Uncertainty Modeling Framework for Constraint-Based Elementary Scenario Detection in Vision Systems

Carlos Fernando Crispim-Junior(✉) and Francois Bremond

INRIA Sophia Antipolis, Valbonne, France
carlos-fernando.crispim_junior@inria.fr

Abstract. Event detection has advanced significantly in the past decades relying on pixel- and feature-level representations of video-clips. Although effective those representations have difficulty on incorporating scene semantics. Ontology and description-based approaches can explicitly embed scene semantics, but their deterministic nature is susceptible to noise from underlying components of vision systems. We propose a probabilistic framework to handle uncertainty on a constraint-based ontology framework for event detection. This work focuses on elementary event (scenario) uncertainty and proposes probabilistic constraints to quantify the spatial relationship between person and contextual objects. The uncertainty modeling framework is demonstrated on the detection of activities of daily living of participants of an Alzheimer's disease study, monitored by a vision system using a RGB-D sensor (Kinect®, Microsoft©) as input. Two evaluations were carried out: the first, a 3-fold cross-validation focusing on elementary scenario detection (n:10 participants); and the second devoted for complex scenario detection (semi-probabilistic approach, n:45). Results showed the uncertainty modeling improves the detection of elementary scenarios in recall (*e.g.,* In zone phone: 84 to 100 %) and precision indices (*e.g.,* In zone Reading: 54.5 to 85.7%), and the recall of Complex scenarios.

Keywords: Uncertainty Modeling · Ontology · Event Detection · Activities of Daily Living · Older People

1 Introduction

Event detection has been significantly advancing since the past decade within the field of Computer vision giving birth to applications on a variety of domains like safety and security (*e.g.,* crime monitoring [9]), medical diagnosis and health monitoring [23][5], and even as part of a new paradigm of human-machine interface in gaming and entertainment (Microsoft© Kinect®). Event detection methods in computer vision may be categorized in (adapted from Lavee *et al.* [11]): classification methods, probabilistic graphical models (PGM), and semantic models; which are themselves based on at least one of the following data

L. Agapito et al. (Eds.): ECCV 2014 Workshops, Part II, LNCS 8926, pp. 269–282, 2015.
DOI: 10.1007/978-3-319-16181-5_19

abstraction level: pixel-based, feature-based, or event-based. Artificial Neural Networks, Support-Vector Machines (SVM), and Independent Subspace Analysis (ISA) are examples of classification methods. For instance, Le *et al.*[12] have presented an extension of the ISA algorithm for event detection, where the algorithm learned invariant spatio-temporal features from unlabeled video data. Wang *et al.* [21] have introduced new descriptors for dense trajectory estimation as input for non-linear SVMs. Common examples of PGMs approaches are Bayesian Network (BN), Conditional Random Fields, and Hidden Markov Models (HMM). BNs have been evaluated at the detection of person interactions (e.g., shaking hands) [16], left luggage [13], and traffic monitoring [9]. Kitani *et al.* [8] has proposed a Hidden Variable Markov Model approach for event forecasting based on people trajectories and scene features. Despite the advances, PGMs have difficulty at modeling the temporal dynamics of an event. Izadinia and Shah [7] have proposed to detect complex events from by a graph representation of joint the relationship among elementary events and a discriminative model for complex event detection.

Even though the two previous classes of methods have considerably increased the performance of event detection in benchmark data sets, as they rely on pixel-based and feature-based abstractions they have limitations in incorporating the semantic and hierarchical nature of complex events. Semantic (or Description-based) approaches use descriptive language and logical operators to build event representations using domain expert knowledge. The hierarchical nature of these models allow the explicit incorporation of event and scene semantic with much less data than Classification and PGM methods.

