Human Activities: Handling Uncertainties Using Fuzzy Time Intervals

M. S. Ryoo[1,2] and J. K. Aggarwal[1]
[1]*Computer & Vision Research Center, The University of Texas at Austin, USA*
[2]*Electronics and Telecommunications Research Institute, Korea*
{*mryoo, aggarwal*}*@ece.utexas.edu*

Abstract

Persons may perform an activity in many different styles, or noise may cause an identical activity to have different temporal structures. We present a robust methodology for recognition of such human activities. The recognition approach presented in this paper is able to handle person-dependent and situation-dependent uncertainties and variations of human activity executions. Our system reliably recognizes human activities with such execution variations, by semantically measuring the similarity between the observations generated by an activity execution and its optimal structure. The system detects fuzzy time intervals associated with low-level gestures of a person, and matches them hierarchically with the representation of the activity that the system is maintaining. Our system is tested for eight types of simple human interactions such as 'pushing' and 'shaking hands', as well as complex recursive interactions like 'fighting' and 'greeting'. The results show that the performance of our system is superior to that of the previous systems using deterministic time intervals.

1 Introduction

Human activity recognition is an important and active research area in the field of computer vision. Automated recognition of gestures or simple single-person actions such as walking and sitting has been particularly successful [14, 2, 5, 12, 7]. Recently, recognition of high-level and complex activities (e.g. fighting and stealing) is gaining increasing amount of interest, because of its applications in other surveillance systems, sports play analyses, and human-computer interaction systems.

Description-based approaches are a particular class of hierarchical recognition methodologies designed for analysis of high-level activities [9, 13, 3, 10, 11]. The motivation behind description-based approaches is to recognize human activities by maintaining the activities' temporal structure. Using time intervals and temporal predicates [1] to represent the structure of each activity, previous approaches have obtained successful results on recognizing high-level human activities, by searching for visual inputs that satisfies the activities' structure. Description-based approaches are able to overcome the limitations of previous statistical and syntactic approaches [4, 6] on recognizing concurrently organized activities.

However, even though the above description-based approaches have been successful in recognizing high-level activities, they had limitations on handling structural uncertainties and variations in activity executions. Actual executions of an activity are person-dependent and situation-dependent, and their temporal structures may vary. For example, even though two persons must exchange multiple punching or kicking consecutively (right after another) in ideal cases of 'fighting', temporal gaps between punchings always exist and their durations may vary. Most of the previous description-based systems recognized activities only when their temporal relationships (i.e. temporal structure) are strictly satisfied, ignoring the variations.

We present a reliable human activity recognition methodology which handles the structural variations of an activity. When a new observation (i.e. video) containing an execution of an activity is provided, our system measures how semantically similar a given observation is to the optimal structure of the activity. This similarity measure is not deterministic but is designed to consider uncertainties of the activities' structures.

Overall process of the system is as follows. At each occurrence of gestures, we associate a fuzzy [15, 16] time interval. In contrast to a deterministic time interval used by previous approaches [13, 3, 10, 11], a fuzzy interval is able to describe a possible range of its starting time and that of its ending time as well as the confidence value associated with time frames within

the ranges. Once fuzzy intervals are calculated, a dynamic programming algorithm that we have designed and presented here is applied to measure the similarity between the detected fuzzy intervals and the structure of the activity specified in the representation. Our algorithm searches for the time points in ranges that satisfy the temporal structure specified in the activity representation while maximizing the fuzzy membership values. A logistic regression technique has been used to estimate the similarity function.

We first present a general overview of description-based human activity recognition approaches in Section 2. In Section 3, we introduce the concept of fuzzy time intervals. Our algorithm to recognize activities under variations is presented in Section 4. Section 5 shows our experimental results on recognition of human-human interactions, and Section 6 concludes the paper.

2 Description-based activity recognition

A description-based activity recognition approach is an approach that uses a representation of a human activity describing its temporal, spatial, and logical structure to recognize the activity. The representation can be viewed as a definition of the activity. Only when time intervals associated with sub-events show similar structure to the definition of the activity, the system should deduce that the activity occurred. For example, in the case of 'pushing' interactions between two persons, humans consciously or unconsciously know that one has to 'stretch his/her arm' and the other has to 'fall back' as a consequence. We are also aware that the person pushing has to 'touch' the other while pushing is happening. By making the system to maintain such descriptive knowledge on 'pushing' and to search for observations that satisfy it, the system is able to recognize the activity.

