Visual Analytics of Nonverbal Behavior to Evaluate Collaborative Group Engagement

Matus Gasparik*
FHNW University of Applied Sciences and Arts Northwestern Switzerland

Carolin Bronowicz†
FHNW University of Applied Sciences and Arts Northwestern Switzerland

Susanne Bleisch‡
FHNW University of Applied Sciences and Arts Northwestern Switzerland

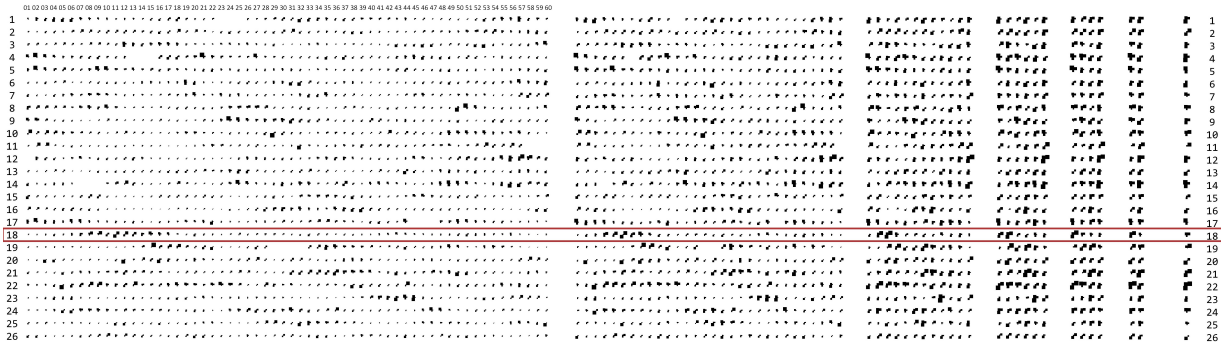

Figure 1: Visualization of the normalized total head movement of three group members as 3-dimensional glyphs. The 26 minutes of group work (one row per minute) are aggregated at different time intervals (columns fltr: 1sec, 2sec, 5sec, 10sec, 15sec, 30sec, 1min). The 18th minute of group work, at all temporal granularities, is highlighted in red.

Abstract

Despite rapid advances in AI, computer vision, and the availability of off-the-shelf tools, analyzing and understanding the dynamics of nonverbal behavior (NVB) remains a significant challenge, especially in the analysis of collaborative group engagement [3]. Research areas such as Social Signal Processing aim to leverage computational methods to automatically extract NVB from high-volume, multimodal video, audio, and language data, but with moderate success. These automated approaches rely heavily on large, high-quality training datasets and often face issues related to predicted constructs' theoretical soundness and context-specific validity. A promising alternative is Visual Analytics (VA), which integrates human reasoning with computational methods for data interpretation. This poster explores a methodological approach using VA to extract and analyze NVB in collaborative learning. We employ state-of-the-art computer vision techniques to generate high-resolution time series of facial, hand, and body landmarks from video recordings of small student groups collaboratively solving computer-based tasks. These landmarks are then processed into meaningful NVB signals and visualized to enable exploration and analysis. We also introduce visual-mapping strategies to address the challenges posed by high-dimensional data and the information loss introduced by aggregation. Finally, we demonstrate the potential and limitations of VA to support the analysis of both individual and dyadic NVB, highlighting temporal patterns in head movement and mutual orientation (facing direction) within small-group interactions.

Index Terms: Visual analytics, nonverbal behavior, collaborative group engagement, video-based body landmarks.

*e-mail: matus.gasparik@fhnw.ch
†e-mail: carolin.bronowicz@fhnw.ch
‡e-mail: susanne.bleisch@fhnw.ch

1 Introduction

Understanding nonverbal behavior (NVB) in group settings is crucial for various domains, especially education. Traditional methods often rely on manual video annotation, which is time-consuming, coarse in resolution, and prone to subjectivity [3]. Advancements in computer vision technologies have enabled more scalable and objective data collection, but often at the cost of interpretability, as these methods tend to compress rich behavioral dynamics into narrow, predefined constructs. In this poster, we explore an alternative approach based on Visual Analytics (VA), which combines computational data processing with interactive visualization to support open-ended exploration [1] of complex, high-dimensional NVB data. Rather than aiming to detect specific behaviors automatically, the VA framework facilitates the discovery of emergent patterns through visual inspection, providing a flexible and theory-agnostic method for analyzing NVB in collaborative group settings.

2 Methodology

The data was collected from face-to-face collaborative learning sessions involving groups of three students working around a shared table for different digital tasks involving the use of a computer. Depending on the tasks, the sessions lasted between 25 and 100 minutes and were recorded using a single video camera positioned approximately 1.5 meters from the group (see [3]). To extract NVB signals, the MediaPipe Holistic library [2] was used, which detects facial, hand, and body landmarks for each frame. Videos were sampled at 5 Hz to maintain a sufficient temporal resolution for capturing subtle movements such as micro-expressions and head turns. Each participant's region of interest was processed separately using a segmentation pipeline based on the YOLOv8 computer vision model architecture, which ensured consistent person identifiers even in static scenes. The extracted landmark data, including facial meshes and body joint positions, were saved alongside timestamps and anonymized metadata. The facial landmarks were transformed into metric 3D coordinates, enabling accurate tracking of head poses and orientation.

