

Haar Features for FACS AU Recognition

Jacob Whitehill

*Department of Computer Science
University of the Western Cape, South Africa
whitehill@cs.stanford.edu*

Christian W. Omlin

*Department of Mathematics & Computing Science
University of the South Pacific, Fiji
omlin_c@usp.ac.fj*

Abstract




We examined the effectiveness of using Haar features and the Adaboost boosting algorithm for FACS action unit (AU) recognition. We evaluated both recognition accuracy and processing time of this new approach compared to the state-of-the-art method of classifying Gabor responses with support vector machines. Empirical results on the Cohn-Kanade facial expression database showed that the Haar+Adaboost method yields AU recognition rates comparable to those of the Gabor+SVM method but operates at least two orders of magnitude more quickly.

1. Introduction

Automatic facial expression recognition has applications to human-computer interaction, interactive computer games, and psychological research. It is also a crucial component of any computer system designed to recognize a signed language in real time. As part of a larger project on the integration of signed with spoken communication, we are studying machine-learning algorithms for the recognition of facial expressions. The well-known Facial Action Coding System (FACS) by Ekman and Friesen [4] provides the framework. FACS defines expressions in terms of the presence or absence of 44 elementary muscle movements, called *action units* (AUs). Example AUs are shown in Table 1.

Two general approaches exist for the automatic recognition of facial expressions. *Feature-point* systems track the locations of various landmarks on the face (e.g., pupils, nostrils). The feature vectors of such systems are computed as some function of the positions

Table 1. Examples of AUs

AU 1 (Brow)	AU 5 (Eye)	AU 15 (Mouth)
		

Pictures courtesy of Carnegie Mellon University Automated Face Analysis Group.

and relative distances between the points. *Appearance-based* systems, on the other hand, process color information of face patches to form their feature vectors.

One of the most successful approaches to expression recognition is to apply Gabor filters to extract features and then use support vector machines to classify them into AUs (e.g., [1], [3]). While recognition rates are high (over 90%), this approach is both inefficient in memory usage and slow due to the high redundancy of the Gabor representation.

In this paper, we investigate an appearance-based approach to facial expression recognition which is based on Haar features and the Adaboost boosting algorithm. This combined approach was employed by Viola and Jones in [11] for face detection and has demonstrated both high recognition accuracy and fast run-time performance. In contrast to Gabor features, Haar features can be extracted quickly without a Fourier transform. Moreover, since the boosting algorithm implicitly performs feature selection, the number of extracted Haar features is far less than the number of Gabors, which saves memory. To our knowledge, our paper is the first to study the suitability of the Haar+Adaboost approach for recognizing FACS AUs.



Figure 1. Examples of Haar wavelets in a true Haar decomposition superimposed onto a face image. Width, height, and (x,y) positions of all wavelets are aligned at powers of 2.

2. Haar Features for Object Detection

Recent computer vision research has demonstrated that the Haar wavelet is a powerful image feature for object recognition. The two-dimensional Haar decomposition of a square image with n^2 pixels consists of n^2 wavelet coefficients, each of which corresponds to a distinct Haar wavelet. The first such wavelet is the mean pixel intensity value of the whole image; the rest of the wavelets are computed as the difference in mean intensity values of horizontally, vertically, or diagonally adjacent squares. Figure 1 shows three example Haar wavelets superimposed onto a face image. The Haar coefficient of a particular Haar wavelet is computed as the difference in average pixel value between the image pixels in the black and white regions.

The two-dimensional Haar decomposition is exactly complete, i.e., the Haar decomposition of an image with n^2 pixels contains exactly n^2 coefficients. Each wavelet is constrained both in its (x,y) location and its width and height to be aligned on a power of 2. For object recognition systems, however, these constraints are sometimes relaxed in order to improve classification results. Papageorgiou, et al [9] modified the wavelet decomposition so that the wavelet basis is shifted at 4 times the normal density of the conventional Haar transform. The resulting set of “quadruple-density” Haar coefficients allows object recognition at a finer resolution than would be possible using the standard approach. Viola and Jones in [11] constructed a face detector by using a modified version of the true Haar decomposition which includes a new “Haar” wavelet containing three subregions (instead of 1, 2, or 4).

