Computational Analysis of Nonverbal Communication Cues in Group Settings

Rui Frazão[1]
ruifilipefrazao@ua.pt
Samuel Silva[12]
sss@ua.pt
Sandra Soares[3]
sandra.soares@ua.pt
António J. R. Neves[12]
an@ua.pt

[1] DETI
University of Aveiro, Aveiro, Portugal
[2] IEETA
University of Aveiro, Aveiro, Portugal
[3] CINTESIS.UA
WJCR.UA
Dept. of Education and Psychology
University of Aveiro, Aveiro, Portugal

Abstract

Human communication is a major field of study in psychology and social sciences. Topics such as emergent leadership and group dynamics are commonly studied cases when referring to groups. Group settings experiments are usually studied in conversational and collaborative tasks environments in order to study the communication process in small groups. Former study methods involved human analysis and manual annotation of other's behaviors in communication settings. Later studies try to replace time consuming and failure prone annotations by resorting to computational methods.

For that purpose, we propose a multimodal approach capable of using a broad range of nonverbal communication in a complementary way in order to allow the quantification of nonverbal aspects from video data. This paper presents a framework capable of contributing to a direct increase in human knowledge about the human communication process, involving data transformation processes in order to transform raw feature data into humanly understandable meanings.

1 Introduction

Communication is a natural, omnipresent process in human lives. Since birth, humans use signals, sounds, movements and expressions to communicate with others. Human communication is defined as the process of human being's interaction to other's behaviors. When one thinks about human communication, verbal communication is what comes to mind first. This type of communication is done verbally and relies on the use of words and phrases to convey meanings. However, information passed during the communication process lies not only on what the sender and receiver are transmitting verbally but also through their behavior.

For a long period of time, the human communication subject has been studied mainly by the psychology and social sciences areas, manually classifying behaviors and annotating datasets without computational methods involved in the process. Nowadays, there are computational methods capable of extracting features from human communication situations, thus allowing a deeper analysis of the communication process.

The research reported in this paper aims to contribute to the study and understanding of the human communication phenomenon. The main objective is to create a framework grounded on a critic analysis of current literature, on how to develop a computational system capable of quantifying nonverbal aspects from video data, following a multimodal approach, by analyzing different nonverbal features simultaneously in a complementary way, and thus broadening the analysis of the communication process context, contributing to richer information, as the message is eventually only understood in full when all its parts are considered.

2 Background

Researchers have defined nonverbal communication by identifying characteristics that constitute it [4]. The set of signals transmitted via a particular medium or channel is called "Code". The various codes in combination form the structure of nonverbal communication as known today. This codes are often defined by the human sense or senses they stimulate and/or the carrier of the signal [2]. Table 1 enumerates the different types of codes along with some of the features they represent.

Code Type	Code Name	Features
Visual	*Kinesics*	Facial expressions; Head movements; Eye behavior; Gestures; Posture; Gait
Auditory	*Vocalics* or *Paralinguistics*	Dialect; Pitch; Tempo; Dysfluencies; Intonation
Body	*Attractiveness*	Appearance; Adornments; Olfatics
Contact	*Proxemics*, *Haptics*	Space; Distance; Touch
Time	*Chronemics*	Timing
Place	*Artifacts*	Environment Objects

Table 1: Types of nonverbal communication codes and corresponding code names and constituent features.

As the data needed to analyze human communication is multimodal by nature, following a multimodal approach tends to achieve better results compared to single modalities [7], yet, the processing of multidimensional data also constitutes a problem due to the need of large computational power and optimized algorithms.

Multimodal studies can follow either a complementary or redundant analysis. Complementary approaches focus on broadening the analysis of information emitted from multiple communication channels, while redundant tend to seek validation to an assumption or conclusion, using different information from various communication features.

3 Computational Approach to the Study of Nonverbal Communication

For a specific use case, a multidisciplinary research group, having members from engineering and psychology backgrounds collected a custom dataset from an experiment conducted by the Department of Psychology of the University of Aveiro regarding the influence of conflict in group collaborative tasks. As such, the experiment consists in two tasks, where only in the latter, the conflict is induced.

