Facial Expression Biometrics Using Tracker Displacement Features

Sergey Tulyakov[1], Thomas Slowe[2], Zhi Zhang[1], and Venu Govindaraju[1]
[1]Center for Unified Biometrics and Sensors
University at Buffalo, NY, USA
{tulyakov,zhizhang,venu}@cedar.buffalo.edu
[2]CUBRC, Buffalo, NY, USA
slowe@cubrc.org

Abstract

In this paper we investigate a possibility of using the face expression information for person biometrics. The idea of this research is that person's emotional face expressions are repeatable, and face expression features can be used for person identification. In order to avoid using person specific geometric or textural features traditionally used in face biometrics, we restrict ourselves to the tracker displacement features only. In contrast to previous research in facial expression biometrics, we extract features only from the pair of face images, neutral and the apex of emotion expression, instead of using the sequence of images from the video. The experiments, performed on two facial expression databases, confirm that proposed features can indeed be used for biometrics purposes.

1. Introduction

The way person behaves, and, in particular, the way person expresses some emotions, can serve as an indicator to person's identity. Previously Eckman [3] identified 18 types of different smiles. The type of smile expressed by the person is partly influenced by the current emotional state of the person and environment. At the same time we can speculate that cultural background of the person and constant psychological traits can determine the frequencies of each type appearance and their strength. If it is so, then the type of the smile can be used for biometric person identification.

Some psychological studies seem to confirm the individual differences in emotions, and in particular in smiles. Krumhuber et al. [7] provide an extensive overview of recent research on differences in smiles and perceptions of smiles in the field of psychology. As it is discussed there, the perception of smile strength can be influenced by the general belief or expectation of a person. These results imply that psychological studies involving human test subjects might not give objective picture of smile or expression individuality. On the other hand, it was noted before that women express smile more frequently than men [1] and such frequency can be measured rather objectively.

In our research we are interested in separating the expressions of different individuals automatically. Besides proving the individuality of facial expressions, our goal is to achieve an algorithm identifying the persons based on their expressions. Since expressions might not provide enough discriminating power, our research might be considered as a soft biometrics [5].

The previous research into facial expression biometrics is very limited. Schmidt and Cohn [9] and Cohn et al. [2] describe the system performing biometric person authentication using the changes in action unit appearances. Such system measures the intervals between different phases of facial action units [4] and their sequence order. The action units and time intervals are measured with the help of special hardware. The reported performance is comparable with the performance of commercial face biometrics system. Another interesting research is presented in [8], where authors investigate if emotion expression is hereditary. The experiments conducted with born-blind subjects seem to indicate that face emotion expression of family members is correlated. The used features included sequences of 43 types of face movements.

In both above approaches long sequences of facial movements were used to produce a statistically significant person identification algorithm. In our research we investigate whether the information obtained from single images can serve for the same purpose. Similar to these works we concentrated on the expressions associated with some emotions. We assumed that the same emotion of the same person is expressed as a similar set of action units of the similar intensity. Thus, using a single image of the person at the height of emotion expression might be sufficient for person discrimination.

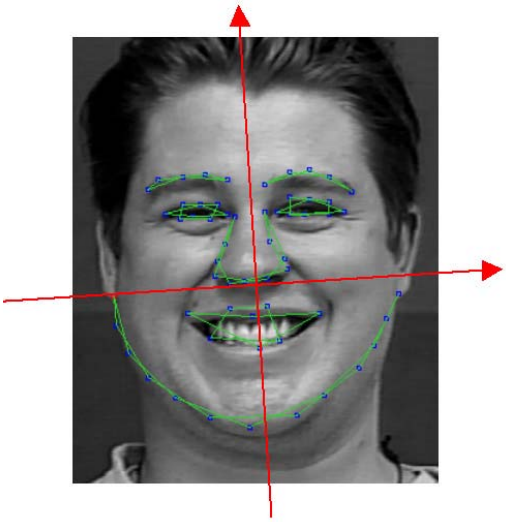

Figure 1. Tracker points and extracted reference frame used for normalization.

