Facial Behavior as a Soft Biometric

Abhay L. Kashyap
University of Maryland, Baltimore County
1000 Hilltop Circle, Baltimore, MD 21250
abhay1@umbc.edu

Sergey Tulyakov, Venu Govindaraju
University at Buffalo
113 Davis Hall, Buffalo NY 14260-2500
{tulyakov,govind}@buffalo.edu

Abstract

The human face forms an important interface to convey nonverbal emotional information. Facial expressions reflect an individual's reactions to personal thoughts or external stimuli. These can act as valuable supplementary biometric information to automated person identification systems. In this study, video segments of individuals were FACS coded to quantify facial expressions. The Action Unit (AU) frequencies, considered both individually and in specific combinations, served as features for person identification. The experiments confirm that these features of the facial behavior are well suited for biometric person identification. Considering both facial asymmetry and Action Unit combinations resulted in a significant improvement in the identification efficiency. Additionally, we observed the convergence in the identification process with the increase of the training data. Thus, given sufficient training data, facial behavior can serve as a reliable biometric modality.

1. Introduction

In recent times, there has been a significant amount of work towards finding new biometrics modalities [6]. Face and related modalities have received significant attention because of the low cost of acquisition and its increasing use in multimodal biometric systems.

One of the research objectives in face processing is the recognition of facial expressions. Such recognition can be utilized for improving human-computer interfaces, detecting any abnormalities in human behaviors such as the presence of deceit [14,3], drowsiness during driving [13], or pain experience [10]. In our research, we are interested in utilizing facial expressions for human identification.

Based on the behavioral traits of an individual, behavioral biometrics serve as supplementary soft biometric information used especially in multi-modal identification systems. Popular biometric modalities which fall into this category are gait, signatures, keystroke patterns, speech patterns etc. Most behavioral biometric modalities suffer from a lack of uniqueness and permanence properties. Uniqueness is defined by how well the biometric separates individuals from each other and permanence indicate how well the biometric resists aging [9]. Signatures, gait etc. are all shown to be coarse with respect to identifying individuals and are subject to change with time much faster than conventional biometrics such as fingerprints. In exchange, these soft biometrics offer inexpensive data acquisition and expect little co-operation from the users. With the increase of multimodal biometric systems, soft biometrics fit right in, furthering the efficiency of unimodal biometric systems.

Facial behavior can be viewed as observable patterns that underlie occurrences of facial expressions over time. Different facial expressions, their frequency of occurrence, length of these expressions and intervals between them contribute towards the facial behavior of a person. Although facial expressions representing common emotions in humans are nearly universal [8], facial behavior on the other hand captures their dependencies with time and each other. However, these sequences of facial expressions are hard to observe precisely and they cannot serve as independently sufficient discernible factors when identifying individuals. Hence facial behavior can be considered as a soft biometric since it does not provide for strong discrimination power by itself.

Facial expressions reflect the emotional state of a person and are hence influenced by the environment. Factors such as the topic of conversation and the degree of formalism associated with it can directly influence a person's emotions. For example, the facial behavior of a person can vary significantly between a humorous conversation with a friend and a stressful interview. Past emotional events and physical factors such as illness can have an indirect effect on the emotional state of a person. These factors influence a person's mood and can result in emotions being carried over onto subsequent conversations. These emotional changes can add a bias to certain facial expressions when measured over that period. However, just like any other soft biometric, we can have some safe assumptions where in only a normal emotional state is considered in a normal and consistent situation.

In our work, the experimental video segments comprise interviews where subjects are asked the same set of

978-1-4673-0397-2/12/$31.00 ©2012 IEEE

questions and hence account for a consistent situation. Facial Action Coding system (FACS) [7] was used to quantify facial expressions. Interviews of 20 individuals were FACS coded and the video segments were compared against different individuals. It was found that facial behavior serve as identity indicators consistent with previous studies. Facial asymmetry along with common AU combinations added to the identification efficiency. Although it was not very significant since most common emotions are universal with respect to their induced Action Units combinations [11] and rarely occur independently of each other. The identification efficiency was increased with larger video segments and with a sufficiently large training data, they can serve as a good soft biometric.

