Probabilistic Corner Detection for Facial Feature Extraction

Edoardo Ardizzone, Marco La Cascia, and Marco Morana

Università degli Studi di Palermo
DINFO - Dipartimento di Ingegneria Informatica
Viale delle Scienze - Ed.6 - 3° piano - 90128 Palermo, Italy
{ardizzon,lacascia,marcomorana}@unipa.it

Abstract. After more than 35 years of resarch, face processing is considered nowadays as one of the most important application of image analysis. It can be considered as a collection of problems (i.e., face detection, normalization, recognition and so on) each of which can be treated separately. Some face detection and face recognition techniques have reached a certain level of maturity, however facial feature extraction still represents the bottleneck of the entire process. In this paper we present a novel facial feature extraction approach that could be used for normalizing Viola-Jones detected faces and let them be recognized by an appearance-based face recognition method. For each observed feature a prior distribution is computed and used as boost map to filter the Harris corner detector response producing more feature candidates on interest region while discarding external values. Tests have been performed on both AR and BioID database using approximately 1750 faces and experimental results are very encouraging.

Keywords: Face detection, face recognition, features extraction, CBIR.

1 Introduction

After more than 35 years of resarch, face processing is considered nowadays as one of the most important application of image analysis and understanding. Even though automatic recognition of faces has reached satisfactory results on well constrained tasks, it is still a challenging problem.

Face processing can be considered as a collection of problems, i.e., *face detection*, *facial feature extraction*, *pose estimation*, *face validation*, *recognition*, *tracking*, *modelling* and so on, each of which can be treated separately.

Face recognition often represents the subsequent step of face detection and face normalization processes. Face detection aims to find the image position of a single face so it is usually the first step in any automated face processing system. Appearance-based approaches could then be used to compare detected faces against a database of known individuals in order to assign them an identity. Face normalization is required to support face recognition by normalizing a face for position so that the error due to face alignment is minimized.

P. Foggia, C. Sansone, and M. Vento (Eds.): ICIAP 2009, LNCS 5716, pp. 461–470, 2009.

Some face detection (e.g., Viola-Jones face detector [1]) and face recognition (e.g. eigenfaces [2]) techniques have reached a certain level of maturity, however feature extraction still represents the bottleneck of the entire process.

In this work we present a novel facial feature extraction method to normalize Viola-Jones detected faces and let them be recognized by an appearance-based face recognition approach, e.g. Turk and Pentland eigenfaces.

We started by analyzing the Viola and Jones face detector and we noticed that the use of rectangle features creates some structure on facial features distribution over the detected faces. Thus, all faces are extracted in similar way and each feature locates inside a specific region. Thus, for each feature a prior distribution is computed and used as *boost map* to filter the Harris corner detector response so that thresholding produce a finer corner detection on interest region while discarding other values. Each corner can then be tested using SVMs to detect the presence of a facial feature.

The paper will show the following structure: an analysis of related work will be given (Sect. 2). The Sect. 3.1 will give an overview of Harris corner detector, while the proposed approach is described in Sect. 3.2. Experimental results are shown and discussed in Sect. 4. Conclusions will follow in Sect. 5.

2 Related Work

Automatic face processing for recognition [3] involves at least three different subtasks: *face detection*, *feature extraction*, *face recognition* and/or *verification*.

Up to the early '90s, most face detection algorithms were focused on images with single frontal face and simple backgrounds. A survey of these approaches was written by Samal and Iyengar [4].

Face recognition has received more attention especially in the last 10 years. Recent works based on face appearance train the detection system on large numbers of samples and perform really better than early template matching methods. State-of-the-art face detection tecniques can detect different faces in many poses and in cluttered backgrounds. A relevant survey of early face recognition methods was written by Yang et al. [5].

In this paper we propose a feature extraction technique based on Viola-Jones [1] face detector (VJFD), that is the most stable and used face detector both in academic and commercial systems. This is due to three key contributions: the first is *integral image* computation of *rectangle features* that allows for very fast feature evaluation; the second is an efficient classifier, based on AdaBoost [6], which selects a small number of critical visual features from a larger set; the third is an efficient method for discarding background regions which are unlikely to contain the object of interest.

