# Video event recognition with Fuzzy Semantic Petri Nets *

Piotr Szwed

**Abstract** Automated recognition of complex video events poses challenges related to: selection of formalisms for efficient event modeling and analysis, mapping semantic high-level concepts used in specifications on information extracted from video sequences, as well as managing uncertainty associated with this information. We propose Fuzzy Semantic Petri Nets (FSPN) as a tool aimed at solving the mentioned problems. FSPN are Petri nets coupled with an underlying fuzzy ontology. The ontology stores assertions (facts) concerning object classification and detected relations. Fuzzy predicates querying the ontology are used as transition guards. Places in FSPN represent scenario steps. Tokens carry information on objects participating in a scenario and have weights expressing likelihood of a step occurrence. FSPN enable detection of events occurring concurrently, analysis of various combinations of objects and reasoning about alternatives.

**Key words:** video event recognition, fuzzy Petri nets, fuzzy ontology

## 1 Introduction

Event recognition is a challenging problem, especially in areas, where observed lower level features are inherently affected by noise or uncertainty.

---

Piotr Szwed
AGH University of Science and Technology, Mickiewicza Av. 30, 30-059 Kraków
e-mail: pszwed@agh.edu.pl

* This is the draft version of the paper presented at the **ICMMI 2013**, **International Conference on Man-Machine Interactions**, October 22–25, 2013 The Beskids, Poland. The paper was published in **Advances in Intelligent Systems and Computing**, Volume 242, 2014 and is available at: http://link.springer.com/chapter/10.1007%2F978-3-319-02309-0_47

This, in particular, pertains to automated high-level video event recognition [9], where meaningful aspects of video sequences are extracted with complex multi-stage algorithms introducing inevitable errors. Hence, modeling languages used to specify events of interest and supporting tools tools should cope with uncertainty of input data. Moreover, specifications, to be meaningful and manageable, should preferably be decoupled from low level extraction methods and use semantic description of features.

The proposed Fuzzy Semantic Petri Nets (FSPN) were conceived as a tool for video event modeling and recognition. However, they are general enough to be applied to other domains. FSPN are Petri nets coupled with an underlying fuzzy ontology, which constitute an abstraction layer allowing to transform observed features, e.g. sizes of detected objects, their speed and positions into a logical description using terms defined in a controlled vocabulary.

In case of video event recognition system the ontology content is updated for each video frame by making appropriate assertions on objects and their relations. The ontology can be queried with unary or binary predicates returning fuzzy truth values from $[0, 1]$. Predicates are used in guards of transitions in FSPN controlling in that way flows of tokens. Tokens carry information on objects participating in an event and are equipped with fuzzy weights indicating likelihood of their assignment to places. In turn, places correspond to subevents (scenario steps).

The paper is organized as follows: next Section 2 reports approaches to video events specification and analysis. It is followed by Section 3, which describes the fuzzy ontology. FSPN are defined in Section 4. An example of a scenario specification and results of detection are given in Section 5. Section 6 provides concluding remarks.

## 2 Related works

Recognition of video events has been intensively researched over last fifteen years. A large number of methods is reported in recent surveys: [9] and [1]. At least two groups of approaches can be indicated. The first group includes methods using state-based models, in which transitions are attributed with probability factors learned from annotated video, e.g. Neural Networks, Hidden Markov Models and Dynamic Bayesian Networks.

The second group is comprised of methods based on descriptions of events prepared in high level languages, either textual [14], or graphical as Situation Graph Trees [12] and Petri nets [7, 3, 10]. The methods falling into this category are considered *semantic*, as specifications are prepared by experts, who give meaningful names to events, engaged objects, actions and conditions. Descriptions are often hierarchical: complex events can be expressed as graphs of subevents. In some approaches scenarios and their ingredients: types of participating objects and relations are defined as ontologies [5, 2].

Petri Nets (PN) are applied in the field of event detection in two modes [9]. In the first mode of *object PN* tokens represent objects, places object states and transitions events of interest. Such approach was applied in surveillance of traffic [7] and people [4]. In the second mode of *plan PN* places correspond to subevents building up a plan. Presence of a token in a place indicates that a particular event assigned to the place is occurring. The latter approach was applied to people surveillance [3].

To our knowledge, none of the Petri net based methods of video event recognition used fuzzy Petri nets. In [10] stochastic Petri nets were applied. The disadvantage of the method is a necessity to learn probability factors attributed to transitions from annotated video sequences.