2. Previous Research

Facial expressions reflect the reactions of an individual to internal and external stimuli. They have been shown to be stable over a significant period of time and to be fairly unique with respect to different individuals to serve as sufficient identity indicators by J. F. Cohn et al. [4]. In that research, relatively long segments of videos were used, particular facial action unit combinations (those which were present during smiling) were considered, and additional features derived from the geometric displacements of tracker points were utilized during recognition.

Facial expressions can be characterized by the changes in the positions of landmark points of the face relative to the positions of such points in a neutral expression face, and the features extracted from such changes have been used to identify individuals with a considerable amount of success [12]. That work was limited in using only static images of the person in two states – neutral and the apex of facial expression, which might be the reason for low recognition rate observed.

Facial behavior involves monitoring and measuring facial expressions over a period of time sufficient enough for it to be used as a biometric. One way to encode the facial expressions over time has been proposed by P. Ekman in the form of Facial Action Coding system (FACS) [7]. The developments in facial processing algorithm have allowed the automated action unit recognition from videos [2], and such algorithms could be used to obtain dynamic expression information for identification purposes.

In contrast to [4], we are using only action unit information in the experiments and consider a bigger variety of possible action units. The shorter video segments are used in our research as well, but due to the nature of collected data, the relatively stressful interview questions, the participants were expressing a high number of facial action units and it was possible to observe the appearance of same action units in both training and testing data.

3. Facial Behavior

For facial behavior to be used as a biometric, by definition, it has to be measured somehow. Measuring facial expressions over a period of time is quite different from measuring features from a single static image. To facilitate this, the FACS coding system is used to quantify the data.

Facial Action Coding System (FACS) is a system developed by Dr. Paul Ekman to classify human facial expressions in an objective manner [7]. FACS uses the muscular anatomy of the face to classify expressions and it can be used to code any anatomically possible human facial expression. FACS uses *Action Units (AU)* to describe the specific movements of muscles independently. Many facial expressions occur as combinations of different Action Units. When facial symmetry is considered, each AU gets a suffix of L or R indicating Left or Right of the face where the AU appears. Every AU has its associated intensities quantified on a scale from A-E with E indicating the maximum intensity and A indicating the slightest indication of the AU's existence. The intensity ratings are not discrete quantifiers but rather merge into each other. The AU set is divided into 5 major groups based on the area of the face they act on and the direction of the muscle movement. Upper face AUs, Lower face AUs further divided into horizontal, vertical, oblique and orbital movements, comprise most of the AUs. FACS also uses *Action Descriptors (AD)* to indicate movements which do not use any specific muscle for movement. Table 1 show the Action Units used in this experiment along with their descriptions. Table 2 show the AU combinations used.

Table 1. Action Units and their descriptions, L/R – Left/Right capture the asymmetry of the AU on the face.

Action Unit	Description
AU 1	Inner eyebrow corners are raised
AU 2/2L/2R	Outer eyebrow corners are raised
AU 4	Eyebrows are lowered and pulled together
AU 5	Upper Eyelids are raised
AU 6	Cheeks are raised
AU 7	Eyelids are pulled together
AU 9	Nose is crinkled
AU 10	Upper lip is raised
AU 12/12L/12R	Lip corners are pulled up obliquely
AU 14/14L/14R	Lip corners are tightened
AU 15	Lip corners are pulled down
AU 16	Lower lip is pulled down
AU 17	Chin boss is pulled up
AU 18	Lips are pulled together
AU 20	Lips are pulled horizontally

