IDIAP RESEARCH REPORT

IDIAP

Martigny - Valais - Suisse

Recognition of Asymmetric Facial Action Unit Activities and Intensities

Beat Fasel [1] and Jürgen Lüttin [2]

IDIAP–RR 99-22

December 1999

published in
Proceedings of ICPR 2000, Barcelona, Spain

Dalle Molle Institute for Perceptual Artificial Intelligence • P.O.Box 592 • Martigny • Valais • Switzerland

phone +41 − 27 − 721 77 11
fax +41 − 27 − 721 77 12
e-mail secretariat@idiap.ch
internet http://www.idiap.ch

1 Beat.Fasel@idiap.ch
2 Juergen.Luettin@idiap.ch

IDIAP Research Report 99-22

Recognition of Asymmetric Facial Action Unit Activities and Intensities

Beat Fasel and Jürgen Lüttin

December 1999

published in
Proceedings of ICPR 2000, Barcelona, Spain

Abstract. Most automatic facial expression analysis systems try to analyze emotion categories. However, psychologists argue that there is no straight forward way to classify emotions from facial expressions. Instead, they propose FACS (Facial Action Coding System), a de-facto standard for categorizing facial actions independent from emotional categories. We describe a system that recognizes asymmetric FACS Action Unit activities and intensities without the use of markers. Facial expression extraction is achieved by difference images that are projected into a sub-space using either PCA or ICA, followed by nearest neighbor classification. Experiments show that this holistic approach achieves a recognition performance comparable to marker-based facial expression analysis systems or human FACS experts for a single-subject database recorded under controlled conditions.

1 Introduction

In the seventies, Ekman and Friesen [3] identified six basic emotions: happiness, sadness, fear, disgust, surprise and anger. These can be discerned relatively easy, however, general emotion display is heavily dependent on culture, age, gender and context. Facial displays are also influenced by cognitive and physiological processes. FACS [3] is a meta representation that allows to describe facial expression prior to interpretation. It decomposes the face into so called Action Units (AUs), face regions that were chosen to distinctively represent underlying muscle actions and that are able to accurately describe changes of facial expressions. There are 46 basic AUs which account for changes in facial expressions, 8 AUs were defined to code the head and another 6 AUs for the eye positions. Our facial expression analysis system was trained to classify 12 different AUs, 9 single AUs and 7 AU combinations out of the total 46 basic AUs. Further on, FACS assigns to each Action Unit either three or five intensity levels. Throughout the tests we present here, 5 levels of intensity were used. Several automatic facial expression analysis systems were presented that can distinguish facial expression intensities [9][1][8][4], but only a few are able to do so in a FACS coding framework [6][7]. Kaiser and Wehrle [6] refined the FACS coding rules by introducing an asymmetric representation of AU intensities. Their so called FEAT coding allows to distinguish different facial action intensities on either the left or the right hand side of a face.[1] We show further below that our system is able to classify asymmetric facial action and intensities by the use of the latter coding scheme.

1.1 Related Work

Automatic FACS classification has been investigated only by a few research teams [7][2][6]. Kaiser and Wehrle's system [6] achieves human expert level performance but depends on markers. Lien and Cohn's [7] approach is mainly based on holistic and local optical flow methods and employs HMMs that allow for sequence classifications employing expression dynamics. Optical flow methods on the other hand can only work reliably with the absence of iconic changes such as eye and mouth closure as well as transient facial features caused by furrows. Facial expression intensities were determined with three methods, namely by measuring the displacement of optical flow based feature points, density measurements of high gradient components in segmented areas as well as with weight vectors that were obtained by PCA compressed dense optical flow. Bartlett et al. [2] used both difference image and optical flow approaches for the extraction of facial expressions. They obtained best results with holistic difference image approaches. We employ a similar facial feature extraction strategy but don't segment the test face into a lower and an upper part as this may lead to errors due to unprecise alignment. Instead, whole face images where processed.

