AUTOMATIC ANALYSIS OF THE ANGER EXPRESSIONS FROM THE FACE

A. Pattin, P. F. Whelan
Machine Vision Laboratory
School of Electronic Engineering
Dublin City University (D.C.U.), Ireland

***Abstract*— This paper describes a simple method for the detection of anger expressions from the face. The other basic expressions are not involved in this work. A single SVM with linear kernel is trained on neutral and angry examples from the D.C.U. face database, a dataset gathered for the purpose of the project. The face is extracted with a boosted cascade classifier similar to Viola and Jones' object detector and processed by a restricted bank of 6 Gabor filters. The Gabor filter coefficients are the features used to train the classifier.**

Two approaches are addressed and compared: a single analysis of the whole face or a combination of the two face "hemispheres" separately analyzed. While the separate analysis is consistent to the model of the human visual perception of the expressions, it is outperformed by the single analysis approach. Both of the approaches perform badly on a complex generalization test, showing that the dataset is too restricted. On a more simple validation procedure 79% of detection accuracy is achieved with the single analysis method.

***Index Terms*— Cascade Classifier, Face Recognition, Gabor Filters, Gesture Recognition, Image Sequence Analysis, Support Vector Machines,**

I. INTRODUCTION

In a communication, a message is defined by its context, its content and its aim. While humans express content with words only, they employ other ways to convey the context and the aim of a message. A few examples of them are the voice tone or the body language. The high communicative power of the face makes of it one the major vector of "unspoken language" [11]. Being able to perform an automatic analysis of the expressions from the face would highly improve the human-machine interaction but also be a major innovation in a wide range of field, such as security monitoring, psychiatry, research on pain and depression, interactive teaching and telecommunications [1][2]. Due to these promising applications, the facial expression analysis has grown as an active research topic. However, many aspects are still to be resolved for a robust detection in a real-world application [11]. The diversity of the face view is one of the major issues to be overcome. Uneven illuminations, glasses or facial hair are likely to disrupt the analysis of the expression [11].

To cope with these issues, a lot of solutions are proposed following many different approaches. The analysis is based on different kinds of facial features inspired by anatomic knowledge (fiducial points) or by complex models (Active Shape Models, Hidden Markov Models) [11]. They are extracted by either tracking facial motion (difference images, feature point tracking, optical flow, motion models) or measuring the deformation of the face (based on images or on models) [11].

This project is mostly inspired from the work of G. Littlewort *et al.* described in [1]. They aim to detect the six universal expressions (anger, disgust, joy, sadness, fear, surprise) with a bank of 63 SVMs trained with Gabor coefficients selected with Adaboost. The face is first extracted from the frame using a version of Viola and Jones face detector. It is then processed by a bank of Gabor filters, of which the most representative coefficients are selected by Adaboost to train the bank of classifiers. Each SVM is trained on a different pair of expressions. Several approaches are described and compared, one of them achieving a classification accuracy rate of 86,3% on images of the Cohn Kanade dataset. [1] is also one of the few papers to discuss about the computation efficiency of the selected approach. The project described in this document is similar to [1] but also simpler as it is focused on the anger expression. A facial expression video set described in section II.1 was created to support this project. Different approaches were tested and compared, they are described in section II and their performances are detailed in the section III.

II. DESCRIPTION OF THE SYSTEM

1. THE DATASETS

In machine learning, descriptions of complex patterns are achieved by providing a comprehensive set of examples. A pattern with a high diversity like the facial expression requires a large dataset with a wide variety of subjects. Due to the increasing interest in facial expressions, many databases are built and shared online among the research community. This project involves two of them: the Cohn-Kanade (CK [3] and CK+ [4]) facial expressions database3 and the Center for Vital Longevity Face Database [5] of the University of Texas at Dallas4 . Another dataset, called the D.C.U. face database, was constituted for this project.

The D.C.U. face database is a video dataset. The recordings feature students performing an anger expression from a neutral face. It gathers 66 videos of 18 subjects, 10 males and 8 females. Building my own dataset allows to flout the constraints imposed by the previous work and to have a better understanding of the data content for the test phase. On the other hand, this is a long and delicate task. This dataset were used for training and testing the developed program.

The Cohn-Kanade dataset is a state of the art database of facial expressions commonly used within the research community [3][4], including in [1]. It includes in total 123 subjects performing different posed expressions in front of a recording camera. The 6 basic expressions are included in the dataset, but all the subjects do not perform all of them. In this project, this dataset is used to test the algorithm with an external dataset and to compare the influence of other expressions than anger (see section V).

