Online and Unsupervised Face Recognition for Humanoid Robot: Toward Relationship with People

Lijin Aryananda

Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: The ability to distinguish among different individuals is crucial for all social animals. Observations of behavior within dolphin male coalitions indicate that dolphins are able to not only recognize friends, but also keep track of their previous behavior to predict future action. This work addresses the questions of how one may implement such social competence in a humanoid robot. Information about various people's identities (appearance, behavior, etc) that we acquire through our daily social experience is so rich and complex that manually encoding them into a database for the robot to memorize is very limiting. In this project, we focus on extending current face technology to allow the robot to opportunistically learn about individuals and their characteristics in an online and unsupervised manner through embodied social interaction.

Motivation: The notion of people as distinct individuals plays a very important role in our daily social life. If a robot has the ability to recognize and remember people it interacts with, it will be able to learn about characteristics of each individual and treat them uniquely as individuals. This leads to complex social behavior, such as cooperation, dislike, loyalty, and affection. As proposed by [3], if robots have long-term contact with humans, it may be desirable to have them develop individual relationships, which is exactly the aftermath of this social dynamic. Moreover, the ability to distinguish among people allows the robot to build toward more complex social competencies where the idea of people as distinct individuals is crucial, including theory of mind and social referencing.

Previous Work: Face recognition is the most frequently explored modality in person identification technology and has been implemented using various approaches [8]. [7] presents an online supervised learning method for face recognition, allowing machines to learn directly from sensory input streams. Despite different constraints, the issue of direct coupling between the face recognition system to sensory input is very relevant to our work, due to the requirement of an embodied setting. [4] proposed an image and video indexing approach, using a neural network based face detector, pseudo two-dimensional Hidden Markov Models and a k-means clustering algorithm. In contrast to this work, we are facing additional constraints due to noisy environment and real time constraint.

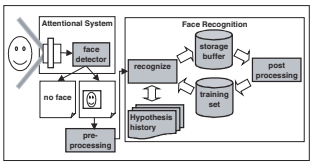

Figure 1: The schematic of the online and unsupervised face recognition system. The system receives video stream as input and learn to recognize individuals in an online and unsupervised manner.

Approach: As an initial step, we have implemented an online and unsupervised face recognition system, based on the eigenface approach [8] As shown in figure 1, visual input from the camera is passed as input to the face detector system [9]. Face detector sends information regarding location of faces found within the robot's visual range, if any,

[8] We would also like to acknowledge usage and include the copyright notice of Turk and Pentland [5] (Copyright 1992, Massachusetts Institute of Technology. All Rights Reserved).

[9] This software was developed by Paul Viola and Michael Jones[6]

for further pre-processing, where the facial image is cropped, scaled, normalized, and masked. Kismet's attention system acts to direct computational and behavioral resources toward salient stimuli and to organize subsequent behavior around them [2]. The face recognition system is simply activated when a face is detected within the visual field. If the face is the most salient stimulus at the time, the cameras will track it, maintaining it within the visual field.

Given an input stream of potential face images of people interacting with the robot, the system has to perform two tasks: discard non-face images and determine which individual the input face belongs to, if known. The system operates in two stages: pre-training (when no training set is available initially) and online training (after an initial training set is formed). In the pre-training phase, the system silently observes interactions, collects batches of input data, and attempts to acquire a rough estimate of each input's distances to the face space and also its closes neighbor. Clustering is then performed using a simple algorithm: if A is close enough to B and A belongs to cluster x, then B is placed into cluster x. Once a cluster reaches a large enough size, it is placed in the training set as a new class. In the online training phase, the system incrementally improves its knowledge of the face space. If an input image is classified as an unknown, it is then collected in batch for post-processing, where the same clustering procedure is applied. If a large enough cluster is found, it goes through another clustering process where each image in the cluster is tested against the existing training set. If the cluster is found to be close enough to a known individual, its images are added into the corresponding class. If the cluster is too far from any known classes, its images are added into the training set as a new individual.

Impact: This work is an extension of an effort to explore socially situated forms of learning for humanoid robots, where the users teach Kismet, our robot, as they would another person [1]. The online and unsupervised face recognition system learns to recognize new individuals and incrementally improve its training images over time. Thus, the more one interacts with the robot, the more likely it is for the system to obtain their characteristic images. However, if one only spends a little time playing with Kismet or just happens to pass by, they would be represented poorly in the database and thus, not easily recognized. This will lead to complex social dynamics, such as attachment to caregivers, dislike of certain people, etc. Teaching the robot would also be easier once the robot is 'familiarized' with the caregivers.

Future Work: Our immediate plan is to further improve the existing system and implement the ability to correlate simple contextual information, such as the robot's emotional state during the presence of certain individuals along with their faces. Moreover, we anticipate that speaker recognition, sound localization, and gaze detection will increase person identification performance.

Research Support: This work was funded by DARPA as part of the "Natural Tasking of Robots Based on Human Interaction Cues" project under contract number DABT 63-00-C-10102.

References:

[1] C. Breazeal. Sociable machines: Expressive social exchange b etween humans and robots. Technical report, Sc.D. dissertation, Department of Electrical Engineer ing and Computer Science, MIT, 2000.

[2] C. Breazeal and B. Scasselati. A context-dependent attention s ystem for a social robot. *Proceedings of the Sixteenth International Joint Conf erence on Artificial Intelligence*, pages 1146–1151, 1999.

[3] K. Dautenhahn. Getting to know each other - artificial soc ial intelligence for autonomous robots. *Robotics and Autonomous Systems 16*, pages 333–356, 1995.

[4] S. Eickeler, F. Wallhoff, U. Iurgel, and G. Rigoll. Content-based indexing of images and videos using face detection and recognition methods. *IEEE Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2001.

[5] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience, Vol. 3 No. 1*, pages 71–86, 1991.

[6] P. Viola and M. Jones. Robust real-time object detection. Technical Report CRL 2001/01, Cambridge Research Laboratory, 2001.

[7] J. Weng, C. Evans, and W. Hwang. An incremental learning method for face recognition under continuous video stream. *Fourth International Conference on Automatic Face and Gesture Recognition*, 2000.

[8] W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips. Face recognition: A literature survey. 2000.