Automatic Mouth Detection for Self-Feeding

Najafa Islam[1,2], Amir Mohammad Amiri[1,3], Jodi Forlizzi[4], Shivayogi V. Hiremath[1,5]

1. Department of Physical Therapy, Temple University, Philadelphia, Pennsylvania, USA
2. Department of Bioengineering, Temple University, Philadelphia, Pennsylvania, USA
3. Department of Biomedical Engineering, Widener University, Chester, Pennsylvania, USA
4. Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA
5. Department of Electrical and Computer Engineering, Temple University, Philadelphia, Pennsylvania, USA

najafa.islam@temple.edu, amir.amiri@temple.edu, forlizzi@cs.cmu.edu, Shiv.Hiremath@temple.edu

Automatic mouth detection can assist in controlling a robotic system with self-feeding of individuals with disability. To address this need we developed and evaluated algorithms that: 1) detect and track the mouth of an individual in real-time, and 2) classify if the mouth is open or closed. A k-nearest neighbors (KNN) clustering algorithm was used to classify and recognize the mouth's posture. The KNN algorithm classified image frames using features extracted by four methods including a histogram of oriented gradients, Harris-Stephens algorithm, maximally stable extremal regions, and local binary patterns. The results of this study indicated a high classification accuracy (~87%) using 10-fold cross validation for three participants without disability. The study has shown that the algorithms can detect the mouth postures of a person in near real-time (<1s) while they have a robot-assisted meal in a social setting.

We evaluated a meal assistance robot (MAR) called Meal Buddy system from Richardson Products Inc. [1]. The MAR has four degrees of freedom and can self-feed an individual with disability. We used a Kinect 2 vision system from Microsoft Inc. to track human facial movements and interaction between the user and the MAR [2]. Kinect's facial recognition feature was used to capture participants' head, mouth, and the interactions between the participant and MAR. Data was collected from three investigators without a disability. The choice of collecting this data in individuals without disability was to make sure that the pilot evaluation is successful and safe for individuals with disability. Each participant was required to take two spoonsful of food out of each bowl (3 bowls in total) for two trials. The participants sat upright in a wheelchair, in front of a table that had the MAR, and faced the Kinect sensor that recorded both the participant and MAR. The participants were asked to act normal as if they were having a meal and asked to socialize with another person around them throughout the feeding time to incorporate the social aspects of having a meal. After all the data was collected, two of the investigators coded each frame as mouth open, closed, or other. We developed a MATLAB (MathWorks Inc.) script that uses an inbuilt function called Cascade Object Detector, which is based on the Viola-Jones algorithm [3], to detect the eyes, nose, and mouth of a person [4]. The object detector uses multiple phases of pre-trained classifiers to process regions of an image for targeted facial feature recognition. In each phase, the intensity of more complex binary classifiers increases, improving the accuracy of the algorithm to refuse sections of the image that do not contain the targeted feature. In addition to the real-time detection of the mouth, features were extracted to classify the mouth movements (open vs. closed).

The K-Nearest Neighbor (KNN) algorithm was used to classify the image into mouth open or closed. For this study, we used three nearest neighbors to detect mouth open or closed. Our first phase of the study was successful as we were able to successfully detect and box the mouth area for real-time data capture when the participants were seated in front of the camera (average: 97%). However, the real-time mouth detection during a meal with a social conversation for the three participants was 73%, 67%, and 52%, respectively. Further analysis, of the individual frames indicated that the variation was due to the participant talking to the other individuals around the table while the MAR fed them. The recording of the scene and frame annotations indicated that individuals who had lower real-time mouth detection had higher durations of social interaction (talking and laughing). A KNN algorithm was able to use 13 features obtained from four image processing methods to classify the mouth posture as open or closed with a high accuracy (87%). Tenfold (k-fold) cross-validation methodology was used to assess the classification accuracy. The misclassification for the mouth open and closed images was 11% and 15%, respectively.

Limitations of this study include small sample size, testing in individuals without a disability, and data collected in a laboratory setting. Even though the participant number is small we were able to collect kinematic data at moderate sampling rates (30 frames per second), which can be used in real-world cost-effective applications. The choice of collecting data in individuals without a disability was to make sure that the pilot evaluation is safe for future studies in individuals with disability. The long-term goal of this research is to develop self-feeding assistive technology that is effective in feeding and sensitive to social aspects of having a meal.