Ceusters *et al.* [3] proposes the use of Ontological Realism to provide semantic knowledge to high-level events detected by a multi-layer hierarchical and dynamical graphical model in a semi-supervised fashion (human in the loop). Zaidenberg *et al.* [22] have evaluated a constraint-based ontology language for group behavior modeling and detection in airport, subways, and shopping center scenes. Cao *et al.* [2] have proposed an ontology for event context modeling associated to a rule-based engine for event detection in multimedia monitoring system. Similarly, Zouba *et al.* [23] have evaluated a video monitoring system at the identification of activities of daily living of older people using a hierarchical constraint-based approach. Oltramari and Lebiere [15] presents a semantic infra-structure for a cognitive system devoted for event detection in surveillance videos.

Although Semantic models advantage at incorporating domain expert knowledge, the deterministic nature of their constraints makes them susceptible to noise from underlying components - *e.g.,* people detection and tracking components in a pipeline of computer vision system - as they lack a convenient mechanism to handle uncertainty. Probabilistic reasoning has been proposed to overcome these limitations. Ryoo and Aggarwal [17] [18] have proposed hallucination concept to handle uncertainty from low-level components in a context-free grammar approach for complex event detection. Tran and Davis [19] have proposed Markov logic networks (MLNs) for event detection in parking lots. Kwak *et al.* [10] have proposed the detection of complex event by the combination of primitive events using constraint flows. Brendel et al [1] propose probabilistic event logic to extend an interval-based framework for event detection; by adopting a learned weight to penalize the violation of logic formulas.

We present a uncertainty modeling framework to extend the generic constraint-based ontology language proposed by Vu *et al.* [20] by assessing the probability of constraint satisfaction given the available evidence. By combining both frameworks we allow domain expert to provide event models following a deterministic process, while probabilistic reasoning is performed in second plan to cope with the uncertainty in constraint satisfaction. In this paper we focus on handling uncertainty of elementary events.

2 Uncertainty Modeling Framework

Uncertainty may come from different levels of the event modeling task; from failures on the low-level components which provided input-data for the event detection task (*e.g.,* sudden change in person estimated dimension) to the model expressiveness at capturing the real-world event. For instance, constraint violation may be due to person-to-person differences in performing an event (event intra-class variation). In both cases it may be desirable that the event model be still detected even with a smaller probability.

We propose here a framework to handle uncertainty on elementary events. The framework may be decomposed on: event modeling, uncertainty modeling, and inference. In event modeling step domain experts use the constraint-based video event ontology proposed in [20] to devise event models based on attributes of tracked physical objects (*e.g.,* a person) and scene semantics (*contextual objects*). In uncertainty modeling step we learn the conditional probability distributions about the constraints using annotation on the events and the event models provided by domain experts. The inference step is performed by the temporal algorithm of Vu *et al.* [20] adapted to also compute event probability. The probability computation sub-step infers how likely a model is given the available evidence based on pre-learned conditional probabilities about the evaluated constraints.

2.1 Video Event Ontology

The constraint-based framework is composed of a temporal scenario (event) recognition algorithm and a video event ontology for event modeling. The video event ontology is based on natural terminology to allow end users (*e.g.,* medical experts) to easily add and change event models of a system. The models take into account *a priori* knowledge of the experimental scene, and attributes of objects (herein called Physical Objects, *e.g.,* a person, a car, etc.) detected and tracked by the vision components. *A priori* knowledge consists of the decomposition of a 3D projection of the scene floor plan into a set of spatial zones which carry semantic information about the monitored scene (*e.g.,* zones like "TV", "armchair", "desk", "coffee machine"). The temporal algorithm is responsible for the inference task, where it takes as input low-level data from underlying vision components, and evaluates whether these objects (or their properties) satisfy the constraints defined in the modeled events. An event model is composed of (up to) five parts [20]:

- **Physical Objects** refer to real-world objects involved in the detection of the modeled event. Examples of physical object types are: mobile objects (*e.g.,* person, or vehicle in another application), contextual objects (equipment) and contextual zones (chair zone).
- **Components** refer to sub-events of which the model is composed.
- **Constraints** are conditions that the physical objects and/or the components should hold. These constraints could be logical, spatial and temporal.
- **Alert** define the level of importance of the event model, and
- **Action** is an optional clause which works in association with the Alert type describes a specific course of action which should be performed in case the event model is detected,(*e.g.,* send a SMS to a caregiver responsible to check a patient over a possible falling down).

