EVENT ANALYSIS USING QSRs & QUANTITATIVE FEATURES

Tayyub J.- University of Leeds
sc12jbmt@leeds.ac.uk

UNIVERSITY OF LEEDS

Abstract

Activity understanding and recognition is a major field in computer vision today. Many techniques have been developed to recognise activities from visual data. Our approach focuses on creating and combining descriptive features that would comprehensively describe an activity. We make use of qualitative and quantitative features and propose that the combination of the many outperforms any single one.

Proposed Approach

Most of the past work focuses on producing descriptive features from video input to learn activity models which would describe a particular activity in a discriminating way [2, 3]. These features commonly strengthen describing some aspects of activities very well but at the same time compromise other aspects. In our work, we make use of qualitative features, e.g. QSRs[1], and quantitative features[3], e.g. distances, velocity etc., to represent the spatial and temporal descriptions of visual data. Qualitative features capture abstract properties of an activity within the scene, such as interactions of objects or durations of interactions. These are sufficient to represent most activities, however, some activities require further representation than just the qualitative description. We make use of quantitative feature that act at a finer level to capture more concrete variations of activities.

References

[1] AG., Cohn, Qualitative spatial representation and reasoning: An overview in Fundamenta Informaticae Journal 2001.

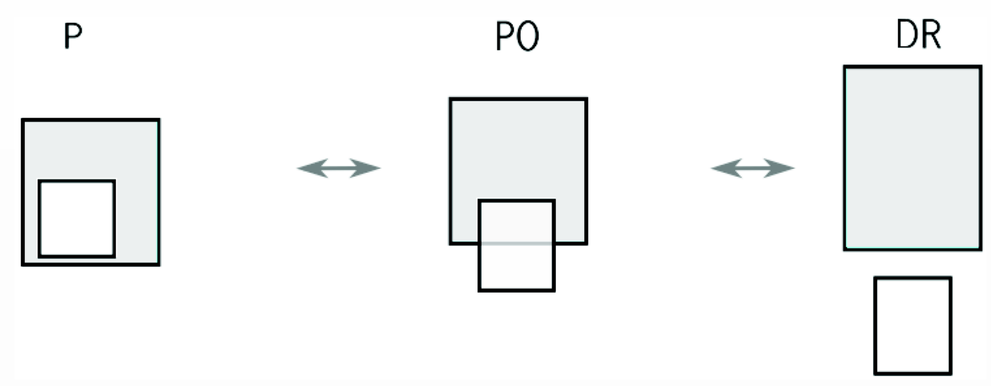
[2] Muralikrishna., Sridhar, AG., Cohn, DC., Hogg, Unsupervised Learning of Event Classes from Video, in AAAI, 2010

[3] A., Behera, DC., Hogg, AG., Cohn, Egocentric activity monitoring and recovery. In Computer Vision ACCV, 2012

