Providing Knowledge-Based Predictions for Dynamic Scene Analysis

Extended Abstract

Bernd Neumann
FB Informatik, Universität Hamburg

Workshop on Dynamic Scene Recognition from Sensor Data
Toulouse, France, June 1997

Providing Knowledge-Based Predictions for Dynamic Scene Analysis

Bernd Neumann
FB Informatik, Universität Hamburg
neumann@informatik.uni-hamburg.de

Extended Abstract

Providing predictions about the temporal development of a scene must be considered a central task both in natural and computer vision. By correctly guessing what will be seen at the next moment, the current situation can be better understood, ressources can be saved, proper actions can be taken earlier etc. For a vision system, prediction may provide constraints which can simplify the overall task considerably.

This contribution deals with predictions based on explicit knowledge about typical occurrences or events such as throwing a ball, overtaking a car, eating a boiled egg etc. Different from prediction formalisms at the signal level (e.g. Kalman filtering), knowledge-based prediction essentially works by comparing available data with possible models given in some knowledge representation language. If a model matches the data, it may be used to predict the rest of the data.

Various representations have been proposed and investigated for event models, including logic-based expressions, frame-based data structures and Bayes Nets. In view of the inferencing power which can be gained from logic-based representations, [Neumann and Schröder 96] argue in favor of description logics but show that important mechanisms for hypothesis generation are still lacking. In this contribution, an alternative approach based on a constraint satisfaction formalism is presented.

Constraints have been shown to be a generally useful formalism for knowledge representation including computer vision applications [Mackworth 96]. In our approach, presented in detail in [Kockskämper 97], constraints are used to model the temporal structure of events. Different from the system of temporal relations introduced in [Allen 83], our basic temporal relations refer to individual time marks at the beginning and end of intervals. Unary relations restrict time marks to an interval, binary relations are constructed from the primitive relation $t1 + c \leq t2$, where c is an offset. By restricting ourselves to convex relations [Nökel 91] (where the set of intervals satisfying the relation is continuous), we obtain a system of constraints which can be checked for consistency in linear time.

The temporal constraint satisfaction mechanism is at the core of a system for incremental event recognition which is used to provide predictions for dynamic scene analysis. The relevant constraints are contained in event models which capture the knowledge about coherent occurences in a scene. An event model is a generic description of an aggregate of primitive or composite events with time marks for the beginning and end of each event and temporal constraints between the time marks. Hence an event model is the temporal analog to traditional spatial models in vision which consist of parts and a spatial structure.

A primitive event is a predicate over some time interval which expresses some constancy which can be computed from the scene data. For example,

```
(parallel vehicle-trajectory street-axis par-B par-E)
```

would be a primitive event, beginning with the time point `par-B` and ending with the time point `par-E`, where the trajectory of a vehicle is parallel to a street axis. The idea is that primitive events are computed bottom-up for a low-level description of a dynamic scene. Early work on motion description provides an interesting background for the choice of low-level primitives [Badler 75, Tsotsos et al. 80, Neumann 89]. In our view it is important to note (i) that the set of primitives can be limited to qualitative measurements of simple time-varying scene properties like distances, angles and areas, and (ii) that primitive events may be chosen pragmatically depending on the measurements which are easily available.

Consider an example scenario taken from [Kockskämper 96] where a driverless transport vehicle (dtv) repeatedly transports loads to a certain location. The composite event model `transport-load` (see box) which in turn is based on several other composite and primitive event models (`enter-room`, `unload, exit-room` etc.) can be used to describe this behaviour and link it to dynamic scene properties.

```
Predicate:      transport-load
                     :is-a eventmodel
                     :local-name tl
Arguments:      (?room ?pos (?dtv :is-a stacker)
                     (?station :part-of ?room))
Time marks:          (tl-B tl-E)
Component events:    (er :is-a (enter-room ?room ?dtv er.B er.E))
                     (fs :is-a (free-station ?station fs.B fs.E))
                     (ul :is-a
                     (unload ?dtv ?station ?pos fs.B fs.E-B))
                     ex :is-a (exit-room ?room ?dtv ex-B ex-E)))
Temporal relations:  ((tl.B + 10 ≤ tl.E)
                     (tl.E - 12 ≤ tl.B)
                     (er (before) ul)
                     (ul (before) ex)
                     (ul (starts-within) fs)
                     (tl.B = er.B)
                     (tl.E 0 ex.B))
```