## 3 Fuzzy ontology

Ontologies are often described as unions of two layers: terminological ($TBox$) and assertional ($ABox$). The $TBox$ defines concepts and types of relation including: taxonomic relations between concepts, *object properties* and *datatype properties*. The $ABox$, in turn, gathers facts about individuals and existent relations. In Description Logic, being a counterpart of ontology languages, concepts and relations can be expressed by means of unary and binary predicates, e.g.: $Person(x)$ – $x$ is a member of the class $Person$, $isWalking(x)$ – a boolean datatype property $isWalking$ of an individual $x$ or $isClose(x, y)$ – an object property between two individuals $x$ and $y$. In this case an $ABox$ can be treated as a particular model of formulas.

For *fuzzy ontologies* and corresponding Fuzzy Description Logics the ontology relations are extended by adding weights being real numbers from $[0, 1]$. They can be used to express uncertainty, e.g. with respect to class membership or relation occurrence. A formalization of fuzzy ontology language including fuzzy classes, roles (object properties) and datatype properties can be found in [6] and [11].

Fig. 1 gives an example of a fuzzy ontology content. Concepts, like *Person* and *SmallObject* are depicted as ovals, individuals ($a$ and $b$) as diamonds, the boolean literal *true* as rounded diamond and asserted relations as rectangles surrounding their names and weight factors.

Minimal requirements for a fuzzy ontology to be used with FSPN are related to supported queries. An ontology component should provide functionality for enumeration of classes, relations and defined individuals, testing fuzzy predicates and calculating values of logical formulas.

For queries in form of logical formulas it is required to support conjunctions of predicates. Their values can be calculated using various fuzzy logic norms. We use the *Gödel t-norm*, which calculates the minimum over a set of values attributed to compound statements. Let us take as an example the formula: $f(x, y) = Person(x) \wedge isWalking(x) \wedge SmallObject(y) \wedge isClose(x, y)$, which can be read as: *a person passes by a small object*. Its value depends on variable binding, i.e. for the ontology content in Fig. 1 $f(a,b) = 0.67$ and $f(b,a) = 0.21$ (according to Gödel t-norm).

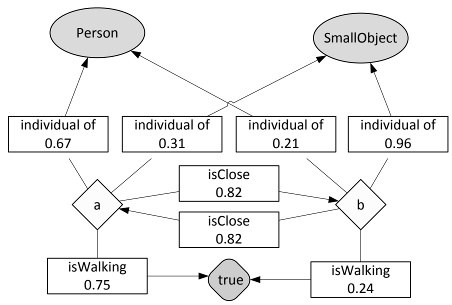



**Fig. 1** Content of a fuzzy ontology

Assertions on relations in the *ABox* are made with special functions called *evaluators*. They examine external model and calculate fuzzy weights of predicates. In opposition to approach proposed in [11] evaluators are external entities beyond the ontology. In many cases they have a form of membership functions described by line segments, as in Fig. 2.

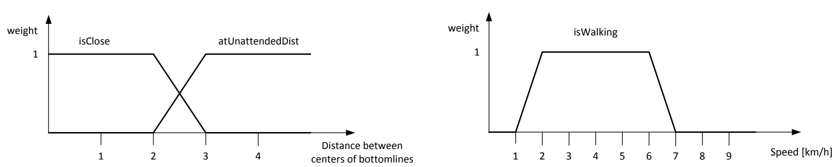



**Fig. 2** Membership functions used by *evaluators*

## 4 Fuzzy Semantic Petri Nets

Definition of FSPN is comprised of three concepts: Petri net structure, binding and fuzzy marking.

**Definition 1 (Petri net structure).** Petri net structure $PN$ is a tuple $(P, T, F, Preds, G, L, H)$, where $P$ is a set of places, $T$ is a set of transitions, $P$ and $T$ are satisfying $P \cap T = \emptyset$ and $P \cup T \neq \emptyset$. $F \subseteq P \times T \cup T \times P$ is a set of arcs (flow relation), and $Preds$ is a set of unary and binary predicates. $G\colon T \to 2^{Preds}$ is a guard function that assigns sets of predicates to transitions. $L\colon P \to \mathbb{N} \cup \{0\}$ is a function assigning lower bound to a place;

this value defines how long a token should stay in a place to be allowed to leave it. $H\colon P \to \mathbb{N} \cup \{\omega\}$ assigns upper bound to a place. The symbol $\omega$ represents infinity.