However, many face recognition systems need facial features location to normalize detected faces avoiding degradation in recognition performance.

Early approaches focused on template matching to detect global features as eyes and mouth [7], while more recent models, i.e., ASM, AFM, AAM, offer more robustness and reliability working on local *feature point* position. Active Shape Model (ASM) [8] extends Active Contour Model [9] using a flexible statistical model to find feature point position in ways consistent with a training set. Active Appearance Model (AAM) [10] combines shapes with gray-level appearance of faces.

Face recognition still attracts researchers from both humanistic and scientific worlds. Zhao et al. [3] classify face recognition methods in *holistic matching methods*, *feature-based matching methods* and *hybrid methods.* In this work we use assume that face recognition step is based on a principal compenent analysis (PCA) technique: eigenfaces [2]. The reason for this choice is that eigenfaces is one of the most mature and investigated face recognition method and it performs well while normalizing faces with respect to scale, translation and rotation. PCA is applied on a training set of face images and eigenvectors (called eigenfaces) are computed. Thus, every face image can be represented as a vector of weights obtained by projecting the image into the "face space" and each new image is verified and identified calculating its distances to face space and to each known class respectively. Experiments performed by Turk and Pentland [2] report approximately 96% correct classification over lighting variation, while performance drops dramatically with orientation (85%) and size(64%) changes. For this reason face normalization is needed.

Berg et. al [11] proposed a *rectification* procedure to move each face image to a canonical frame by identifying five facial feature points (corners of the left and right eyes, corners of the mouth, and the tip of the nose) and then applying an affine transformation. Geometric blur feature [12] is used as input of five SVMs and each point in the entire image is tested to identify features. This approach gives good results, however $M \times N$ points need to be tested, where $M \times N$ is the image size.

In next section we propose a feature extraction method to localize facial features using a prior on location for each feature point and Harris corner detector [13].

3 Methods

3.1 Harris Corner Detector

Several interest point detection techniques have been proposed and evaluated [14], however the Harris corner detector is still one of the most used due to low numerical complexity and invariance to image shift, rotation and lighting variation.

Harris approach relies on the fact that at some image points, *corners*, the image intensity changes largely in multiple directions. Corners are captured by considering the changes of intensity due to shifts in a local window.

Let I is a gray-scale image; consider taking a window W and shifting it by $(\Delta x, \Delta y)$, the auto-correlation function[15] E is defined as,

$$E\left(x,y\right)=\sum_{W}\left(I\left(x_i,y_i\right)-I\left(x_i+\Delta x,y_i+\Delta y\right)\right)^2 \tag{1}$$

where (x_i, y_i) are the points in the gaussian window centered on (x, y).

Approximating $I\left(x_i+\Delta x,y_i+\Delta y\right)$ by Taylor expansion,

$$I\left(x_i+\Delta x,y_i+\Delta y\right)\approx I\left(x_i,y_i\right)+\left(I_x\left(x_i,y_i\right)I_y\left(x_i,y_i\right)\right)\begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix} \tag{2}$$

where $I_x = \frac{\partial I}{\partial x}$ and $I_y = \frac{\partial I}{\partial y}$ denote partial differentiation in x and y, we obtain

$$E\left(x,y\right)=\begin{bmatrix}\Delta x\ \Delta y\end{bmatrix}M\left(x,y\right)\begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix} \tag{3}$$

The matrix $M\left(x,y\right)$ captures the local intensity structure of the image and angle brackets denote summation over W.

Corner detection can be done by analyzing the eigenvalues of M for each point in the image, however this computation is computationally expensive. Harris suggested a measure based on the determinant and trace of M to find corners avoiding eigenvalue decomposition of M

$$\begin{aligned}R_H &= \alpha\beta - k(\alpha+\beta)^2\\ &= Det(M) - kTr^2(M)\end{aligned} \tag{4}$$

where α and β are the eigenvalues of M.

