# Interpreting hand-over-face gestures

Marwa Mahmoud and Peter Robinson

University of Cambridge

**Abstract.** People often hold their hands near their faces as a gesture in natural conversation, which can interfere with affective inference from facial expressions. However, these gestures are valuable as an additional channel for multi-modal inference. We analyse hand-over-face gestures in a corpus of naturalistic labelled expressions and propose the use of those gestures as a novel affect cue for automatic inference of cognitive mental states. We define three hand cues for encoding hand-over-face gestures, namely hand shape, hand action and facial region occluded, serving as a first step in automating the interpretation process.

## 1 Introduction

Nonverbal communication plays a central role in how humans communicate and empathize with each other. The ability to read nonverbal cues is essential to understanding, analyzing, and predicting the actions and intentions of others. As technology becomes more ubiquitous and ambient, machines will need to sense and respond to natural human behaviour. Over the past few years, there has been an increased interest in machine understanding and recognition of people's affective and cognitive states, especially based on facial analysis. One of the main factors that limit the accuracy of facial analysis systems is hand occlusion.

Hand-over-face gestures, a subset of emotional body language, are overlooked by automatic affect inferencing systems. Many facial analysis systems are based on geometric or appearance facial feature extraction or tracking. As the face becomes occluded, facial features are either lost, corrupted or erroneously detected, resulting in an incorrect analysis of the person's facial expression. Figure 1 shows a feature point tracker in an affect inference system [11] failing to detect the mouth borders in the presence of hand occlusion. Only a few systems recognise facial expressions in the presence of partial face occlusion, either by estimation of lost facial points [2, 17] or by excluding the occluded face area from the classification process [6]. In all these systems, face occlusions are a nuisance and are treated as noise, even though they carry useful information.

This research proposes an alternative facial processing framework, where face occlusions instead of being removed, are combined with facial expressions and head gestures to help in machine understanding and interpretation of different mental states. We present an analysis of hand-over-face gestures in a naturalistic video corpus of complex mental states. We define three hand cues for encoding hand-over-face gestures, namely hand shape, hand action and facial region occluded and provide a preliminary assessment of the use of depth data in detecting hand shape and action on the face.

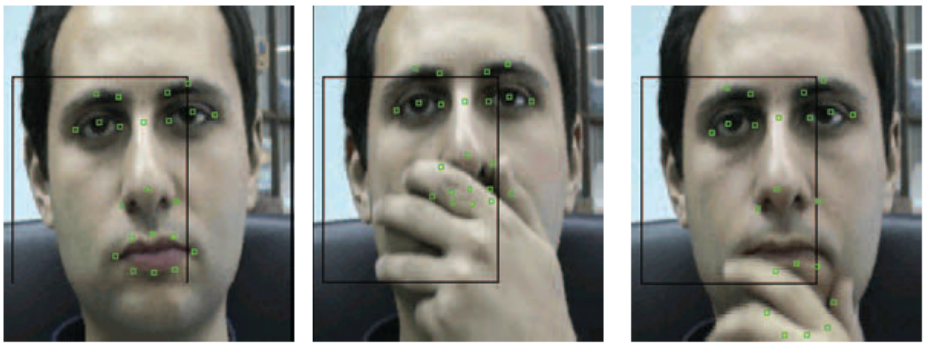

Fig. 1: In existing facial expression recognition systems hand-over-face occlusions are treated as noise.

## 2 Why hands?

From kinesics, the study and interpretation of non-verbal behaviour related to movement, the movement of the body, or separate parts, conveys many specific meanings. Ekman and Friesen [7] developed a classification system identifying five types of body movements; in most of them, hand gestures constitute an important factor and they contribute to how emotions are expressed and interpreted by others. Human interpretation of different social interactions in a variety of situations is most accurate when people are able to observe both the face and the body. Ambady and Rosenthal [1] have observed that ratings of human understanding of a communication based on the face and the body are 35% more accurate than the ratings based on the face alone.