2.1. Feature Selection

The set of Haar features used by Viola and Jones is many times overcomplete. While this allows very fine-grained inspection of an image, it also increases the training time and can reduce generalization perfor-

mance. For these reasons, the Viola-Jones approach uses the Adaboost boosting algorithm as a means of feature selection by constructing a weak classifier out of each Haar feature. Specifically, a threshold-based binary classifier is created from each Haar feature so that the weighted training error is minimized. During each round of boosting, the single best weak classifier for that round is chosen (corresponding to a particular Haar feature). The final result of boosting is a strong classifier whose output is computed as a thresholded linear combination of the weak classifiers. The Viola-Jones face detector has demonstrated that this classification method is both fast and effective for recognition.

3. Related Work

3.1. Haar Features

To our knowledge, only one system has been developed to-date which uses Haar wavelets for facial expression recognition. Wang, et al [12] use Haar-like wavelets derived from integral images to classify 7 prototypical facial expressions. As in Viola and Jones’ work [11], they create one weak classifier for each Haar-like feature and use Adaboost to select features. Instead of using threshold-based weak classifiers that output discrete values in $\{-1, 1\}$, however, their system uses lookup-tables that map ranges of feature values onto class confidences in $[-1, 1]$ for each emotion category. Using the multi-class, confidence-based Adaboost algorithm, Wang et al achieve 92.4% recognition accuracy on a database of 206 frontal facial expressions. This result is superior to the 91.6% accuracy which they measured when using a SVM with RBF kernel on the same set of features. However, the statistical significance of this 0.8% difference was not assessed. In terms of execution speed, their Adaboost-Haar method clearly outperforms the SVM-based approach: the Adaboost method is 300 times faster [12].

3.2. AU Recognition

Both appearance-based and feature point-based expression recognition systems have achieved state-of-the-art accuracy. Tian, et al [10] developed a feature point-based, multi-state model of 7 upper- and 11 lower-face AUs. Using neural networks as classifiers, they achieved over 95% accuracy in each group of AUs. Donato, et al [3] compared a variety of appearance-based methods and achieved 96% accuracy on 12 AUs, both with Gabor filters and independent component analysis. Bartlett, et al [1], in more recent work, used Gabor filters, support vector machines, and

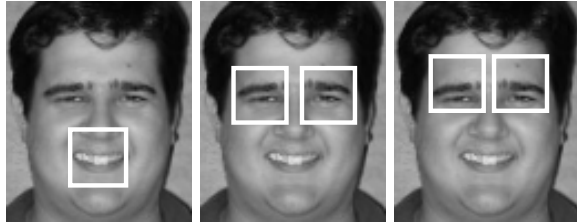


Figure 2. The local face regions of the mouth (left), eye (middle), and brow (right) regions from which features were selected for each AU classifier.

hidden Markov models to detect AUs 1, 2, and 4 with up to 90% accuracy.

4. System Design

Our AU recognition system consists of four stages: image normalization, face region segmentation, feature extraction, and AU classification. We describe each stage below.

4.1. Image Normalization

On each image, the positions of the eyes and mouth were manually located. All images were rotated and scaled such that the coordinates of the eyes and mouth were constant over all images. The face width was set to 64 pixels; the inter-ocular distance was set to 24 pixels; and the y-distance between the eyes and mouth was 26 pixels.

4.2. Face Region Segmentation

In order to reduce the length of time necessary for the lengthy Adaboost-based feature selection process, we designed our system to recognize AUs from local subregions of the face instead of the whole face window. Performing this segmentation greatly reduces the size of the set of all possible features from which a few can be selected. After rotating, cropping, and scaling the face, we selected square regions 24 pixels in width around the mouth, each eye, and each brow. Figure 2 shows the face regions that were cropped from each image.

4.3. Feature Extraction

For each AU, we used Adaboost to select 500 Haar features for classification. Features for classifying mouth AUs were selected only from the corresponding mouth region. Features for the eye AUs were extracted

both from the left and the right eye regions; a similar approach was taken for the brow AU classifiers. Figure 3 shows example Haar features that were actually chosen for AU recognition during the feature selection process. The Viola-Jones “integral image” method (see [11] for details) was used to extract features from images.