A point $c\left(x,y\right)$ is then detected as corner if $R_H\left(x,y\right)$ is an 8-way local maximum.

3.2 Feature-Based Corner Detection

Harris corner detector performs well on different types of images, however it is not sufficient for facial feature extraction.

We want to obtain a set of points C_P that contains, among others, the true facial features so that each point in C_P can be tested using SVMs to detect facial features.

We started by analyzing Viola-Jones face detector and we noticed that the use of rectangle features creates some structure on facial features distribution over the detected faces. The reason for this is that each Viola-Jones face region is selected using the rectangle features [1] response to facial features (i.e. eyes, nose, mouth) position. Thus, all faces are extracted in similar way and each feature locates inside a well defined region, as shown in Fig. 1.

In order to reduce the computational cost and increase the rate of success of feature classification using SVMs, all points to be tested should represent true feature candidates. However Harris output is a "general-purpose" set of points, therefore many useless corners are detected and necessary ones are frequently missed. The proposed method is based on feature points distribution over

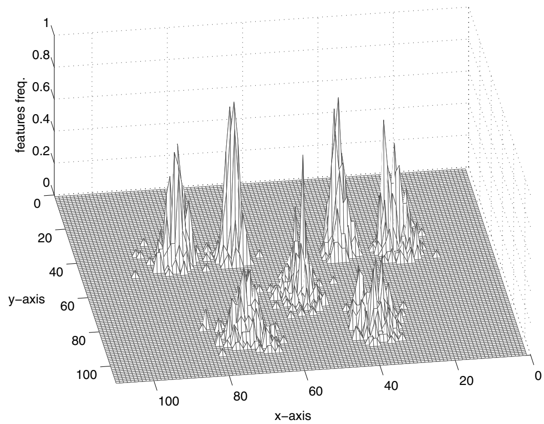


Fig. 1. 3-D distribution of 7 facial feature points over 400 Viola-Jones size-normalized faces (110x110 pixels)

size-normalized Viola-Jones detected faces. For each feature j a prior distribution B_j is used as *boost map* to filter Harris response:

1. reducing the number of corners outside the region of interested
2. increasing the number of corners inside the region of interested

Considering a training set of N face images of size $W \times L$ detected by VJFD, for each feature j the boost map B_j is given by:

$$B_j\left(x,y\right) = \frac{1}{N}\sum_{i=1}^{N} b_{ij}\left(x,y\right) \tag{5}$$

where $1 \leq x \leq W$, $1 \leq y \leq L$ and

$$b_{ij}\left(x,y\right) = \begin{cases} 1 \;\; if \;\; \left(X_{ij} = x\right) and \left(Y_{ij} = y\right) \\ 0 \qquad\qquad otherwise \end{cases} \tag{6}$$

Each point $B_j\left(x,y\right)$ represents the frequency with which observed features of coordinates $\left(X_{ij}, Y_{ij}\right)$ fall in (x,y).

To reduce the dependence from training data, each B_j is approximated by a Thin Plate Spline (TPS) function [16]. Harris response R_H is then filtered using the corresponding *boost map* to obtain the feature-based corner detection.

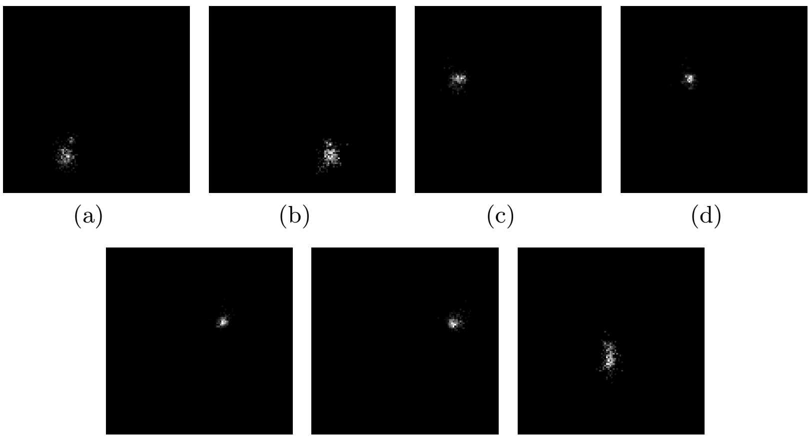
$$R_j = B_j\left(x,y\right) R_H \tag{7}$$

A feature point candidate $c_j\left(x,y\right)$ is finally detected as corner if $R_j\left(x,y\right)$ is an 8-way local maximum.