REFERENCES

[1] Meal Buddy: Robotic Assistive Feeder. Richardson Products Inc., March 29 2018.

[2] Kinect for Windows, Microsoft Inc., March 29 2018.

[3] Viola, P. and Jones, M. Rapid object detection using a boosted cascade of simple features. Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference 1 2001, I.

[4] Detect objects using the Viola-Jones algorithm, MathWorks Inc., April 4, 2018.

Automatic Mouth Detection for Self-Feeding

Najafa Islam[1,2], Amir Mohammad Amiri[1,3], Jodi Forlizzi[4], Shivayogi V. Hiremath[1,5]

1. Department of Physical Therapy, Temple University, Philadelphia, Pennsylvania, USA
2. Department of Bioengineering, Temple University, Philadelphia, Pennsylvania, USA
3. Department of Biomedical Engineering, Widener University, Chester, Pennsylvania, USA
4. Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA
5. Department of Electrical and Computer Engineering, Temple University, Philadelphia, Pennsylvania, USA

Background

- According to the National Spinal Cord Injury (SCI) Statistical Center, the number of people with SCI is estimated to be around 358,000 in 2017, with 17,700 new cases of SCI occurring each year [1].
- Individuals with SCI who have a severe disability rely on caretakers for performing activities of daily living such as eating and drinking.
- Automatic mouth detection can assist in controlling a robotic system with self-feeding of individuals with disability.
- In this study we show that machine learning algorithms can be used to detect the mouth postures of a person in near real-time while they have a robot-assisted meal in a social setting.

Objectives

- We developed and evaluated algorithms that:
  - Detect and track the mouth of an individual in real-time (Figure 1)
  - Classify if the mouth is open or closed during eating in a social setting.