AU 22	Lip are funneled
AU 23	Lips are tightened
AU 24	Lips are pressed up against each other
AU 28	Lips are sucked into the mouth
AD 32	Lip bite
AD 37	Lip wipe

Table 2. Action Unit combinations used

AU 1+2/2L/2R+4	AU 1+2/2L/2R	AU 1 + 4
AU 6+7+12/12L/12R	AU 14/14L/14R+17	AU 6 + 7
AU 14/14L/14R+17+24	AU 14/14L/14R+24	AU 15 + 17

Every facial expression is broken down into its constituent Action Units. So, for a video segment, the sequence of the set of AUs which quantify expressions at regularly spaced frames, represent facial behavior of a particular individual which is a discrete approximation of the continuous changes on the face. Since few expressions last for extremely small intervals, it is safe to discretize the video into intervals small enough to capture an instance of majority of the expression.

4. Experiments

4.1. Dataset

The dataset was a collection of interviews in a controlled environment, aimed at deceit detection used by Bhaskaran N. et al. [3]. The interview comprises a set of objective as well as abstract questions regarding the possible theft of a check. The questions are aimed to establish a baseline for a subject with respect to facial expressions and also induce emotional responses aimed towards identifying deceit. These provide for a good mix of facial expressions from different subjects while providing a stable and uniform environment. A subset comprising 20 of these interviews, of which 11 were female and 9 were male, was used for this experiment. All the video segments offer a zoomed in, full frontal view of the subject's face under uniform lighting. The average length of the videos is 238 seconds at 30 frames per second.

For our experiments, to discretize the continuous video segments, every video was divided into equally spaced frames. The interval was chosen to be large enough so as to eliminate redundant frames, which do not contain any significant change from the previous frame and at the same time small enough to capture most of the expressions. Only micro expressions last for around 1/15th to 1/25th of a second and an attempt to capture these, rather rare expressions would require inclusion of a large number of redundant frames. So an interval of 10 frames, which is 1/3rd of a second, is chosen so that at least an instance of almost all macro expressions is captured.

These extracted images were manually FACS coded in terms of the Action Units listed in Table 1. Since the videos dealt with were interviews, there was a lot of speech involved. Speech, by itself, induces some action units, primarily AUs of the lips. These muscle movements are required to produce different sounds during speech and hence carry little or no information which reflects the emotional state of a person. So these particular AUs were ignored. However other AUs involved during speech like those from smiles, frowns etc. or any AU which was not purely responsible for modulating the air flow out of the mouth to facilitate speech, was considered. This ensured that nearly every AU potentially stemming from emotions were captured and the collected AUs contained mostly emotional AUs.

The FACS coding also facilitates to capture intensity of different action units. They range from A, representing the slightest presence of an AU, through B, C, D, till E, representing the maximum intensity of a particular AU. These intensity levels merge into each other and it becomes harder to accurately capture the intensity levels of AUs for different individuals. For our experiments, we do not consider the intensity level of AUs and monitor only its presence or absence.

4.2. Experimental Setup

For a given FACS coded set of frames from a video segment, the AU statistic was obtained as a total number of frames containing that AU. Firstly, all frames were separated into neutral frames, those which contained no AU, and expressive frames, those which contained at least one AU. Next, the total number for frames for each AU or a combination of AUs was computed. To accurately capture AU combinations, if a frame contained, for example a combination of AU 6 and 12, then the frame was tagged only under the combination 6+12, and not under their constituents, AU 6 and AU 12. The neutral frames, represented by AU0, along with the individual AUs and their combinations form the feature set. Lastly, this feature set was normalized to maintain uniformity across different segment lengths so that each feature represents a proportion of its influence in the segment. The neutral frames were divided by the total number of frames in the given video segment, representing the neutrality ratio. The AUs were divided by the total number of expressive frames in the video segment.

The score was computed as a total distance between two feature sets. For any given two feature sets, the absolute difference between corresponding features was summed to obtain the total score. These scores represent the dissimilarity in the facial behavior of two individuals.

