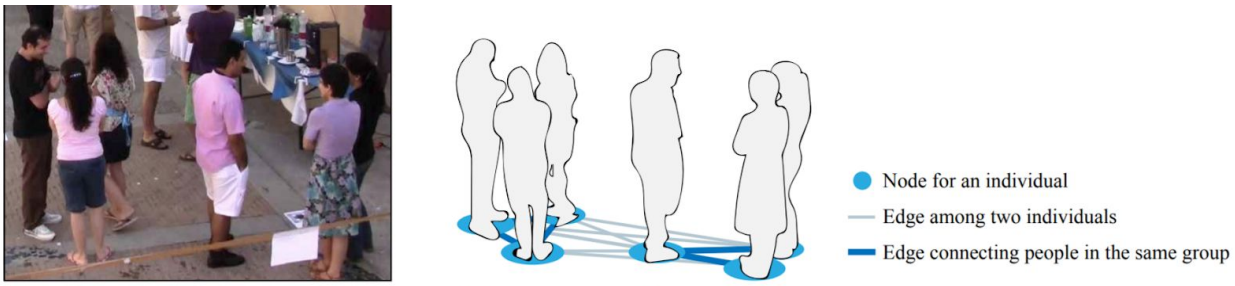
CPSC 290 Project Proposal
Abhijit Gupta [avg32]
Advisor: Dr Vazquez
January 16, 2020

Improving Social Awareness and Group Detection through Deep Learning

Background:

Automatic group detection using computer vision is critical to surveillance systems, socially-aware mobile systems, interactive displays, and more. In the field of human-robot interaction, group detection is necessary for optimal verbal and non-verbal behavior [1]. Group conversation can be visually recognized by analyzing the location and orientation of all individuals. In this model, each individual is a node on a graph, with edges connecting people in the same group [2]. Figure 1 below gives an example input and output of this process.



(a) Scene from the Coffee Break dataset (b) Interaction Graph

Figure 1: Sample Input and Output [2]

Currently, explicitly modeled heuristics are often used to group people by spatial arrangement, e.g., [3]. But these methods do not generalize well to more complex scenarios due to the rigid assumptions made. With increasingly available high-quality data inputs, powerful deep learning methods can improve the detection of conversational groups in comparison to prior approaches. The current state-of-the-art technique uses a Deep Affinity Network (DANTE) and the Dominant Set (DS) clustering algorithm to create group clustering predictions [2]. The data analysis pipeline is shown in Figure 2 below.

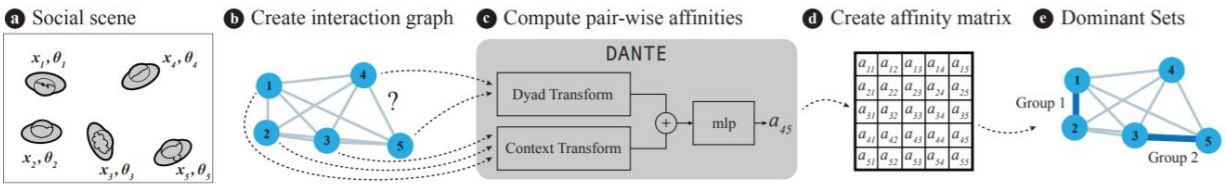


Figure 2: DANTE Pipeline [2]

In Figure 2 (a), the scene is captured and spatial feature vectors for each individual are extracted. An interaction graph is created in (b) and DANTE uses both a Dyad and a Context Transform with several multilayer perceptrons (c) to compute weights that represent the likelihood that two individuals are part of the same conversational group based on their spatial features. These weights are then organized into an affinity matrix, showin in (d). Finally, Dominant Sets (e) is applied to the matrix to determine the predicted group structure. The DANTE algorithm is preferred for its deep learning components and ability to contextualize both individual interactions and the overall social scene. Compared to previous benchmarks tested on publicly available datasets, DANTE resulted in greater precision and recall with statistical significance [2]. We plan on extending this approach in this work.

Proposed Project:

In this project, I would collaborate with Dr. Vazquez and the Interactive Machines Group to further improve the DANTE algorithm. Currently, the implementation is only able to handle six people simultaneously, reducing its effectiveness in crowded locations. In theory, though, the design of the model should be able to reason about more people in a scene. We will improve the implementation to remove this artificial restriction from the way that the model is implemented.

We also are looking into expanding the feature set with body language information from the Kinect Azure to improve classification accuracy. I will also be updating the publicly-shared ROS nodes that allow for real-time usage of this application with the interactive Shutter tabletop robot. This semester, I hope to pursue all these avenues to improve the speed, accuracy, and robustness of the model.

Timeline:

Over the 12 week course, I intend to make steady progress towards the final goal of an improved group clustering implementation. A tentative timeline is shown below.

1. January 15 - 22: Create Project Proposal and get course approved.
2. January 23 - 28: Complete IRB training. Familiarize myself with the current codebase
3. January 29 - February 12: Update DANTE to handle more people
4. February 12 - February 23: Collect dataset with new features
5. February 23 - March 8: Label and organize data so that it's amenable to learning
6. March 23 - April 12: Update and train DANTE on expanded feature set
7. April 13 - April 24: Update ROS node with new model and create documentation

Although I do not expect to adhere strictly to the timeline due to natural slowdowns and/or breakthroughs, this planning will help ensure that I stay on track throughout the semester.

Deliverables:

My deliverables for this project will be the improvements to the DANTE project in scalability and accuracy. Particularly, I will expand the program to handle more than six individuals, improve the feature space, and assist in collecting an improved dataset for training.

References:

[1] Marynel Vázquez, Elizabeth J Carter, Braden McDorman, Jodi Forlizzi, Aaron Steinfeld, and Scott E Hudson. 2017. Towards robot autonomy in group conversations: Understanding the effects of body orientation and gaze. In Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction. ACM, 42–52

[2] Mason Swofford, John Peruzzi, Nathan Tsoi, Sydney Thompson, Roberto Martin-martin, Silvio Savarese, and Marynel Vázquez. 2018. Improving Social Awareness Through DANTE: A Deep Affinity Network for Clustering Conversational Interactants. J. ACM 37, 4, Article 111 (August 2018), 22 pages. https://doi.org/10. 1145/1122445.1122456

[3] Francesco Setti, Chris Russell, Chiara Bassetti, and Marco Cristani. 2015. F-formation detection: Individuating freestanding conversational groups in images. PloS one 10, 5 (2015), e0123783