The physical object types depend on the domain of application. Two disjoint default types are presented, Mobile and Contextual Objects, with one extensions each, respectively, Person and Contextual Zone. Mobile is a generic class which defines the basic set of attributes for any moving object detected in the scene (*e.g.,* 3D position, width, height, depth). Person is an extension of Mobile class whose attributes are body posture and appearance signature(s). Contextual Object (CO) type refer to *a priori* knowledge of the scene. Contextual zone is an extension of CO commonly used to define a set of vertices in the ground plane which corresponds to a region with semantic information (*e.g.,* eating table, tv, desk) for an event model. Contextual objects may be defined at the deployment of the system by the domain experts or by launching an object detection algorithm for scene description at system installation, and specific times where object displacement is identified. Physical object types can be expanded accordingly to describe all types of objects in the scene.

Constraints define conditions that physical object properties and/or components must satisfy. They can be non-temporal, such as spatial (person->position *in* a contextual zone; or displacement(person1) >1 m) and appearance constraints (person1->AppearanceSignature = person2->ApperanceSignature); or temporal to capture specific duration patterns or time ordering between a model sub-events (components). Temporal relation are defined following Allen's interval algebra (*e.g., before, and, meet, overlaps*). Fig. 1 describes the model $Person\ changing\ from\ zone 1\ to\ zone\ 2$; which is defined in terms of a temporal relationship between two sub-events: *e.g.,* $c1$, $Person\ in\ zone\ 1$ before $c2$, $Person\ in\ zone\ 2$.

The ontology hierarchically categorizes event models according to their complexity as (in ascending order):

- **Primitive State** models property(ies) and/or relationship among physical object(s) constant on a time interval (person posture, or person inside a contextual zone).

```
CompositeEvent(Person changing from zone1 to zone 2,
    PhysicalObjects( (per:Person), (z1: Zone), (z2: Zone) )
    Components (
    (c1: PrimitiveState Person_in_zone_1 (p1,z1)
    (c2: PrimitiveState Person_in_zone_2 (p1,z1)
    )
    Constraints( (c1 before c2) )
    Alert( NOTURGENT )
)
```

Fig. 1. Person changing from zone 1 to zone 2

- **Composite State** refers to a composition of two or more primitive states.
- **Primitive Event** models a change in a value of physical object property (*e.g.,* person changes from sitting to standing posture), and
- **Composite Event** refers to the composition of two previous event models which should hold a temporal relationship (person changes from sitting to standing posture before person in corridor zone).

2.2 Uncertainty Modeling for Elementary Scenarios

For uncertainty modeling purposes we divided the constraint-based ontology event models into two categories: elementary and composite scenarios. The term scenario is used to differentiate the modeling and inference tasks. Elementary Scenario have a direct correspondence to the primitive state type of the ontology, and the Composite Scenario represents all other ontology event types (Primitive Event, Composite States and Composite Events). This simplification is performed since these ontology event categories were devised to help domain experts at devising models in a modular fashion and then reduce model complexity and increase its re-usability. But, none difference exists for the inference algorithm while processing these event categories besides to the hierarchy depth of the sub-events they define a relationship for.

The uncertainty modeling framework is based on the following concepts:

- **Elementary Scenario**(ES) is composed of physical objects and constraints. This scenario constraints are only related to instantaneous values (*e.g.,* current frame) of physical object(s) attribute(s).
- **Composite Scenario**(CS) is composed of physical objects, sub-scenarios (components) and constraints; where the latter generally refer to composition and/or temporal relationships among model sub-scenarios.
- **Constraint** is a condition that physical object(s) or sub-scenarios must satisfy, and refer to the constraint types presented on the constraint-based ontology section.
- **Attributes** correspond to the properties (characteristics) of real world objects measured by the underlying components of the event detection task (*e.g., vision system*).

