ACTIVITY RECOGNITION USING RELATIONAL GRAPH LEARNING

Duckworth P., Cohn A. G., Hogg D. C. - University of Leeds
{scpd, a.g.cohn, d.c.hogg}@leeds.ac.uk

UNIVERSITY OF LEEDS

Abstract

This research aims to develop a largely unsupervised learning system from visual data, to recognise new event classes and detect events occurring during long-term observation. The learning system will use qualitative spatio-temporal relationships (QSRs) between observed objects to represent a video sequence in a graph structure [4].

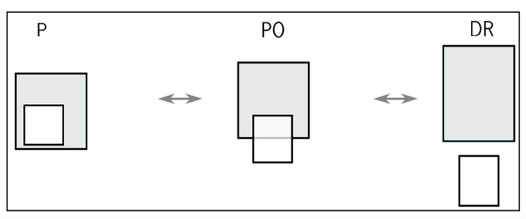
We apply our learning system within a robotics framework. The robot will observe and learn from a variety of human populated environments over a long period of time. This is the main goal of the EU funded robotics project STRANDS [5].

QSRs

Spatial and temporal information represented qualitatively is natural, efficient and similar to the information available to humans. This is the idea behind using a set of qualitative relations to represent visual data.

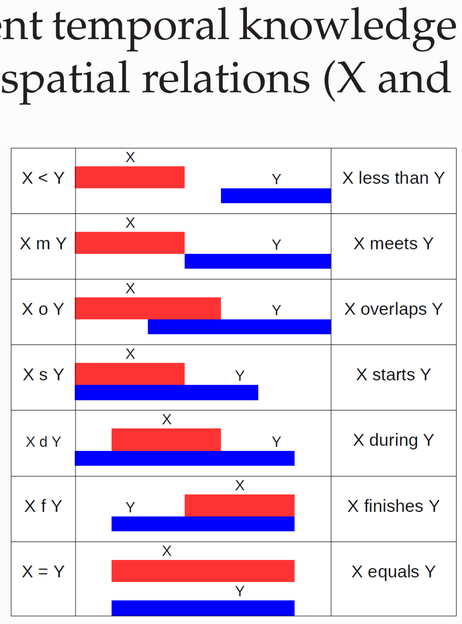
Spatial Relations

Region Connection Calculus [1] represents topological relations between pairs of objects. RCC3 has three distinct relations: Part of (P), Partially Overlap (PO) and Discrete (DR):



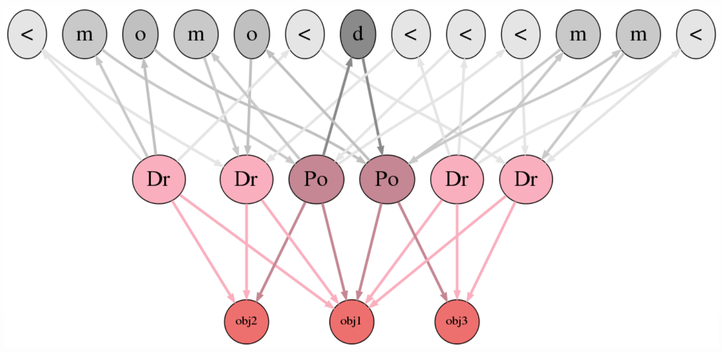
Temporal Relations

Allen's Temporal logic [2] is used to represent temporal knowledge about a pair of spatial relations (X and Y):



Graph Formulation

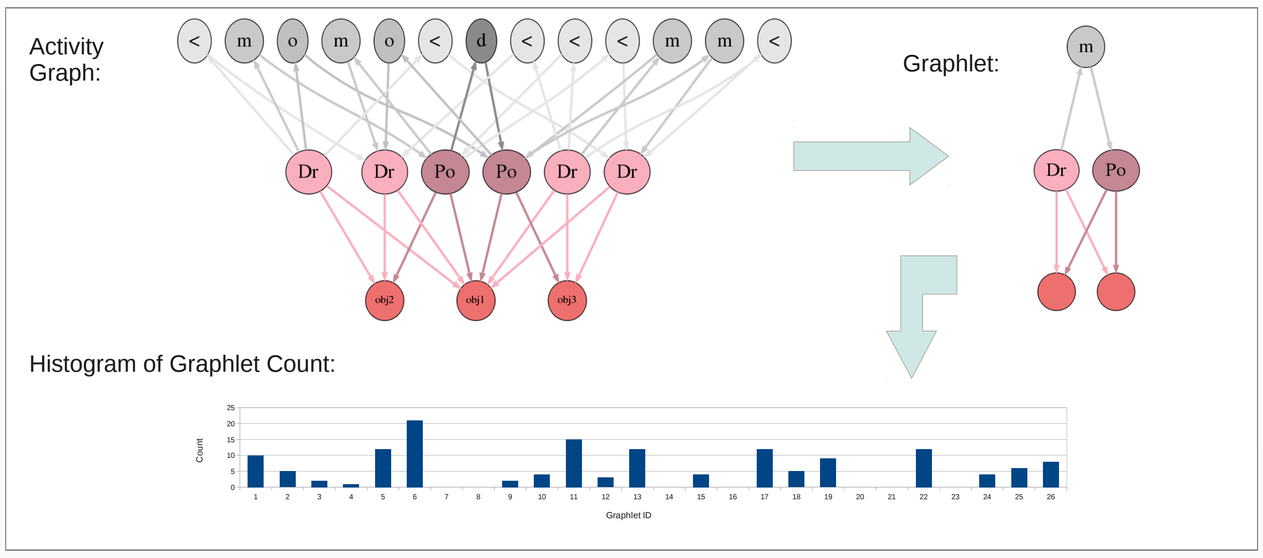
From a video stream, regions of interest in visual data are detected and tracked. An *Activity Graph* is created and represented as a 3-layered graph. This captures all the spatio-temporal relations between the object tracks:



- Temporal relation nodes
- Spatial relation nodes
- Object nodes

Similar sub-graph structures in an Activity Graph can be interpreted as qualitatively similar interactions between objects [4]. A spatio-temporal representation of an event is learnt from the Activity Graph using statistical evidence that events are more frequently observed than noise [3]. This implies that frequent sub-graphs can be considered as event classes.

Event class models are learnt using a set of graphlets which are mined from the Activity Graph. These graphlets are used to compute a similarity measure between sub-graphs. A histogram of the frequency of each graphlet occurrence is generated and passed into a classifier (e.g. a SVM).



References

[1] J. Chen, A. G. Cohn, D. Liu, S. Wang, J. Ouyang, Q. Yu, A Survey of Qualitative Spatial Representations in The Knowledge Engineering Review, 2013.
[2] J. F. Allen. Maintaining knowledge about temporal intervals in Communications of the ACM, 1983.
[3] J. Fernyhough, A. G. Cohn, D. C. Hogg, Building Qualitative Event Models Automatically from Visual Input, in ICCV, 1998.
[4] M. Sridhar, A. G. Cohn, D. C. Hogg, Unsupervised Learning of Event Classes from Video, in AAAI, 2010.
[5] http://strands-project.eu

Acknowledge: The EU FP7 funded STRANDS (600623) and RACE (287752) projects in the area of Cognitive Vision.