16–811: Project Report

Quantifying Gestural Mimicry in Kinect-taped Communication

Prerna Chikersal (pchikers@andrew.cmu.edu)

1 Problem Definition:

The Microsoft Kinect SDK v2 can track 25 point skeletons for up to 6 people. It has both standing and seated tracking modes, and returns (x,y,z) positions as well as rotations for each of the joints shown in figure 1. Given the joint positions (i.e. (x,y,z) coordinates) of two people engaged in a conversation in time (30 frames per second), we aim to calculate the similarity between their gestures, in order to estimate the amount of gestural mimicry in a social interaction.

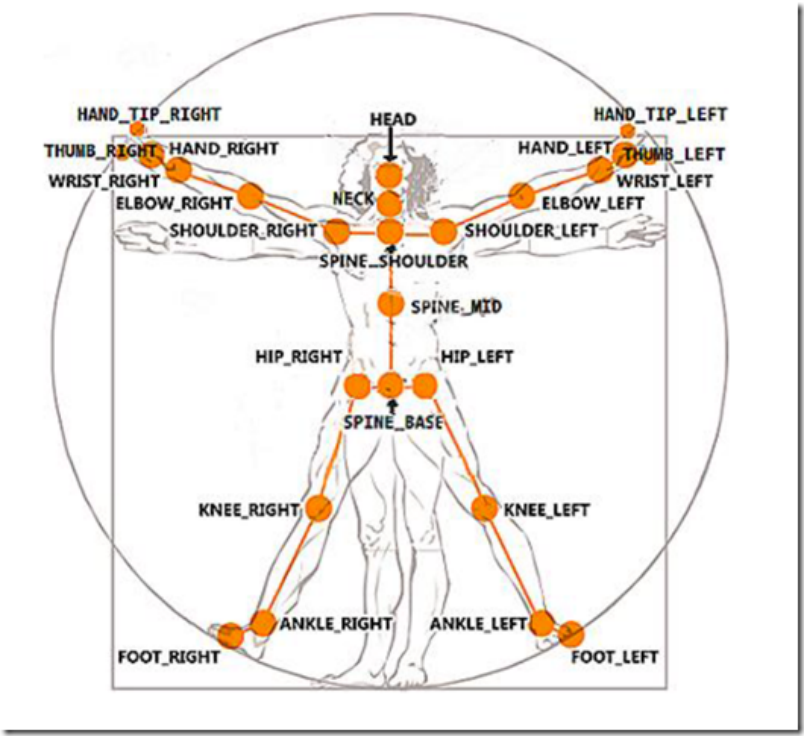

Figure 1: Joints returned by Kinect

2 Motivation

Previous works such as [1] and [2] show that behavioral mimicry during a social interaction leads to stronger rapport formation. My current research involves automatically measuring rapport in a social interaction using various modalities such as Kinect Skeletal Tracking

(for gestural mimicry), video (for facial mimicry and speech-related features like tone and prosody), electrodermal activity (for physiological synchrony), EEG (for neurological synchrony), and eye gaze (for eye contact and engagement).

3 Methods

This section describes the methods we tried to compute a measure for gestural mimicry.

4 Similarity of Motion using Lagged Cross-correlation

Lagged cross-correlation takes two curves, and outputs how similar they are at different units of lag. However, though it gives us the lag at which two curves are the most similar, it zeroes out all the areas where the two curves don't overlap at that lag. To counteract this problem, we will take one of the curves, and rotate the curve in place until it best matches the other curve. Then, to compare the two new curves, we will use Spearman's and Pearson's correlations. Pearson's is better suited for linear relation- ships in data, whereas Spearman's rho is more accurate for non-linear correlation and less affected by outliers.

We get a curve for the motion of each joint, and compute lagged cross-correlation between the curves for each joint of the two people to get similarity for each joint. We average similarities of all joints to get overall average similarity of motion.

Note that this method is not a very effective measure for measuring gestural similarity or mimicry due to many reasons, such as:

1. It measures average similarity of motion of all joints, which is not the same as gestural mimicry. Not all motion indicates a gesture. For gestural mimicry, it would make sense to first extract gestures from the motion of the joints and then match them across participants.
2. Similarity is averaged across all joints. This does not seem right, as all joints may or may not contribute equally to the gestural mimicry score.
3. Correlation is calculated on a joint-by-joint basis where each joint is compared to the corresponding joint of the other participant (Right hand of person 1 to right hand of person 2, left hand of person 1 to left hand of person 2, and so on). This is also not theoretically correct, since gestural mimicry can also include mirroring. For example, person 1 might use his/her left hand to mimic the gestures of person 2's right hand.

The method described in the next section overcomes limitations 1 and 2.