We adopt the activity representation syntax developed in [10] to describe the structure of an activity formally. Figure 1 (a) shows an example representation of the activity 'push', describing the temporal relationship among time intervals associated with the sub-events. What we present throughout the paper is a new robust activity recognition methodology, which matches observations with representations while handling structural variations of activities. In order for the recognition system to be reliable and flexible, even when an observation is not strictly identical to the representation (e.g. Figure 1 (b)), the system must match it with the representation and measure how similar it is, as illustrated in Figure 1.

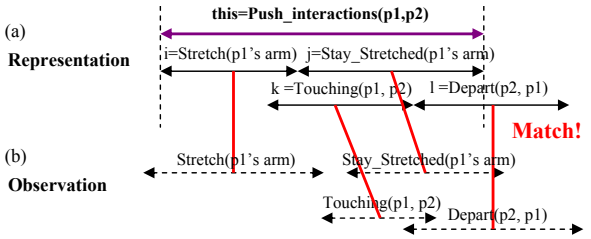

Figure 1. An example matching of 'push'.

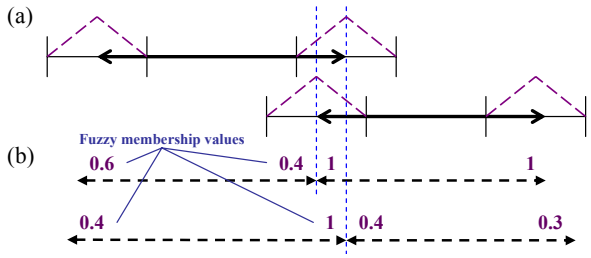

Figure 2. Example fuzzy intervals.

3 Fuzzy time intervals

In this section, we introduce the concept of 'fuzzy time intervals', which is designed to capture uncertainties and variations in activity executions. A time interval is a pair of starting time and ending time, which describes the time associated with an occurring activity or sub-event. Previously, most of approaches have used deterministic time intervals with strictly fixed starting and ending time to describe detected sub-events, and analyzed their relationships to recognize activities [13, 3, 10, 11]. Our system associates 'fuzzy intervals' for detected actions; we adopt the concept of fuzzy sets, and describes each starting time or ending time of actions as a fuzzy range of time rather than a time point. Each frame (i.e. discrete time points) within the range will have corresponding fuzzy set function value describing how confident the system is on the fuzzy interval. Fuzzy intervals are not only associated with atomic actions, but also associated with high-level actions and interactions for the hierarchical recognition.

Figure 2 shows example fuzzy intervals and possible time intervals described by fuzzy intervals. Figure 2 (a) shows an example of two fuzzy intervals whose ranges overlap slightly. Figure 2 (b) illustrates examples of possible time intervals extracted from the fuzzy intervals, which are selected to satisfy the temporal relationship $meets$. A starting or ending point of each time interval has an associated fuzzy set value describing how confident the system is on the time interval.

Figure 2 clearly illustrates why fuzzy time intervals are more desirable than traditional deterministic inter-

vals on handling variations of an activity. Assume that a structure of an activity is described as time intervals of two sub-events occurring in a sequence (i.e. $meets$). When two overlapped intervals shown in Figure 2 (a) are detected (instead of sequential ones) due to an execution variation, systems using the deterministic time intervals whose starting times and ending times are fixed to local maximums fail to recognize the activity. On the other hand, as illustrated in Figure 2 (b), the fuzzy intervals contain time intervals that $meets$ with a certain confidence, thereby enabling the recognition.

A fuzzy interval describes a set of possible time intervals. Among sets of possible time intervals described by fuzzy intervals of sub-events, there may exist particular choices that make the temporal constraints of an activity to be satisfied, as seen from Figure 2 (b). The goal is to make the system search for such selections while maximizing fuzzy values associated with intervals, so that the overall similarity using fuzzy intervals can be measured. We present the detailed algorithm to measure such similarity in the following section.

In principle, our recognition methodology is able to cope with any function as a fuzzy membership function associated with starting or ending point of a time interval. We have chosen a triangular function that is commonly used in fuzzy logic to be the fuzzy function of starting or ending point of an atomic-level action. Based on the training data, a variance of a starting or ending time of each atomic actions has been measured, and the height and the width of the triangle function have been empirically decided. The fuzzy function of higher-level activities are calculated hierarchically, as a consequence of the recognition process.

