HAL
open science

Automatic facial action detection using histogram variation between emotional states

Thibaud Sénéchal, Kevin Bailly, Lionel Prevost

To cite this version:

Thibaud Sénéchal, Kevin Bailly, Lionel Prevost. Automatic facial action detection using histogram variation between emotional states. Proc. of Int'l Conference on Pattern Recognition (ICPR10), 2010, Istanbul, Turkey. pp.3752-3755. hal-03181842

HAL Id: hal-03181842

https://hal.science/hal-03181842

Submitted on 25 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Automatic facial action detection using histogram variation between emotional states

Thibaud Senechal, Kevin Bailly, Lionel Prevost
ISIR, Universite Pierre et Marie CURIE, Paris
thibaud.senechal@isir.upmc.fr, kevin.bailly@isir.upmc.fr, lionel.prevost@isir.upmc.fr

Abstract

This article presents an appearance based method to detect automatically facial actions. Our approach focuses on reducing features sensitivity to identity of the subject. We compute from an expressive image a Local Gabor Binary Pattern (LGBP) histogram and synthesize a LGBP histogram approaching the one we would compute on a neutral face. Difference between these two histograms are used as inputs of Support Vector Machine (SVM) binary detectors associated with a new kernel: the Histogram Difference Intersection (HDI) kernel. Experimental results carried out for 16 Action Units (AUs) on the benchmark Cohn-Kanade database can be compared favorably with two state-of-the-art methods.

1. Introduction

The Facial Action Coding System (FACS) developed by Ekman and Friesnen [3] has proven to be the most objective and comprehensive coding system and is useful to psychologists and animators. Using FACS, human coders can manually code nearly any anatomically possible expression, decomposing it into specific Action Units (AUs).

A facial image is often the confluence of many factors like the facial expression, identity of the subject, illumination etc. We propose here an appearance based method that minimizes influence of these factors except expression. Local Gabor Binary Pattern (LGBP) histograms are chosen as features because they are very robust to illumination changes and misalignment [10]. Moreover, the use of histograms results in the loss of spatial information which really depends on identity. An original method is proposed using the difference between histograms computed from expressive and neutral faces to reduce even more sensitivity to identity. Relation between neutral and expressive states is applied during the training stage while the neutral state (when non available) is deduced from the expressive one during the testing stage.

Finally a new kernel is associated with Support Vector Machine (SVM) detectors (one per AU): the Histogram Difference Intersection (HDI) kernel which is well-suited for these features.

The paper is organized as follows. Section 2 describes LGBP features and histograms. Section 3 explains the histogram intersection kernel and the proposed HDI kernel. In particular, we prove that the HDI kernel is a Mercer's kernel. Section 4 reports experimental results with different features and kernels and compares to other methods. Section 5 concludes the paper.

2. Facial expression coding

First, facial images are cropped automatically using our eyes detector [8] and resized to 128x128 pixels. To represent a facial expression we compute a LGBP histogram on the expressive facial images (when the expression intensity is maximal). We also compute LGBP histograms on neutral facial images. We use as features either expressive LGBP histograms or histogram differences between neutral and expressive.

2.1 Local Gabor Binary Pattern (LGBP) histograms

LGBP have already been explored for face recognition [10] and emotion recognition [7]. This coding is adapted to our task as it is able to detect small local changes occurring in the face texture.

The facial image is convolved with Gabor filters. We use three spatial frequencies $\nu = (1/2, 1/4, 1/8)$ and six orientations $\theta = (k\pi/6, k \in \{0 \dots 5\})$ for a total of 18 Gabor filters. As we only keep magnitude value, it results in 18 Gabor magnitude pictures. Then, we

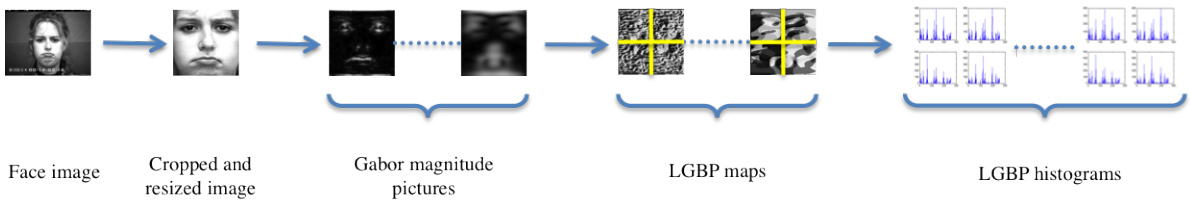

Figure 1. Overall architecture to obtain Local Gabor Binary Pattern histograms.

apply the local binary pattern operator to each of these magnitude pictures. This operator codes a pixel of an image by thresholding its 3×3 neighborhood with its value and considering the result as a binary number.