The U.T. face database of Dallas was created in the need of a dataset that represents all the age groups [5]. It includes colour images of 352 female and 226 male subjects with neutral faces [5]. Other expressions are included, but with fewer examples. The dataset represents a good cross section of the face diversity with subjects of different ages and ethnicities. In addition, the examples are recorded in conditions similar to the DCU dataset: they are displayed in front of a simple background in a frontal view, with a uniform illumination. This dataset was particularly fitted to test the face detector described in section II.C.

2. THE ANGRY EXPRESSION AND THE CHOSEN ACTIONS UNITS

There is no unique description of an angry face of which more than 60 variations have been listed. However, they all involve some of the following Action Units (A.U.): 4,5,7, 23 [6] shown in the figure 1.

Typical anger facial deformations are lowered eyebrows forming wrinkles on the forehead. Upper eyelids are raised; lower eyelids are tensed and straightened. The mouth is whether closed with the lips pressed together, tensed and thinned, or opened showing teeth. A boss might appear on the chin.

Fig. 2 Typical Action Units involved in the anger expression. Figures taken from [7]

The analysis of the angry face shows that two regions of interest emerge:

- the upper part of the face, including the forehead, the root of the nose and the eyebrows and the eyes,
- and the lower part with the mouth and the chin

This is consistent with the results of psychophysical experiments showing that the perception of facial expressions follows a "late integration model" [1] in which these two parts of the face are scanned separately. In [1], G. Littlewort and her team improved their results of 1% by implementing a separate analysis of the two parts of the face.

Therefore in this project the detection of facial angry expression is attempted following two strategies:

- The single analysis of the whole face
- And the separate analysis of the two "hemispheres" of the face.

3. EXTRACTING THE FACE WITH A BOOSTED CASCADE CLASSIFIER BASED ON HAAR-LIKE FEATURES

The face is automatically extracted in each frame with a cascade classifier inspired from Viola and Jones object detector [8]. To ensure a robust detection, the image is scanned at different scales with a set of Haar-like features. They are fast to compute, but generate a huge amount of data to process. An efficient way to reduce the number of features is to select the most relevant candidates with a boosting algorithm. They are passed through a series of classifiers of

[3] Official website: http://www.pitt.edu/~emotion/index.html

[4] More information about the database are available online: http://agingmind.utdallas.edu/facedb

increasing complexity called a "cascade", leaving the finest analysis to the end where only a few candidates remain.

The face detector used in this project is the implementation available in OpenCV libraries. It was tested on the Center for Vital Longevity Face Database of the University of Texas at Dallas and the DCU angry faces database that are described in section II.A. The detection accuracy rates are higher than 98% for both of the databases without any false detection.

These results are expectable due to the fact that the databases are recorded in controlled environments: the faces are shown with a frontal view, a simple background, and uniformly lit.

The analysis operates in an efficient way but takes an average of 500 ms by frame on a 2 GHz Pentium processor, preventing a real time analysis.

Another concern is the face extraction accuracy. Even though the false alarm rate of the detector is very low, it is of high importance to extract only the face of the subject that most of the time does not fit the shape of a rectangle. A solution reported in [1] is the use of an eye detector to relocate the location of the detected face. The eye detector implemented in the OpenCV libraries was tested for this purpose. The results obtained were bad, so it is not included in the project.

Fig. 3 Two outputs of the face detector. The extraction example on the left is invalid. The right example shows a valid extraction.

In the separate analysis approach, the face is extracted using the same detector. It is then split in two parts using predefined coordinates that have proved to extract the facial features without any background on the considered dataset. The two images are rescaled to 24x48 pixels.

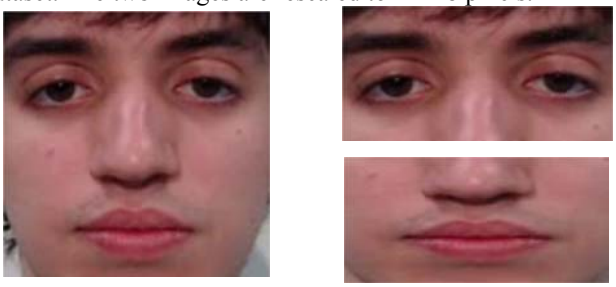

Fig. 4. A comparison of the extracted regions of interest in the two approaches.

The figure above shows that the second approach allows to get rid of most of background extracted along the cheeks.