4 Recognition algorithm

In this section, we present an algorithm to recognize human activities using fuzzy intervals. We first present a methodology to recognize activities by measuring similarities between its structure and observations (i.e. detected fuzzy intervals of sub-events). A hierarchical similarity measurement is presented next.

The problem of recognizing an activity based on detection results of sub-events can be formulated as follows: Given fuzzy intervals associated with each sub-event, the goal is to search for a valid combination of time intervals within the ranges of fuzzy intervals that maximize the fuzzy values (i.e. confidence) while satisfying the temporal constraints of the activity. If the assigned fuzzy values are high enough, the system is able to deduce that given fuzzy intervals are similar to the activity's structure and conclude that the activity occurred. In order to integrate fuzzy values associated

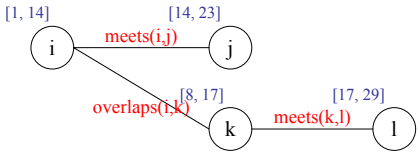

Figure 3. An example temporal graph.

with sub-events to calculate the overall confidence of the occurring activity formally, we have used a logistic regression technique. Confidence of the activity is computed as a weighted sum of starting and ending times' fuzzy values fitted into the logistic function.

Let $(v_1, ..., v_n)$ be time intervals within the ranges of fuzzy intervals of n sub-events, and $(x_1, ..., x_n)$ be their fuzzy membership values. Then, overall fuzzy confidence of the activity, L, is measured as

$$L = max(L(x_1, ..., x_n)) = logit^{-1}(max(F(x_1, ..., x_n))) \quad (1)$$

where

$$F(x_1, ..., x_n) = b + a_1 \cdot (x_1^s + x_1^e) + ... + a_n \cdot (x_n^s + x_n^e) \quad (2)$$

where x_k^s indicates the fuzzy value of the starting time of v_k and x_k^e indicates that of ending time. The function $logit$ is defined as $logit(p) = ln(p/(1-p))$, and $a_1, ..., a_n$ and b are constant weight values which need to be trained.

The system is required to maximize the $F(x_1, ..., x_n)$ function while meeting temporal constraints posed for $(v_1, ..., v_n)$. There exist various temporal constraints that the time intervals have to satisfy depending on the representation of the activity. Most trivial constraint is that the starting time of a sub-event can not exceed its ending time. Representation of the activity also specifies other constraints using temporal predicates. For example, if the representation contains $before(v_1, v_2)$, then the ending time of v_1 must be strictly less than the starting time of v_2. Choosing time point 9 for v_1^e and 5 for v_2^s leads to a contradiction, regardless their fuzzy values.

In order to compute $max(F(x_1, ..., x_n))$ while satisfying the constraints, we have developed a dynamic programming algorithm. We first convert temporal representation of an activity into an undirected acyclic graph representation (i.e. tree) where each node corresponds to a time interval and each edge specifies that two intervals are required to satisfy a particular relationship. An edge is labeled with the relationship that needs to be satisfied between the two nodes (e.g. $during(v_1, v_2)$). Multiple graphs may be constructed from disjunctive normal form (DNF) of the representation. Figure 3

shows an example temporal graph of the interaction 'push' mentioned in Figure 1.

We formulate the recursive equation as:

$$G_k(t) = a_k \cdot (x_k^s + x_k^e) + \sum_{all\ v_c} max_{\{t'\}} G_c(t') \quad (3)$$

where v_c are child nodes of v_k, and t' are time intervals that satisfies temporal relations with the interval t. $G_k(t)$ specifies the maximum weighted sum of possible assignments for x_k and its descendant nodes, if the interval t is assigned for x_k. Therefore, the similarity measure $L(x_1, x_2, ..., x_n)$ are enumerated as follows:

$$\begin{aligned} max(L(x_1, ..., x_n)) &= logit^{-1}(max(F(x_1, ..., x_n))) \\ &= logit^{-1}(max_{\{t\}} G_r(t)) \end{aligned} \quad (4)$$

where node v_r is the root node of the tree.