- **Observation** corresponds to the amount of evidence on a constraint or a scenario model.
- **Instance** refers to an individual detection of a given scenario.

Fig. 2 presents a description for the elementary scenario *Person in zone Tea*. This scenario is based on the physical objects *Person* and the semantic zone *zoneTea*. For instance, *zoneTea* would be polygon drawn on the floor - close or around the table where the kitchen tools to prepare tea are commonly placed - *a priori* defined by a domain expert during system installation or automatically detected by the system. The model has two constraints: the logic constraint that the target zone is *zoneTea*; and a spatial constraint called *In* which verifies whether the person position lies inside the given zone. Fig. 3 illustrates an example of a scene where semantic zones were manually drawn on the floor plane where contextual objects are located.

```
ElementaryScenario(Person_in_zone_Tea,
    PhysicalObjects( (per:Person), (zT: Zone) )
    Constraints(
       (per->Position In zT->Vertices)
       (zT->name = "zoneTea")
       (displacement(per->Position) < stopConstant)
    )
)
```

Fig. 2. Elementary Scenario Person in zone Tea

2.3 Computation of Elementary Scenario Uncertainty

The uncertainty of an Elementary Scenario is formalized as function of the framework confidence on the satisfaction of the Elementary Scenario constraints. Equation 1 presents an formalization of Elementary Scenario Uncertainty using Bayes Rule.

$$P(E_i|C_i) = \frac{P(C_i|E_i) * P(E_i)}{P(C_i)} \tag{1}$$

where,

- $P(E_i|C_i)$: Conditional Probability of Event E_i given its observed constraints C_i;
- $P(C_i|E_i)$: Probability of constraints which intervene on E_i at the current frame; and
- $P(E_i)$: Prior Probability of Event.

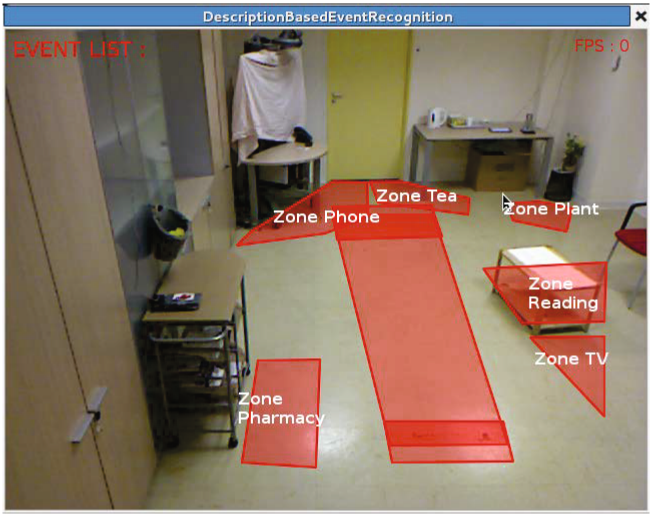

Fig. 3. Scene semantic zones

The conditional probability of event E_i given its set of observed constraints C_i is given by the multiplication of the individual conditional probabilities of its constraints. We assumed all constraints contribute equally to the event model detection and are conditionally independent (see Equation 2).

$$P(C_i|E_i) = \prod_{c_{i,j} \in C_i}^{N_j} P(c_{i,j}|E_i) \tag{2}$$

where $C_{i,j}$:

– Conditional probability of Constraint j of given event i.

To avoid computing $P(C_i)$ which can become costly as the number of constraints increase, we opted to use the non-normalized probability of $P(E_i|C_i)$ as described in Equation 3.

$$\tilde{P}(E_i|C_i) = P(E_i) \prod_{c_{i,j} \in C_i}^{N_j} P(c_{i,j}|E_i) \tag{3}$$

In its final form the proposed formula for elementary scenario uncertainty (Equation 3) addresses small violations of constraints from noise coming from underlying components and due to event intra-class variations.