4.3. Experiment I

The first experiment was aimed to measure the influence of using facial asymmetry and AU combinations. Every video segment was split into exactly two halves. The first segment represented the training data and the second segment represented the test data. In three different trials, the first trial selected only individual AUs ignoring their combinations and asymmetry. For example, AU 2L, where L signifies left part of the face, was tagged under AU 2 when AU asymmetry was ignored. AU combinations like AU 1+2 were tagged under their constituents, AU 1 and AU 2. The second trial considered individual AUs along with AU asymmetry while ignoring AU combinations. This was to study the influence of facial symmetry independently as a factor in increasing biometric information since a lot of AU combinations are induced from nearly universal emotions. The third considered both AU asymmetry and all their combinations as shown in Table 2. The exact matches were found to be 50%, 50% and 55% respectively among 20 individuals. The baseline is 5% which would be identifying the individual with no biometric knowledge, i.e. random selection. This indicates that facial behavior data added biometric information about a person's identity to increase match rate from 5%, with no knowledge, to 50-55%, with facial behavior data. The ROC curves of the three are shown in Figure 1. These curves indicate that facial symmetry as a factor, adds to the biometric information and AU combinations coupled with facial symmetry give more biometric information.

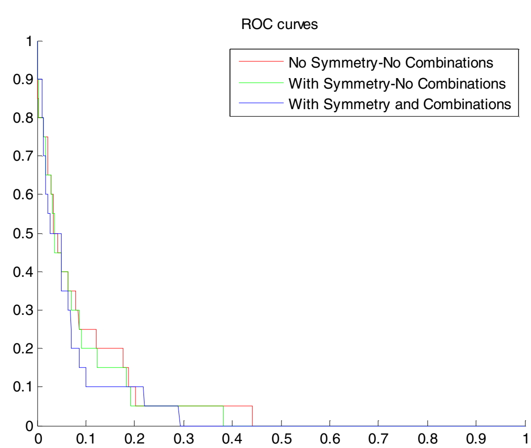

Figure 1: ROC curves for different AU configurations.

4.4. Experiment II

The second experiment was to observe the influence of increased training data. So, for three different trials, video segments of length 1/8, 1/4 and 1/2 of the total video were considered. In these trials, a complete feature set comprising AUs and their combinations along with asymmetry was considered. Two random segments of the mentioned lengths were compared against each other to obtain a score as described earlier. The ROC curves of the three are shown in Figure 2. The curves indicate that facial behavior is significantly dependent on the considered video segment length. Shorter video length brings out the influence of a particular situation on the emotional state of the person. A longer video segment can offer a more diverse set of emotional states when across different types of situation. This would provide a more accurate estimate of a person's facial behavior. This convergence can be seen among the curves of increasing length.

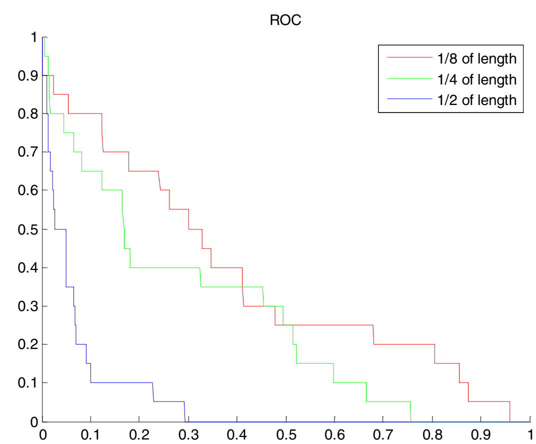

Figure 2: ROC curves for different training data length.

5. Conclusions

Facial behavior was found to carry significant amount of biometric information consistent with previous studies. With the automation of Action Unit detection, facial behavior can serve as a valuable supplementary behavioral biometric in multimodal identification systems. Data acquisition is relatively cheap with the requirement of just a camera, although there is a significant amount of dependence on the environment a person is in, which can affect the emotional state of a person and hence dilute critical identifiers with universal emotional Action Unit combinations.