As a result, by solving the recursive equation using the dynamic programming algorithm, we are able to calculate the maximum L, which is the confidence of the detection. Furthermore, we are able to calculate the fuzzy interval associated with the detection. By calculating the argument maximum while computing the maximum, we also are able to compute the exact time intervals of sub-events that make the fuzzy value of the activity to be the maximum. This implies that the system is able to calculate the starting time and ending time of the special time interval 'this', which is always associated with the defining activity itself. Ranges are associated with the detected starting and ending time of 'this', making the interval to be fuzzy. The overall complexity of the algorithm is $O(m^2)$, where m is the average number of intervals within ranges per node.

We have developed a hierarchical algorithm which analyzes human activities from bottom (i.e. atomic-level actions) to top (i.e. high-level interactions). At the bottom level, the system detects atomic-level actions (e.g. arm stretching) using low-level recognition techniques such as hidden Markov models (HMMs) from Park and Aggarwal [8], and associates fuzzy intervals to describe their starting and ending time. Higher-level activities are recognized based on fuzzy intervals associated with their sub-events, which are atomic-level actions and/or other activities composed of their own sub-events. With the fuzzy interval calculation method presented above in this section, fuzzy intervals of an activity are computed based on those of sub-events, enabling the recognition of high-level activities.

5 Experiments

We have evaluated the performance of our system using fuzzy time intervals, while comparing it with the previous systems [10, 11] using deterministic intervals. Eight types of relatively simple interactions between humans (*approach*, *depart*, *point*, *shake-hands*, *hug*, *punch*, *kick*, and *push*), as well as complex recursive interactions of *fighting* and *greeting* have been tested by the systems. We have used the dataset used in [10, 11], which contains sequences of continuous executions of activities in 320*240 resolutions at 15 fps. Complex fighting-related sequences containing a total of 53 simple and recursive activities have been newly added. As a result, a total of 161 activity executions have been tested for both systems. HMMs for gesture recognition and logistic regression weights, $a_1, \cdots, a_n$ and b, have been estimated based on a separate training set.

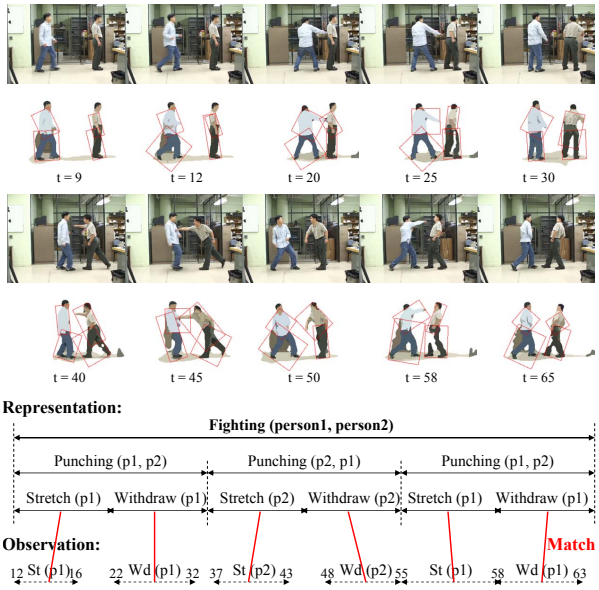

Figure 4. Example experimental results of the recursive activity 'fighting', composed of three consecutive 'punching' interactions.

The experimental results clearly illustrate that the recognition accuracy of our system is better than that of the previous system. Table 1 compares true positive rates obtained from two systems whose false positive rates are similar. The result confirms that the use of fuzzy time intervals helps reliable recognition of activities from noisy videos with structural execution variations. Figure 4 shows a successful recognition result of our system tested on a *fighting* interaction composed of three punching interactions, which the previous deterministic systems failed recognition due to its structural variation. Figure 5 shows a recognition result of *pushing*. Even though the structure of the gesture recognition results was slightly different from the representa-

System	Simple	Recursive	Total
Ours	0.920	0.783	0.907
Previous	0.862	0.522	0.814

Table 1. Recognition accuracy.

