

A Formal Fuzzy Framework for Representation and Recognition of Human Activities

Suphot Chunwiphat[†], Patrick Reignier[‡] and Augustin Lux[‡]

[†]Department of Electronic Engineering Technology, College of Industrial Technology
King Mongkut's University of Technology North Bangkok
1518 Pibulsongkram Road, Bangsue, Bangkok 10800, Thailand

[‡]LIG – PRIMA – INRIA Rhône-Alpes,
655 avenue de l'Europe, Montbonnot, 38334 Saint Ismier cedex, France
spcp@kmutnb.ac.th , {Patrick.Reignier, [Augustin.Lux](mailto:Augustin.Lux@inrialpes.fr)}@inrialpes.fr

Abstract This paper focuses on the problem of human activity representation and automatic recognition. We first describe an approach for human activity representation. We define the concepts of roles, relations, situations and temporal graph of situations (the context model). This context model is transformed into a Fuzzy Petri Net which naturally expresses the smooth changes of activity states from one state to another with gradual and continuous membership functions. Afterward, we present an algorithm for recognizing human activities observed in a scene. The recognition algorithm is a hierarchical fusion model based on fuzzy measures and fuzzy integrals. The fusion process nonlinearly combines events, produced by an activity representation model, based on an assumption that all occurred events support the appearance of a modeled scenario. The goal is to determine, from an observed sequence, the confidence factor that each modeled scenario (predefined in a library) is indeed describing this sequence. We have successfully evaluated our approach on the video sequences taken from the European CAVIAR project¹.

1 Introduction

As one of the most active research areas in computer vision, human activity analysis is currently receiving a great interest in computer vision research community. This is due to its promising applications in many areas such as visual surveillance, human machine interaction, content-based image storage and retrieval, video con-

¹ European CAVIAR project/IST 2001 37540: <http://homepages.inf.ed.ac.uk/rbf/CAVIAR/>

ferencing, etc. One of the major problems in such systems is how the system can produce the high-level semantic interpretation of human activities from the low-level numerical pixel data.

This paper focuses on the representation and recognition of human activities for a generic human activity interpretation system. We propose formalism for context aware observation to describe and model human activities. Then the activity models will be transformed into a graphical model, a fuzzy Petri net, for analyzing the activities in a mathematical way. Finally, we present a hierarchical fusion model, based on fuzzy measures, for recognizing the human activities.

The rest of this paper is organized as follows. Related work is discussed in Section 2. Section 3 describes architecture for human activity modeling. Section 4 presents our technique for representing human activities. A hierarchical fusion model for activity recognition is proposed in Section 5. Section 6 presents experimental results. Section 7 summarises the paper and discusses future work.

2 Related Work

Madabhushi [6] presented a Bayesian network approach to recognize human activity by making an assumption that the change of position of human head is related with some human action. A more complicated Bayesian network is proposed by [5]. Their framework contains a chained hierarchical representation that describes scenarios from general properties of the moving objects. One practical difficulty in applying Bayesian networks is that they typically require initial knowledge of many probabilities.

Yamato [10] proposed a method based on Hidden Markov Models (HMMs) for recognizing different tennis strokes. Recently, Duong [3] proposed the Switching Hidden Semi-Markov Model (S-HSMM), which is a two-layered extension of HSMM, to recognize human activities of daily living (ADL). However, their work only focused on detecting a more subtle form of abnormalities, which are only abnormalities in the state duration, not in the state order.

Ghanem [4] presented an interactive system for querying events on surveillance video. The Petri nets are provided with primitive events detected from video streams and are used as complex filters to recognize composite events. Although Petri nets provide a nice graphical representation and support the representation of sequentiality, concurrency and synchronization of events. However, the firing process of transitions in Petri nets has a binary characteristic. Moreover, Petri nets themselves lack a mechanism for aggregating values issued by their places.

Shi et al. [8] presented a model, called Propagation Networks (P-Nets), for representing and recognizing partially ordered sequential activities. To recognize activities, they proposed a discrete particle filter based search algorithm, called D-Condensation. However, many activities form a network rather than a sequence and may often exhibit loops. The recognition algorithm is difficult to be adapted for dealing with such activities.

3 A Formalism for Context Aware Observation

The core component of our architecture is a situation model [2]. This framework organizes the observation of interaction using a hierarchy of concepts: scenario (context), situation, role and relation. A *scenario* is a composition of situations that share the same set of roles and relations. A *role* is an agent that performs certain action, while a *relation* describes a connection among objects that play the roles in the situation. Thus, *situations* are a form of state defined over observations.