Although researchers focus on facial expressions as the main channel for social emotional communication, de Gelder [4] suggests that there are similarities between how the brain reacts to emotional body language signals and how facial expressions are recognized. Hand-over-face gestures are not redundant information; they can emphasize the affective cues communicated through facial expressions and speech and give additional information to a communication. De Gelder's studies reveal substantial overlap between the face and the hand conditions, with other areas involved besides the face area in the brain. When the observed hand gesture was performed with emotion, additional regions in the brain were seen to be active emphasizing and adding meaning to the affective cue interpreted. In situations where face and body expressions do not provide the same meaning, experiments showed that recognition of the facial expression was biased towards the emotion expressed by the body language [5].

There is ample evidence that the spontaneous gestures we produce when we talk reflect our thoughts - often thoughts not conveyed in our speech [9]. Moreover, gesture goes well beyond reflecting our thoughts, to playing a role in shaping them. In teaching contexts, for example, children are more likely to profit from instruction when the instruction includes gesture - whether from the

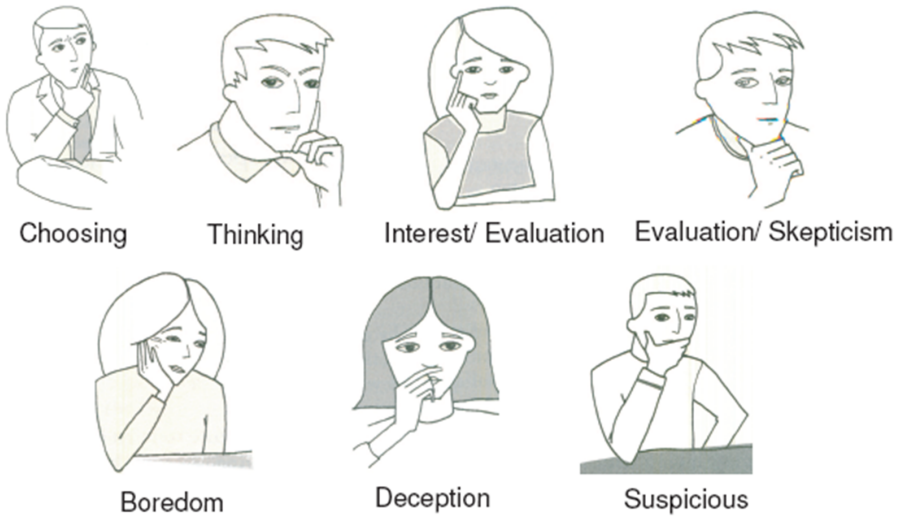


Fig. 2: The meaning conveyed by different hand-over-face gestures according to Pease and Pease [15]

student or the teacher - than when it does not. Teachers who use gestures as they explain a concept are more successful at getting their ideas across, and students who spontaneously gesture as they work through new ideas tend to remember them longer than those who do not move their hands [3]. Gesture during instruction encourages children to produce gestures of their own, which, in turn, leads to learning.

## 3 Hand-over-face gestures in natural expressions

In *The Definitive Book of Body Language*, Pease and Pease [15] attempt to identify the meaning conveyed by different hand-over-face gestures, as shown in Figure 2. Although they suggest that different positions and actions of the hand occluding the face can imply different affective states, no quantitative analysis has been carried out.

Studying hand-over-face gestures in natural expressions is a challenging task since most available video corpora lacked one or more factors that are crucial for our analysis. For instance, MMI [14] and CK+ [12] don't have upper body videos or hand gestures, while BU-4DEF [16] and FABO [10] datasets contain only posed non-naturalistic data. That was one of the motivations for building Cam3D, which is a 3D multi-modal corpus of natural complex mental states. The corpus includes labelled videos of spontaneous facial expressions and hand gestures of 12 participants. Data collection tasks were designed to elicit natural expressions. Participants were from diverse ethnic backgrounds and with varied


Fig. 3: Different hand shape, action and face region occluded are affective cues in interpreting different mental states.

fields of work and study. For more details on Cam3D, refer to Mahmoud et al. [13].