2.4 Probabilistic Constraints

The uncertainty of a scenario model or its conditional probability given the evidence is addressed by associating each of its constraints to a Probability Density

Function (PDF) responsible for quantifying how likely the constraint would be satisfied given the available evidence. The use of PDFs provide a modular and flexible way to model and change the uncertainty process that governs the conditional probability distribution of a constraint given the available evidence - e.g., by modeling the variation of the low level data the constraint is conditioned on during the targeted event execution - and allowing us to avoid the fully specification of the set of assignments of a conditional probability table. Moreover, different constraints may use different PDFs according to the low-level data, and the PDF may be easily changed without any other changes to the event model.

Besides to selecting the fitting PDF to a given constraint it is also important to how we evaluate the constraint goal in a probabilistic fashion. In the case of the spatial operator In its deterministic version is susceptible to different sources of uncertainty: firstly, from the estimated position of the person which may be influenced by noise from low-level computer vision components; and secondly, from the semantic zone *zoneTea* - *a priori* defined by an expert - which may not accommodate the complete floor surface where people may stand to prepare tea. Its probabilistic counter-part should quantify how likely is the person position to be inside the zone of interest given these sources of noise. We here propose two probabilistic alternatives to the deterministic constraint In: the Center *In* and the Border *In*.

- The Center *In* is fully based on a PDF with respect to the relative distance between the centroid of the person - projected onto the floor - and the central position of the given semantic zone.
- The Border *In* is a hybrid implementation which provides maximum probability (100 %) when the person is anywhere inside the semantic zone, and a probability proportional to the distance of the person to the closest zone edge otherwise.

To model the conditional probability distribution of the distances between the person position and the semantic zone we have used Equation 4. Briefly, this equation converts the observed distance among objects into the corresponding value in an uniform Gaussian distribution using expected parameters pre-learned per semantic object. The corresponding value is then applied to an exponential function to obtain the probability of the constraint given the evidence, *e.g.*, a specific low-level data value for elementary scenario. The resulting PDF provides a probability curve with maximum value around the mean parameter and a monotonically decreasing behavior is observed as the observed value distances from the mean.

$$P(C_{i,j}) = \exp(\frac{1}{2} * (\frac{observed\_value - \bar{x}}{s})^2) \tag{4}$$

where, $\bar{x}$: learned mean of constraint value, and s: standard deviation of $\bar{x}$

2.5 Learning Constraint Conditional Probabilities

The conditional probability distribution of the elementary constraints were obtained by a learning step based on the event models provided by domain experts - using the constraint-based ontology - and annotated RGB-D recordings of the targeted events. The learning step was performed as follows: firstly, an event detection process was performed using the deterministic event models. Each time the deterministic *In* was evaluated the relative distance used by the probabilistic counterparts was stored independent of whether the current constraint is satisfied. Secondly, using the event annotation we collect the distance values frequently assumed by the In variants when elementary scenario annotation is present for the given RGB-D recording. Thirdly and finally, we computed statistics about the the collected values of the attribute the constraint was conditioned on. By performing the learning step using event models combined with event annotation (both provided by domain experts) we aim at capturing the Conditional Probability Distribution (CPD) of the constraints according to the event model semantics and maybe reduce the semantic gap between the event model and the real-world event.

Elementary Scenarios are assumed to be equally probable as their evidence is mainly related to a single time unit (e.g., a frame). The Temporal aspect of scenario models such as instance filtering is currently performed by a threshold method which removes low-probability events. The influence of previous instances probabilities into the evaluated time unit will be evaluated in the future in conjunction with uncertainty modeling at Composite Scenario level (Composite Event).