The temporal relations include qualitative relations like `before` and `starts-within` which can be expressed by primitive inequalities, e.g.

starts-within <=> $I1.B \geq I2.B+\Delta t$, $I1.E+\Delta t \leq I2.E$

Δt is the unit of the time point algebra used in this approach [Vila 94].

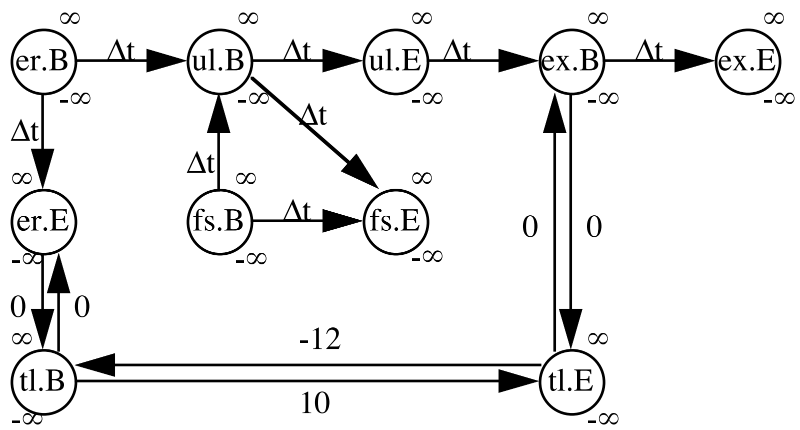
Event recognition is the process of checking whether the primitive events computed for a dynamic scene satisfy an event model. An event recognition algorithm which recognizes events post-mortem, i.e. after their termination, has been presented in [Neumann 89]. However, in order to allow prediction, event recognition has to be performed incrementally. An incremental recognition and prediction procedure also developed in my group [Kockskämper et al. 94] will now be described.

Event prediction consists of two tasks which must be carried out for each time increment:

- Test whether a potential event begins and add such event to the list of current events.
- Test whether a current event can be continued. Remove any event which cannot be continued or is completed.

The list of current events is the basis for prediction. Each event on the list is only partially instantiated due to component events which have not yet happened. The prediction which can be derived from the partially instantiated event models specifies expected events within the constraints resulting from the instantiated parts and the event model.

A concise representation of the temporal constraints of a predicted event can be given in terms of the time net which is used for constraint propagation in the event prediction procedure. Below is the time net for an unload event before instantiation.

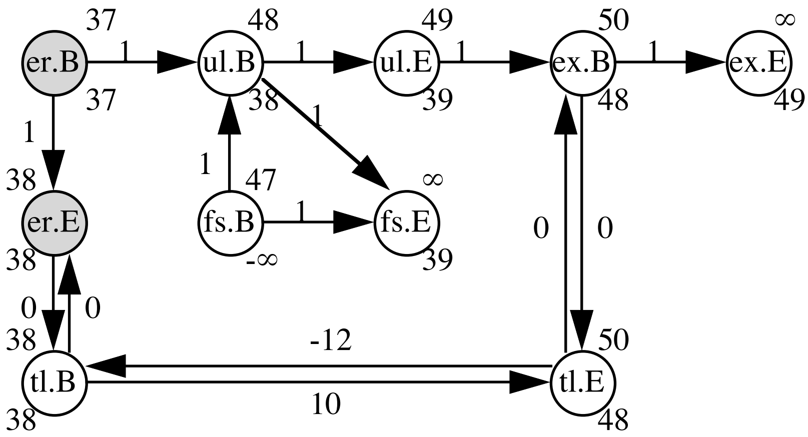


Time net before instantiation

Each node corresponds to a time mark in the event model. Unary constraints of a time mark are entered below and above a node (initially typically -∞ and ∞). Binary constraints of the form $t1 + c \leq t2$ are represented by edges between two time marks along with a number marking the offset c in the inequality. Δt is the temporal increment between observations and is used as an offset when strict inequality must be expressed.