Another approach based on a mouth and eyes detectors for the two parts of the face extraction was tested. It was proved to be unreliable mainly because of the inaccuracy of the mouth detector implemented in OpenCV. In addition, this approach is twice longer as it involves two multi scale detection.

4. Feature extraction with a bank of Gabor Filters

Once the face is extracted, it is resized into a 48x48 pixels and processed by a bank of Gabor Filters. Gabor filtering is a robust extraction technique for local features, invariant to small motion and deformation [2]. A Gabor function is the combination of a 2D Gaussian kernel and a sinusoidal plane wave oriented following the direction θ. The filtering process highlights the edges of the image that are perpendicular to the given direction θ. The scale of the analysis is defined via the standard deviation of the Gaussian kernel. Several filters are used to extract features at different scales and with different orientations. It is referred as a bank. The rule of thumb is a 40 filters bank of 8 different orientations at 5 different scales, as reported in [1]. The output of such a set of filters is a Gabor mosaic of 40*48*48 = 92160 coefficients that are computed for each frame. In [1], G. Littlewort and her team use Adaboost to select the features training the classifiers. Using a statistical boosting algorithm in this case is not possible because the dataset is too limited. Instead, an investigation based on the entropy of the Gabor mosaics has shown that most of the variation between the anger and the neutral frames can be extracted by a restricted set of filters.

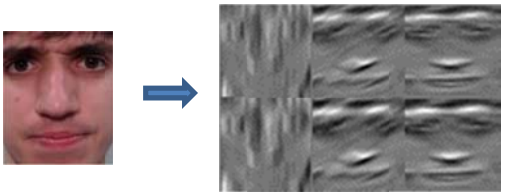

Fig. 5. The facial features are extracted with a restricted bank of Gabor filters

Two wavelengths (4 and $4\sqrt{2}$) and three orientations (0 rad, $\frac{4\pi}{8}$ and $\frac{5\pi}{8}$) are retained. It can be noticed that the angles considered are horizontal or vertical only. An explanation to this is that the anger expression results from a horizontal stretching of the face (tightening of the lips and the lid, horizontal muscle bulge on the forehead, frown resulting in eyebrows flattening) and a vertical contraction (tightening of the eyelid, eyebrows lowering).

In the separate analysis of the two parts of the face, the regions of interest are processed in the same way.

5. Training the Support Vector Machines (S.V.M.)

The Gabor filters coefficients are used to train a C-type SVM with a linear kernel.

In [1], G. Littlewort et al. evaluated the performances of different SVM kernels for the purpose of facial expressions classification (Linear, polynomial, Laplacian and Gaussian). The linear kernel and unit-width Gaussian reach the best performances, but the linear kernel is selected because it is the fastest option [1], [9].

A C-type SVM allows the differentiation of n classes (with $n \geq 2$) with possible imperfect separation [10]. The best hyperplane separating the n clusters of data is chosen in an attempt to minimize the number of misclassified examples and maximize the margin between the hyperplane and the closest examples [10]. The advantage is given to one of the approach in function of the regularization parameter C value. For a large value of C, the optimization minimizes the number of misclassified examples, even if it leads to smaller margins to the hyperplane. The opposite approach is taken for a small value of the regularization parameter.

In this project, the SVM is trained with different values of C to find the optimal value minimizing the test set error [9]. A cross validation estimate gives the classification error for each new value of the regularization parameter. The training examples are a neutral and an angry expression at its highest intensity for each video of the D.C.U. face dataset

In the second approach, a SVM is trained for each part of the face with the same type of examples. The classification decision is based on the addition of the classification value of the two parts of the face.

III. Results Obtained

Four different tests were performed to measure the performance of the two chosen approaches.

6. Test with an external dataset

First, a simple test was performed to measure the performance of the single analysis approach on an external dataset. The testing sample was composed of image of neutral and angry faces from 21 subjects selected from the Cohn-Kanade dataset [3], [4].

The algorithm, trained with the D.C.U. face database, scanned the images and predicted the expressions displayed by the subjects. The very bad results of the experiment motivated another test with the same testing sample. This time, the classification values of the neutral frames were compared subject by subject to the values of the angry ones.

The results of the two tests are shown in the table below.

The first test show bad results: only 6 out of 21 subjects are correctly classified. Most of the expressions are misclassified, but the classification values of the angry frames tend to be higher that the neutral ones. The average difference between the two classification values is 0.7904, while typical classification values are included between -2 and 2. This consideration shows that the algorithm differentiates the two expressions but cannot recognize them without a reference.

