recognition is an important method for the automatic event detection and recognition application, such as, surveillance system, patient monitoring system, and etc. Previously, many researchers proposed many human activity representation and recognition strategies. For instance, some researches tried to learn pattern of activity within the scene [1,2,3], some proposed methods using the interaction model between agents to describe activity [4], while some interesting works used hand-crafting model to recognize a particular activity [5,6]. However, in general these strategies use a similar concept that is matching an unknown sequence with the references to recognize a particular activity. Several techniques were introduced for solving the activity recognition problems, such as, Hidden Markov Model (HMM), Dynamic Time Warping (DTW), and Finite State Machine (FSM).

Hidden Markov Model is a kind of stochastic state machine. Since an activity can be represented as a sequence of actions, it can be described by the HMM representation via training process by given the observation sequences. Eventually, HMM cannot apply to recognize the activity where its action sequence is uncertain. For example, if the current activity is interrupted by another, the state become unpredictable that make HMM inapplicable. This technique has been found in many researches such as [10,11,12].

Dynamic Time Warping is a template-based dynamic programming matching technique. The advantage of DTW is that the reliable time alignment between reference and test patterns is provided. The disadvantage of using DTW is the heavy computational problem that requires determining the optimal time alignment path. DTW has been used in several activity recognitions [7,8,9].
Finite State Machine is a predefined state machine with specific transitions. Both states and transitions are used for describing the specific problem, defined by its nature mostly from observation. FSM is lightweight, human-readable and easy to parse. However, FSM may easily fail in the presence of noise. It is applied as recognizer in various methods [15,16,17,18,19].

In this paper, we propose a human activity recognition method based on FSM model. The basic actions (standing, walking, sitting, bending, laying) with their properties (location, movement) for each person in the interested area, as an action stream, are recognized using the predefined FSM based on rational activity. We focus to recognize four general activities that are: (1) walk through the scene (2) observation (3) rest (4) browse. We found that in some activities, such as, observation, its behavior can be described randomly by the stream of walking and standing states that finally make it unpredictable. So, the HMM is not the quite suitable technique for our test activities. Moreover, the DTW have a major weakness on the heavy computation costs that not allows the recognition in real time. Finally, FSM is chosen for its lightweight and capacity to deal with our activities.

2 SYSTEM OVERVIEW

In this section, we present our proposed activity recognition process that describes a complex activity by a stream of basic actions with its properties. Figure 1 shows the system overview. The system is divided into 3 main parts: (1) Image Calibration, (2) Action extraction and features calculation, (3) Activity Recognition. The system run sequentially from (1) to (3) as the flowchart detailed as follows:

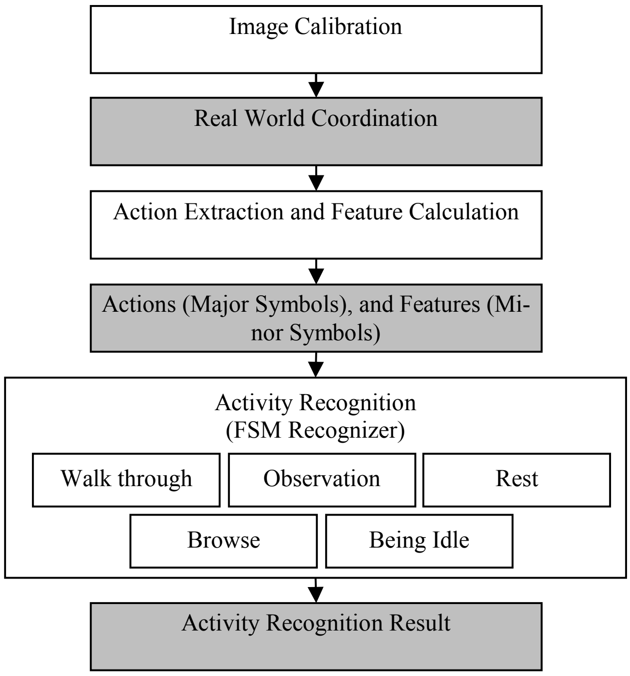


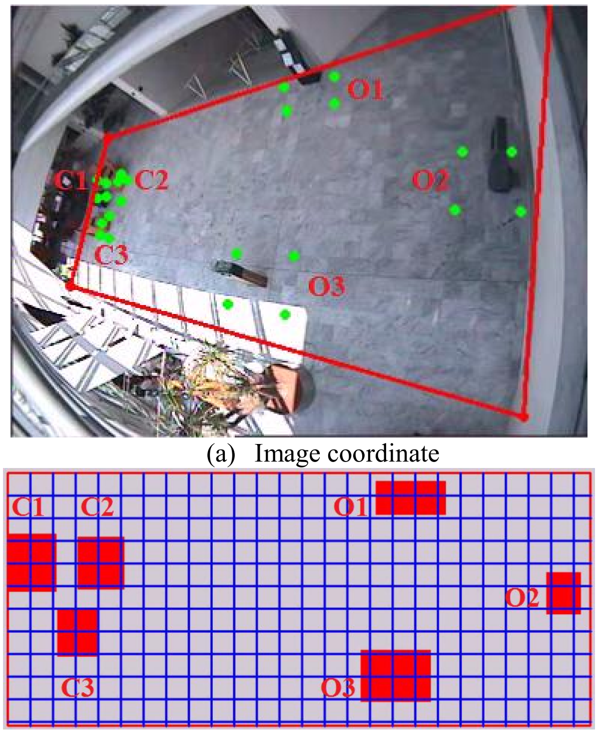
Fig. 1. System overview

3 IMAGE CALIBRATION

In this part, the human location detected in the image frames from video data is transformed from the pixel-based coordinate to a reference coordinate, related to the real-world location. This enables us to use data from any sensors referred to the same international unit system, such as, centimeter, that is very useful for the real-life activity definition, and for forming recognition condition, which is easily understandable from the real-life perception.

In this step, a perspective transformation is applied. The four points from the considering image with its corresponding points from the real-world coordinate are selected for calculating the transformation matrix. The Fig. 2 (a) shows the original image with our interested areas (red rectangles), corresponding the six locations associated to our recognition testing cases. Fig 2 (b) shows the transformed locations (red rectangle) in the real-world coordinate. Each blue square represents the 30x30 cm^2 square in the real-world scale.

In our experimentation, three objects (O1, O2 and O3), and three seats (C1, C2 and C3) are considered in the interested area. These objects and locations must be transformed to the referenced real-world location. Any interactions between people and objects in the ROI area using in our recognition process can compute the features in real-world scale, such as, movement distance, velocity, and acceleration.



(a) Image coordinate

(b) Real-world coordinate

Fig. 2. (a) Image coordinate (b) Real-world coordinate

4 ACTION EXTRACTION AND FEATURE CALCULATION

This section explains how to define actions and its additional features. We represent each feature as a symbol that can be separated into two groups: major symbol and minor symbol. Firstly, the major symbol is defined from the basic actions: standing, walking, sitting, bending, and lying. For each person who does an activity will produce a stream of these actions over time: one person/action/frame. In this paper, we use the action symbol defining from the ground truth. The major symbols are described in Table 1.

Table 1. Major symbols and meaning

Symbol	Meaning
a_en	A person appear in interested area
a_st	Act standing action
a_wk	Act walking action
a_si	Act sitting action
a_bn	Act bending action
a_ly	Act laying action
a_ex	A person disappear from interested area

Major symbol consist of five basic actions and two appearing statuses (enter and exit from interested area).

Secondly, the minor symbols as the additional features are defined from the movement of person, which are divided into seven groups: (1) action time period (start with t_) (2) movement direction (start with d_) (3) direction variation (start with dv_) (4) velocity (start with v_) (5) acceleration (start with ac_) (6) object interaction (start with oi_) (7) current location (start with lo_). The minor symbols are detailed in the Table 2.

Table 2. Minor symbols

Symbols	Meaning	Symbol	Meaning
t_l	Take the action with short time period	dv_l	Have little movement direction variation
t_m	Take the action with middle time period	dv_m	Have middle movement direction variation
t_h	Take the action with long time period	dv_h	Have large movement direction variation
t_un	Unable to specify time period for first time appearing	dv_un	Unable to specify movement direction variation for first time appearing
d_n	Move to the north	v_l	Low velocity
d_s	Move to the south	v_m	Middle velocity
d_w	Move to the west	v_h	High velocity
d_e	Move to the east	v_un	Unable to specify velocity for first time appearing
d_nw	Move to the north-west	ac_ l	Low acceleration
d_ne	Move to the north-east	ac_m	Middle acceleration
d_sw	Move to the south-west	ac_h	High acceleration
d_se	Move to the south-east	ac_un	Unable to specify acceleration for first time appearing
d_un	Unable to specify movement direction for first time appearing and no movement action	lo_se	Person stay on seat
oi_b	Take the object out from video scene	lo_bd	Person stay in bed
oi_l	Leave the object in video scene	lo_fl	Person stay on floor
oi_no	No interaction with objects	lo_sp	Person stay near some special object that can browse or have interaction with it such as signboard or ATM