**Definition 2 (Binding).** Let $V$ be set of variables and $I$ a set of objects. *Binding* $b$ is defined as a partial function from $V$ to $I$. A variable $v$ *is bound* for a binding $b$, iff $v \in \operatorname{dom} b$. A set of all bindings is denoted by $B$.

**Definition 3 (Fuzzy marking).** A set of fuzzy tokens $FT$ is defined as $FT = B \times \mathbb{R} \times (\mathbb{N} \cup \{0\}) \times (\mathbb{N} \cup \{0\})$. Components of a token tuple $(b, w, c, \tau) \in FT$ are the following: $b \in B$ denotes a binding, $w \in [0, 1]$ is a fuzzy weight, $c \geq 0$ is a counter storing information, how long the token rests in a place and $\tau$ is a time stamp. *Fuzzy marking* for a Petri net $PN = (P, T, F, Preds, G)$ is defined as a function that assigns sets of fuzzy tokens to places $FM : P \to 2^{FT}$.

The defined above FSPN can be considered a subclass of Colored Petri Nets (CPN) proposed by Jensen [8]. Fuzziness is introduced to the model by equipping tokens with weights, what can be easily achieved in CPN.

A difference between FSPN and CPN lies in their behavior. FSPN are not intended to analyze such issues as concurrency and conflicts, but to perform a kind of fuzzy reasoning and classification of sequences of events. Hence, they process multiple tokens in one step and differently handle conflicts, what is illustrated in Fig. 3. In FSPN all conflicting transitions fire generating tokens with various weights, what allows to reason on scenario alternatives and their likelihoods.

A single execution step of FSPN is comprised of three basic stages:

1. *Firing all enabled non-initial transitions and generating new tokens.* During this stage for each input token and a transition the guard is calculated. If the guard contains free variables, they are bound to objects in the ontology *ABox*. Then the guard value (activation level) is aggregated with the input token weight and assigned to a new token. This behavior is illustrated in Fig. 3.a and Fig. 3.b.
2. *Removing and aggregating tokens.* It is assumed, that creation of a new token consumes a part of input token weight. If this value falls below a certain threshold, the input token is removed (Fig. 3.c). Also in this step, multiple tokens sharing the same binding and assigned to the same place are aggregated, as in Fig. 3.e and Fig. 3.f.
3. *Firing initial transitions.* New tokens are introduced into the net, by firing initial transitions (not having an input place). For each initial transition variables appearing in its guard are bound to objects, then the guard value is calculated and used as a weight of new tokens. A threshold (0.2) preventing from creation of tokens with a small weight is used.

The semantics of Petri nets proposed in this paper is closer to mentioned in Section 2 *plan PNs*, as tokens represent combination of objects participating

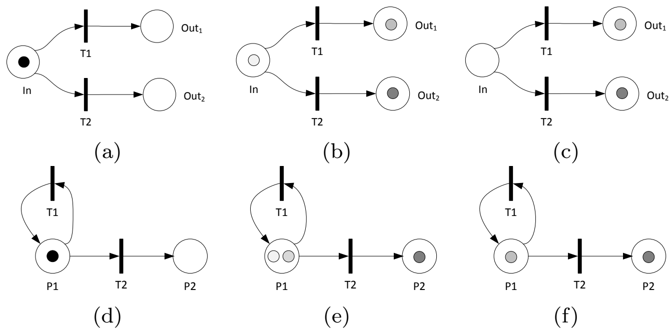



**Fig. 3** Conflict, removal and aggregation of tokens. Weights of tokens are marked with color intensity.

in scenarios. There are, however, some salient differences. 1) FSPNs do not require learning. They rely on fuzzy weights returned by carefully designed and tested evaluators. 2) In probabilistic PNs discussed in [3] in case of a conflict (e.g. two enabled transitions sharing input place with a single token) only one transition with a higher learned probability would fire, whereas in our model they both can be executed and produce two tokens with weights aggregating the weight of the input token and transition guards. This allows to reason concurrently about scenario alternatives. Moreover, a weak initial likelihood of a scenario branch can be amplified by future events. 3) In our approach all enabled transitions are executed in a single parallel step.

## 5 Initial experiments

An application of FSPN to event recognition is discussed for an abandoned luggage scenario proposed as a benchmark for PETS 2006 workshop [13]. The scenario is comprised of the following steps: (1) a man enters the scene and remains still, (2) he takes off a rucksack and puts it on the floor, (3) then moves away and leaves the luggage unattended and, finally, (4) disappears. In the published specification, the luggage is considered unattended if the distance from its owner is greater than 3 m.