2. Tracker Displacement Features

The main difficulty that we had in our research was to produce features which are unique to face expressions and do not contain information traditionally used for face biometrics. For example, we did not want to use geometric distances between different points in face, or texture information which might give strong information about the person. Such information might be irrelevant for face expressions, and our results would greatly skewed if such features are used.

Thus we decided to restrict ourselves in using only displacement distances between tracker points. The used tracker positions are described in [10] and illustrated in Figure 1. Though the tracker positions can be extracted automatically [10], we found out that algorithm does not work quite reliably for single frame images or very short video sequences that we used. Thus, we used the help of human expert to find the tracker points.

To find displacements we had to use two images of a person - one neutral emotion expression and one at the apex of emotion. The distances between corresponding tracker points are indicative on the movement of the face, but not of the face geometrical form.

In order to account for face global transformation in the image frame, we searched for a frame of reference shown in Figure 1. The vertical axis and the direction of horizontal axis were found by using regression on the sums of corresponding left and right tracker points. The location of horizontal axis was found by averaging vertical positions of some stable points - tracker points of eyes and nose. The positions of tracker points were also scaled to account different distances from face to the camera.

After normalization procedures in both neutral and emotion images, the displacements between corresponding tracker positions was recorded. The final result is 116 features corresponding to x and y coordinates of 58 tracker points. These features seem to not contain any geometric face information, but only the information about emotion expression. Though it is possible that some type of face geometry is favorable to specific movements of tracker points, we regard this very improbable.

Two images, neutral and emotion, are used to construct a single feature vector of a person having particular emotion expression. The genuine biometric match consists in comparison of two such vectors of the same person having same emotion, so four images (two neutral and two emotional) are needed to derive a single match score. The impostor biometric match is produced by having two different people expressing same emotion. Again, four images are required to produce a matching score. The matching score is an Euclidean distance between two 116-dimensional feature vectors.

In our research we used only emotions of joy and sadness. These emotions constituted the largest part of considered databases. Though for biometric experiments only images from only one emotion would have sufficed, we also wanted to verify if our features can separate emotion expressions. For this purpose we also performed experiments comparing the matching distances of two cases: the distance between feature vectors of different persons having same emotion and the distance between feature vectors of different persons having different emotions.

3. Experiments

We used two face expression databases in our experiments. The first database is publicly available Cohn-Kanade Facial Expression Database [6], and the second one is the database of face images extracted by authors from a television show "'Big Brother"'. Below we describe separately the setup and the results of experiments performed on each database.

3.1. Cohn-Kanade Database

Cohn-Kanade database [6] contains sequences of 100 persons performing 8 different emotional or non-emotional facial expression changes. All the images are frontal face images, which ideally suits our feature extraction method. Though the sequences contain few frames of video of a gradual change of person's expression, in these experiments we used only the first (neutral) and the last (apex of emotion expression)frames.

The only disadvantage for using this database was that there were too few sequences of the same person doing same emotion expression. The human expert was able to identify 13 sad and 9 happy emotion pairs of expression sequences from 17 different persons. Note, that some persons did two pairs of the sad and happy emotions. So the total number of genuine match scores was 22. For impostor matching scores we produced all possible combinations of same emotion sequence produced by two different persons. As a result we had 312 sad emotion pairs and 144 happy emotion pairs, which gave us 456 pairs, or impostor matching scores.

The statistics of these matching score sets is presented in table 1. Same person and same emotion score set corresponds to genuine biometric match scores. Different persons and same emotion corresponds to the impostor biometric match scores. The average matching distances (μ) for genuine scores are a little lower than for impostor scores, as we predicted. But the standard deviation (σ) for both sets of scores is rather high, and there is a concern that this biometric matching might not be reliable.

Score set	μ	σ
Same person, same emotion	.8078	.2394
Different persons, same emotion	.8688	.2253
Different persons, different emotions	.9813	.2143

Table 1. Means and standard deviations for different matching score sets.