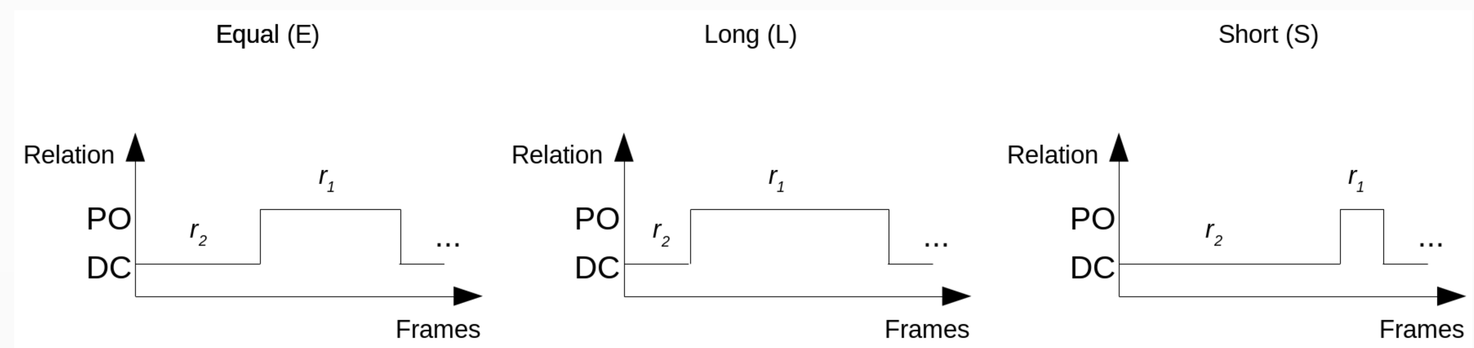
[4] http://strands-project.eu

Qualitative Features

Spatial Relations: We make use of RCC3 calculus, which represents interactions between a pair of elements in one of three relations. These are Part of, Discrete and Partially Overlap.[1]



Temporal Features: These features capture the duration of the interactions. It extends on the QSRs to provide temporal information for each interaction. The duration is calculated as the ratio between two consecutive QSR relations. The resulting metric is discretise into either Equal, Long and Short.

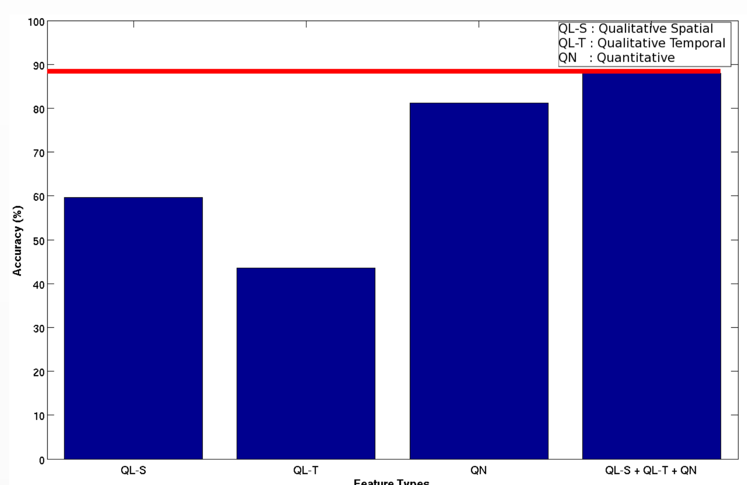


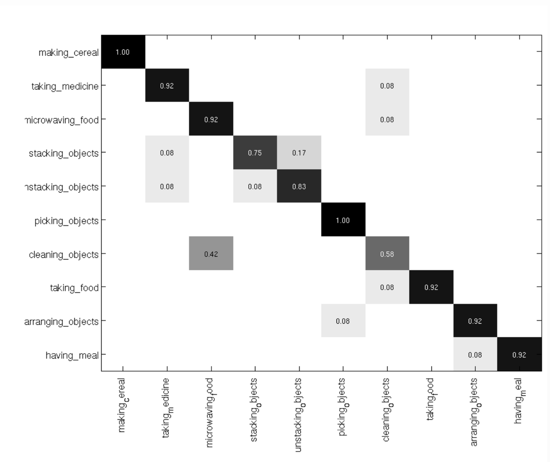
Quantitative Features

Distance Features: We compute the distances between each element and every other element in the scene in each frame. For example, from each joint of a human skeleton to every other joint and the objects in the visual scene. Features such as the mean and standard deviation from all distance data are then extracted.
Velocity: Distance features are then extended to capture the rate of change in distances. These are useful to discriminate fast from slow motion within the activity.[3]
Direction of Motion: The direction of motion is calculated using distances. DoM is useful to distinguish reverse activities such as stacking and unstacking.

Experiments

We have experimented our method on the CAD 120 dataset. It comprises of 10 high-level activities. It is seen that the combination of different types of features works much better than individual types. The results are presented.





Conclusion & Future Work

We have shown that using qualitative and quantitative features capture much more information from visual data compared to any one alone. We aim to extend our approach to focus on detection and recognition of sub-level activities using similar techniques. More specifically, we aim to perform sub-level activity using high-level activity detections for a more confident classification.