We have analysed hand-over-face gestures and their possible meaning in spontaneous expressions. By studying the videos in Cam3D, we argue that hand-over-face gestures occur frequently and can also serve as affective cues. Figure 3 presents sample frames from the labelled segments of hand-over-face gestures.

Hand-over-face gestures appeared in 21% of the video segments (94 segments), with 16% in the computer-based session and 25% in the dyadic interaction session. Participants varied in how much they gestured, some exhibited a lot of gestures while others only had a few. Looking at the place of the hand on the face in this subset of the 94 hand-over-face segments, the hand covered upper face regions in 13% of the segments and lower face regions in 89% of them, with some videos having the hand overlapping both upper and lower face regions. This indicates that in naturalistic interactions hand-over-face gestures are very common and that hands usually cover lower face regions, especially chin, mouth and lower cheeks, more than upper face regions.

### 3.1 Coding of hand gestures

Looking for possible affective meaning in those gestures, we introduced a preliminary coding of hand gestures. we encoded hand-over-face gestures in terms of three cues: hand shape, hand action and facial region occluded by the hand. These three cues can differentiate and define different meaningful gestures. Moreover, coding of hand-over-face gestures serves as a first step in automating the process of interpreting those gestures.

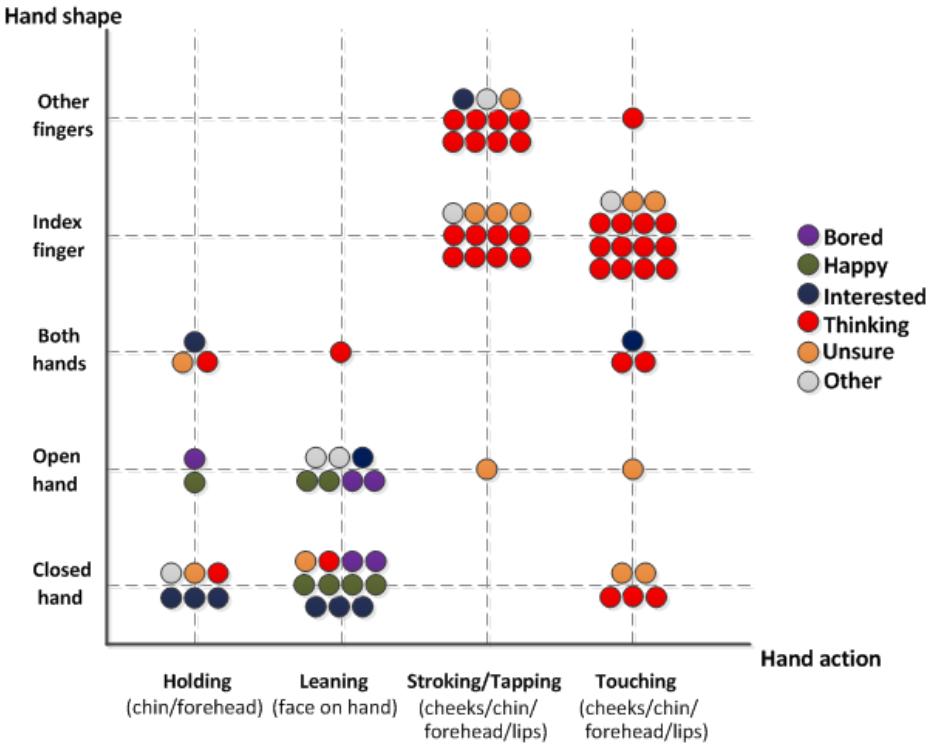


Fig. 4: Encoding of hand-over-face shape and action in different mental states. Note the significance of the index finger actions in cognitive mental states.