We refers at B_j as *boost map* since it boost the values in R_H according to the observed distribution of feature j. The values in B_j perturb the structure of R_H so that Harris thresholding produce a finer corner detection on interest region

(a) (b) (c) (d)

Fig. 2. Harris corner detection (a) and (c) compared with boost map-based facial feature detection (b) and (d). Detected corners are marked with x while a circle denotes the true feature position.



(a) (b) (c) (d)

(e) (f) (g)

Fig. 3. Boost maps for right (a) and left (b) corners of the mouth, external (c-f) and internal (d-e) corners of the left and right eye, tip of the nose (g)

while discarding other values. TPS approximation allows to generalize training data while preserving the characteristics of the observed distributions. For this reason experimental results are very promising even using different training and test datasets.

4 Experimental Tests and Results

To enable detailed testing and *boost map* building we used two datasets manually annotated by Tim Cootes' staff [17]. In order to detect feature position for face normalization, 7 feature points have been selected from the 22 facial features available from the AR and BioID face database annotation. Face detection on both AR (359 labelled images) and BioID (1412 labelled images) datasets has been performed and 7 *boost maps* (Fig. 3) have been computed observing the position of the corners of the left and right eyes, corners of the mouth, and the tip of the nose on a 400 images subset.

To validate our method, several tests were conducted on the 1771 faces detected by VJFD. Our goal is to obtain feature candidates as close as possible to true feature point, so that for each feature we computed the corrispondent *boost map*-filtered Harris response. We then compared the known position of each feature with the nearest corner detected by Harris and proposed detector. Three test sessions have been run using different couples of training and test data:

- Test A: 359 images from AR database to build the *boost map* and 1412 BioID images as test set,
- Test B: 400 images from BioID database to build the *boost map* and the remaining 1012 images as test set,
- Test C: 400 images from BioID database to build the *boost map* and 359 AR images as test set.

Results of Test A, B and C are shown in Table 1, Table 2 and Table 3 respectively. Each column contains the number of images in which detected corners falls at distance m from the true feature position using Harris (H_m) and proposed *boost map*-based detector (B_m). We tested for distance $m = 0$, $0 < m \leq 1$, $1 < m \leq 2$, $2 < m \leq 3$, $m > 3$, evaluating the ratio B/H for each m.

Each row contains results for the right (R) and left (L) corners of the mouth (*mR,mL*), external (E) and internal (I) corners of the left and right eyes(*eER, eIR, eLR, eIR*) and tip of the nose (n).

Positive ratio ($B/H > 1$) is obtained for $0 < m \leq 2$, that is the proposed approach performs better than Harris detector finding more corners in the radius of 2 pixels from the considered feature point.

Table 1. Test A - AR training (359 images) and BioID testing (1412 images). Results for Harris (H) and proposed approach (B) using 7 feature points.

	H_0	B_0	B/H	H_1	B_1	B/H	H_2	B_2	B/H	H_3	B_3	B/H	H_n	B_n	B/H
mR	62	168	2,71	813	804	0,99	466	421	0,90	60	18	0,30	11	1	0,09
mL	157	144	0,92	716	783	1,09	479	482	1,01	46	3	0,07	14	0	0
eER	29	127	4,38	77	286	3,71	413	755	1,83	468	216	0,46	425	28	0,07
eIR	47	83	1,77	109	228	2,09	493	707	1,43	382	364	0,95	381	30	0,08
eIL	18	117	6,50	63	210	3,33	366	673	1,84	482	261	0,54	483	151	0,31
eEL	27	92	3,41	130	269	2,07	641	719	1,12	390	268	0,69	224	64	0,29
n	9	71	7,89	143	259	1,81	374	711	1,90	349	230	0,66	537	141	0,26

Previous tests indicate system performance referring to the number of images in which a corner is found at distance m from the true feature position, while Fig. 4 shows previous values normalized to the number of corners detected by Harris (N_H) and proposed (N_B) method for each test set.