As soon as an event gets instantiated, say, by observing an `enter-room` beginning at some time 37 and ending at time 38 (assuming $\Delta t=1$), the unary constraints [-∞ ∞] of er.B and er.E are reduced to the intervals [37 37] and [38 38], respectively, and the new values are propagated through the net., tightening other unary constraints.

At this point, the time net represents the predicted event in terms of the temporal constraints which are valid for all components of the event. Note that the unary constraints may also be used to represent uncertain observations, e.g. the occluded beginning or end of an event. Uncertainty boundaries may be propagated just as certain time points.



Partially instantiated time net

Obviously, the more an event model is instantiated, the more one expects its successful completion. Conversely, when little to no evidence is available for an event, one would not predict it although it may be possible. In order to differentiate between different degrees of certainty, additional information about the probabilistic structure of events must be introduced. This is not part of the model presented here. Instead, all events on the list of current events according to the event prediction procedure sketched out above are considered for prediction. Hence each predicted event has at least one instantiated part. Another simple alternative is to mark the component event whose instantiation would allow a sufficiently certain prediction of the whole event.

The complexity of constraint propagation is O(n), where n is the number of nodes of a time net, for each propagation. This favourable property can be achieved by eliminating pathological circles in a preprocessing step executed once for all event models. Given a large number of event models which must be tracked simultaneously and a rich scene description in terms of primitive events, complexity considerations may become very important, hence the use of constraint systems with larger complexity may become prohibitive.

We have used our incremental event prediction system in a simulated robot environment for path finding and in monitoring applications.

References

Allen 83
Maintaining Knowledge About Temporal Intervals
J.F. Allen
in: Communications of the ACM 26 (11), 832-843, 1983

Badler 75
Temporal Scene Analysis: Conceptual Descriptions of Object Movements
N.I. Badler
Report TR 80, Dep. of Computer Science, University of Toronto, 1975

Kockskämper et al. 94
Extending Process Monitoring by Event Recognition
S. Kockskämper, B. Neumann, M. Schick
in: Proc. ISE-94, 455-460, 1994

Kockskämper 97
Vorgangsmodelle - Ein Ansatz zur Repräsentation und Analyse zeitabhängigen Verhaltens bei der Überwachung und Diagnose technisacher Systeme
S. Kockskämper
Diss. FB Informatik, Universität Hamburg, 1997

Mackworth 96
Quick and Clean: Constraint-Based Vision for Situated Robots
A.K. Mackworth
in: Proc. ICIP-96, Vol. III, 189-792, 1996

Neumann 89
Description of Time-Varying Scenes
B. Neumann
in: Semantic Structures, D. Waltz, Ed., Lawrence Erlbaum, 1989

Neumann and Schröder 96
How Useful is Formal Knowledge Representation for Image Interpretation?
B. Neumann, C. Schröder
in: Proc. Workshop on Conceptual Descriptions from Images, ECCV-96, 1996, 58 - 59

Nökel 91
Temporally Distributed Symptoms in Technical Diagnosis
K. Nökel
Lecture Notes in Artificial Intelligence, Springer, 1991

Tsotsos et al. 80
A Framework for Visual Motion Understanding
J.K. Tsotsos, J. Mylopoulos, H.D. Covvey, S.W. Zucker
IEEE PAMI-2, 1980, 563-573

Vila 94
A Survey on Temporal Reasoning in Artificial Intelligence
L. Vila
AI Communications 7 (1), 4-28, 1994