Finally the 18 Local Gabor Binary Patterns maps are divided into $4 \times 4 = 16$ non-overlapping regions r. Then one histogram of 256 bins is computed for each region. A face i is coded by one vector H_i which is the concatenation of $16 \times 6 \times 3 = 288$ histograms $h_i^{r\theta\nu}$ computed for each region r, orientation θ and spatial frequency ν resulting in $288 \times 256 = 73728$ features per facial image.

2.2 Histogram difference

In the case where we have at our disposal the neutral face of a subject, we just compute the neutral LBGP histogram H_i^n. Then we get the histogram difference:

$$\Delta H_i = H_i - H_i^N \tag{1}$$

Without this neutral image of a test subject, we can synthesize its histogram using the neutral face of all the training subjects. We compute a projection matrix P (using the Singular Value Decomposition algorithm) projecting data on an eigen-space built with LGBP histograms of training subjects neutral face:

$$\widetilde{\Delta H}_i = H_i - P^T P H_i \tag{2}$$

3. Kernels for facial action classification

We have chosen Support Vector Machine (SVM) as our classifier because it has been widely and successfully used in binary class problems.

The kernel function of SVM is an important factor for the generalization performance. Previous works on expression recognition have often focused on linear, polynomial or Gaussian Radial Basis Function (RBF) kernel [5] even in histogram-based approaches [7]. We present here the well-known histogram intersection kernel and a new kernel that can handle histogram difference: the Histogram Difference Intersection (HDI) kernel.

3.1 Histogram intersection kernel for histogram comparison

The histogram intersection kernel has proven to be very efficient in the histogram-based approach for object recognition [6]. This kind of kernel has not been explored yet for facial expression recognition. It is computed by:

$$K^{hi}(H_i, H_j) = \sum_n min(H_i(n), H_j(n)) \tag{3}$$

3.2 Histogram Difference Intersection (HDI) kernel for histogram difference comparison

We define here a new kernel, the HDI kernel:

$$K(\Delta H_i, \Delta H_j) = \sum_{r,\theta,\nu} \frac{\sum_n minabs(\Delta h_i^{r\theta\nu}(n), \Delta h_j^{r\theta\nu}(n))}{\sqrt{\sum_n |\Delta h_i^{r\theta\nu}(n)| \cdot \sum_n |\Delta h_j^{r\theta\nu}(n)|}} \tag{4}$$

with

$$minabs(x, y) = \begin{cases} min(|x|, |y|) & if\ x \cdot y > 0 \\ 0 & else \end{cases} \tag{5}$$

This is basically the intersection histogram kernel adapted for negative values and non-constant sum of histogram. The HDI kernel measures the similarity between two histogram differences. In order to have a higher score if both histograms vary in the same way and with the same amplitude, this measure is normalized by the sum of absolute values of each histogram. Thus, two histogram differences with small values can have a high score (close to the maximum 1). Note that if we use the HDI kernel and the histogram intersection kernel for histogram comparison, they lead to same values up to a multiplicative factor.

3.3 Is the HDI function a kernel ?

To assure that the SVM training will be a convex optimization problem, we have to prove

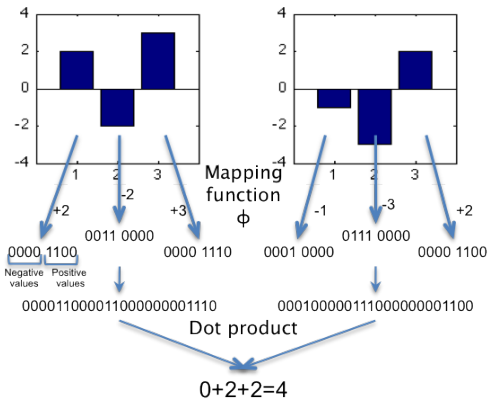

Figure 2. Calculation of $k_{ij}^{r\theta\nu}$ and feature space associated in the case of 3 bins histogram difference with values in [-4 4]