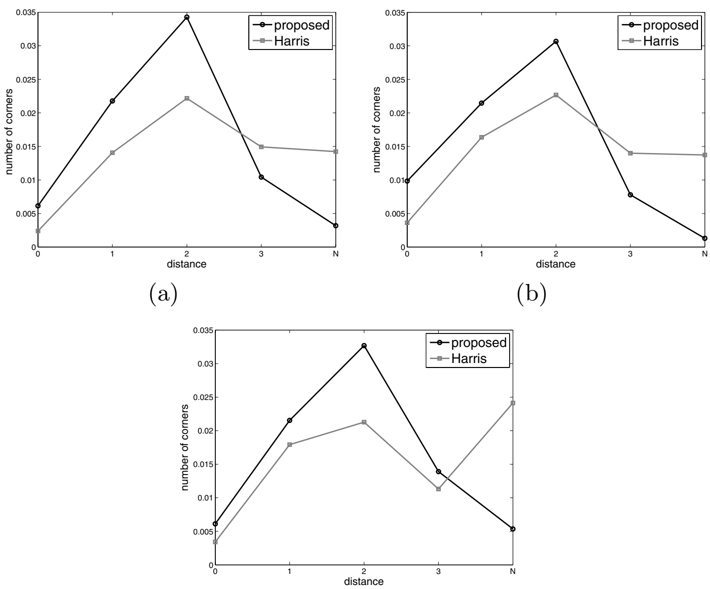
Experimental results showed that our *boost map*-based facial feature detector performs generally better than general purpose Harris corner detector. Even using different training and test data results are stable showing that each *boost map* attains an adequate level of generalization apart from used training data.

Table 2. Test B - BioID training (400 images) and BioID testing (1012 images). Results for Harris (H) and proposed approach (B) using 7 feature points.

	H_0	B_0	B/H	H_1	B_1	B/H	H_2	B_2	B/H	H_3	B_3	B/H	H_n	B_n	B/H
mR	88	194	2,20	618	550	0,89	274	262	0,96	28	6	0,21	3	0	0
mL	141	201	1,42	535	632	1,18	307	177	0,58	26	2	0,09	3	0	0
eER	37	129	3,50	82	199	2,43	338	500	1,48	323	150	0,47	232	33	0,14
eIR	49	103	2,11	120	183	1,52	369	481	1,30	243	198	0,82	232	47	0,20
eIL	31	115	3,67	83	172	2,08	274	509	1,86	328	190	0,58	297	26	0,09
eEL	16	134	8,38	115	177	1,54	421	553	1,31	287	139	0,48	173	9	0,05
n	2	106	45,29	94	228	2,42	300	574	1,92	174	90	0,52	442	14	0,03

Table 3. Test C - BioID training (400 images) and AR testing (359 images). Results for Harris (H) and proposed approach (B) using 7 feature points.

	H_0	B_0	B/H	H_1	B_1	B/H	H_2	B_2	B/H	H_3	B_3	B/H	H_n	B_n	B/H
mR	21	33	1,57	201	207	1,03	121	100	0,83	12	7	0,58	4	12	3,00
mL	20	37	1,85	206	188	0,91	122	129	1,06	7	4	0,57	4	1	0,25
eER	23	28	1,22	27	65	2,41	78	204	2,62	57	48	0,84	174	14	0,08
eIR	4	13	3,25	8	26	3,25	23	109	4,74	84	165	1,96	240	46	0,19
eIL	2	6	3,00	12	26	2,17	31	108	3,48	99	136	1,37	215	83	0,39
eEL	18	37	2,06	34	72	2,12	140	184	1,31	51	53	1,04	116	13	0,11
n	22	39	1,77	89	96	1,08	170	198	1,16	54	26	0,48	24	0	0



(a) (b)

(c)

Fig. 4. Results for Harris and proposed approach normalizing to average number of corners detected for each image.(a) Test A: $N_H = 145796$, $N_B = 130375$. (b) Test B: $N_H = 100636$, $N_B = 99655$. (c) Test C: $N_H = 32194$, $N_B = 31563$.

