ARTICLE IN PRESS

Measuring dynamic micro-expressions via feature extraction methods

Wen-Jing Yan*, Yu-Hsin Chen

Institute of Psychology and Behavior Sciences, Wenzhou University, Wenzhou 325035, China

ARTICLE INFO

ABSTRACT

Historically, the assessment of nonverbal behaviors such as facial expressions and gestures have relied heavily on observations made with the naked eye. Previous studies have shown that the manual analysis of micro-expressions can be both arduous and time consuming, and often suffers from a low level of reliability. However, new technology may bring forth new opportunities to refine this type of investigation. This research applied three feature extraction algorithms from the field of computer science to assist in measuring subtle dynamic movements in the human face. The Constraint Local Model (CLM) algorithm was employed to detect the faces and track the feature points; coordinates changing position across the frame indicated local facial activity. Based on these features points, facial regions of interest (ROIs) were identified for further analysis by Local Binary Pattern (LBP) and Optical Flow (OF). The LBP algorithm was used to extract texture information from the ROIs and measure correlations among frames; OF was used to analyze facial actions based on the main direction and distance of movement. The results of the three methods were then compared with manual coding. The proposed methods, especially OF, proved to be both robust and reliable.

© 2017 Published by Elsevier B.V.

1. Introduction

For behavioral scientists, reliable methods of measuring behavior are essential to clean, objective, replicable results. Historically, the assessment of nonverbal behaviors has primarily relied upon in-person observation. However, several new apparatuses have been developed in the past decade to aid researchers in quantifying these types of actions. One example is the motion capture system, which is able to measure, calculate, and record the kinematic information of very subtle body movements with high spatial and temporal resolution. Another example is electromyography (EMG), which measure subtle muscle movements (even when they are not necessarily accessible to the human eye). Though such technology can be helpful in assisting researchers to better investigate these types of actions, they require a certain level of cooperation from the individual; reflective markers or electrodes must be placed on the subject in order for the apparatus to work. Methods of measuring nonverbal behaviors that require individuals to wear or equip devices are generally referred to as intrusive methods of observation; such methods may influence the inherent behaviors researchers are attempting to observe, and thus the results obtained may have less

ecological validity. As such, less intrusive methods that originate from the field of computer science, such as computer vision and feature extraction, have garnered significant attention (such as with emotion, gait, gesture, and face-recognition research). The present research attempts to investigate whether feature extraction techniques might be used by researchers as a means of better assessing the very subtle facial movements called micro-expressions.

Micro-expressions are generally defined as very brief and subtle facial movements that reveal an emotion a person may be trying to conceal [1–3]. It has been claimed that inspectors well-trained in reading micro-expressions can reach a high level of accuracy in lie detection [4], even more so than by observing other nonverbal cues. Ekman [5] argued that micro-expressions are the most promising approach to deception detection. Hence, the analysis of micro-expressions has become a focus of both the press and scholars alike. Beyond lie detection, the study of micro-expressions has intrinsic scientific value because it provides a window into individuals' genuine emotions. However, though micro-expressions offer both theoretical and practical applications [3,6], only a handful of studies have been conducted to better understand their characteristics (see, for example [3,7]).

1.1. Concerns regarding the manual analysis of micro-expressions

There are several reasons why so few academic studies have been conducted to investigate micro-expressions. Despite more

* Corresponding author at: Rm. 9A-207, South Campus, Wenzhou University, Wenzhou 325035, China.
E-mail addresses: eagan-ywj@foxmail.com (W.-J. Yan), yhandrewc@gmail.com (Y.-H. Chen).

and more databases on the topic becoming publicly available (e.g., Polikovsky's database [8], SMIC [9], CASME [10], and CASMEII [11]) in recent years, the amount of scholarly research has yet to increase substantially; this is most likely due to difficulties in assessing and conducting analyses of the micro-expressions themselves [6]. Such investigations usually rely on a scoring system called the Facial Action Coding System (FACS), which taxonomizes human facial movements by their appearance on the face, and gives an objective method for describing them in terms of component actions [12]. To date, FACS remains the most widely used method for analyzing facial movements and many recent studies [13,14] have use the it to measure facial behaviors.

Yet researchers have expressed a number of concerns. Manual coding of this fashion is laborious, time-consuming, and strenuous, especially with regards to the more subtle facial movements. For instance, one previous study stated that comprehensively coding a one minute video typically takes over two hours [15]. For very subtle facial expressions (in which the intensity is lower than the lowest level represented in the FACS manual), manual coding is even more demanding. The situation becomes further complicated when researchers investigate micro-expressions, because several prerequisites (see [3]) have to be met in order to obtain a sample. Micro-expressions are currently recognized and defined by their duration [3]; to calculate this element, researchers must manually count video frames and ensure they fall within a range of 0.5s; otherwise, the facial action cannot be classified as a micro-expression. Spotting the beginning (onset) and ending (offset) frames of these subtle facial movements is both time consuming and labor intensive.