that the HDI function is a Mercer's kernel. Let us notice that the matrix $k^{r\theta\nu}$ defined by $k_{ij}^{r\theta\nu} = \sum_n minabs(\Delta h_i^{r\theta\nu}(n), \Delta h_j^{r\theta\nu}(n))$ is a semi-definite positive matrix by showing that this is an inner product in a suitable feature space as shown in Fig. 2. This is a proof equivalent to the one used in [1] to demonstrate that the histogram intersection function is a kernel. Adaptation to our case is straightforward.

Then kernel matrices $k^{r,\theta,\nu}$ are normalized by dividing each line i and column j by $\sqrt{\sum_n |\Delta h_i^{r\theta\nu}(n)|}$ and $\sqrt{\sum_k |\Delta h_j^{r\theta\nu}(n)|}$ respectively, resulting in semi-definite positive matrices. Finally the HDI matrix is computed by summing all these normalized semi-definite positive matrices. And because the sum of kernels is a kernel, the HDI function is a kernel.

4 Experimental results

4.1 Experimental setup

We use in our experiment the AU-Coded face expression image CohnKanade database [4]. The database contains 486 sequences of 97 subjects starting with the neutral expression and ending with the expression apex. We use the last image of each sequence as target/non-target according to their AU-label and the first image of the sequence as neutral face.

We present results for AUs that occur at least thirty time in the database which is already quite a small number of samples to train SVM detectors. We detect 7 upper face AUs (AU1, AU2, AU4, AU5, AU6, AU7 and AU9) and 9 lower face AUs (AU11, AU12, AU15, AU17, AU20, AU23, AU24, AU25, AU27).

We test generalization to new subjects using leave-one-out cross-validation. We exclude all the images of one subject from the database to train the SVM and use them for test. The overall percent of correct detections depends on the ratio of targets to non targets. We use the area under the ROC [2] because it is a more reliable performance measure. By using the distance to the hyperplane of each sample and varying a decision threshold, we plot hit rate (true positives) against false alarm rate (false positives). The area under this curve is equivalent to the percent of correct detections in a 2-alternative forced choice task, in which the system must choose which of the two images contains the target.

4.2 Detection using neutral face

Results in this section are obtained by computing histogram variations ΔH_i for each sequence i. In this way, features are less dependent to the subject identity. Several kernels have been tested using ΔH_i as input: linear, polynomial, Gaussian Radial Basis Function (RBF) kernel (tuned on the test database) and the HDI kernel. We have also used the histogram intersection kernel K^{hi}. We subtract the neutral face LGBP histogram from the unknown expression LGBP histogram in the implicit feature space ϕ^{hi} of the histogram intersection kernel:

$$\begin{aligned} K_{ij} =& < \phi^{hi}(H_i) - \phi^{hi}(H_i^N), \phi^{hi}(H_j) - \phi^{hi}(H_j^N) > \\ &= K^{hi}(H_i, H_j) - K^{hi}(H_i, H_j^N) - K^{hi}(H_i^N, H_j) \\ &\quad + K^{hi}(H_i^N, H_j^N) \end{aligned} \tag{6}$$

First part of Table 1 reports area under the ROC for those kernel. Best results are obtained for the HDI and the histogram intersection kernel and do not need any tuning. The Gaussian RBF kernel leads to decent results too but was tuned directly on the test database resulting in biased performances.

4.3 Detection without the neutral face

The drawback of the previous approach is the necessity to have at our disposal a neutral face. In this section, we present results where only one expressive image is needed to detect AUs when we test generalization to new subjects. Two approaches have been investigated.

The first is a classical approach where only expressive images are used to train the SVM detectors. In this case we do not exploit the relation between neutral and expressive faces of the same subject to learn our detectors.