Figure 4 shows the distribution of the mental states in each category of the encoded hand-over-face gestures. For example, index finger touching face appeared in 12 *thinking* segments and 2 *unsure* segments out of a total of 15 segments in this category. The mental states distribution indicates that passive hand-over-face gestures, like leaning on the closed or open hand, appear in different mental states, but they are rare in cognitive mental states. This might be because those gestures are associated with a relaxed mood. On the other hand, actions like stroking, tapping and touching facial regions - especially with index finger - are all associated with cognitive mental states, namely *thinking* and *unsure*. Thus, we propose the use of hand shape and action on different face regions as a novel cue in interpreting cognitive mental states.

### 3.2 Hand detection using 3D data

Automatic detection of the hand when occluding the face is challenging because the face and the hand usually have the same colour and texture and the hand can take different possible shapes. The recent availability of affordable depth sensors (such as the Microsoft Kinect) is giving easy access to 3D data. 3D

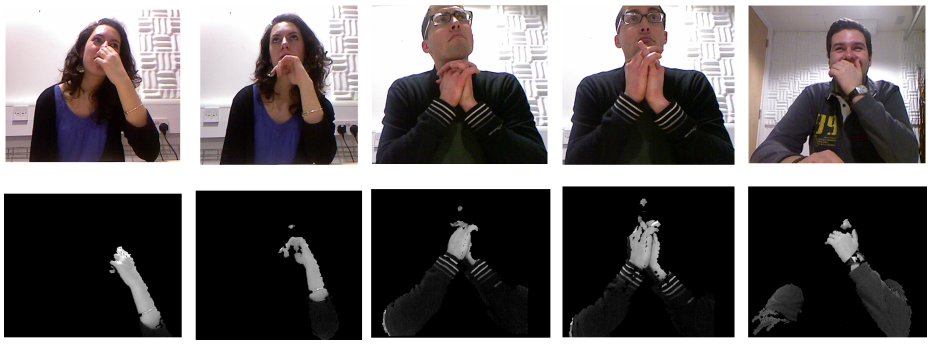

Fig. 5: Simple background subtraction based on depth threshold shows that depth images provide details of hand shape that can be utilised in automatic recognition defining different hand cues.

information can be used to improve the results of expression and gesture tracking and analysis. Cam3D provides a depth image for each frame in the labelled video segments, which enabled us to investigate the potential use of depth information in automatic detection of hand over face.

We used simple thresholding technique on the depth images in order to visualise how the hand can be segmented over the face using depth. Figure 5 presents preliminary results of this simple manual thresholding of the depth image. First, we define the depth value of the face in the depth image, using major facial points like the mouth and the nose. Then, we perform background subtraction based on the depth value of the face. Finally, we add the colour values for the segmented pixels to differentiate between hands or other objects in front of the face.

Initial results show that depth images provide details of hand shape that can be utilised in automatic recognition defining different hand cues. We are currently working in combining depth information with computer vision techniques for automatic detection of hand-over-face cues.

## 4 Conclusion and future work

We have presented an alternative facial processing framework, where face occlusions instead of being removed, are studied to be combined with facial expressions and head gestures to help in machine understanding and interpretation of different mental states. We analysed hand-over-face gestures in naturalistic video corpus of complex mental states and defined a preliminary coding system for hand cues, namely hand shape, hand action and facial region occluded. Looking at the depth images, we noticed the potential of using depth information in automatic detection of hand shape and action over the face. Our future work can be summarised in the following sections.

### 4.1 Analysis of more hand-over-face gestures

The lack of available multi-modal datasets slows down our work in analysing the meaning of hand-over-face gestures. Cam3D currently has 94 video segments of labelled hand gestures, which is not enough for studying all possible meaning of hand gestures. More than one category in our coding matrix had less than two videos, which might not be representative of the whole category. Future work includes collecting more data and adding more videos to the corpus. Studying hand-over-face gestures in more videos of natural expressions, we expect to enhance our coding schema and discover more encoded mental states associated with other hand-over-face gestures.