It leads to the second test for which a much better detection accuracy rate of 90.5% is reached.

In the first test, most of the misclassifications are due to the same labelling of the neutral and angry faces, while the classification values follow the right trends. A solution to this issue would be to extend the training dataset to provide more examples.

Direct detection accuracy	28,6%
Detection Accuracy With neutral reference	90,5%
Average classification distance between the neutral and angry examples	0,7904

Table 1: Results from the anger recognition test with training samples extracted from the CK and CK+ dataset [3], [4].

7. Validation procedure with the video set

The second test is much more complex, and was run with the two approaches. The algorithms were tested on videos of the D.C.U. face database following the Leave One Out Cross Validation procedure. In this procedure, all the subjects of the dataset except one are used to train the classifier. A video of the subject left over is used to test the classifier. This process is repeated with a different subject for the test as many times as there are subjects in the dataset. The global error rate is computed as the average of all the misclassified frames divided by the total number of frames of the video set. The transient state between two expressions is not included in the test phase.

Detection Accuracy Rate	Direct Classification	Classification with neutral reference	Classification with average value of the first ten frames
Average on the whole dataset with the single analysis	42,66%	42.61%	37.63%
Average on the whole dataset with the combined analysis	30,82%	25.77%	43.41%

Table 2: Results from the anger recognition test with

training samples extracted from the CK and CK+ dataset [3], [4].

As the table 2 shows above, both of the tested methods performed badly with accuracy rates of 42,66% and 30,82%. Surprisingly, the second approach combining the analysis of the two parts of the face exhibits the worse results. The separate analyses of the face interfere, leading to an incorrect prediction. In particular, the classification of the lower part of the face has proven to be unreliable. This is due to the fact that it is easier to provide good training examples of the upper part of the face. A clear distinction between the neutral and angry state is harder to perform with the lower part of the face. The solution to this issue would be to investigate more complex decisions schemes to combine the two analyses, such as the MLR (multinomial logistic ridge regression) reported in [1].

As the results of the direct classification method are bad, the validation procedure was run again, using this time, the classification value of the first frame as the neutral reference. Therefore, every frame with a classification with a classification higher than this reference is interpreted as an angry frame. This approach is motivated by the fact that for some subjects, the shape of the eyebrows suggests the characteristic frown of the anger expression. As every testing video starts with a neutral face, the first frame can be used as a reference. The results with this approach are even worse, with accuracy rates of respectively 42,61% and 25,77%.

Another attempt is to use the average of the first 10 frames classification value as the neutral reference. This allows to reduce the probability that the reference value is an outlier. This strategy shows also bad results, with respectively 37,63% and 43,41% of detection accuracy.

These results show that the validation procedure is too complex and that the detection can hardly be performed on new subjects. The inability of the algorithm to generalize is due to the restricted size of the dataset, not broad enough to cope with the high diversity of the human face.

8. SECOND VALIDATION PROCEDURE WITH THE VIDEO SET

The second validation procedure is a simpler version of the Leave One Out Cross Validation procedure suggested from the bad results of the previous validation procedure. This time, all the videos of the dataset but one are used for training, the video left behind being used for the test phase. Each of the 18 subjects of the dataset features in at least three training videos. The testing sample is therefore a new sample, but from a subject that the classifier has been trained with. This procedure is repeated as many times as there are subjects in the dataset.

Detection Accuracy Rate	*Correct Classification*	*False Negative*	*False Positive*
Average rate on the whole dataset with the single analysis	*79%*	*12,23%*	*8.76%*

Table 3: Results from the anger recognition test with training samples extracted from the CK and CK+ dataset [3], [4].

This procedure shows much better results with an average detection accuracy rate of 79%. For three testing samples a detection accuracy of 100% is even achieved. The results in table 3 also show that the algorithm misclassifies more often the angry faces (12,23% of the frames) than the neutral ones (8.76%).

9. EXPERIMENT WITH A SIMILAR UNIVERSAL EXPRESSION

Most of the state of the art solutions developed to automatically detect facial expressions attempt to recognize the six universal expressions reported by P. Ekman: anger, sadness, happiness, surprise, fear and disgust [1],[6], and [11].

In this project, the outcomes of the analysis are either anger or neutral, but the neutral actually refers to “all the expressions but anger”. This solution was selected because it is too complex to provide a comprehensive sample of all the facial expressions that differ from anger. At this stage, it has then been assumed that no other expression would interfere in the analysis. However, this is too restrictive for a real world application. It is important to know the influence of the others expressions in the detection process.