The input data is composed of three camera sensors pointed to a table where four subjects sit and perform LEGO construction tasks. To limit the image processing to the relevant image regions reducing computational times and unwanted artifacts, the image data is clipped to a custom-defined region of interest, and created a method to match subjects in different perspectives.

Considering the nonverbal codes and its features and the available raw pose and facial features extracted from the video data, it is possible to discard the chance of including auditory and time features, as the former requires audio data and the latter is not relevant as each task must be done within a specific time frame, which is monitored by the experiment staff and thus would not provide any additional information.

There are two main elements to extract: Pose and Facial Landmarks. There are several state-of-the-art methods for human pose and facial keypoints extraction. We use OpenPose[3] and DensePose[5] for pose extraction and OpenFace[8] for facial landmarks extraction.

After the feature extraction phase, it is be possible to match nonverbal cues to behaviors and understand the meaning of those behaviors. By identifying relevant features, an individual analysis of every element's features is done, followed by the combination of each of their corresponding feature vectors in the following step.

Having extracted the feature vectors and correspondent meanings, it was necessary to study how to quantify the nonverbal cues. Determining metrics that apply to the specific use case is also of value to the task and can be calculated by transforming and/or combining the extracted vectors. This is a major step in understanding the correlation between behaviors and the human communication process. Some of this information can also be presented overlaid on the original image data in order to better analyze and understand the behavior during the communication phenomenon.

Visual features

Visual features are mainly linked to posture and positioning of the involved subjects. Such features involve: Facial features, which allow the analysis of head movement and direction and emotion related data, and can be directly related to the level of interest of a subject in a determined task [7, 8]. In order to quantify emotion, a naive approach based on combining the activation of facial muscles was followed [9]; Body expansiveness, which is usually correlated to the perception of influence, power and dominance, can be quantified as the occupied area both horizontally and vertically by the polygon involving the furthest pose keypoint; Group activity, as it is intended to understand how is group energy affected by each experiment condition and if, as a consequence of the existence of conflict in a group, habits tend to vary and how as this can be an indicator of subjects' disengagement. As a way of quantifying the activity, two approaches were taken: Motion Energy Image as proposed by [1, 7], and analysis of the keypoint movement between frames.

Body features

Body features are not easy to quantify based on nonverbal behavior. Studies show strong correlation between physical attractiveness, body and face symmetry, and social and cognitive attributes as the most relevant characteristics in the attractiveness field [4, 6]. Not being able to retrieve data covering those aspects, only antropometric information could possibly be used. Although it is shown in literature that there is a correlation between some physical attributes and, perceived competence and leadership in group settings [4], this information alone is not considered relevant to this specific case.

Contact features

Contact and visual codes can be considered highly correlated as features such as occupied space, distance and touch are measured taking posture and gestures into account. Proximity-related features such as overlap and distance between group subjects (intragroup distance) can also be correlated as the closer the group is, the bigger is the overlap. Overlap is calculated based on the subjects' occupied areas intersections, and intragroup distance is given by the distance between subjects. These features are considered representative of group cohesiveness and consequently, the level of engagement in the task. Cohesiveness is an important matter in study of group dynamics, as it is positive involvement behavior [4, 10].

Place features

Artifact related aspects can easily be extracted in this particular use case, as the experiment features the handling of objects displayed in the experiment environment. Such interactions can provide enriched information about the subjects' behavior and engagement in the group tasks. In this specific use case, the interaction with the displayed objects is measured by the subjects' distance to the center of the table.

4 Results and Conclusion

The most important output of the work described in this document is a framework capable of quantifying the described nonverbal aspects in a group setting from video data, with the intent to aid field professionals analyzing a group's behavior, offering overlaid visual information on top of the original data, and generating plots of the quantified features. The dataset used in this work was fully annotated by this framework.

Figure 1, on top, presents the visualization tool developed to display the processed feature data regarding nonverbal communication aspects on top of the original input data, contributing to a easier analysis of such aspects. Here is demonstrated both the overlaid information of subject's keypoints and overlap. At the bottom, a plot illustrating the comparison between groups' intragroup distance along the experiment.

Some aspects that would improve the results obtained in the work would be the use of higher quality image data and the use 3D information in order to be possible to analyze other types of features that are not possible to quantify in a 2D space, such as body orientation.