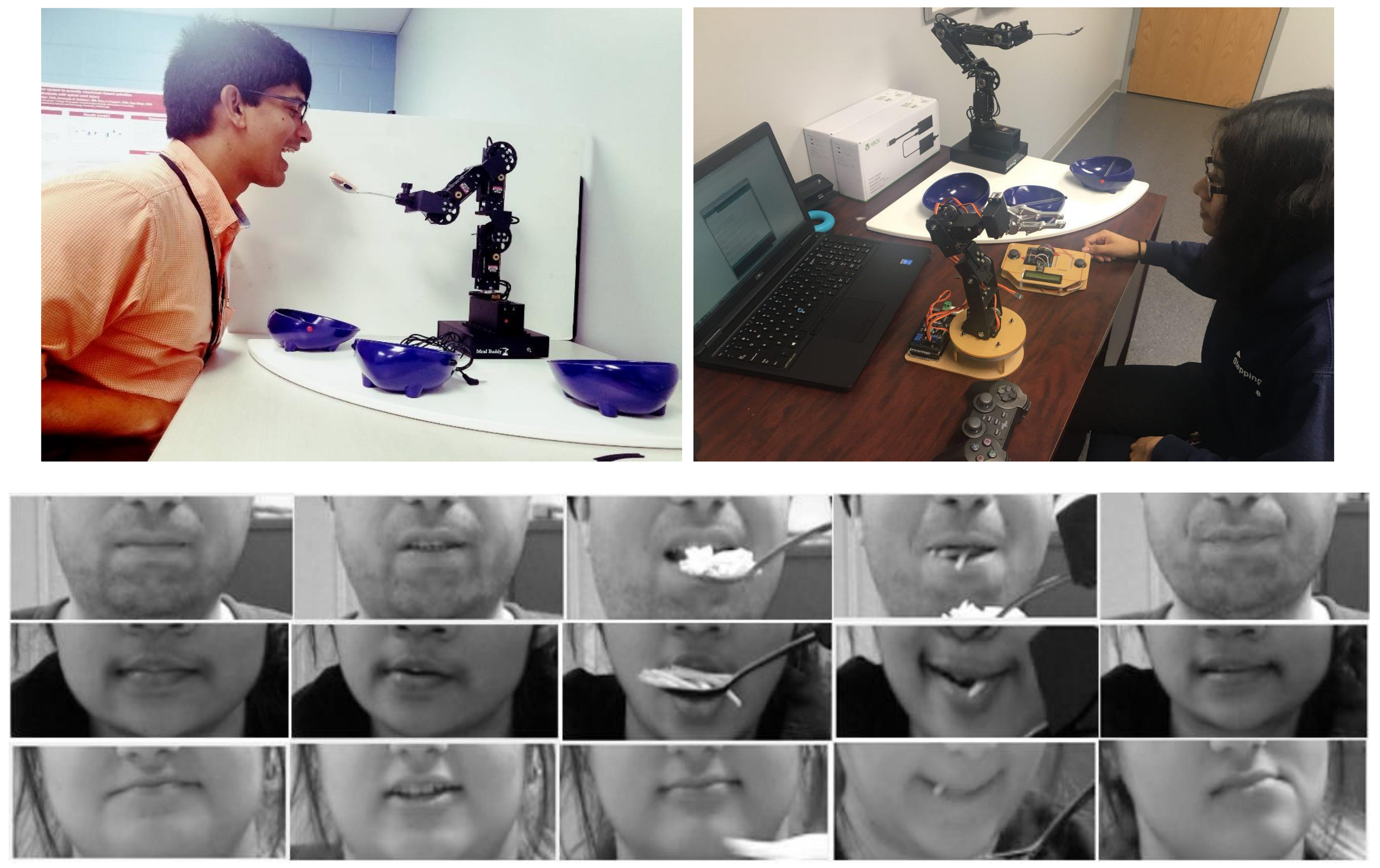

Figure 1 (Top) Two individuals interacting with a meal assistance robot. (Bottom) A sample of 3 participants' mouth detected in real-time during a social meal.

Methods

- Three individuals without a disability participated in the study within a university-based laboratory setting.
- We evaluated a meal assistance robot called Meal Buddy system from Richardson Products Inc.
  - The robot has four degrees of freedom and can self-feed an individual with disability when they activate a switch.
  - The device can be easily calibrated for a new person, home, or a bowl position.
- We used a Kinect 2 vision system from Microsoft Inc. to track interactions of a participant with the robot (self-feeding) and another person (social).
- Each participant was required to take two spoonsful of food out of each bowl (3 bowls in total) for two trials.
- Real-time mouth detection was performed using Cascade Object Detector, which is based on the Viola-Jones algorithm [2], to detect the eyes, nose, and mouth of a person [3].
- In addition to the real-time detection of the mouth, features were extracted to classify the mouth movements (open vs. closed):
  - Histogram of Oriented Gradients (HOG)
  - Detecting Corners using Harris-Stephens Algorithm
  - Maximally Stable Extremal Regions (MSER)
  - Extracting Local Binary Patterns (LBP).