In the derivation of the used feature vectors, we assumed that these features are designed to well separate emotional expressions. In order to verify this, we performed a testing involving a set of matching scores between sequences of different people expressing different emotions. By mixing together 13*2 sad emotion sequences and 9*2 happy emotion sequences, we obtained 468 such pairs. The statistics of this score set is shown in Table 1 as well. The separation between means (different persons and same or different emotion) is higher than in previous comparison.

In order to verify the statistical significance of the difference between our matching score sets, we used Wilcoxon rank-sum test. This test verifies a null hypothesis that two sample sets originate from the same distribution versus the hypothesis that they originate from different distributions. The p-values of this test are presented in Table 2.

Score set separation	p-value
Biometric person identification	.2234
Emotion identification	5.5e-16

Table 2. p-values of the Wilcoxon Rank-Sum test on corresponding score sets.

The emotion identification means that we compare two distributions - matching scores from different persons with same emotion and matching scores from different persons with different emotions. The biometric person identification means that we compare distributions of matching scores from same persons with same emotion and matching scores from different persons with same emotion. The small p-value for emotion identification (5.5e-16) indicates that the difference between corresponding score distributions is statistically significant, and that we indeed have good features for emotion identification. On the other hand, high value of p-value for biometric person identification is more than traditionally used threshold of 5%, which indicates that corresponding distributions might not be statistically different.

Though the statistical significance can be achieved by increasing the number of testing samples, these samples are rather hard to obtain. Instead, we attempted to reduce a noise in used features by considering their projection into smaller dimensional space using principal component analysis (PCA). The PCA projection is derived using feature vectors obtained from all image sequences (pairs of neutral and emotional faces). Table 3 contains results of these experiments. We listed only statistics related to biometric person identification distributions.

Projection dimension	μ_{gen}	σ_{imp}	μ_{gen}	σ_{imp}	p-values
N=20	.72	.27	.79	.24	.191
N=10	.63	.25	.70	.25	.113
N=5	.50	.24	.57	.27	.147
N=2	.37	.23	.44	.28	.189

Table 3. Score statistics and p-values of the Wilcoxon Rank-Sum test on biometric score sets after PCA transform.

As we can see from the p-values of rank-sum test, we get a better separation between genuine and impostor scores than before. Though the numbers are still higher than 5%, we see a consistency among means of the distances - the genuine distance scores are always lower than impostor distance scores, which gives a positive support into hypothesis that scores come from different distributions.

3.2. Big Brother 3 Database

The experiments on Cohn-Kanade database were not able to provide a statistical proof that tracker displacement features are really useful for biometric person identification. As we noted, one problem with the test is that we have too few sample of one class (genuine matching scores). So, we expanded our experiments using additional image database extracted from Big Brother 3 television show episodes. The advantage of using this database is that each person participating in the show has multiple video appearances over a long period of time (1-3 months) and expressed emotions are more natural. The disadvantage of this database as compared to Cohn-Kanade database is that video sequences might not contain neutral expressions. As a consequence, neutral and emotional expression images were extracted from different video sequences and matched in pairs randomly in order to obtain displacement features.

Using the help of human expert we obtained 46 sad, 38 happy expression images from 3 persons, as well as 4 neutral expression images per person. Each emotional image was used only once, but neutral expression images were used multiple times in pairs. Such setup was rather necessary since it turned out that videos contain much more emotional expression sequences than neutral face sequences. We tried to select emotional expression images from different video sequences in order to avoid dependencies in matching scores.

Since each person has many images with the same emotional expression, we are able to obtain a much bigger number of genuine matching scores for our tests: 488 sad genuine match pairs and 236 happy genuine match pairs for a total of 724 pairs, or genuine matching scores. The numbers for impostor matching: 547 sad emotion pairs and 467 happy emotion pairs for a total of 1014 pairs, or impostor matching scores.

Since PCA feature selection showed an improvement in performance for Cohn-Kanade database, we used it also here to obtain 10 features instead of original 116. The match score corresponds to th Euclidean distance between two feature vectors in a projected space. The statistics of resulting genuine and impostor score set is presented in Table 4.

