

The paper represents an approach to infer the interaction between human, object and environment by given context information. The context information includes the human pose, human intention and environment constraints. In other words, it defines that a behavior is determined by certain objects and actions under a particular environment. Grounding an affordance can be modeled as probabilistic model with given context information from training video data. Given the observations semantic affordance models the most likely active affordance. Spatial affordance captures the information of contacts points with corresponding object. It indicates which part of the object will be interacted with the human activity and the distribution of interested area over the whole object. Temporal affordance models the trajectory of moving an object and human joints and hands. Different trajectory estimations are used to captures the characters of different motion patterns. By given context information of human pose, intention and environment, the learn model can outperform the other methods in predicting the trajectory. Predicting trajectory is meaningful for robot by transfer trajectory into partitioned commands. Also the probabilistic model has potential to enable robot to do simple activity by reading the static information from the environment. However, the limit is obvious. By the assumption, the only one affordance is given at a given time upon a given object and the environment is static. In the real world, it is hard to keep the environment static. Also the model is not robust to handle complicated situations when multiple objects show with multiple interacted active affordances.