A FSPN modeling the scenario is given in Fig. 4. Places $P_1 - P_4$ correspond to defined above steps. Bounds in curly braces specify how long a token should rest in a place to state that a step was achieved. Specification of guards reference classes and relations defined in the fuzzy ontology. Values of predicates can be established by evaluators that analyze sizes and positions of objects assigned to variables: $x$ (a person) and $y$ (a small object). Variable $x$ is bound at the beginning of the scenario. Binding of variable $y$ is carried out dynamically, as the transition linking places $P_1$ and $P_2$ is fired.

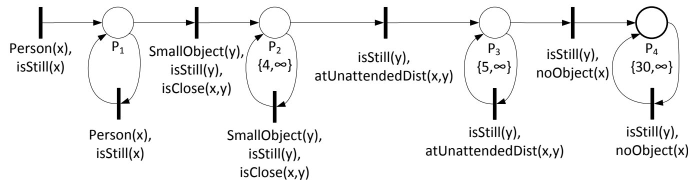
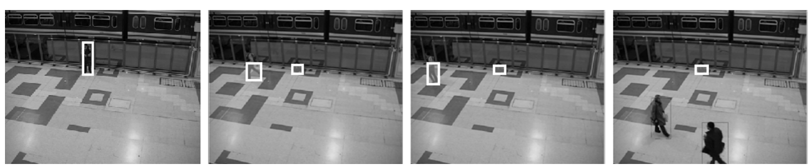



**Fig. 4** Fuzzy Semantic Petri Net representing the luggage left scenario

We have implemented an event recognition framework following the described approach. The system entirely written in Java is comprised of a lightweight fuzzy ontology component, a pluggable set of evaluators, an FSPN execution engine, as well as a GUI displaying video content with accompanying semantic information pertaining to recognized events. The system takes at input a video sequence annotated with object tracking data.

Results of a correct scenario recognition are shown in Fig. 5. Subsequent images correspond to places in the Petri net in Fig. 4.



$P_1$ $P_2$ $P_3$ $P_4$

**Fig. 5** Event recognition steps for the luggage left scenario. Images marked as $P_1$ – $P_4$ correspond to places in FSPN.

Analogous experiments were conducted for two event recognition tasks: graffiti painting and detecting violation of a surveillance zone by people walking in a certain direction. Tests for the abandoned luggage and graffiti painting events yielded 100% correct results (true positives). For a zone violation the recognition ratio was about 76%. Detailed analysis revealed that in this case the lower performance was caused by tracking problems (lost of identity in case of occlusion and in some cases invalid segmentation).

It should be observed that black-box recognition tests are related to the whole processing chain, i.e. detection, object tracking and high-level event interpretation. FSPN are intended to be applied at the last stage. The effectiveness of recognition depends on three factors:

1. Correct tracking (in particular, reliable identity assignment to detected objects);
2. FSPN design: a selection of subevents and their sequences;
3. Quality of evaluators, i.e. functions used make fuzzy assertions in the ontology. In general, it is expected that correctly implemented evaluators

should yield stable subevents: both as regards durations and amplitudes of weights.

At present we are more concerned with the two last factors. To put forward, the problem consists in analyzing video samples and selecting features that should be used in a logical specification to reason about occurrences of high-level events.

To facilitate the evaluation of a FSPN at the *design* time, the framework collects analytic information related to weights of tokens and their flows. Fig. 6 presents in form of a Gannt chart values of tokens assigned to places $P_1$–$P_4$ for the FSPN in Fig. 4 at consecutive frames. For the purpose of presentation their values were shifted by adding 2, 4 and 6 for tokens in $P_2$, $P_3$ and $P_4$. Hence, each elevation above a baseline represents a subevent occurrence. Compound subevents that do not lead to recognition of the scenario can be observed at frames 67–89 and 353–419. The expected and successfully recognized event occurrence is developed within the frames 463–782. It should be mentioned that the input tracking information was prepared by processing every third frame from a 25 FPS video clip, thus frame numbers should be multiplied by 3 to reflect the features of the original material.