4.4. Classification

Each feature in the set of 500 Haar features for each AU was fed to the corresponding weak classifier, which outputs a label in $\{-1, 1\}$. The Adaboost-based strong classifier then outputs the final classification label for that AU based on whether the weighted sum of the weak classifiers’ outputs exceeds the strong classifier’s threshold. See [5] for details.

5. Experiment

For our performance comparison of the Gabor+SVM and Haar+Adaboost methods, we use the Cohn-Kanade AU-Coded Facial Expression Database [8]. In particular, we labeled the mouth and eye positions of 580 images and used this image subset for training and validation. We evaluate classification performance on all AUs for which at least 40 positively labeled images were present. In our data set, these are: 1, 2, and 4 (brow AUs); 5, 6, and 7 (eye AUs); and 15, 17, 20, 25, and 27 (mouth AUs).

5.1. Feature Extraction

The Haar features were extracted as described in Section 4.3. For the Gabor features, each image was converted into a Gabor representation using a bank of 40 Gabor filters. Five spatial frequencies (spaced in half-octaves) and eight orientations (spaced at $\pi/8$) were used. Feature vectors were calculated as the complex magnitude of the Gabor jets, and vectors were then subsampled by a factor of 16 and normalized to unit length as in [3].

5.2. Classification

Each trained classifier detected the presence or absence of one AU, regardless of whether it occurred in combination. We did not attempt to account for non-additive AU combinations.

Ten-fold cross-validation was employed to test the generalization performance. The set of human subjects was partitioned into ten disjoint groups of approximately equal size. Each cross-validation fold contained all the images of a particular group of subjects. None

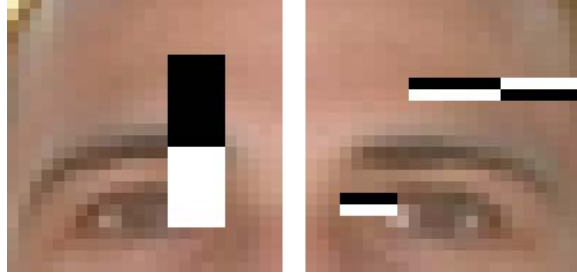


Figure 3. The first three Haar features chosen for AU 1, superimposed on the combined left and right brow regions. Feature 1 (left-most) is Type 2, Feature 2 (middle) is Type 2, and Feature 3 (right-most) is Type 4.

of the validation folds contained the same human subject. We calculated mean accuracies over the ten test folds. When comparing recognition accuracy between two facial segmentations, we performed matched-pairs *t*-tests in order to assess the statistical significance of any difference in mean performance.

6. Results

6.1. Accuracy

Table 2 lists the percentage of images classified correctly for each AU for both the Haar+Adaboost and Gabor+SVM classifiers. Whenever there was a statistically significant difference (for 95% confidence) in performance, we listed the best classifier of the two under the “Best” column. When the difference was insignificant, we listed an = sign. As is shown in the table, the conventional Gabor+SVM approach achieved higher recognition accuracy only for AU 7 (lid tightener). The Haar+Adaboost method, on the other hand, yielded higher accuracy for AUs 1, 2, and 6. Averaged over all 11 AUs classified, the Haar+Adaboost method was close to 1% more accurate than the Gabor+SVM classifier.

7. Run-time Performance

We also compared the run-time performance of both strategies, both in terms of feature extraction and feature classification.

7.1. Feature Extraction

For the FFT implementation necessary for the extraction of the Gabor features, we used the popular li-

Recognition Accuracy (% Correct) of Haar+Adaboost versus Gabor+SVM methods.

AU #	Method		
	Gabor+SVM	Haar+Adaboost	Best
<i>Brow AUs</i>			
1	77.99	82.83	H+A
2	88.29	93.26	H+A
4	86.65	85.23	=
<i>Eye AUs</i>			
5	94.08	94.39	=
6	87.80	93.39	H+A
7	93.86	88.31	G+S
<i>Mouth AUs</i>			
15	94.96	95.66	=
17	90.67	89.51	=
20	96.51	97.27	=
25	96.49	97.85	=
27	98.16	98.11	=
Avg	91.41	92.35	

Table 2. Recognition accuracies of the Haar+Adaboost and Gabor+SVM classification methods. The metric for comparison was the percentage of images classified correctly for the presence or absence of each AU.

brary *FFTW* (the Fastest Fourier Transform in the West) [6]. For basic image manipulation, we employed the simple and efficient *TiP* library (Tools for Image Processing) [7].