As an example, we consider a simple video, called “Browsing”, from the CAVIAR project. The situation model for this video is described by an occurrence of three situations and the related roles and relations as follows:

Situation 1 (s_1): A person walks toward an information desk.

Role: *Walker* (anyone walking in the scene)

Relation: *Toward* (Walker heading toward an information desk)

Relation: *Close* (small distance between the walker and the information desk)

Relation: *Slow* (speed)

Situation 2 (s_2): The person stops to read some information at the information desk.

Role: *Browser* (anyone immobile)

Relation: *Toward* (direction heading toward an information desk)

Relation: *Very close* (distance being very close to the information desk)

Relation: *Very slow* (speed)

Situation 3 (s_3): The immobile person starts to walk away from the information desk.

Role: *Walker* (anyone walking in the scene)

Relation: *Away* (direction heading away from the information desk)

Relation: *Close* (distance being close to the information desk)

Relation: *Slow* (speed)

The situation network (context model), associated with the scenario is shown in Fig. 1. An arc between two situations is the temporal operator “followed by”. The transition between two situations is triggered by an event. The event corresponds to changes of the roles or relations of the corresponding situations.

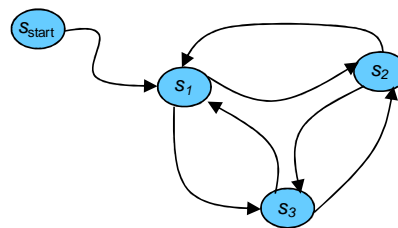


Fig. 1. Situation network of scenario “Browsing”.

In the next section, we propose to transform the situation model into a Petri net. As described above, the transition firing in Petri Nets is the instantaneous change from one state to another. However, in real world situations, the change of human activities from one state to another is not binary. In the next section, a method for relaxing this binary character of the transition firing in Petri nets is presented.

4 Fuzzy Petri Nets Representation

4.1 Fuzzy Transition Condition [1]

Let us consider a fuzzy set A in Fig. 2(a). This fuzzy set can be interpreted as a condition, called *fuzzy condition*, used to describe the concept of “the sensor value is close to 7”. When the fuzzy condition is assigned to a transition in Petri nets, the transition firing will be defined by duration according to the membership function. The duration of a transition firing can be described by the support of fuzzy set as shown in Fig. 2(a). From Fig. 2(b), the occurrence of an event “ x ” on the transition t_1 is associated to the fuzzy condition represented by the fuzzy set A . The firing of t_1 will begin as soon as the support of the condition is reached and it terminates when the event has crossed this support completely. During this firing, we will consider that the token is on both input and output place. The functions on the input and output place will proceed simultaneously. The two corresponding situations are simultaneously active.

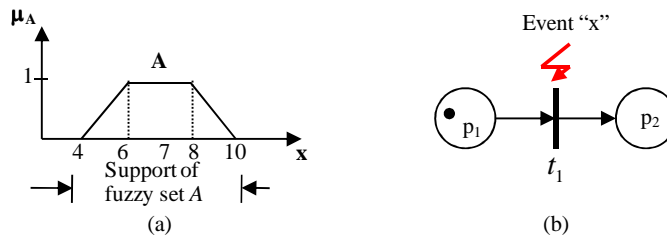


Fig. 2. (a) A Concept “the sensor value is close to 7”. (b) An Occurred event “ x ” at t_1 .

4.2 Human Activity Representation by Fuzzy Petri Nets

To represent a situation model with a Fuzzy Petri Net, situations will be represented by places and transitions by fuzzy conditions. Situations are defined with roles and relations. A role is an acceptance test. We have used a Support Vector Machines approach to automatically learn from objects’ properties the person’s roles (we have used the LIBSVM library from Chang and Lin whose software is available at <http://www.csie.nut.edu.tw/~cjlin/libsvm>). A relation is a predicate on entities selected by roles.

In Fig. 3, place s_1 , s_2 and s_3 represent three situations of scenario “Browsing” (see Sect. 3). The signs \rightarrow represent the changes of relation truth value from one situation to other. The fuzzy condition function that describes, for example, the change of speed between s_1 and s_2 can be constructed by creating a link between

“Slow” and “Very slow” of the speed relation, and “Close” and “Very close” of the distance relation. The transition occurs when two membership functions are true simultaneously: “Slow” AND “Very slow”. The logical connective AND can be implemented by the intersection operator (\cap) in fuzzy sets. Examples of membership functions representing the speed relations, “Very slow” and “Slow” in situation s_2 and s_1 are show in Fig. 4(a) and (b). The shaded are shown in Fig. 4(c) represents the transition between “Slow” AND “Very slow”.