Kernel type	Inputs	Upper AUs	Lower AUs
Linear	ΔH_i	95.6%	93.9%
Gaussian RBF	ΔH_i	96.2%	94.9%
3-rd deg poly.	ΔH_i	93.8%	91.0%
Histogram inter.	H_i, H_i^n	96.7%	95.4%
HDI	ΔH_i	**97.3%**	**95.8%**
Linear	H_i	94.7%	93.1%
Gaussian RBF	H_i	95.5%	94.1%
Histogram inter.	H_i	95.8%	94.9%
Linear	$\widetilde{\Delta H}_i$	95.4%	94.1%
Gaussian RBF	$\widetilde{\Delta H}_i$	96.0%	94.9%
HDI	$\widetilde{\Delta H}_i$	**97.0%**	**95.6%**

Table 1. Area under the ROC for different kernels and inputs.

In the second approach, we train our SVM detectors using the histogram difference ΔH_i as in section 4.2 and use $\widetilde{\Delta H}_i$ (which does not require a neutral face) to test on new subjects.

Last two parts of table 1 report results for different kernels using these two approaches. We observe with the first approach that each kernel leads to lower performance using expressive histogram instead of histogram differences. This confirms the fact that histogram difference is less sensitive to inter-subject variations. With the second approach, results are close to those using ΔH_i whatever the kernel. The synthesized neutral histogram is precise enough to remove most part of the sensitivity to identity. Once again, the HDI kernel performs the best.

4.4 Comparison with other methods

Results with HDI kernel using a synthesized neutral face LGBP histogram is compared with two other state-of-the-art methods: [2] and [9]. Despite our efforts to have the same setup, there are small differences that we believe end up being in our disadvantage. In [2], authors report results of AUs detections on the last and the first image (neutral faces in which AU non-detection is easily performed) of the Cohn-Kanade database sequences and on the unpublished Ekman-Hager database. As they do not have the same target/non-target ratio, we compare in Fig. 3a the mean area under the ROC for the 16 AUs our system can detect . In [9] authors use all the images of the Cohn-Kanade database sequences to detect AUs whereas we use only the expression apex images. They only report detection score for 13 AUs. The mean result is compared in Fig. 3b with the mean result we got for the same 13 AUs.

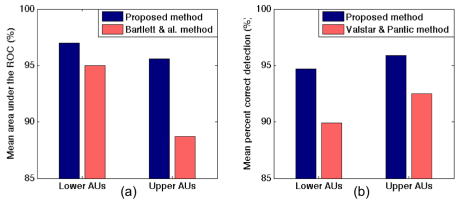

Figure 3. Comparison with state-of-the-art methods.

5 Conclusion

We have presented here an automatic effective AUs detection system. We noticed that LGBP histogram difference leads to better results than expression apex LGBP histogram. To remove the constraint of having at our disposal a neutral face of a subject, we subtracted from the expression apex histogram a synthesized histogram. This latter corresponds to the histogram one should obtain on a neutral face of the subject. The new HDI kernel leads to better results than classical kernels. Generalization results on new subjects within the Cohn-Kanade database outperform those of two state-of-the-art methods. Complexity of the overall system (eye detection, resizing, convolution with Gabor filters, LBP coding and SVM detection) is compatible with real-time applications.

References

[1] A. Barla, F. Odone, and A. Verri. Histogram intersection kernel for image classification. In *ICIP*, 2003.

[2] M. Bartlett, G. Littlewort, M. Frank, I. Fasel, and J. Movellan. Automatic recognition of facial actions in spontaneous expressions. *Journal of Multimedia*, 2006.

[3] P. Ekman and W. Friesen. Facial action coding system (facs): A technique for the measurement of facial actions. *Consulting Psychologists Press*, 1978.

[4] T. Kanade, J. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *FG*, 2000.

[5] G. Littlewort, M. Bartlett, I. Fasel, J. Susskind, and J. Movellan. Automatic system for measuring facial expression in video. *Image and Vision Computing*, 2005.

[6] S. Maji, A. Berg, and J. Malik. Classification using intersection kernel svm is efficient. In *CVPR*, 2008.

[7] S. Moore, R. Bowden, and U. Guildford. The effects of pose on facial expression recognition. In *BMVC*, 2009.

[8] T. Senechal and L. Prevost. Neural network cascade for facial feature localization. In *ANNPR*, 2010.

[9] M. Valstar and M. Pantic. Fully automatic facial action unit detection and temporal analysis. In *CVPR*, 2006.

[10] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang. Lgbphs: A novel non-statistical model for face representation and recognition. In *ICCV*, 2005.