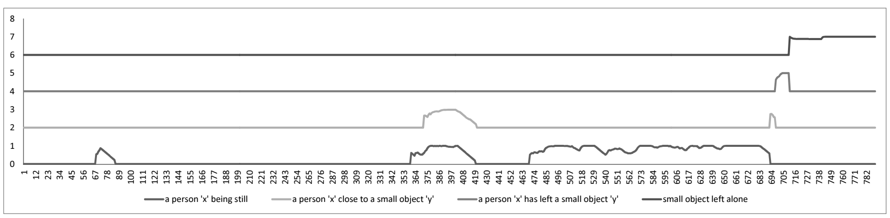



**Fig. 6** Weights of tokens assigned to places at consecutive frames

Time series in Fig. 6 were obtained by observing behavior of a validated final FSPN specification. On the way, a number of experiments has been made, some evaluators causing non-stable events were removed and some, e.g. *isStill*(), were corrected.

## 6 Conclusions

In this paper we address the problem of automatic recognition of video events and introduce Fuzzy Semantic Petri Nets, a tool allowing to specify scenarios and reason about their occurrences. FSPN can be considered a *semantic* event modeling language, both at the level of structure (places correspond to subevents) and used descriptions, as guards reference terms defined in an ontology.

FSPN are coupled with a fuzzy ontology, a logical abstraction layer linked with with underlying model of tracked objects by fuzzy predicates evaluators. Fuzziness is a mean to manage uncertainty of input, but also vagueness of terms used in events specification.

An advantage of FSPN is their capability of detecting concurrently occurring events, in which participate various combinations of objects, analyze scenario alternatives and their likelihoods. Petri nets state (marking) gives general overview of the situation, of *what's going on*. A presence of a token in a place can be reported as *semantic output*, e.g. to a surveillance system operator.

## References


1. Aggarwal, J., Ryoo, M.S.: Human activity analysis: A review. ACM Computing Surveys (CSUR) 43(3), 16 (2011)
2. Akdemir, U., Turaga, P., Chellappa, R.: An ontology based approach for activity recognition from video. In: Proceedings of the 16th ACM international conference on Multimedia. pp. 709–712. ACM (2008)
3. Albanese, M., Chellappa, R., Moscato, V., Picariello, A., Subrahmanian, V.S., Turaga, P., Udrea, O.: A constrained probabilistic Petri net framework for human activity detection in video. Multimedia, IEEE Transactions on 10(8), 1429–1443 (Dec 2008)
4. Borzin, A., Rivlin, E., Rudzsky, M.: Surveillance event interpretation using generalized stochastic Petri nets. In: Image Analysis for Multimedia Interactive Services, 2007. WIAMIS'07. Eighth International Workshop on. pp. 4–4. IEEE (2007)
5. Bremond, F., Maillot, N., Thonnat, M., Vu, V.T., et al.: Ontologies for video events. Research report number 51895. Tech. rep., INRIA Sophia-Antipolis (2004)
6. Calegari, S., Ciucci, D.: Fuzzy ontology, Fuzzy Description Logics and fuzzy-OWL. Applications of Fuzzy Sets Theory 4578/2007(D), 118–126 (2007)
7. Ghanem, N., DeMenthon, D., Doermann, D., Davis, L.: Representation and recognition of events in surveillance video using Petri nets. In: Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. Conference on. pp. 112–112 (June 2004)
8. Jensen, K.: Coloured Petri Nets: Basic Concepts, Analysis Methods and Practical Use, vol. 1. Springer (1996)
9. Lavee, G., Rivlin, E., Rudzsky, M.: Understanding video events: A survey of methods for automatic interpretation of semantic occurrences in video. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 39(5), 489–504 (Sept 2009)
10. Lavee, G., Rudzsky, M., Rivlin, E., Borzin, A.: Video event modeling and recognition in generalized stochastic Petri nets. Circuits and Systems for Video Technology, IEEE Transactions on 20(1), 102–118 (Jan 2010)
11. Lukasiewicz, T., Straccia, U.: Managing uncertainty and vagueness in description logics for the Semantic Web. Web Semantics Science Services and Agents on the World Wide Web 6(4), 291–308 (2008)
12. Munch, D., Jsselmuiden, J., Arens, M., Stiefelhagen, R.: High-level situation recognition using fuzzy metric temporal logic, case studies in surveillance and smart environments. In: Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on. pp. 882–889 (Nov 2011)
13. PETS 2006: Ninth IEEE international workshop on performance evaluation of tracking and surveillance - benchmark data (2006), `http://www.cvg.rdg.ac.uk/PETS2006/data.html`, [Online; accessed 04-March-2013]

14. Vu, V.T., Bremond, F., Thonnat, M.: Automatic video interpretation: A novel algorithm for temporal scenario recognition. In: International Joint Conference on Artificial Intelligence. vol. 18, pp. 1295–1302. Lawrence Erlbaum Associates Ltd (2003)