World Academy of Science, Engineering and Technology
International Journal of Electrical and Computer Engineering
Vol:14, No:09, 2020

# A Semantic and Concise Structure to Represent Human Actions

**Authors :** Tobias Strübing, Fatemeh Ziaeetabar

**Abstract :** Humans usually manipulate objects with their hands. To represent these actions in a simple and understandable way, we need to use a semantic framework. For this purpose, the Semantic Event Chain (SEC) method has already been presented which is done by consideration of touching and non-touching relations between manipulated objects in a scene. This method was improved by a computational model, the so-called enriched Semantic Event Chain (eSEC), which incorporates the information of static (e.g. top, bottom) and dynamic spatial relations (e.g. moving apart, getting closer) between objects in an action scene. This leads to a better action prediction as well as the ability to distinguish between more actions. Each eSEC manipulation descriptor is a huge matrix with thirty rows and a massive set of the spatial relations between each pair of manipulated objects. The current eSEC framework has so far only been used in the category of manipulation actions, which eventually involve two hands. Here, we would like to extend this approach to a whole body action descriptor and make a conjoint activity representation structure. For this purpose, we need to do a statistical analysis to modify the current eSEC by summarizing while preserving its features, and introduce a new version called Enhanced eSEC or (e2SEC). This summarization can be done from two points of the view: 1) reducing the number of rows in an eSEC matrix, 2) shrinking the set of possible semantic spatial relations. To achieve these, we computed the importance of each matrix row in an statistical way, to see if it is possible to remove a particular one while all manipulations are still distinguishable from each other. On the other hand, we examined which semantic spatial relations can be merged without compromising the unity of the predefined manipulation actions. Therefore by performing the above analyses, we made the new e2SEC framework which has 20% fewer rows, 16.7% less static spatial and 11.1% less dynamic spatial relations. This simplification, while preserving the salient features of a semantic structure in representing actions, has a tremendous impact on the recognition and prediction of complex actions, as well as the interactions between humans and robots. It also creates a comprehensive platform to integrate with the body limbs descriptors and dramatically increases system performance, especially in complex real time applications such as human-robot interaction prediction.

**Keywords :** enriched semantic event chain, semantic action representation, spatial relations, statistical analysis
**Conference Title :** ICRV 2020 : International Conference on Robotics and Vision
**Conference Location :** Lisbon, Portugal
**Conference Dates :** September 16-17, 2020