The minor symbol features are computed from location changing from frame to frame in the real-word coordinate. The unit is cm/second. However, these minor symbols are defined by floating point number, such as, distance, velocity, and acceleration, which cannot apply to our state machine recognition system. From the observation, we found that its values can be represented to the Gaussian distribution. Then, the value is bounded into three values: (1) low value (lower than -1σ) (2) middle value (between -1σ and 1σ) (3) high value (higher than 1σ). The minor symbol features are detailed in the Fig. 3.

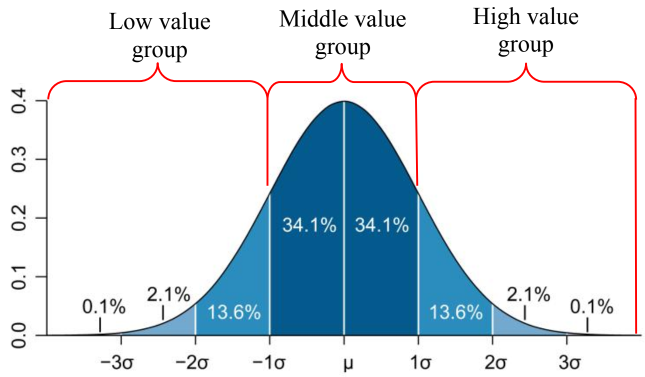


Fig. 3. Movement features clustering by Gaussian distribution

5 ACTION RECOGNITION

In this section, we defined the Finite State Machine for recognizing the five activities: walk through the scene, observation, rest, browse and idle. The definition of FSM for each activity is defined from its observed behavior, which is directly related to the stream of major and of minor symbols.

5.1 Walk through the scene

This activity means that a person walks through the interested area without stop or having interaction with any objects in the scene. FSM recognizer for this activity consist of three major symbols, which is Enter(a_en), Walk(a_wk) and Exit(a_ex). The minor symbols used in this activity are: low direction variation(dv_l), and middle direction variation (dv_m), which means that the person must walk through the scene with low changing direction. See Fig. 4 for the finite state machine.

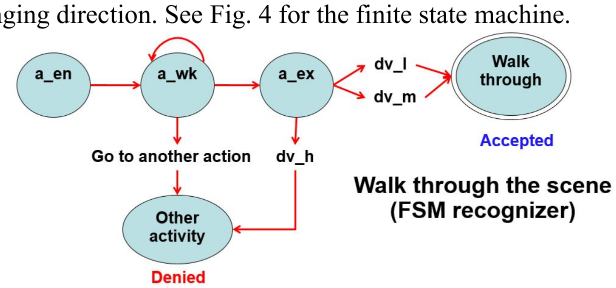


Fig. 4. FSM recognizer: Walk through the scene

5.2 Observation

This activity means that a person repeats the following action pattern, walking-standing-walking, with the large direction variation inside the interested area. So, the observation activity should have a high value of direction variation dv_h. However, the sit (a_si), lay (a_ly) and bend (a_bn) action states are not included in this FSM recognizer for preventing the undefined activities that may have much more meaning than observation activity. The Finite State Machine is shown in Fig. 5.

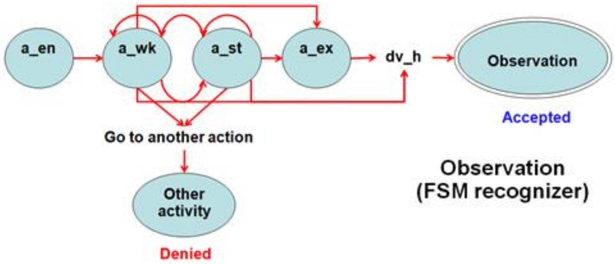


Fig. 5. FSM recognizer: Observation

5.3 Rest

This activity can be defined as two sub-activities: normal rest and abnormal rest.

a) Normal rest activity means that a person sits or lays on the normal rest area, such as, seat or bed.

b) Abnormal rest activity means that a person sits or lays on the unusual rest area, such as, corridor.