It was found that facial asymmetry added to the identification efficiency and it was furthered when Action Unit combinations were considered. Convergence, with respect to identification efficiency, was observed with the increase in training data length indicating the stronger influence of the situational factors on shorter segments. So with a large video sample of an individual, possibly across different situations as the training data, facial behavior can serve as a good soft biometric.

6. Acknowledgement

Authors express a gratitude to Mark G. Frank for the provided experimental video dataset.

References

[1] Bartlett, M., Littlewort, G., Whitehill, J., Vural, E., Wu, T., Lee, K., Ercil, A., Cetin, M. Movellan, J. “Insights on spontaneous facial expressions from automatic expression measurement.” in Giese,M. Curio, C., Bulthoff, H. (Eds.) *Dynamic Faces: Insights from Experiments and Computation*. MIT Press.

[2] Bartlett, M.S., Littlewort, G.C., Lainscsek, C., Fasel, I., Frank, M.G., Movellan, J.R. (2006). “Fully automatic facial action recognition in spontaneous behavior.” in *7th International Conference on Automatic Face and Gesture Recognition, p. 223-228*.

[3] Bhaskaran, N.; Nwogu, I.; Frank, M.G.; Govindaraju, V.; , "Lie to Me: Deceit detection via online behavioral learning," *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on* , vol., no., pp.24-29, 21-25 March 2011.

[4] J. F. Cohn, K. Schmidt, R. Gross, and P. Ekman. Individual differences in facial expression: Stability over time, relation to self-reported emotion, and ability to inform person identification. In *Fourth IEEE International Conference on Multimodal Interfaces (ICMI’02)*, page 491, 2002.

[5] Cohn JF. “Foundations of human computing: facial expression and emotion” In: Huang TS, Nijholt A, Pantic M, Pentland A, editors. *Artificial intelligence for human computing*.

[6] Delac, K.; Grgic, M.;, "A survey of biometric recognition methods," *Electronics in Marine, 2004. Proceedings Elmar 2004. 46th International Symposium*, vol., no., pp. 184- 193, 16-18 June 2004.

[7] P. Ekman and W. V. Friesen. *Facial Action Coding System.* Consulting Psychologists Press, Palo Alto, 1978.

[8] Ekman, P., Sorenson, E.R., Friesen, W.V.: Pancultural elements in facial displays of emotion. Science 164 (1969) 86-88

[9] Jain AK. “Biometric recognition: how do I know who you are?” *Proceedings of the IEEE 12th Signal Processing and Communications Applications Conference 2004*.

[10] Littlewort, G.C., Bartlett, M.S., Lee, K.: Faces of pain: automated measurement of spontaneous facial expressions of genuine and posed pain. Proceedings of the 9th international conference on Multimodal interfaces. ACM, Nagoya, Aichi, Japan (2007) 15-21

[11] Russell, James A., “Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies.” Psychological Bulletin, Vol. 115(1), Jan 1994, 102-141.

[12] Tulyakov, S.; Slowe, T.; Zhi Zhang; Govindaraju, V.; "Facial Expression Biometrics Using Tracker Displacement Features," *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on,* vol., no., pp.1-5, 17-22 June 2007.

[13] Vural, E., Cetin, M., Ercil, A., Littlewort, G., Bartlett, M., Movellan, J.: “Drowsy driver detection through facial movement analysis.” Proceedings of the 2007 IEEE international conference on Human-computer interaction. Springer-Verlag, Rio de Janeiro, Brazil (2007)

[14] Zhang, Z., Singh, V., Slowe, T.E., Tulyakov, S., Govindaraju, V. “Real-time Automatic Deceit Detection from Involuntary Facial Expressions.” Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on (2007) 1-6