1.2 Database

Our database consists of posed facial expressions performed by a FACS expert that was recorded by Kaiser and Wehrle [6]. It contains image sequences with frontal faces onto which distinctive markers were placed for the use with an automatic facial expression recognition system. These markers were removed for our experiments using morphological operators. The database was recorded under controlled conditions, i.e. the face retains a constant position through all sequences without scale or rotation changes. Furthermore there were no lighting variations. There are totally 48 facial expression sequences showing 16 distinct facial actions, each repeated three times and thus allowing for a training, evaluation and test set, see Figure 1. Images are of size 288×384, cropped down to 230×160 and were recorded at a framerate of 22-24 frames per second. Test were performed on the whole database and on the first 9 series only, as these show single FACS AU activities. Our evaluation sets encompass 182 training and 45 evaluation images of single FACS AU activities (series 1-9), respectively 322

[1] The FEAT coding scheme describes FACS intensities with letters ranging from *a*-*e* for both sides of the face. *a* stands for minimum and *e* for maximum intensity. Example: *AU 12 cb* means FACS Action Unit 12 activated with medium intensity on the right-hand side and weak intensity on the left-hand side. Note that the left and right-hand side relate to the test subject's view.

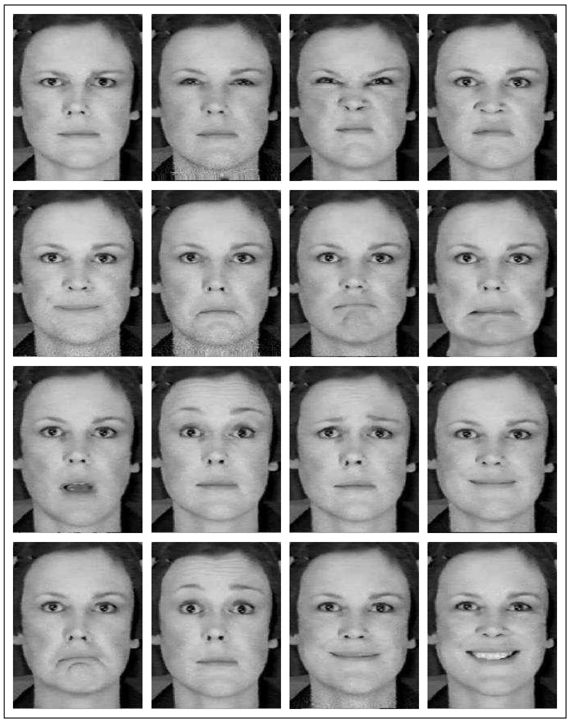

Figure 1: Our database contains 16 FACS AU displays: AU4, AU7, AU9, AU10, AU14, AU15, AU17, AU20, AU26 and following combinations AU1+AU2, AU1+AU4, AU6+AU12, AU15+AU17, AU1+AU2+AU4, AU6+AU12+AU15 and AU6+AU12+AU26 (from top left to bottom left). Each sample shows a key frame with maximal AU intensity.

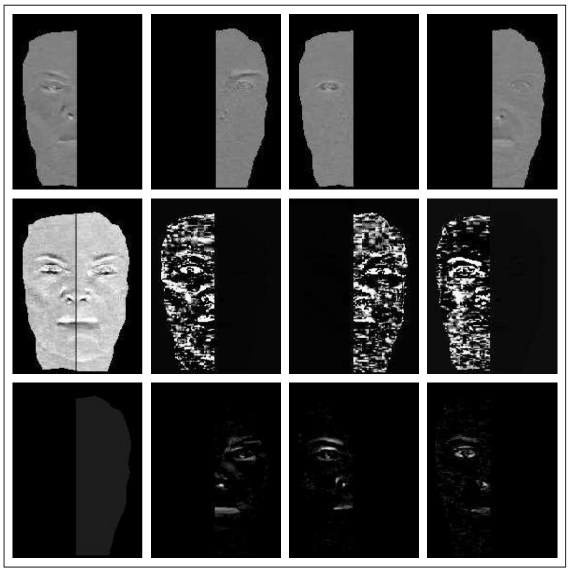

Figure 2: Shown are sample face difference images in the first row, ICA components in the second row and PCA components in the last row.

training and 80 evaluation images for the whole database (series 1-16). Totally 45 test images were used to evaluate the single FACS AU series and 90 to evaluate series 1-16. As most facial actions were performed symmetrically, we artificially created asymmetric facial action images by masking one side of the face. This way all possible asymmetric facial actions for the subset of FACS AUs analyzed in this paper were represented in the database.