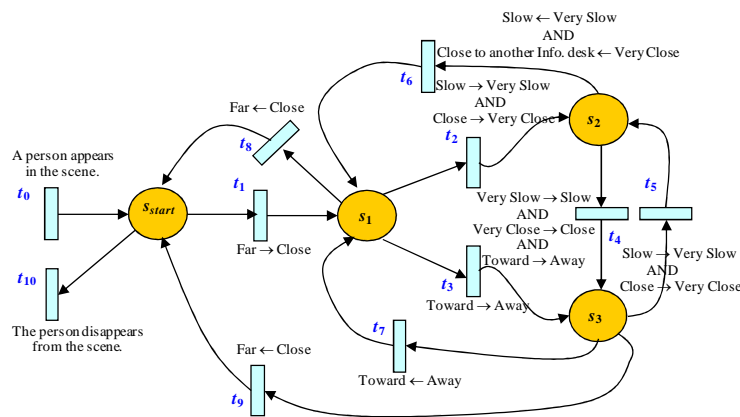


Fig. 3. A fuzzy Petri net model for the scenario “Browsing”

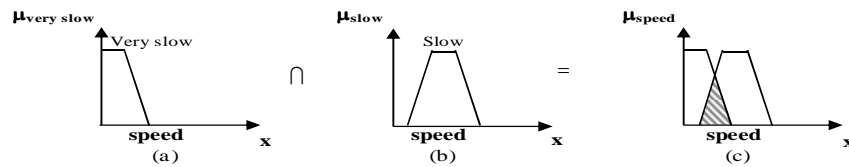


Fig. 4. (a) Fuzzy term “very slow”. (b) Fuzzy term “slow”. (c) Fuzzy condition attached to t_2 .

The fuzzy condition describing the transition between “Close” and “Very close” for the distance relation is built using the same approach. Finally, we can show the firing rule attached to transition t_2 in Fig. 3 as follows:

IF [(**Speed** is *Slow*) AND (**Speed** is *Very slow*)] AND
 [(**Distance** is *Close*) AND (**Distance** is *Very close*)]
THEN Fire the transition t_2

The complete representation of the scenario “Browsing” by using fuzzy Petri nets shown in Fig. 3, places can be considered as a classifier. Namely when a token is inserted into a place, functions that support roles and relations in the situations will calculate the values of roles and relations for confirming the state of the situation: it is interpreted as the confidence value for the recognition of this situation. A simple sequence of situations as well as their confidence values, obtained

by the evolution of a fuzzy Petri net model, can be shown in the forms of $s_1, s_2, \dots, s_n \Rightarrow h_1, h_2, \dots, h_n$. Where n is the number of situations in a model. s_1, s_2, \dots, s_n is a sequence of situations provided by the evolution of a fuzzy Petri model. h_1, h_2, \dots, h_n is a set of confidence values associated to s_1, s_2, \dots, s_n .

5 Scenario Recognition Algorithms

A Fuzzy Petri Net is a scenario model. When using real perception data, all the situations might not be perceived. Each time a situation of a scenario model is recognized, it can be interpreted as a new evidence that the corresponding scenario is recognized. The problem now is how we can fuse all those evidences to estimate how the scenario corresponds to what is currently observed.

5.1 Fuzzy Measures and Fuzzy Integrals

Let $X = \{x_1, x_2, \dots, x_n\}$ be a finite set and let $P(X)$ denote the power set of X . A fuzzy measure on a set X is a function $g: P(X) \rightarrow [0, 1]$ such that

$$g(\emptyset) = 0, g(X) = 1 \text{ and } g(A) \leq g(B), \text{ if } A \subset B \text{ and } A, B \in P(X)$$