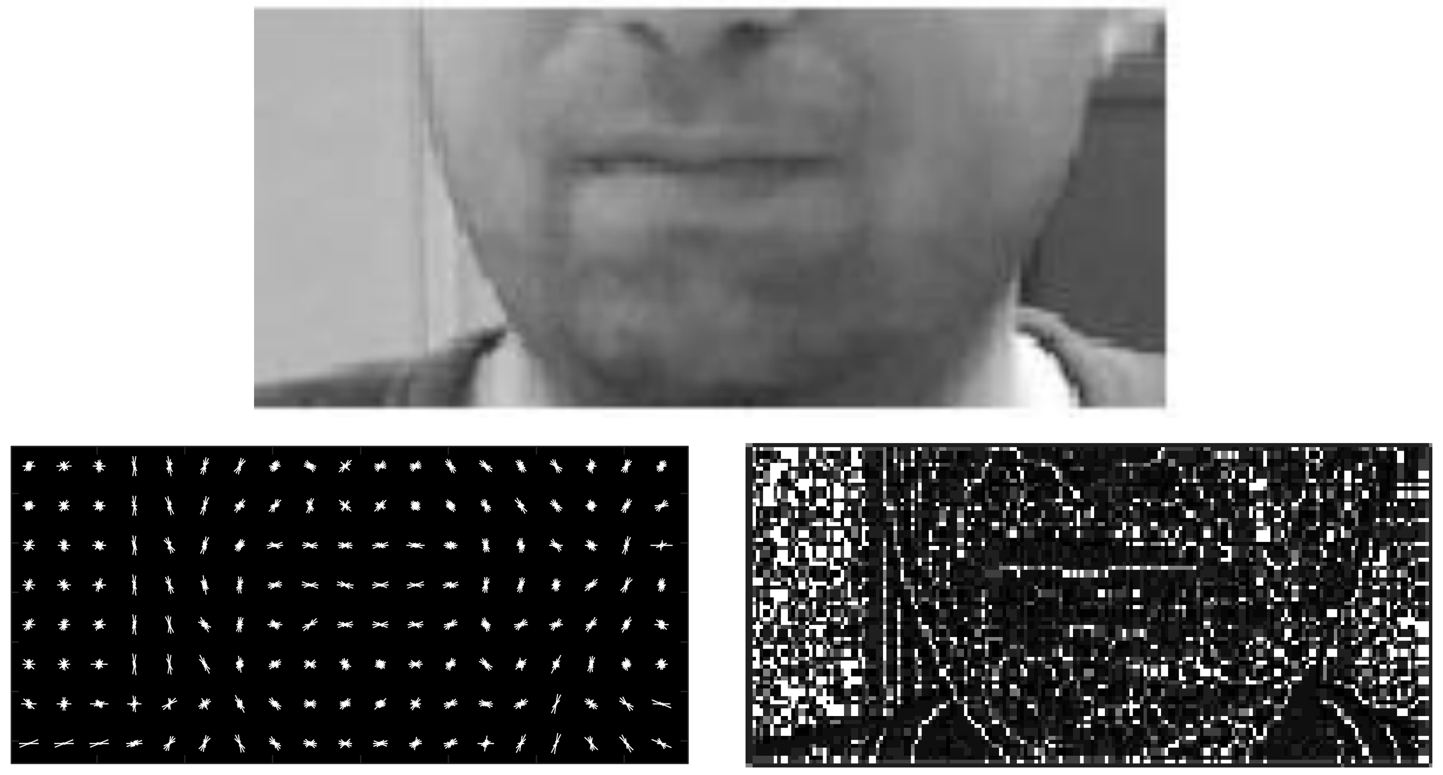

Figure 2: (Top) Original Image of a participant's mouth. (Bottom Left) HOG feature extracted from the original image. (Bottom Right) LBP feature extracted from the original image.

Results

- Several machine learning algorithms were used to classify and recognize an individual's mouth posture.
- The K-Nearest Neighbors (KNN) clustering algorithm was able to use 13 features to classify a open or closed mouth with high accuracy (87%).
- The Receiver Operating Characteristic (ROC) curve for the KNN classification algorithm resulted in an area under the curve of 0.93, indicating an excellent performance.

	Predicted (%)		
Actual	Open	Closed	Correct
Open	320 (85%)	121 (11%)	85%
Closed	60 (15%)	955 (89%)	89%
Overall accuracy for 1456 samples			87%

Table 1: Classification Result of K-NN Algorithm.

Model	Accuracy	AUC	Training Time
Decision Tree	82.6%	0.85	0.8s
Support Vector Machine	83.0%	0.89	1.3s
KNN	87.0%	0.93	3.5s
Decision Trees	82.4%	0.88	2.4s
Linear Discriminant	81.0%	0.88	0.7s

Table 2: Comparison of Accuracy, Area Under Curve (AUC) and Training Time for different classifiers.

Conclusion

- The study showed that machine learning algorithms can detect mouth postures of a person in near real-time while they have a robot-assisted meal in a social setting.

References

[1] National Spinal Cord Injury Statistical Center. Spinal cord injury (SCI) facts and figures at a glance. 2018
[2] Viola, P. and Jones, M. Rapid object detection using a boosted cascade of simple features. Computer Vision and Pattern Recognition. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference.
[3] MathWorks Inc. Detect objects using the Viola-Jones algorithm, 2018.

Acknowledgments

- We would like to thank Richardson Products Inc. for kindly allowing us to use their Meal Buddy System.

IEEE Signal Processing in Medicine and Biology Symposium 2018