Score set	μ	σ
Same person, same emotion	2.5733	1.8032
Different persons, same emotion	3.2252	1.6661
p-value of Wilcoxon rank-sum test	2.67e-026	

Table 4. Score statistics of Big Brother database matching score sets.

As we can see from the results we get statistically significant separation between genuine and impostor score sets. Though there is still a significant overlap between genuine (same person, same emotion) and impostor (different persons, same emotion) score sets (judging from means and standard deviations), due to the bigger size of test sets we can confirm that considered features are indeed useful for biometric person identification. Note, that the matching scores are generally much larger than the matching scores obtained in Cohn-Kanade database, which might be explained by lower quality images, bigger distance from face to the camera and face rotations. Face rotations producing non-frontal face images might have been the biggest culprit for general matching distance increase, since our feature extraction method did not account for such rotations.

The ROC curves for proposed face expression based biometrics are shown on Figure 2 for both used databases. The curves show similar performance of a little less than

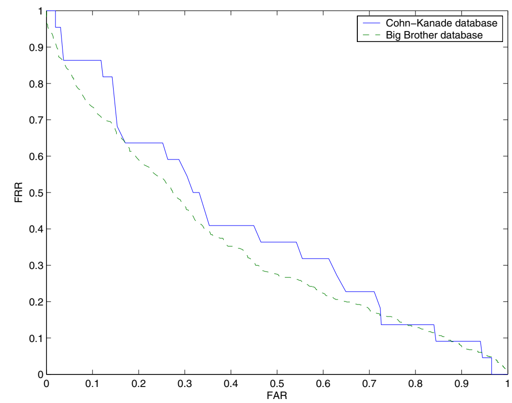

Figure 2. ROC performance curves for face expression biometrics on two used databases.

0.4 equal error rate. Note, that the curve for experiments on Cohn-Kanade database has visible steps corresponding to few genuine matching pair samples. Though proposed method has only little discriminative ability, it still can be used along with other biometrics methods, e.g. traditional face biometrics. Also, as previous research suggests, using it for video sequences can result in further improvement.

4. Conclusion

In this paper we considered using emotion expression face dynamics for the purposes of biometric person identification. If we not only enroll a person's face into the biometric database, but also the way person expresses emotions, for example, smiles, such information might be useful for identifying the person. In our experiments we used the displacements of tracker points as our features. We assume that such features only convey the emotional expression information, but not traditional face biometrics information. The statistical testing we performed showed that such features can indeed be used for person identification.

References

[1] N. J. Briton and J. A. Hall. Gender-based expectancies and observer judgments of smiling. *Journal of Nonverbal Behavior*, 19(1):49–65, 1995.

[2] 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.

[3] P. Ekman. *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*. W.W. Norton, New York, 1985.

[4] P. Ekman and W. V. Friesen. *The facial action coding system*. Consulting Psychologists Press, Palo Alto, CA, 1978.

[5] A. Jain, S. Dass, and K. Nandakumar. Can soft biometric traits assist user recognition? In A. K. Jain and N. K. Ratha,

editors, *Biometric Technology for Human Identification. Proceedings of the SPIE*, volume 5404, pages 561–572, 2004.

[6] T. Kanade, J. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, pages 46–53, 2000.

[7] E. Krumhuber, A. S. R. Manstead, and A. Kappas. Temporal aspects of facial displays in person and expression perception: The effects of smile dynamics, head-tilt, and gender. *Journal of Nonverbal Behavior*, 31(1):39–56, 2007.

[8] G. Peleg, G. Katzir, O. Peleg, M. Kamara, L. Brodsky, H. Hel-Or, D. Keren, and E. Nevo. From the cover: Hereditary family signature of facial expression. *PNAS*, 103(43):15921–15926, 2006.

[9] K. L. Schmidt and J. F. Cohn. Dynamics of facial expression: Normative characteristics and individual differences. In *Multimedia and Expo, 2001. ICME 2001. IEEE International Conference on*, pages 547 – 550, 2001.

[10] M. B. Stegmann. Analysis and segmentation of face images using point annotations and linear subspace techniques. Technical report, Technical University of Denmark, DTU, 2002.