We performed experiments for two different image sizes: 24x24 and 64x64. The smaller window size is suitable for classifying facial expression from individual local regions of the face (e.g., mouth); the larger window size is appropriate when analyzing the face as a whole. For Haar feature extraction, 500 selected features were computed. For Gabor features, we applied a standard filter bank of 5 frequencies and 8 orientations and extracted Gabor responses at all points in each filtered image. The execution times were measured on a Pentium IV 1.8 GHz machine and averaged over 1000 rounds of extraction; results are shown in Table 3. The results show that, for 24x24 images, Haar feature extraction is approximately 80 times faster than Gabor feature extraction. For 64x64 images, the Haar features can be extracted nearly 160 times more quickly.

7.2. Classification

Using the same parameters as in section 7.1, we compared the running times of the Adaboost strong

Full Gabor versus Selected Haar Extraction Times

Feature Type	Resolution	Extraction Time
Haar	24x24	0.11msec
	64x64	0.31msec
Gabor	24x24	8.8msec
	64x64	49.3msec

Table 3. Execution times of feature extraction for Gabor features versus selected Haar features.**Adaboost versus SVM Classification Times**

Classifier	Classification Time
Adaboost	0.02msec
SVM (Linear)	21.17msec
SVM (RBF)	93.97msec

Table 4. Classification execution times of an Adaboost strong classifier versus a linear SVM.

classifier with a support vector machine. We used the `libsvm` library [2] for the SVM implementation. Execution times are shown in Table 4. As illustrated by the running times, the Adaboost strong classifier is 3 orders of magnitude faster than the SVM.

8. Summary

In this paper we evaluated both the recognition accuracy and run-time performance of using Haar features and Adaboost to classify FACS AUs. Compared to the standard Gabor+SVM approach, the Haar+Adaboost method achieved similar recognition accuracy, but performed several orders of magnitude more quickly. We believe that Haar features, as well as other image features that can be extracted from the Viola-Jones integral image, will lead to continued improvements in FACS AU recognition.

References

- [1] M. Bartlett, G. Littlewort, B. Braathen, T. Sejnowski, and J. Movellan. A prototype for automatic recognition of spontaneous facial actions. In S. Becker and K. Obermayer, editors, *Advances in Neural Information Processing Systems, Vol 15*. MIT Press, 2003.
- [2] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [3] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski. Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(10):974–989, 1999.
- [4] P. Ekman and W. Friesen. *Manual for the Facial Action Coding System*. Consulting Psychologists Press, 1977.
- [5] Y. Freund and R. E. Schapire. A short introduction to boosting. In *Journal of Japanese Society for Artificial Intelligence*, 1999.
- [6] M. Frigo and S. G. Johnson. The design and implementation of FFTW3. *Proceedings of the IEEE*, 93(2):216–231, 2005. special issue on "Program Generation, Optimization, and Platform Adaptation".
- [7] S. Grigorescu, C. Grigorescu, and A. Jalba. Tools for Image Processing (TiP) Library. No longer available; originally at <http://www.cs.rug.nl/~cosmin/tip/TiP-0.0.1.tar.gz>.
- [8] T. Kanade, J. Cohn, and Y. L. Tian. Comprehensive database for facial expression analysis. In *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00)*, pages 46 – 53, March 2000.
- [9] C. P. Papageorgiou, M. Oren, and T. Poggio. A general framework for object detection. In *Proceedings of the Sixth International Conference on Computer Vision*, 1998.
- [10] Y. L. Tian, T. Kanade, and J. F. Cohn. Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):97–115, 2001.
- [11] P. Viola and M. Jones. Robust real-time object detection. *International Journal of Computer Vision*, 2001.
- [12] Y. Wang, H. Ai, B. Wu, and C. Huang. Real time facial expression recognition with adaboost. In *Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004)*, 2004.