This activity is defined by two major symbols including sit (a_si) and lay (a_ly) with four additional current location symbols: seat (lo_se), bed (lo_bd), floor (lo_fl) and special object (lo_sp). See Fig. 6 for the finite state machine.

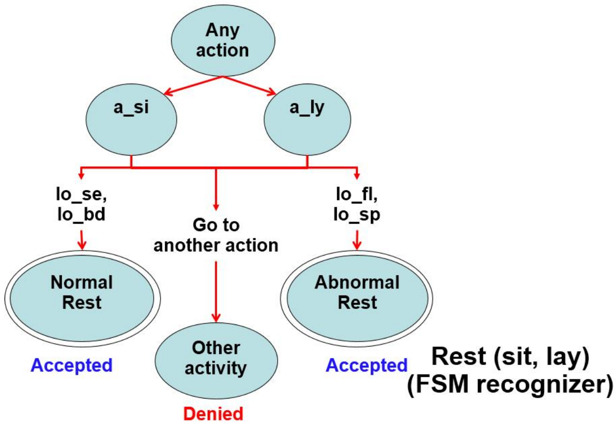


Fig. 6. FSM recognizer: Rest

5.4 Browse

This activity means that a person walks to the special object then stop near it. Browse activity is accepted when a person stand or sit near special objects that can browse or having interaction with it. This activity use the major symbols sit (a_st) and stand (s_st) with the predefined special object location (O1, O2 and O3). See Fig. 7 for the finite state diagram.

5.5 Idle

This activity means that the current sequence of symbols is not matched to any activities above.

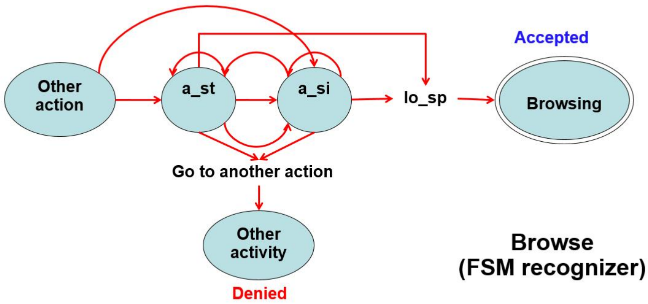


Fig. 7. FSM recognizer: Browse

6 EXPERIMENTAL RESULT AND DISCUSSION

We test our activity recognition process with the well-known dataset from CAVIA dataset. Our experiment uses 13 video-clips including: six videos for browsing, four videos for rest, and four videos for walk. For the overall testing dataset, there are 23 activities, which are divided into 10 walking, five resting, five browsing, and three observations. The recognition result for each activity is detailed in Table 3.

Table 3. Experimental result

		True Activity				
		walk through	normal rest	browse	abnormal rest	observation
Recognized Activity	walk through	10				
	normal rest		1			
	browse			5	1	
	abnormal rest				3	
	observation					1
	idle					2
Accuracy (%)		100	100	100	75	33.33

From experimental result, our activity recognition method provides a good accuracy for walk through the scene, normal rest and browse because the sequence pattern of actions for these activities can be well defined by our FSM. However, the accuracy may depend directly on the basic action identification; if the actions are correctly classified, then the activity would be properly recognized.

For the case of observation activity, it uses the high value of direction variation feature as acceptance condition. We found that the error may occur when a person make an observation activity with short walk distance during a period of time. Short movement will ignore the direction variation value, which leads to the wrong recognition result.

For the case of abnormal rest activity, we found that when a sit occur near to a special object, there have an ambiguity in recognition process, which may consider it as a browse activity. This shows that our method lack of some crucial features, such as, face direction.

Overall, our method has ability to recognize activities with a good accuracy rate, but need more additional features in the browse and rest activities.

7 CONCLUSION

In this paper we proposed an activity recognition system using the FSM. We studied and defined the specific characteristic of six human activities from the stream of five basic actions and its corresponding properties, which are necessary for the recognition of complex activities using the FSM. Our proposed method can recognize activities at 86.96% in average on using the well-known dataset. As a future work, in some activities we need more additional features for better configuring the recognition model.