3 Evaluation

The proposed framework has been evaluated at modeling the uncertainty of activities of daily living of participants of a clinical protocol for Alzheimer's disease study. Two evaluations were performed, firstly on the detection of elementary scenarios, and secondly on the detection of complex events by using uncertainty framework for elementary scenarios as basis for the deterministic complex event models. The latter evaluation intends to assess the improvement brought to the detection of high-level scenario by low-level uncertainty modeling. For both evaluations contextual objects were defined *a priori* by domain experts and mostly refer to static furniture in the scene.

Concerning the learning step necessary to obtain the parameters for the constraint conditional probabilities, in the first evaluation the parameters were computed following the rules of the 3-fold cross-validation procedure. For the second evaluation, the 10 videos involved in the 3-fold cross-validation procedure were used for the learning procedure, and the complex detection performance was evaluated on a set of recordings of 45 participants new to the system, which were only annotated in terms of Composite Events.

3.1 Data Set

Participants aged 65 years and over were recruited by the Memory Center of Nice Hospital. Inclusion criteria of the Alzheimer Disease (AD) group are: diagnosis of AD according to NINCDS-ADRDA criteria and a Mini-Mental State Exam (MMSE) score above 15. AD participants who have significant motor disturbances (per the Unified Parkinson's Disease Rating Scale) are excluded. Control participants are healthy in the sense of behavioral and cognitive disturbances. Experimental recordings used a RGB-D camera (Kinect®, Microsoft©).

The clinical protocol is divided into three tasks: directed tasks, semi-directed tasks, and discussion with the clinician task. The directed tasks (10 minutes) are divided on two sub-tasks: physical directed- and vocal directed-tasks. In the semi-directed task (15 minutes) the participants are asked to undertake a set of Instrumental Activities of Daily Living in a Hospital observation room furnished with home appliances [6]. The participants enter the room alone with a list of activities to perform and are advised to leave the room only feeling all the required tasks are accomplished.

For this framework evaluation we have focused only on the semi-directed task. The list of semi-directed activities is composed as follows:

- Read 1 article and answer three questions,
- Turn on the TV,
- Establish the account balance,
- Pay the phone bill (check writing),
- Answer the phone,
- Call the psychologist to confirm the appointment afterwards,
- Find on a bus map the line that takes you to the train station,
- Prepare the drug box for tomorrow according to the prescription,
- Water the plant,
- Prepare a hot tea.

3.2 RGB-D Monitoring System

The framework for uncertainty modeling was evaluated using a RGB-D sensor-based monitoring system, built on the event detection framework proposed by Vu *et al.* [20], and later evaluated on the detection of daily living activities of older people by Crispim-Junior *et al.* [5] using a 2D-RGB camera as the input sensor.

The evaluation monitoring system can be composed into three main steps: people detection, people tracking, and event detection. People detection step is performed by a depth-based algorithm proposed in Nghiem *et al.* [14], since we have replaced the 2D-RGB camera by a RGB-D sensor. The depth-based algorithm performs as follows: first, background subtraction is employed on the depth image provided by the RGB-D camera to identify moving regions. Then, region pixels are clustered in objects based on their depth and neighborhood information. Finally, head and shoulder detectors are employed to detect people amongst other types of detected objects.

The set of people detected by the previous algorithm is then evaluated by a multi-feature tracking algorithm proposed in Chau *et al.* [4], which employs as features the 2D size, the 3D displacement, the color histogram, and the dominant color to discriminate among tracked objects.

Event detection step has as input the set of tracked people generated in the previous step and *a priori* knowledge of the scene provided by a domain expert. This step was evaluated for two different components for comparison purposes: the proposed framework for uncertainty modeling, and the deterministic event modeling framework proposed by Vu *et al.* [20] and evaluated by Crispim-Junior *et al.* [5]. Both components frameworks used the same underlying components.

3.3 Performance Measurement

The framework performance on event detection is evaluated using the indices of Recall (Rec.) and Precision (Prec.) described in Equations 5 and 6, respectively in comparison to ground-truth events annotated by domain experts.

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

where TP: True Positive rate, FP: False Positive rate and FN: False Negative rate.

