

# CMSC 828T Activity Recognition from Group Context

Swaminathan Sankaranarayanan, Kota Hara

February 20, 2014

In this project, we will work on activity recognition of people in images and videos. In traditional activity recognition, each person is analyzed independently from others. For instance, if we just look at a single person standing, we do not know whether he/she is talking with others, being in a queue or waiting for someone. However, if we take into account other people in the scene as a context, we will be able to identify more detailed activity of the person. For instance, if there are multiple people forming a line, facing towards the same direction, it is likely that they are in a line. If there are two people facing each other, most likely they are talking each other. Thus, our goal is to identify each person's activity with considering a context from other people.

Our system needs the following steps to obtain an input to the high level activity recognition tasks. First, people in each frame are detected using a standard pedestrian detector. Then the orientation of each person is estimated by a trained estimator. Each person's location is represented as a point on a 2d plane using a camera parameters and the size of the detection. Tracking is done by considering the moving distance. These spatial and temporal information of pedestrians are used to analyze the activity of the people.

For the activity recognition task, we take a learning-based approach, where we have an access to a database of videos with activity class label associated to each person. The detection, orientation estimation and tracking is done on them. We will come up with algorithms which utilize the training data to estimate the activity in testing videos.

In this project, we assume that people in a group (i.e., people talking each other ) have the same activity label. Thus, before processing the test video, we cluster people in the scene into different groups, each of which is conducting independent activity. The clustering can be done based on the distance between each person.

One of the approaches would be based on nearest neighbor matching. From the training dataset, we mathematically represent each group activity by capturing each person's movement as well as the group interaction. With such representation, we define a proper distance between two instances of the group activity. This distance should be smaller if two activities are same and larger if not. The group activity recognition is done by finding the instance in the training database which is the closest to the given test instance. We will represent group activity as a point on a special Euclidean group, which is a Riemannian manifold and define a distance as a geodesic distance on the manifold.