Following the definition, Sugeno [9] introduced the so-called g_λ -fuzzy measure satisfying the following additional property

$$g(A \cup B) = g(A) + g(B) + \lambda g(A)g(B) \quad (1)$$

for all $A, B \subset X$ and $A \cap B = \emptyset$, and for some fixed $\lambda > -1$. The value of λ can be found from the boundary condition $g(X) = 1$ by solving the following equation

$$\lambda + 1 = \prod_{i=1}^n (1 + \lambda g^i) \quad (2)$$

where g^i is called a fuzzy density value. Let $A_i = \{x_i, x_{i+1}, \dots, x_n\}$. When g is a g_λ -fuzzy measure, the values of $g(A_i)$ can be computed recursively [9]. Murofushi [7] proposed the so-called Choquet fuzzy integral which can be determined in the following form

$$\int_X h(x) \circ g(\cdot) = \sum_{i=1}^n [h(x_i) - h(x_{i+1})] g(A_i), \quad (3)$$

where $h(x_1) \geq h(x_2) \geq \dots \geq h(x_n)$ and $h(x_{n+1}) = 0$. In this paper, $h(x_i)$ represents the confidence value produced by each situation.

5.2 A Hierarchical Fusion Model for Evidence Combination

Consider a scenario composed of three situation sources, i.e. $S = \{s_1, s_2, s_3\}$ together with density values (degrees of importance of the situations) $g^1 = 0.14, g^2 =$

0.45 and $g^3 = 0.12$. The fuzzy measure on the power set of S can be calculated by Eq. (1) and shown in Table 1.

Table 1. Fuzzy measure on the power set of S.

Subset X	$g_{1.953}(X)$
\emptyset	0
$\{s_1\}$	0.14
$\{s_2\}$	0.45
$\{s_3\}$	0.12
$\{s_1, s_2\}$	0.71
$\{s_1, s_3\}$	0.29
$\{s_2, s_3\}$	0.67
$\{s_1, s_2, s_3\}$	1.0

The power set of S enumerates situations that can occur from a combination of situation sources. From Table 1, we can consider that the fuzzy measure, $g(X)$, is the values for confirming the occurrence of combined situations. Thus, we use these values as density values and determine them to the situations that arise from a combination of situation sources. The hierarchical model for the combination of three

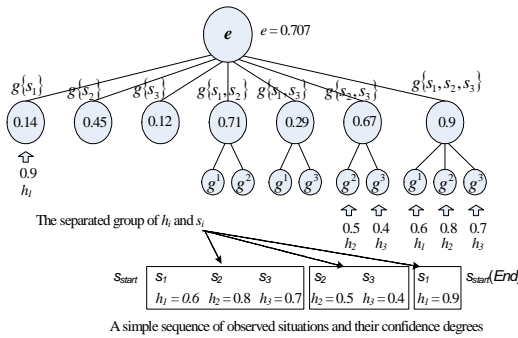


Fig. 6. Hierarchical model for evidence combination.

situation sources is shown in Fig. 6. The nodes at the middle level will be determined as the degrees of importance of combined situations using the values of fuzzy measure in Table 2. We reduce the degree of importance of the sequence $\{s_1, s_2, s_3\}$ from 1.0 to 0.9 in order to prevent the effect of source that is over confident. To combine the evidence, a sequence of occurred situations, s_i , and its confidence values, h_i , are first separated according to the subset of the power set of S (see Fig. 6). Then the separated group of h_i will be aggregated by using Eq. (3) at the bottom level of the model. Afterward the aggregated values will be conveyed up to the nodes at the middle level and aggregated by Eq. (3) again. Finally, the aggregated value at the middle level will be propagated to the top level for representing the confidence value of the scenario, e .

6 Experimental Results

We have created four different models of activities for the European Caviar project: “Browsing”, “Leaving bag behind”, “Two people fighting” and “People

meeting walking together and splitting up”. Table 2 only shows the results produced by the “Browsing” model when all sample videos are feed as an input.

The model produces the best confidence value when the most relevant video is presented to the model. We can also classify the videos pertinent to the “Browsing” scenario from the confidence value that the model produces on each video. The overall confidence value is rather high on the video “Person leaving bag but then pick it up again” because the activities contained in these video can also be interpreted as “Browsing” (the person leaves a bag near a desk, stays for a while and then goes away with the bag). Other videos may consist of activities that support the “Browsing” scenario partly.

7 Conclusions and Future Work

In this paper, we have first defined the concepts and terms for human activities modelling. Then, we have presented a fuzzy Petri net based-model in which the human activities are represented and analyzed in a graphical way. With fuzzy transition functions, the change of activity situations proceeds gradually and smoothly. To recognize an observed scenario, we have presented a hierarchical fusion model that nonlinearly combines evidence based fuzzy measures. The experimental results confirm that our proposed framework is optimal to represent and recognize human activities based on the following reasons: 1) our representation model allows a flexible and extendable representation of human activities, 2) the importance of information sources is taken account, which makes the process of evidence combination more consistent with real world situations. In the future, we plan to develop sophisticated temporal constraints in order to build a more generic model for representing scenarios in complicated situations.