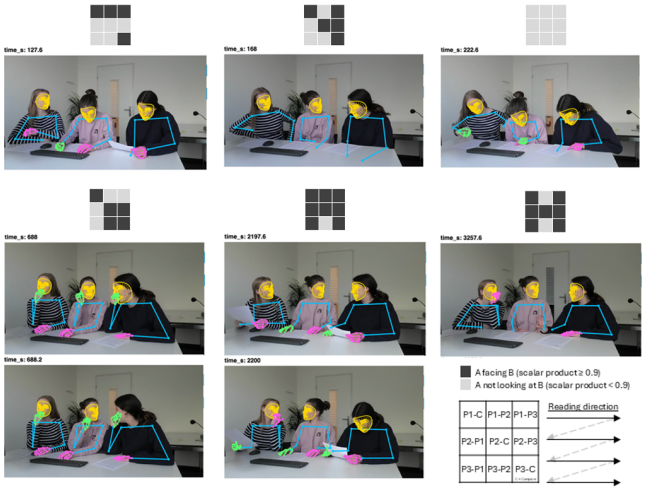

Figure 2: Examples of the 3x3 glyph arrays showing the binarized maximum MOIs during a 5-second segment, along with one (top row of figures) or two (two bottom rows of figures) characteristic frames from this period that illustrate the pattern of the 3x3 glyph array. The example frames are overlaid with the extracted body landmarks (colored lines and points).

Several derived signals were calculated from the body landmarks to support behavioral analysis: The head pose was described using translational and rotational parameters based on rigid-body transformations of the face mesh. To quantify mutual attention and orientation, a Mutual Orientation Index (MOI) was computed for each pair of participants as well as between participants and the computer screen.

To maintain participant privacy, no video frames were stored. Instead, the system supports optional resynchronization of landmark data with video when needed for validation of findings. The processed data was visualized using custom interactive tools. We used glyph-based encodings, such as glyph arrays or multidimensional area-based glyphs, to represent multi-dimensional time-series data in compact visual formats. These visualizations allowed to explore both individual and dyadic NVB patterns by aggregating data in different time intervals and interactively connecting representations.

3 Results

The raw body landmark data provided high temporal resolution but it was too dense for meaningful pattern analysis without aggregation. Different options of time aggregation were implemented but only regular segmentation was acceptable to the domain specialists. Exploring the data at different temporal granularities (see Fig. 1) seems to provide insight into persistent and temporary group activities. Density maps summarize combined translational and rotational movements, providing intuitive overviews of interaction behavior over time. The MOI, an abstraction derived from face pose data, is visualized using a 3×3 grid for each time segment. In these arrays, diagonal elements indicated attention toward the computer screen, while off-diagonal elements showed mutual gaze or orientation between participants. Binary color coding and hierarchical layout enabled the identification of key patterns over time, and user-adjustable thresholds allowed for flexible interpretation (see Fig. 2).

4 Discussion

The exploration of derived signals and different visualizations of them illustrates the strength of Visual Analytics in exploring rich, temporally detailed NVB data. Unlike black-box machine learning models for the extraction of defined signals, an interactive visual approach offers interpretability by keeping the human in the loop and allowing researchers to iteratively process and visualize data. However, for the visual analysis of domain-specific data, the domain experts play a crucial role. They are able to provide the context required to make sense of the visible patterns. The setup of this project between domain specialists and visualization researchers agreed on the need to extend on existing behavioral constructs as they are commonly used in manual encoding of group interactions and the potential value of visual analytics to do so. But the domain specific value was not fully exploited. Visualization alone cannot solve every analytical challenge. The quality and granularity of the underlying data define what patterns can be meaningfully observed. Inaccurate or noisy signals may lead to misinterpretations if visualizations are not grounded in data preprocessing and feature extraction. Our hybrid strategy that combines targeted computer vision methods with interactive visualization tools seemed promising. However, custom visual representations like glyph arrays require domain experts to learn new interpretive strategies and to consider emerging, potentially unexpected, relevant NVB patterns. Moreover, NVB is inherently contextual and culturally variable. While we focused on gaze and movement, many NVB phenomena, such as emotional expression or impression management, rely on subjective interpretation and cultural norms. Therefore, while VA enhances insight into NVB, it must be paired with strong data foundations and domain expertise to ensure interpretive accuracy.

In exploring how NVB can be meaningfully analyzed based on high-resolution multimodal data, the question arises whether insights are primarily gained by designing suitable interactive visualizations, or rather in statistical modeling to interpret the data. However, all approaches depend to some degree on the quality and structure of the data they represent and the willingness and contextual knowledge of the domain experts to interpret findings. Future work will continue to examine how best to extract meaningful interpretations from complex NVB data of collaborative group work.

Acknowledgments

This work has been funded by the Swiss National Science Foundation within the National Research Program NRP 77 (Project 407740_187258) Next generation learning: Investigating and enhancing collaborative group engagement quality to support learning groups. We thank all the study participants for their time and engagement. Without them, this work would not have been possible. All data is from participants who have given their informed consent about the use and publication of the collected data, including pictures.

References

[1] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. Visual analytics: Definition, process, and challenges. In A. Kerren, J. T. Stasko, J.-D. Fekete, and C. North, eds., *Information Visualization*, vol. 4950, pp. 154–175. Springer, 2008. doi: 10.1007/978-3-540-70956-5_7 1

[2] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. Yong, J. Lee, and others. Mediapipe: A framework for perceiving and processing reality. In *Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR)*, vol. 2019, 2019. doi: 10.48550/arXiv.1906.08172 1

[3] L. Paneth, L. T. Jeitziner, O. Rack, K. Opwis, and C. Zahn. Zooming in: The role of nonverbal behavior in sensing the quality of collaborative group engagement. *International Journal of Computer-Supported Collaborative Learning*, 19(2):187–229, jun 2024. doi: 10.1007/s11412-024-09422-7 1