### 4.2 Automatic detection of hand cues

Automatic detection of hand shape, action and facial region occluded will include exploring computer vision techniques in hand detection that are robust to occlusion, as well as further analysis of Cam3D depth images. Automatic detection of hand cues is a step towards automatic inference of their corresponding mental states.

### 4.3 Automatic coding of hand gestures

One of the possible applications of this research is to provide tools for developmental psychologists who study gesture, and language in child development and social interactions to be able to objectively measure the use of gestures in speech and in communication instead of manual watching and coding, such as in the work done by Goldin-Meadow [8]

### 4.4 Multimodal inference system

Ultimately, our vision is to to implement a multi-modal affect inference framework that combines facial expressions, head gestures as well as hand-over-face gestures. This includes looking at integration techniques, such as: early integration or feature fusion versus late integration or decision fusion. Moreover, we aim at answering questions like: how the face and gesture combine to convey affective states? When do they complement each other and when do they communicate different messages?

## 5 Acknowledgment

We would like to thank Yousef Jameel Scholarship for generously funding this research. We would like also to thank Tadas Baltrušaitis for his help in the analysis of Cam3D corpus.

## References


1. Ambady, N., Rosenthal, R.: Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. Psychological bulletin 111(2), 256 (1992)
2. Bourel, F., Chibelushi, C., Low, A.: Robust facial expression recognition using a state-based model of spatially-localised facial dynamics. In: IEEE Automatic Face and Gesture Recognition (2002)
3. Cook, S., Goldin-Meadow, S.: The role of gesture in learning: do children use their hands to change their minds? Journal of Cognition and Development 7(2), 211–232 (2006)
4. De Gelder, B.: Towards the neurobiology of emotional body language. Nature Reviews Neuroscience 7(3), 242–249 (2006)
5. De Gelder, B.: Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. Phil. Trans. of the Royal Society B 364(1535), 3475 (2009)
6. Ekenel, H., Stiefelhagen, R.: Block selection in the local appearance-based face recognition scheme. In: Computer Vision and Pattern Recognition Workshop. pp. 43–43. IEEE (2006)
7. Ekman, P., Friesen, W.: The repertoire of nonverbal behavior: Categories, origins, usage, and coding. Semiotica 1(1), 49–98 (1969)
8. Goldin-Meadow, S.: Hearing gesture: How our hands help us think. Belknap Press (2005)
9. Goldin-Meadow, S., Wagner, S.: How our hands help us learn. Trends in cognitive sciences 9(5), 234–241 (2005)
10. Gunes, H., Piccardi, M.: A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior. In: International Conference on Pattern Recognition. vol. 1, pp. 1148–1153. IEEE (2006)
11. el Kaliouby, R., Robinson, P.: Real-time vision for human computer interaction, chap. Real-Time Inference of Complex Mental States from Facial Expressions and Head Gestures, pp. 181–200. Springer-Verlag (2005)
12. Lucey, P., Cohn, J., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In: Computer Vision and Pattern Recognition Workshop. pp. 94–101. IEEE (2010)
13. Mahmoud, M., Baltrusaitis, T., Robinson, P., Reik, L.: 3D corpus of spontaneous complex mental states. In: Affective Computing and Intelligent Interaction (ACII) (2011)
14. Pantic, M., Valstar, M., Rademaker, R., Maat, L.: Web-based database for facial expression analysis. In: IEEE Conf. Multimedia and Expo. p. 5. IEEE (2005)
15. Pease, A., Pease, B.: The definitive book of body language. Bantam (2006)
16. Sun, Y., Yin, L.: Facial expression recognition based on 3D dynamic range model sequences. In: ECCV. pp. 58–71. Springer-Verlag (2008)
17. Tong, Y., Liao, W., Ji, Q.: Facial action unit recognition by exploiting their dynamic and semantic relationships. IEEE Transactions on Pattern Analysis and Machine Intelligence pp. 1683–1699 (2007)