Table 2. Confidence values produced by the “Browsing” model.

Input Videos	“Browsing” Model												(e)
	Sequence of Situations												
	s ₁	s ₂	s ₃	s ₁	s ₂	s ₃	s ₁	s ₂	s ₃	s ₁	s ₂	s ₃	
Person browsing back and forth	0.53	0.88	0.89	0.42	0.71	0.87	-	-	-	-	-	-	0.6322
Person browsing and reading for a while	0.68	0.86	0.84	-	-	-	-	-	-	-	-	-	0.7184
Person browsing and reading with back turned	0.48	0.71	0.93	-	-	-	-	-	-	-	-	-	0.5959
Person leaving bag by wall	0.53	-	0.78	0.65	-	0.94	0.19	-	0.32	-	-	-	0.14
Person leaving bag at chairs	0.42	-	0.93	-	-	-	-	-	-	-	-	-	0.1406
Person leaving box	0.38	-	0.89	-	-	-	-	-	-	-	-	-	0.1287
Two people meet, fight and run away *													
Person's ID 6	0.14	-	0.81	-	-	-	-	-	-	-	-	-	0.0626
Person's ID 7	0.42	-	0.43	-	-	-	-	-	-	-	-	-	0.1224
Two other people meet, fight and run away	0.54	-	0.84	-	-	-	-	-	-	-	-	-	0.1676
Two people meet, fight and chase each other	0.38	-	0.9	-	-	-	-	-	-	-	-	-	0.13
Two people meet, walk together and split*													
Person's ID 1	0.22	-	0.83	-	-	-	-	-	-	-	-	-	0.085
Person's ID 3	0.24	-	0.87	-	-	-	-	-	-	-	-	-	0.0855
Two people meet, walk together	0.44	-	0.92	-	-	-	-	-	-	-	-	-	0.144
Two other people meet, walk together	0.62	-	0.68	-	-	-	-	-	-	-	-	-	0.182
Person leaving bag but then pick it up again	0.16	-	0.93	0.8	0.71	0.89	0.47	0.93	0.91	0.55	-	0.93	0.702

*There is more than one person whose actions provoke the evolution of situations in the “Browsing” model.

References

1. D. Andreu, J. C. Pascal, R. Valette, Fuzzy Petri Net-Based Programmable Logic Controller, *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetic*, 27(6), 1997, pp. 952-961.
2. J. L. Crowley, J. Coutaz, G. Rey, P. Reignier, Perceptual Components for Context Aware Computing, In *Proceedings International Conference on Ubiquitous Computing*, 2002, pp. 117-134.
3. T. V. Duong, H. H. Bui, D. Q. Phung, S. Venkatesh, Activity Recognition and Abnormality Detection with the Switching Hidden Semi-markov Model. In *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, pp. 838-845.
4. N. M. Ghanem, Petri Net Models for Events Recognition in Surveillance Video, PhD thesis, University of Maryland, USA, 2007.
5. S. Hongeng, F. Bremond, R. Nevatia, Bayesian Framework for Video Surveillance Application, In *Proceedings 15th International Conference on Pattern Recognition*, 2000, pp. 164-170.
6. A. Madabhushi, J. K. Aggarwal, A Bayesian Approach to Human Activity Recognition, In *Proceedings 2nd International Workshop on Visual Surveillance*, Fort Collins, 1999, pp. 25-32.
7. T. Murofushi, M. Sugeno, An Interpretation of Fuzzy Measures and the Choquet Integral as an Integral with respect to a Fuzzy Measure, *Fuzzy Sets System*, 29(2), 1989, pp. 201-227.
8. Y. Shi, D. Huang, Y. Minnen, A. F. Bobick, I. A. Essa, Propagation networks for recognition of partially ordered sequential action, In *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004, pp. 862-869.
9. M. Sugeno, Fuzzy Measures and Fuzzy Integrals - A Survey, In: M.M. Gupta, G.N. Saridis, B.R. Gaines (Eds.), *Fuzzy Automata and Decision Processes*, Amsterdam, 1977, pp. 89-102.
10. J. Yamato, J. Ohya, K. Ishii, Recognizing Human Action in Time-Sequential Images using Hidden Markov Model. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, 1992, pp. 379-385.