The last test described in this report involves the expression of disgust which is really close to anger (see Fig. 6).

The testing set is composed of neutral and disgusted frames of 50 subjects selected from the CK dataset [3], [4]. The classifier predicting the nature of the expression has been trained on the D.C.U. face dataset.

As expected, 72% of the disgusted expressions are misclassified as anger expressions. This is due to the similarity of the two expressions and to the nature of the training.

Anger misclassification	72%
Average classification distance between the neutral and disgust examples	0.8998

Table 4: Results from the anger recognition test with training samples extracted from the CK and CK+ dataset [3], [4].

The margin between the neutral and disgust is 0.8998, which is even higher than the one observed with the test section III.A. It means that the separation of the two classes is even more marked with the disgusted faces.

Fig. 6. From the two figures, it can be hard to tell which face displays the disgusted expression from the angry one. Two strong characteristics allow to detect the disgusted face (on the left): the furrows displayed on the root of the nose and the two vertical furrows going from the mouth to the extremities of the lips.

To deal with different expressions, the classifier must be trained with all of them, the considered expression being labeled as one class, and all the other expressions as the second class. As the classifier in this case has only been trained with neutral and angry faces, it can hardly perform well with another expression such as the disgust.

IV. CONCLUSION

This paper describes the different approaches developed for the automatic detection of angry faces. In particular, the single analysis method is compared to a late integration model combining the two parts of the face. These solutions do not perform well on new subjects (respectively 42.66 and 30.82% of detection accuracy) mainly because of the restricted size of the database used. The influence of another expression such as disgust is studied and shows that with the present training (only neutral and angry faces) the classifier cannot distinguish the anger from the disgust.

REFERENCES

[1] G. Littlewort and I. Fasel, "Fully automatic coding of basic expressions from video," *INC MPLab Tech. Rep.*, p. 6, 2002.

[2] M. Valstar and M. Pantic, "Fully Automatic Facial Action Unit Detection and Temporal Analysis," *2006 Conf. Comput. Vis. Pattern Recognit. Work.*, pp. 149–149, 2006.

[3] T. Kanade, J. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," *Autom. Face Gesture …*, 2000.

[4] P. Lucey, J. Cohn, and T. Kanade, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," *Comput. Vis. …*, no. July, 2010.

[5] M. Minear and D. C. Park, "A lifespan database of adult facial stimuli.," *Behav. Res. Methods. Instrum. Comput.*, vol. 36, no. 4, pp. 630–3, Nov. 2004.

[6] P. Ekman and F. W.V., *EMFACS-7: Emotional Facial Action Coding System*. Unpublished manual, 1984.

[7] Y. Tian, T. Kanade, and J. Cohn, "Chapter 11. Facial Expression Analysis," in *Handbook of Face Recognition*, 2005.

[8] P. Viola and M. Jones, "Robust real-time object detection," *Int. J. Comput. Vis.*, no. February, 2001.

[9] opencv dev Team, "Support Vector Machines - OpenCV documentation." [Online]. Available: http://docs.opencv.org/modules/ml/doc/support_vector_machines.html. [Accessed: 30-Aug-2014].

[10] C. Chang and C. Lin, "LIBSVM: a library for support vector machines," *ACM Trans. Intell. Syst. …*, pp. 1–39, 2011.

[11] A. Pattin, "Automatic Anger Expression Detection from Facial Video Sequence," 2014.

[12 Paul F Whelan (2011) "VSG Image Processing and Analysis (VSG IPA) Toolbox", Technical Report, Centre for Image Processing & Analysis, Dublin City University.

[13] Paul F Whelan (2011) "VSG Image Processing and Analysis (VSG IPA) Matlab Toolbox - Software"", http://www.cipa.dcu.ie/code.html, Centre for Image Processing & Analysis, Dublin City University.

[14] Paul F. Whelan and D. Molloy (2000), "Machine Vision Algorithms in Java: Techniques and Implementation", Springer (London), 298 Pages. ISBN 1-85233-218-2.

[15] Paul F Whelan (2011), "EE544: Computer Vision Course Notes", Centre for Image Processing & Analysis, Dublin City University.

[16] Paul F Whelan (2011), "EE425/EE453: Image Processing and Analysis Course Notes", Centre for Image Processing & Analysis, Dublin City University.