5 Matching Unit Motions to Quantify Gestural Mimicry

This method modifies Wang and Lai's [3] approach and extends it to compute interpersonal gestural similarity or mimicry.

5.1 Segmenting Motion into Unit Motions

We plot the speed-time graph of each joint, and threshold it to get unit motions. We tried the following thresholds: 0.05, 0.10, 0.15, 0.20, and 0.25. Figure 2 shows an example of one such plot.

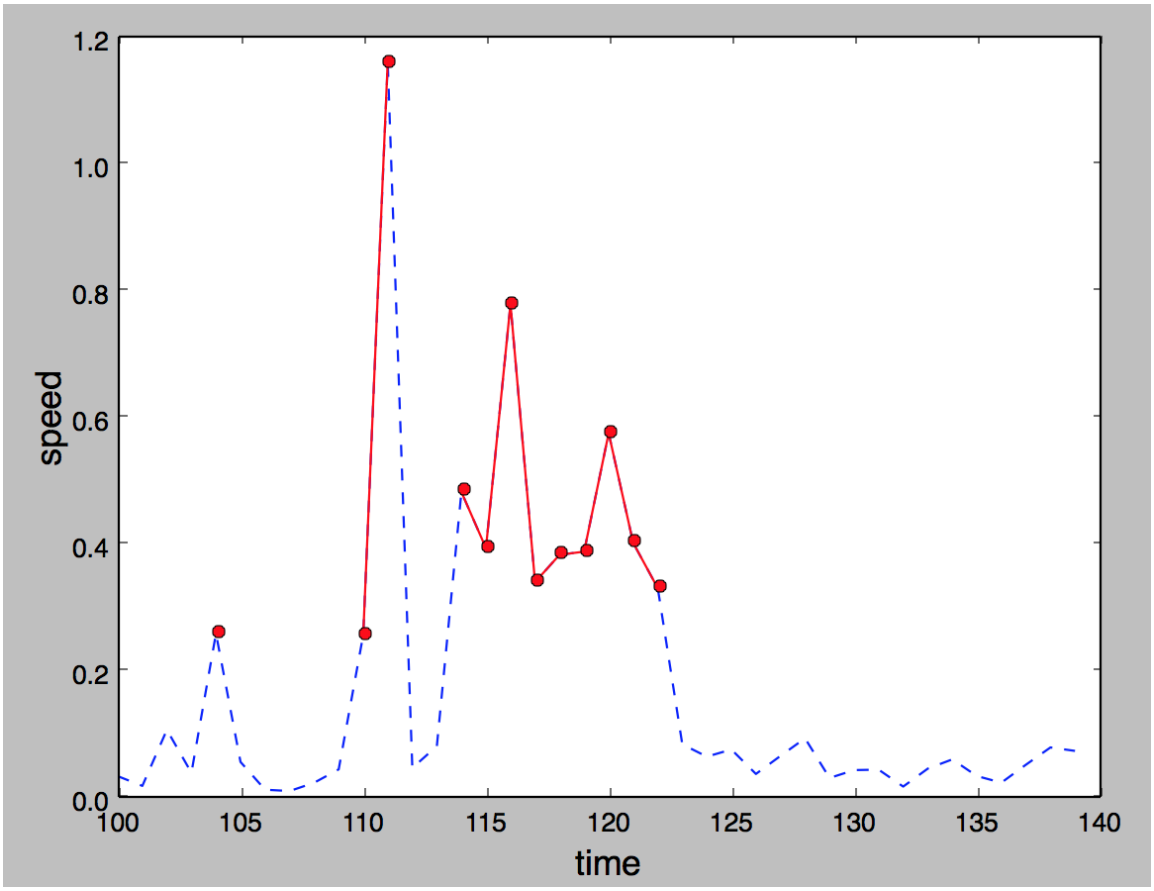

Figure 2: The lines in red are the segmented unit motions from time 104 to 104, 110 to 111, 114 to 122, extracted from the motion of a person's right wrist from time 100 to 140.

5.2 Extracting Features from Unit Motions

Once we have unit motions for each joint of each participant, we extract the 18 features per unit motion, as described below.

Let **time duration** of unit motion be t.