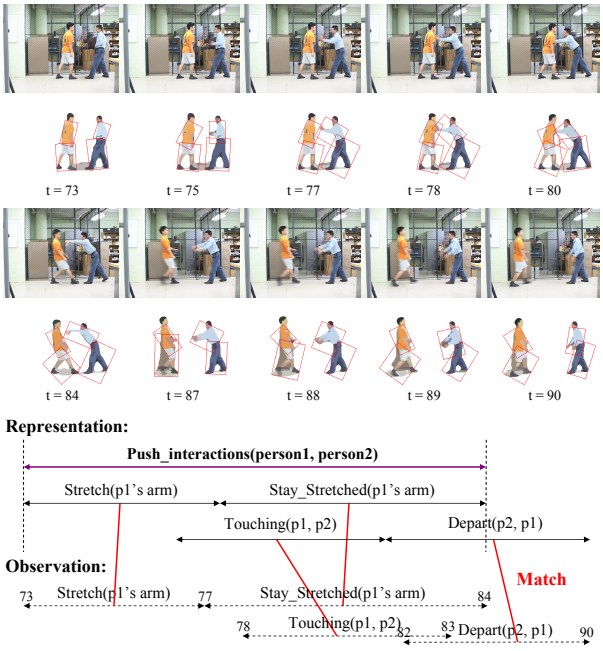

Figure 5. Example experimental results of the 'pushing' interaction.

tion, our system was able to recognize the *pushing* interaction. False positive rates were almost 0 for both systems, since the probability of sub-events satisfying particular relations detected 'by accident' is extremely low.

6 Conclusion

We have presented a reliable recognition methodology that is able to handle uncertainties in human activities' structure. We have introduced the concept of 'fuzzy time intervals', and presented the dynamic programming algorithm to calculate the similarity between the activity and the observations. Experimental results suggest that the ability to handle structural variations enables better recognition of human activities.

Acknowledgment

The research was supported in part by Texas Higher Education Coordinating Board award # 003658-0140-2007.

References

[1] J. F. Allen and G. Ferguson. Actions and events in interval temporal logic. *Journal of Logic and Computation*, 4(5):531–579, 1994.

[2] A. Bobick and J. Davis. The recognition of human movement using temporal templates. *IEEE T PAMI*, 23(3):257–267, Mar 2001.

[3] S. Hongeng, R. Nevatia, and F. Bremond. Video-based event recognition: activity representation and probabilistic recognition methods. *CVIU*, 96(2):129–162, 2004.

[4] Y. A. Ivanov and A. F. Bobick. Recognition of visual activities and interactions by stochastic parsing. *IEEE T PAMI*, 22(8):852–872, 2000.

[5] I. Laptev and T. Lindeberg. Space-time interest points. In *ICCV*, page 432, Washington, DC, USA, 2003. IEEE Computer Society.

[6] N. T. Nguyen, D. Q. Phung, S. Venkatesh, and H. H. Bui. Learning and detecting activities from movement trajectories using the hierarchical hidden markov models. In *CVPR (2)*, pages 955–960, 2005.

[7] J. C. Niebles, H. Wang, and L. Fei-Fei. Unsupervised learning of human action categories using spatial-temporal words. In *BMVC*, 2006.

[8] S. Park and J. K. Aggarwal. A hierarchical Bayesian network for event recognition of human actions and interactions. *Multimedia Systems*, 10(2):164–179, 2004.

[9] C. S. Pinhanez and A. F. Bobick. Human action detection using PNF propagation of temporal constraints. In *CVPR*, page 898, 1998.

[10] M. S. Ryoo and J. K. Aggarwal. Recognition of composite human activities through context-free grammar based representation. In *CVPR (2)*, pages 1709–1718, 2006.

[11] M. S. Ryoo and J. K. Aggarwal. Semantic understanding of continued and recursive human activities. In *ICPR*, pages 379–382, 2006.

[12] C. Schuldt, I. Laptev, and B. Caputo. Recognizing human actions: a local svm approach. In *ICPR*, volume 3, pages 32–36 Vol.3, August 2004.

[13] V.-T. Vu, F. Brémond, and M. Thonnat. Automatic video interpretation: A novel algorithm for temporal scenario recognition. In *IJCAI*, pages 1295–1302, 2003.

[14] J. Yamato, J. Ohya, and K. Ishii. Recognizing human action in time-sequential images using hidden markov model. In *CVPR*, pages 379–385, 1992.

[15] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.

[16] L. A. Zadeh. Toward human level machine intelligence - is it achievable? the need for paradigm shift. *IEEE Computational Intelligence Magazine*, pages 11–22, August 2008.