2 Facial Expression Analysis System

First, a given test face is separated from the background with the aid of a manually chosen mask that retains only the parts of the face relevant for facial expression analysis. Facial features were then extracted by subtracting a reference frame containing a neutral face from the given test images. This procedure is quite robust in our case, since the positions of all face samples are constant. In less controlled recording conditions, methods that accurately align the reference image to the test images are necessary. The subsequent processing of the difference images is critically dependent on precise alignment of position and scale of the faces. Difference images have the advantage of being robust to changes in illumination and skin texture variations between subjects and allow to focus on changes caused by facial expressions. We chose PCA and ICA to produce statistically independent basis images of the centered difference images. This allows for dimensionality reduction by using only the most representatives components. Often, this has the additional advantage of reducing noise contained in components with small eigenvalues (PCA). ICA is a generalization of PCA and separates also high-order dependencies in the input data as it doesn't constrain representation axes to be orthogonal but places them in directions of statistical dependencies in the data. We employed the FastICA algorithm [5] that allows for cubic convergence and simple handling. No step-size parameters are needed as in the case of gradient-based approaches. Figure 2 shows sample difference images (top row) together with resulting ICA (middle row) and PCA (bottom row) components. According to the FACS standard, each frame was coded independently, since AUs are not based on temporal information. The classification task consists in determining the AU for each test image as well as the corresponding intensity of the unit. For this purpose, each reference face sample was attributed a corresponding FACS AU and two intensity levels for the left and the right-hand side of the face.

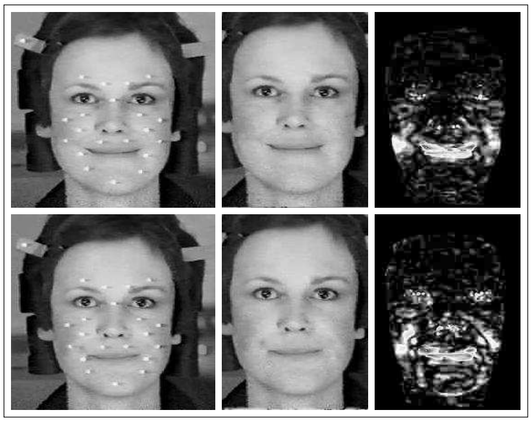

Figure 3: Influence of JPEG compression artifacts. The first column shows the original marked face image of series 12 and 5, the column in the middle the marker-free samples used for our system. In the third column the resulting difference images are given. Most of the noise in the difference images stems from JPEG artifacts that are caused by the lossy compression of both the reference frame and the test frame.

This leads to 36 possible intensity combinations per single AU (with 5 active intensity levels and also including zero intensity). A class was chosen per AU and its corresponding intensity level. Combined FACS AU activities (series 10-16) were represented by a single class. The nearest neighbor classifier was used to determine facial actions. The closest distance of the projection of a test face image onto the ICA / PCA components determined both the active AUs and their intensities. Best candidates were computed independently for each side of the face. Ground truth was provided by the marker-based system implemented by Kaiser and Wehrle [6] that run on the same database using the original marked face images.

3 Experiments and Results

Several test runs were performed using the evaluation set in order to determine the optimal number of PCA and ICA components. Best AU recognition results were obtained with 60 PCA, respectively 75 ICA components for series 1-9 and 105 respectively 120 for series 1-16. For most facial action series, ICA performed better than PCA, especially with regard to the recognition of AU intensities. Note that both PCA and ICA achieve inferior recognition performances for series 6 and 12, mainly due to the resemblance of both classes, see also Figure 3. JPEG artifacts are probably responsible for the loss of information the skin would have provided in this case (dimples in the corners of the mouth). Frames recorded without compression or a higher image resolution could probably remedy the current performance drop on these sequences. Nonetheless, this is an important finding as most image databases are compressed with lossy algorithms. Table 1 shows recognition results obtained for different facial action sequences using either PCA or ICA. We found that most false classifications were obtained due to weak AU activations that were classified as being neutral. AU classification is more difficult with low AU intensities because these frames resemble closely those with neutral faces. Another finding is that intensity miss-classifications were often only one level away from correct classification. A significant better recognition performance was obtained when correct classification was declared in cases where the computed intensity was within +/- one level from ground truth.