Harris method detects more corners than proposed approach, moreover Harris corners are distributed over full image area while we detect corners just in features region as shown in Fig. 2. Thus, boost maps improve the SVMs point classification both reducing the number of points to be tested and increasing the quality/importance of those points.

Each face image is fully processed (i.e., detected, analyzed, recognized) in about 3 seconds using Matlab on a conventional 2,4 GHz Intel Pentium 4, however it does not represent a limit to the efficiency of Viola-Jones approach, since *face detection* is conceptually distinct from other face processing steps. Face detection aims to find the image position of a single face so it is anyway the first, necessary, step in any automated face processing system. The efficiency of Viola-Jones technique is required to quickly detect a face while discarding other regions, however detected faces need to be processed again to perform subsequent tasks, e.g., face recognition, face tracking, face modelling and so on. Thus, even if the proposed technique is not suitable for a real-time system, the computational cost is not prohibitive for online image analysis (e.g. image annotation, classification, recognition, etc.).

5 Conclusions

In this work was addressed the task of detecting facial featuers to normalize detected faces and perform face recogniton. We presented a novel facial feature extraction approach that could be used for normalizing Viola-Jones detected faces noticing that rectangle features creates some structure on features position over the detected faces. For each observed feature a prior distribution has been computed and used as *boost map* to filter the Harris corner detector response producing a finer corner detection on interest region while discarding external values. Experimental results are very promising using both AR and BioID face database. Further consideration could be done testing the proposed approach on natural images (i.e., personal photo album), however ground truth is needed and no more labelled datasets are currently available. This will be subject of future work.

References

1. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. CVPR 2001, vol. 1, pp. 1:I–511–I–518 (2001)
2. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience 3(1), 71–86 (1991)
3. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey. ACM Comput. Surv. 35(4), 399–458 (2003)
4. Samal, A., Iyengar, P.A.: Automatic recognition and analysis of human faces and facial expressions: a survey. Pattern Recogn. 25(1), 65–77 (1992)

5. Yang, M.-H., Kriegman, D.J., Ahuja, N.: Detecting faces in images: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(1), 34–58 (2002)
6. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. In: Vitányi, P.M.B. (ed.) EuroCOLT 1995. LNCS, vol. 904, pp. 23–37. Springer, Heidelberg (1995)
7. Yuille, A.L., Cohen, D.S., Hallinan, P.W.: Feature extraction from faces using deformable templates. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1989. Proceedings CVPR 1989, June 1989, pp. 104–109 (1989)
8. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models—their training and application. Comput. Vis. Image Underst. 61(1), 38–59 (1995)
9. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. International Journal of Computer Vision 1(4), 321–331 (1988)
10. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. In: Burkhardt, H., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1407, pp. 484–498. Springer, Heidelberg (1998)
11. Berg, T.L., Berg, A.C., Edwards, J., Maire, M., White, R.: Names and faces in the news. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004, vol. 2, pp. II–848–II–854 (2004)
12. Berg, A.C., Malik, J.: Geometric blur for template matching. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, vol. 1, pp. 1:I–607–I–614 (2001)
13. Harris, C., Stephens, M.: A combined corner and edge detection. In: Proceedings of The Fourth Alvey Vision Conference, pp. 147–151 (1988)
14. Schmid, C., Mohr, R., Bauckhage, C.: Evaluation of interest point detectors. International Journal of Computer Vision 37(2), 151–172 (2000)
15. Derpanis, K.G.: The harris corner detector (2004)
16. Duchon, J.: Spline minimizing rotation-invariant semi-norms on sobolev spaces. Lecture Notes in Math., vol. 571, pp. 85–100 (1977)
17. Cootes, T.F.: `http://personalpages.manchester.ac.uk/staff/timothy.f.cootes/`