Learning Human Activities and Object Functionalities through Conditional Random Fields

Nakul Gopalan
Department of Computer Science
Brown University
Providence, RI 02912
ngopalan@cs.brown.edu

Jun Ki Lee
Department of Computer Science
Brown University
Providence, RI 02912
jklee@cs.brown.edu

Abstract

Recognizing a human or animal activity is a skill that humans learn early in life. Activity recognition helps humans in predicting next steps of the activitiy in question and also in planning other activities with respect to the recognized activity. Our objective is to recognize human activities using machine learning, to predict next steps of an action, or provide help using robots or other interfaces, or help recognizing problems in the activity being performed. Specifically of interest to us are the human activities being performed on a table-top, where the human can be observed and helped by a robot. We plan to use a dataset from [1] to learn a Graphical Model with temporal and spatial structure, which can then be used to predict the action being performed. As a baseline we have results of a bag of words classification of the 10 activities, which yields a result of 67%.

1 Introduction

2 Background

3 Dataset

We explored data sets with activity labels and corresponding skeletal trajectories. Humans frequently perform activities using objects, especially in table-top scenarios. The dataset from [1] was constructed keeping this relationship in mind. The data consisits of skeletal poses of human subjects performing activities, but along with this the dataset also has object lables, positions and poses, for all objects used. The activities present in the dataset are: cleaning, stacking objects, unstacking objects, eating, making cereal, taking medicine, microwaving food, arranging objects, taking food and picking objects. Along with activity labels we have sub-activity labels, which are latent and not visible in our model.

4 Methods

References

[1] Hema Koppula, Rudhir Gupta, and Ashutosh Saxena. Learning human activities and object affordances from rgb-d videos. *IJRR*, 2013.