We have tested our facial expression analysis system also with regard to image quality degradation. Images corrupted by noise stemming from standard JPEG compression and vertical skipping of face position (several pixels) occurred on the video tape the database was recorded on. Informal experiments showed that difference images are suitable to suppress granular noise effectively whereas

Test Description	Action Unit Activity	Action Unit Activity + Intensity
PCA Series 1-16	74%	**30%**(**55%** +/- one int.) 5 asymmetric levels
ICA Series 1-16	74%	**28%**(**65%** +/- one int.) 5 asymmetric levels
PCA Series 1-9	79%	**35%** (**65%** +/- one int.) 5 asymmetric levels
ICA Series 1-9	83%	**41%** (**70%** +/- one int.) 5 asymmetric levels

Table 1: FACS Action Unit Activity and Intensity Recognition Results. On the upper two rows are given the results using the full database (series 1-16) and on the lower two rows results obtained with series 1-9 (single FACS AU active).

vertical skipping greatly degrades recognition rates.

4 Conclusions

We presented one of the first approaches that can handle asymmetric FACS AU activation and intensity recognition. Our system does not depend on markers and no face segmentation is necessary. Since it applies neither feature tracking nor optical flow, computational requirements are relatively modest during the test phase and the system can therefore run in real-time on a low-end PC / Sun. Our results show that the chosen methods are sufficient for the determination of both AU activations and intensities. We obtained an almost perfect FACE AU activation recognition performance, comparable to marker-based facial expression analysis systems or human FACS experts. Tests were performed on a database that contains only a single subject. The reason for this is that there is currently no publicly available database that would serve our purpose, especially with regard to asymmetric AU intensity coding.

It is difficult to label AU with intensities as FACS doesn't indicate exactly, how to accomplish this. Difference images are suitable for the determination of AU intensities, because no geometrical measurements have to be introduced. Some AUs were discerned with less reliability, probably due to noisy images. More research has to be done to determine the influence of image compression on difference images. The ICA representation approach showed to be an interesting alternative to PCA. On the other hand ICA seems to be rather sensitive to noise in the signal as can be seen in Figure 3. Even week signals can be amplified considerably. Preprocessing in order to eliminate noise is therefore important. This can be achieved with various techniques. We have employed PCA that is also needed to prewhite data before the actual source separation procedure of ICA.

Further work will concentrate on multiple subjects, varying recording conditions and a more sophisticated classifier approach.

5 Acknowledgments

The authors would like to thank Susanne Kaiser and Thomas Wehrle from the University of Geneva for making available their database as well as fruitful discussions concerning FACS in general. Furthermore we are grateful for the hints provided by Perry Moerland on how to handle PCA and ICA representations with signals of high dimensionality. Last but not least thanks go to the SNSF (Swiss National Science Foundation) for funding this project under project number 21-54000.98.

References

[1] A. Lanitis, C.J. Taylor, and T. F. Cootes. Automatic Interpretation and Coding of Face Images using Flexible Models. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 19(7):743–756, 1997.

[2] Marian Stewart Bartlett. *Face Image Analysis by Unsupervised Learning and Redundancy Reduction*. PhD thesis, University of California, San Diego, 1998.

[3] P. Ekman and W. V. Friesen. Manual for the Facial Action Coding System. In *Consulting Psychologists Press, Palo Alto*, 1977.

[4] Irfan A. Essa and Alex P. Pentland. Facial Expression Recognition using a Dynamic Model and Motion Energy. In *ICCV95*, 1995.

[5] Hyvaerinen A. and Oja E. *Independent Component Analysis: A Tutorial*. Helsinki University of Technology, P.O. Box 5400,FIN-02015 Espoo, Finland, 1995.

[6] Kaiser S. and Wehrle T. Automated Coding of Facial Behavior in Human-Computer Interactions with FACS. *Journal of Nonverbal Behavior*, 16(2):67–83, 1992.

[7] James J. Lien, Takeo Kanade, Jeffrey F. Cohn, and Ching-Chung Li. Automated Facial Expression Recognition Based on FACS Action Units. In *IEEE Proceedings of FG'98, Nara, Japan*, April 14-16 1998.

[8] Christine L. Lisetti and David E. Rumelhart. Facial Expression Recognition using a Neural Network. In *Proceedings of the 11 th International Flairs Conference*. AAAI Press, 1998.

[9] Yaser Yacoob and Larry Davis. Computing Spatio-Temporal Representations of Human Faces. Technical report, Computer Vision Labratory, University of Maryland, 1994.