Predicting abnormalities in complex human-object interaction by using object affordance context

Mahmudul Hassan and Anuja Dharmaratne
School of IT, Monash University Malaysia

Abstract— **Detection of anomalous video events in complex Human-Object interactions is a novel and challenging computer vision problem. Researchers have applied various contexts like spatial, temporal, sequential etc. to address this issue. This conceptual paper explores the potentiality of object's affordance as a promising context for recognizing and predicting complex human-object interaction based abnormalities in novel video sequences.**

Keywords- Affordance;Abnormalities,Mutual contexts.

I. INTRODUCTION

In contrast with the views of the orthodox schools of constructive psychology, the concept of affordance (1979), has allowed us to look at the intrinsic and functional attributes of objects. As object's affordance trigger actions, the relational mapping of object class, human and the ambient scene may aid anomaly detection process in Human-Object interactions.

II. APPLICATION OF AFFORDANCE

'Affordance' has been used in saliently solving different aspects of Human-Object interactions (HOI) related challenges in computer vision. Such as: improving the detection process of actions and objects [1], saliently comprehend scenes, anticipate probable human activities [2] etc. Finding further potential applications of object's affordance is an extremely interesting proposition.

III. ABNORMALITY DETECTION USING OBJECT AFFORDANCE

The definition of abnormal actions refers to the course of actions that do not follow the expected pattern. To label a novel action abnormal, we may need to possess a profile of normality. Ironically, this 'normality profile' depends largely on contexts. For example, in a lying action, the 'spatial context-**where**' is important to judge its normality. Lying on bed is normal but lying on the street may be anomalous. Researchers related to anomaly detection in videos have used various contexts like spatial, temporal, sequential (i.e. proper ordering of sub actions) etc. But they have often overlooked 'object's affordance' as a potential context of anomalous HOI detection. Most of our daily activities involve HOI and anomalous events can occur in them. Apparently, these abnormalities depend on the mutual relations between the human, objects and the ambient scene. For example: a knife affords chopping and cutting actions but throwing it may be anomalous. At the same time standing on a tool may be affordable, hence, normal, for an adult but can be potentially dangerous for a child; here, 'age' of the subjects affects the affordance of the tool. Interactions with an object do also rely on physical attributes of the object (material, size, shape etc.), we expect a fragile object to be handled with care, so vicious way of interacting with it, may be considered as unusual. Sometimes, the human pose does not give any clue about the abnormality, i.e. a person drinking from an aerosol (toxic) sprayer, if we would know the affordance of the sprayer, then we could predict the abnormality. Even there may be abnormalities in the scene. For instance, missing grill in front of a fire place (relational affordance of fire place and grill), a turned on (objects state) hob without any utensils, are potential sources of danger. We argue, that these anomalies are hard to recognize by the current algorithms in video based anomaly detection and the object's affordance can be potential tool to find these HOI or object-object interaction based abnormalities. By the inclusion of different attributes to the object such as: (Object's affordance class, material, aspect ratio, object's state, danger index, shape, ambient objects etc.) and to the human (age, gender, disability, aspect ratio, alone/group) we may map and learn the mutual relational contexts between the human, object and scene in a normal setting and try to recognize anomalies in a novel event.

IV. PROPOSED APPROACH

Probabilistic graphical models (PGM) are powerful means of reasoning in the domain of uncertainty. The Human poses and its attributes and the object and its attributes may be considered as different nodes of the PGM. With substantial amount of training videos, the structural relationships between these nodes may be learned. From partial observation of a novel test video, with some available evidences (i.e. human poses, detected objects etc.) inference can be done for most probable actions or for most probable joint assignments of these nodes. The inferred quantities can subsequently be used for anomaly detection by observing the deviation in the actual states of the nodes and the predicted states.

REFERENCES

[1] Gupta, A.; Davis, L.S., "Objects in Action: An Approach for Combining Action Understanding and Object Perception," Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on , vol., no., pp.1,8, 17-22 June 2007

[2] Hema S Koppula, Ashutosh Saxena, "Anticipating Human Activities using Object Affordances for Reactive Robotic Response",. In Robotics: Science and Systems (RSS), 2013.