8 REFERENCES

1. Liu, C., & Yuen, P. C. (2010). Human action recognition using boosted EigenActions. Image and Vision Computing, 28(5), 825–835.
2. Makris, D., & Ellis, T. (2005). Learning semantic scene models from observing activity in visual surveillance.IEEE Transactions on Systems, Man and Cybernetics, 35(3), 397–408.
3. Stauffer, C., & Grimson, E. (2000). Learning patterns of activity using real-time tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8), 747–757.
4. Oliver, N. M., Rosario, B., & Pentland, A. P. (2000). A Bayesian computer system for modeling human interactions.IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8), 831–843.
5. González, J., Rowe, D., Varona, J., & Roca, F. X. (2009). Understanding dynamic scenes based on human sequence evaluation.Image and Vision Computing, 27(10), 1433–1444.
6. Jan, T., Piccardi, M., Hintz, T., (2002). Detection of suspicious pedestrian behavior using modified probabilistic neural network. InProceedings of Image and Vision Computing (pp. 237–241).
7. Htike, Z.Z.; Egerton, S.; Kuang Ye Chow, "Monocular viewpoint invariant human activity recognition," Robotics, Automation and Mechatronics (RAM), 2011 IEEE Conference on , vol., no., pp.18,23, 17-19 Sept. 2011.

8. Shariat, S.; Pavlovic, V., "Isotonic CCA for sequence alignment and activity recognition," Computer Vision (ICCV), 2011 IEEE International Conference on , vol., no., pp.2572,2578, 6-13 Nov. 2011.
9. Gupta, S.K.; Kumar, Y.S.; Ramakrishnan, K. R., "Learning Feature Trajectories Using Gabor Filter Bank for Human Activity Segmentation and Recognition," Computer Vision, Graphics & Image Processing, 2008. ICVGIP '08. Sixth Indian Conference on , vol., no., pp.111,118, 16-19 Dec. 2008.
10. Kang Li; Yun Fu, "ARMA-HMM: A new approach for early recognition of human activity," Pattern Recognition (ICPR), 2012 21st International Conference on , vol., no., pp.1779,1782, 11-15 Nov. 2012.
11. Ping Guo; Zhenjiang Miao, "Multi-person activity recognition through hierarchical and observation decomposed HMM," Multimedia and Expo (ICME), 2010 IEEE International Conference on , vol., no., pp.143,148, 19-23 July 2010.
12. Uddin, M.Z.; Nguyen Duc Thang; Tae-Seong Kim, "Human Activity Recognition via 3-D joint angle features and Hidden Markov models," Image Processing (ICIP), 2010 17th IEEE International Conference on , vol., no., pp.713,716, 26-29 Sept. 2010.
13. Chun Zhu, Weihua Sheng, Motion- and location-based online human daily activity recognition, Pervasive and Mobile Computing, Volume 7, Issue 2, April 2011, Pages 256-269.
14. Hongqing Fang; Lei He, "BP Neural Network for Human Activity Recognition in Smart Home," Computer Science & Service System (CSSS), 2012 International Conference on , vol., no., pp.1034,1037, 11-13 Aug. 2012.
15. Antonio Fernández-Caballero, José Carlos Castillo, José María Rodríguez-Sánchez, Human activity monitoring by local and global finite state machines, Expert Systems with Applications, Volume 39, Issue 8, 15 June 2012, Pages 6982-6993.
16. L. Rodriguez-Benitez, C. Solana-Cipres, J. Moreno-Garcia, L. Jimenez-Linares, Approximate reasoning and finite state machines to the detection of actions in video sequences, International Journal of Approximate Reasoning, Volume 52, Issue 4, June 2011, Pages 526-540.
17. Chun Yuan; Wei Xu, "Multi-object events recognition from video sequences using extended finite state machine," Image and Signal Processing (CISP), 2011 4th International Congress on , vol.1, no., pp.202,205, 15-17 Oct. 2011.
18. Trinh, H.; Quanfu Fan; Jiyan Pan; Gabbur, P.; Miyazawa, S.; Pankanti, S., "Detecting human activities in retail surveillance using hierarchical finite state machine," Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on , vol., no., pp.1337,1340, 22-27 May 2011.
19. Teixeira, T.; Deokwoo Jung; Dublon, G.; Savvides, A., "Recognizing activities from context and arm pose using finite state machines," Distributed Smart Cameras, 2009. ICDSC 2009. Third ACM/IEEE International Conference on , vol., no., pp.1,8, Aug. 30 2009-Sept. 2 2009.