Features 1-3 are based on "Path" of gestural motion, as follows:

$p_x = |x_2 - x_1| + |x_3 - x_2| + ... + |x_n - x_{n-1}|$
$p_y = |y_2 - y_1| + |y_3 - y_2| + ... + |y_n - y_{n-1}|$
$p_z = |z_2 - z_1| + |z_3 - z_2| + ... + |z_n - z_{n-1}|$

Features 4-6 are based on "Displacement" of gestural motion, as follows:

$d_x = |x_n - x_1|$
$d_y = |y_n - y_1|$
$d_z = |z_n - z_1|$

Features 7-9 are based on "Speed" of gestural motion, as follows:

$s_x = p_x/t$
$s_y = p_y/t$
$s_z = p_z/t$

Features 10-12 are based on "Velocity" of gestural motion, as follows:

$v_x = d_x/t$

$v_y = d_y/t$
$v_z = d_z/t$

Features 13-15 are based on "Maximum Distance" of gestural motion, as follows:

$Max_x = max(|x_1| + |x_2| + ... + |x_n|)$
$Max_y = max(|y_1| + |y_2| + ... + |y_n|)$
$Max_z = max(|z_1| + |z_2| + ... + |z_n|)$

Features 16-18 are based on "Sum of directional movement" of gestural motion, as follows:

$\overrightarrow{a_k} = (x_k, y_k, z_k) - (x_{k-1}, y_{k-1}, z_{k-1}) = (p_k, q_k, r_k)$
$k = 2 \sim n$
$cos\alpha_k = \frac{p_k}{|\overrightarrow{a_k}|}$
$cos\beta_k = \frac{q_k}{|\overrightarrow{a_k}|}$
$cos\gamma_k = \frac{r_k}{|\overrightarrow{a_k}|}$
$=> cos_x = \sum_{k=2}^{n} cos\alpha_k$
$=> cos_y = \sum_{k=2}^{n} cos\beta_k$
$=> cos_z = \sum_{k=2}^{n} cos\gamma_k$

5.3 Computing Pairwise Similarity between Unit Motions

To compute similarity between two unit motions, we calculate the cosine similarity between their feature vectors from above.

For each joint type, we get a pairwise cosine similarity matrix that contains the similarity between the unit motions for that joint of person 1 and person 2. Thus we get a pairwise similarity matrix for each joint type. These pairwise cosine similarity matrices are updated in real-time as skeleton tracking data and thus new unit motions come in from newly arrived Kinect frames.

5.4 Computing Interpersonal Gestural Mimicry between Two People

From the pairwise cosine similarity matrices of each joint, we try to count mimicked gestures instead of finding an averaged similarity score for mimicry. This is because mimicry is not clearly defined for each joint in theory, and hence we cannot simply average or calculate a weighted average of similarity across all joints and call that mimicry.

Based on psychological studies, gestures are usually mimicked unconsciously within 10 seconds of each other. Hence, for each joint, unit motions are considered to be mimicked if they occur within 10 seconds of each other, and their cosine similarity is greater than 0.8. We add the counts of mimicry for all joints to get the final mimicry count.

6 Experiments and Preliminary Results

6.1 Dataset

We use two Kinects to collect the skeleton tracking data of two participants engaged in a dyadic interaction. We also ask the participants to answer a survey, and calculate rapport score for ground truth by averaging their answers to two questions related to their perceived rapport. We collected 8 such interactions, and the preliminary results below are based on them.

6.2 Analyzing Correlation between Occurrence of Mimicry and Self-Reported Rapport for the Whole Interaction

To demonstrate that our method works, we attempt to show a positive correlation between the calculated mimicry count and the averaged self-reported rapport score. We get the following results for the methods we implemented:

1. Persons cross-correlation: +0.028
2. Spearmans cross-correlation: 0.000
3. Unit motion method with speed threshold 0.05: +0.061
4. Unit motion method with speed threshold 0.10: +0.101
5. Unit motion method with speed threshold 0.15: +0.228
6. Unit motion method with speed threshold 0.20: +0.380
7. Unit motion method with speed threshold 0.25: +0.300
8. Unit motion method using average of the results obtained with all speed thresholds above: +0.314

7 Conclusion and Future Work

You can see from the above that the lagged cross-correlation methods do not work well, while the methods that matches unit motions (especially speed threshold 0.2, average speed threshold, and speed threshold 0.15) to count mimicry works reasonably well.

However, currently, we still do not take into account laterally inverted mimicry (e.g. left hand of person 1 mimicking right hand of person 2). Also, we count mimicry of each joint separately. In future, we will use geometry-based posed descriptors to take into account mimicry of gestures that involve more than one joint.

References

[1] J. L. Lakin, V. E. Jefferis, C. M. Cheng, and T. L. Chartrand, "The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry," *Journal of nonverbal behavior*, vol. 27, no. 3, pp. 145–162, 2003.

[2] J. L. Lakin and T. L. Chartrand, "Using nonconscious behavioral mimicry to create affiliation and rapport," *Psychological science*, vol. 14, no. 4, pp. 334–339, 2003.

[3] H.-C. Wang and C.-T. Lai, "Kinect-taped communication: using motion sensing to study gesture use and similarity in face-to-face and computer-mediated brainstorming," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3205–3214, ACM, 2014.