4 Results and Discussion

Table 1 presents the performance of the uncertainty modeling framework on elementary scenario (primitive state) detection in a 3-fold cross-validation scheme. The cross-validation scheme used 10 RGB-D recordings of participants of the clinical protocol data set. “Deterministic” stands for the deterministic constraint-based approach. Results are reported as the average performance on the frameworks on the validation sets.

Table 1. Framework Performance on Elementary Scenario Detection on a 3-fold-cross-validation scheme

	Deterministic		Border In		Center In	
IADL	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.
In zone Pharmacy	100.0	71.4	100.0	100.0	100	83.3
In zone Phone	84.0	95.45	92.0	92.0	100.0	100.0
In zone Plant	100.0	81.8	100.0	34.6	100.0	81.8
In zone Tea	93.3	77.7	100.0	36.6	93.3	73.7
In zone Read	75.0	54.5	100.0	38.1	75.0	85.7

$N:\ 10\ participants; 15\ min.\ each;\ Total:\ 150\ min.$

The proposed probabilistic constraints outperformed the deterministic approach on the recall index and on precision index in a few cases such as “In zone reading" and "In zone Pharmacy" with *Center In* constraint. *Border In* constraint presented the highest recall, but the lowest average precision.

Table 2 presents the results of the framework on Composite Event Detection. Here an hybrid strategy is adopted where the uncertainty modeling is used on elementary scenarios and the deterministic constraint-based framework is used on composite event modeling.

Table 2. Framework Performance on Composite Event Detection Level

	Deterministic		Border In		Center In	
IADL	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.
Talk on Phone	88.76	89.77	89.88	70.79	88.76	85.86
Preparing Tea/Coffee	81.42	73.07	95.71	40.36	92.85	55.08
Using Pharmacy Basket	87.75	97.72	89.79	95.65	89.79	97.77
Watering plant	78.57	84.61	100.0	23.14	100.0	28.86

$N:\ 45\ participants;\ 15\ min.\ each;\ Total:\ 675 min.$

The results on complex event detection showed *Center In* and *Border In* had similar performance on recall index outperforming the deterministic approach. *Center In* outperformed *Border In* in the precision index for this test but was still worse than the deterministic approach in most cases. The worse performance in precision index may be attributed to other model constraints which did not have their uncertainty addressed. Based on the results presented we select *Center In* constraint as the probabilistic alternative for the deterministic *In*.

5 Conclusions

We have presented a uncertainty modeling framework to handle uncertainty from low-level data in constraints of elementary scenarios (low-level events). The framework improves the detection performance of elementary scenarios in recall and precision and of composite scenarios in recall.

Further work will extend the framework to model composite scenarios and the uncertainty related to composite and temporal relations among its subcomponents. Moreover, we will also investigate alternatives to allow small deviations from the scenario constraint without the need of performing a supervised learning step.

References

1. Brendel, W., Fern, A., Todorovic, S.: Probabilistic event logic for interval-based event recognition. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3329–3336, June 2011
2. Cao, Y., Tao, L., Xu, G.: An event-driven context model in elderly health monitoring. In: Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing, pp. 120–124 (2009)