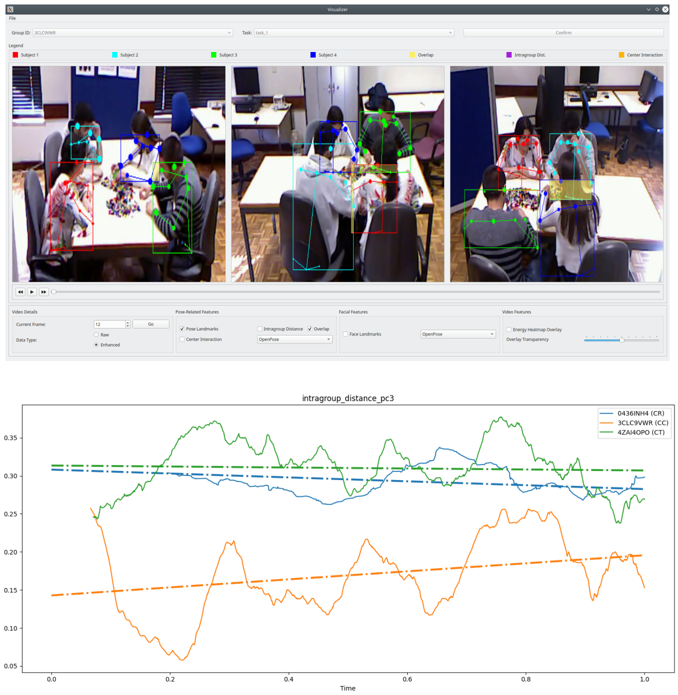

Figure 1: Top: Visualization tool displaying keypoint position and overlap features; Bottom: Comparison between groups intragroup distance along the experiment.

References

[1] Aaron F. Bobick and James W. Davis. The recognition of human movement using temporal templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(3):257–267, 2001. ISSN 01628828. doi: 10.1109/34.910878.

[2] Judee K Burgoon, Laura K. Guerrero, Kory Floyd, Laura K. Guerrero, and Kory Floyd. *Nonverbal Communication*. Routledge, jan 2016. ISBN 9781315663425. doi: 10.4324/9781315663425.

[3] Zhe Cao, Tomas Simon, Shih En Wei, and Yaser Sheikh. OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, pages 1302–1310, 2018. doi: 10.1109/CVPR.2017.143.

[4] Laura K. Guerrero. Interpersonal functions of nonverbal communication. In *Interpersonal Communication*, pages 53–75. DE GRUYTER, Berlin, Boston, 2014. ISBN 9783110276794. doi: 10.1515/9783110276794.53.

[5] Rıza Alp Güler, Natalia Neverova, and Iasonas Kokkinos. DensePose: Dense Human Pose Estimation In The Wild. feb 2018.

[6] J. L. Rennels. *Physical attractiveness stereotyping*, volume 2. Elsevier Inc., 2012. ISBN 9780123849250. doi: 10.1016/B978-0-12-384925-0.00099-7.

[7] Dairazalia Sanchez-Cortes, Oya Aran, Marianne Schmid Mast, and Daniel Gatica-Perez. A nonverbal behavior approach to identify emergent leaders in small groups. *IEEE Transactions on Multimedia*, 14:816–832, 2012. ISSN 15209210. doi: 10.1109/TMM.2011.2181941.

[8] Tadas Baltrusaitis. *Automatic facial expression analysis*. PhD thesis, University of Cambridge, 2005.

[9] Sudha Velusamy, Hariprasad Kannan, Balasubramanian Anand, Anshul Sharma, and Bilva Navathe. A method to infer emotions from facial Action Units. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2028–2031. IEEE, 2011. doi: 10.1109/ICASSP.2011.5946910.

[10] Bin Zhu, Xin Guo, Kenneth E. Barner, and Charles Boncelet. Automatic group cohesiveness detection with multi-modal features. In *ICMI 2019 - Proceedings of the 2019 International Conference on Multimodal Interaction*, pages 577–581. Association for Computing Machinery, Inc, oct 2019. doi: 10.1145/3340555.3355716.