Abstract

Human activity recognition (HAR) deals with recognition of activities or interactions that include humans. Machine learning techniques that learn generic patterns from examples can be used to recognize unseen activities. Human activities by nature are spatio-temporal, involving spatial changes over time. The interacting entities of an activity can be abstracted in a variety of ways, ranging from the detailed silhouette of the entity to the very basic axis-aligned minimum bounding rectangles (MBR). MBRs require less storage and computation while abstracting out noise and video specific details. However, they also abstract away important details for human activities. For a more precise description, which offers a reasonable tradeoff between efficiency, noise elimination, and interaction details, this thesis discusses an *extended object* abstraction of the human body in an activity. Herein, extended objects are defined as a set of components, where each component is approximated by an axis-aligned minimum bounding rectangle. The thesis then explores appropriate representation and reasoning techniques for activities wherein interacting entities are abstracted extended objects.

An activity is characterized by the evolution of spatial relations between entities over time. The thesis presents a geometric framework for computing qualitative spatial relations between extended objects, called *Extended CORE9*. To keep track of the changing spatial relations between the extended objects over time during an activity, a *Temporal Activity Graph* (TAG) representation is described in this thesis. For classification of spatio-temporal activities represented as a TAG, a *Temporal Activity Graph Kernel* is designed. The kernel can be used in a Support Vector Machine to classify human activities. However, a kernel based classifier does not model the underlying structure of a class of activities represented as TAGs. Therefore, this thesis further discusses a grammar-based generative learning mechanism. *Temporal Activity Graph Grammar* is probabilistic context free graph grammar that is defined for modeling activities represented as TAGs. The thesis presents a grammar induction algorithm that induces the rules of the grammar from a set of positive examples.

The thesis reports results of experiments conducted on three different datasets for various stages of development of the representation and reasoning techniques. Experiments towards classification of human activities are performed using - 1) only spatial information obtained using Extended CORE9 without any temporal information, 2) spatio-temporal information encoded using TAGs with a discriminative classifier, and 3) generative activity models in the form of TAG Grammars.