3. Ceusters, W., Corso, J.J., Fu, Y., Petropoulos, M., Krovi, V.: Introducing ontological realism for semi-supervised detection and annotation of operationally significant activity in surveillance videos. In: Proceedings of the 5th International Conference on Semantic Technologies for Intelligence, Defense and Security (STIDS) (2010). http://www.cse.buffalo.edu/jcorso/pubs/stids2010_istare_withresponses.pdf
4. Chau, D.P., Bremond, F., Thonnat, M.: A multi-feature tracking algorithm enabling adaptation to context (2011)
5. Crispim-Junior, C., Bathrinarayanan, V., Fosty, B., Konig, A., Romdhane, R., Thonnat, M., Bremond, F.: Evaluation of a monitoring system for event recognition of older people. In: 2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 165–170, August 2013
6. Folstein, M.F., Robins, L.N., Helzer, J.E.: The mini-mental state examination. Archives of General Psychiatry 40(7), 812 (1983). http://dx.doi.org/10.1001/archpsyc.1983.01790060110016
7. Izadinia, Hamid, Shah, Mubarak: Recognizing Complex Events Using Large Margin Joint Low-Level Event Model. In: Fitzgibbon, Andrew, Lazebnik, Svetlana, Perona, Pietro, Sato, Yoichi, Schmid, Cordelia (eds.) ECCV 2012, Part IV. LNCS, vol. 7575, pp. 430–444. Springer, Heidelberg (2012)
8. Kitani, Kris M., Ziebart, Brian D., Bagnell, James Andrew, Hebert, Martial: Activity Forecasting. In: Fitzgibbon, Andrew, Lazebnik, Svetlana, Perona, Pietro, Sato, Yoichi, Schmid, Cordelia (eds.) ECCV 2012, Part IV. LNCS, vol. 7575, pp. 201–214. Springer, Heidelberg (2012)
9. Kumar, P., Ranganath, S., Weimin, H., Sengupta, K.: Framework for real-time behavior interpretation from traffic video. IEEE Transactions on Intelligent Transportation Systems **6**(1), 43–53 (2005)
10. Kwak, S., Han, B., Han, J.H.: Scenario-based video event recognition by constraint flow. In: CVPR, pp. 3345–3352. IEEE (2011)
11. Lavee, G., Rivlin, E., Rudzsky, M.: Understanding video events: A survey of methods for automatic interpretation of semantic occurrences in video. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews **39**(5), 489–504 (2009)
12. Le, Q.V., Zou, W.Y., Yeung, S.Y., Ng, A.Y.: Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, pp. 3361–3368. IEEE Computer Society, Washington, DC (2011). http://dx.doi.org/10.1109/CVPR.2011.5995496
13. Lv, F., Song, X., Wu, B., Kumar, V., Nevatia, S.R.: Left luggage detection using bayesian inference. In: PETS (2006)
14. Nghiem, A.T., Auvinet, E., Meunier, J.: Head detection using kinect camera and its application to fall detection. In: 2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA), pp. 164–169. July 2012
15. Oltramari, A., Lebiere, C.: Using ontologies in a cognitivegrounded system: Automatic action recognition in video surveillance. In: Proceedings of STIDS 2012 (7th International Conference on "Semantic Technology for Intelligence, Defense, and Security (2013)
16. Park, S., Aggarwal, J.K.: A hierarchical bayesian network for event recognition of human actions and interactions. Multimedia Syst. **10**(2), 164–179 (2004)
17. Ryoo, M.S., Aggarwal, J.K.: Recognition of composite human activities through context-free grammar based representation. In: CVPR (2), pp. 1709–1718. IEEE Computer Society (2006)

18. Ryoo, M.S., Aggarwal, J.K.: Semantic representation and recognition of continued and recursive human activities. International Journal of Computer Vision **82**(1), 1–24 (2009)
19. Tran, Son D., Davis, Larry S.: Event Modeling and Recognition Using Markov Logic Networks. In: Forsyth, David, Torr, Philip, Zisserman, Andrew (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 610–623. Springer, Heidelberg (2008)
20. Vu, V.T., Bremond, F., Thonnat, M.: Automatic video interpretation: A novel algorithm for temporal scenario recognition. In: Proc. 8th Int. Joint Conf. Artif. Intell., pp. 9–15 (2003)
21. Wang, H., Klaser, A., Schmid, C., Liu, C.L.: Action recognition by dense trajectories. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3169–3176, June 2011
22. Zaidenberg, S., Boulay, B., Brmond, F.: A generic framework for video understanding applied to group behavior recognition. CoRR abs/1206.5065 (2012)
23. Zouba, N., Bremond, F., Thonnat, M.: An activity monitoring system for real elderly at home: Validation study. In: 2010 Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 278–285, August 2010