Current manual coding methods deal with video footage of facial expressions via a frame-by-frame approach, employing assessment methods similar to those used to evaluate static images; a process has not yet been developed that allows for a comprehensive manual assessment of the dynamic information present in recorded video material. Several studies have shown the importance of dynamic information to the study of facial expressions. For example, temporal dynamics convey unique information such as unfolding and ending, which can not only be used to judge emotional expressions, but also to provide strong signals of action and intent [16]. It was also found that shorter durations (i.e., the time between onset and offset) and more irregular onset actions are associated with contrasts between politeness and amusement, and in the case of smiles, lower perceived genuineness and spontaneity [17–19]. Quantification of dynamic micro-expressions could lead to similar findings, but first the current methods of assessment must be replaced with a system that better assists researchers in their investigations.

1.2. Utilization of feature extraction methods by facial expression analysis

Computer scientists have developed tools that analyze facial movements [20–25] for biometrics and emotion recognition purposes. The process usually includes the following three steps: face detection in a facial image or video, facial feature extraction based on shape or appearance information, and facial expression or Action Unit (AU) classification. Researchers working on computer vision have focused on accurately classifying different facial expressions [15,26] and AUs [24,27,28]. The patterns of facial movements, which require the quantification of details such as the onset time, duration, velocity, and synchronization of certain regions, is of equal interest to behavioral scientists.

Consequently, there are many computer vision algorithms that could also be used to recognize micro-expressions [29–32], but they are not yet perceived as sensitive enough for use in scholarly research. Therefore, we tested various feature extraction methods and selected those that might be suitable for such use: (1) the Constraint Local Model (CLM), which is a facial landmark localization method mainly based on geometric features; (2) Local Binary Pattern (LBP), which is an appearance-based method; and (3) Optical Flow (OF), which is based on motion information. These three algorithms were compared in order to assess their viability as assessment tools in micro-expression research.

1.2.1. The constraint local model

CLM is a type of point distribution model (PDM) that represents the geometric mean and certain statistical modes of geometric variation inferred from a training set of shapes. It typically involves an exhaustive local search for the best location for each PDM landmark; these landmarks are then constrained to adhere to the PDM's parameterization. CLM's texture sampling method is an improvement over Active shape model (ASM), and compared with Active Appearance Model (AAM), it exhibits much more satisfactory generalizations from limited data and can offer a degree of robustness [33,34].

In CLM, a training patch is sampled around a given feature and normalized such that the mean value of the pixels is zero and the variance is one. Next, the texture patches are vectorized. The training vectors and normalized shape coordinates are then used to construct linear models. Finally, the shape and template texture models are combined using a further Principle Component Analysis (PCA), in order to produce a joint model that has the following form:

$$b = P_c c \text{ where } P_c = \begin{pmatrix} P_{cs} \\ P_{cg} \end{pmatrix} \& b = \begin{pmatrix} W_s b_s \\ b_g \end{pmatrix} \quad (2)$$

where b is the concatenated shape and texture parameter vector with a suitable weighting, W_s, to account for the difference between the shape and texture units; c is a set of joint appearance parameters; and P_c is the orthogonal matrix computed using a PCA that partitions into two separate matrices, P_{cs} and P_{cg}, which together compute the shape and texture parameters given by c.

CLM uses a joint shape and texture appearance model to generate a set of region template detectors. For further information, please refer to [35]. In this research, we tested the source codes provided by Asthana et al.[1] [36], as well as those provided by Jason Saragih,[2] and Yan Xiaoguang.[3] We found Jason Saragih's version to perform best on CASME II.

1.2.2. Local binary pattern

The texture information of a certain region can be used to measure changes in facial movements across time. Previous studies have shown that LBP is a powerful algorithm for texture description [37]. An LBP operator describes the local texture pattern for a pixel, C, in a particular image, by comparing and thresholding its gray values against those of the neighboring pixels. For the center pixel C with P neighboring pixels sampled with a radius R, the LBP value can be calculated as:

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s\left(g_p - g_C\right) 2^p \quad (3)$$

where (x_c, y_c) are the coordinates of the center pixel C, g_C is its gray value, g_p ($p = 0, \ldots, P-1$) is the gray value of the p_{th} neighboring pixel on the radius R, and 2^p is the weight corresponding to the locations of the neighboring pixels used to transform the binary pattern string into a decimal index for the pattern.

---

[1] https://sites.google.com/site/akshayasthana/clm-wild-code.

[2] https://github.com/kylemcdonald/ofxFaceTracker.

[3] https://sites.google.com/site/xgyanhome/home/projects/clm-implementation.