Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 13 (2012) 162 – 170

Procedia Computer Science

Proceedings of the International Neural Network Society Winter Conference (INNS-WC 2012)

A set of criteria for face detection preprocessing

Hossein Ziaei Nafchi[a,]*, Seyed Morteza Ayatollahi

[a]*Nafch Image Processing Research Group, Nafch, Shahrekord, Iran*

Abstract

The goal of this paper is to provide a robust set of preprocessing steps to be used with any face detection system. Usually, the purpose of using preprocessing steps in face detection system is to speed up the detection process and reducing false positives. A preprocessing step should reject an acceptable amount of non-face windows. First proposed criterion is based on linear image transform (LIT) which ignores scanning a number of non-face windows. Second criterion utilizes regional minima (RM) to reject non-face windows. The last one uses a modified adaptive thresholding (ADT) technique to convert input image into a binary representation and perform an exclusion process on the latter form. The proposed criteria have been used in conjunction with a version of Viola-Jones face detector. Experimental results show significant advantage against early exclusion criterion or variance classifier in terms of speed and rejection rate. CMU-MIT and BioID datasets have been used in the experiments.

Keywords: face detection, adaptive thresholding, linear image transform, regional minima, adaboost.

1. Introduction

Detecting faces in image have been studied for two past decades and thousands of methods have been proposed for solving the problem of face detection [1,2]. Among them, boosting based algorithms show high-performance face detection systems [2,3,4], and after utilizing by Viola and Jones [5], widely used in face detection literature. Usually, the speed of boosting based methods is higher than other proposed face detection methods. For example, benchmark algorithms such as Rowley et al [6] and Schneiderman et al [7] are many times slower than VJ method. Other studies with comparable results [8,9,10] also achieved slower process time. The reason is to utilize integral image and a cascade of weak classifiers as a final strong classifier. In the

* Corresponding author. Tel.: +98-913-282-1519; fax: +98-381-222-1902.
E-mail address: hossein_zi@yahoo.com.

1877-0509 © 2012 Published by Elsevier B.V. Selection and/or peer-review under responsibility of Program Committee of INNS-WC 2012
Open access under CC BY-NC-ND license. doi:10.1016/j.procs.2012.09.125

early stages of the cascade, most of the windows will be rejected yielding to high speed face detection system. Also many of the face detection systems uses one or more preprocessing steps. A number of these works are described in the related works section. This shows the need of using preprocessing steps in face detection systems, which can reject the most of the non-face windows.

This paper introduces a set of criteria, which can eliminate an acceptable number of non-face windows while no face is lost. LIT is a criterion which can eliminate 36 to 60% of non-face windows. This approach uses a skip process resulting to scan a small portion of non-face windows. ADT is a robust technique to represent face images. ADT criterion can eliminate 75 to 85% of non-face windows. RM is a robust technique for non-face windows exclusion as it was able to eliminate about 50% of windows in images with different complexity. A cascade containing ADT, RM and the LIT can reject 89 to 92% of non-face windows in images. The proposed criteria are fast enough for the real time applications. First, each criterion is introduced and then the rules resulting in acceptable rejection rate and optimum detection rate are mentioned.

The rest of the paper is organized as follows. In section 2, related works are discussed. A method for non-face windows rejection based on LIT is proposed in section 3. Learning process used during feature selection for RM and ADT criteria is discussed in section 4. The use of RM as a preprocessing step is shown in section 5. Section 6 elaborates proposed ADT method for binary representation of face images and shows how to use it as a preprocessing step. Section 7 deals with the experimental results and performance of each criterion and Section 8 draws a conclusion.

2. Related works

Usually every face detection system uses a preprocessing step to achieve better performance specially to reduce the computational efficiency and number of false positives. Rowley et al [6] used a fast neural network to preprocess the image by choosing the face candidates and a slower network as a final detector. First network has high detection rate but also high false positives. Second network has low detection rate and small number of false positives in comparison with the first network. Elad et al [11] utilized an iterative rejection based classification algorithm to reject non-face windows at the end of the each iteration. They reported that first stage of classification iteration can reject up to 90% of non-face windows. However, they did not reported results of their proposed method for the standard datasets. Usually, the term “Coarse-to-fine” is used in the face detection systems which utilize preprocessing steps [9, 10, 12, 13].

Liu [8] introduced a single response criterion and the early exclusion criterion to reject acceptable rate of the non-face windows to achieve better computational efficiency. These criteria have been used in [9] and early exclusion criterion (EEC) also has been used in [10]. EEC uses a heuristic procedure to eliminate windows, which could not be faces at all. EEC divides a window into five regions as shown in Fig. 1 [9]. EEC first computes the variances in regions E and D respectively. If either values are lower than a predefined threshold, the corresponding window will be excluded immediately. Then m_a and m_b will be computed, where m_a is the average intensity of those pixels which are darker than the average intensity of the region A. m_b is the average intensity of those pixels which are brighter than the average intensity of the region B. A window will be excluded if $m_b < k^* m_a$. Then m_c just like m_a will be computed and the window excluded if $m_b < k^* m_c$.

Because EEC widely used in the previous works [8,9,10], the proposed criteria are compared with EEC in terms of detection rate, false alarm rate and speed. Results can be found in section 7.

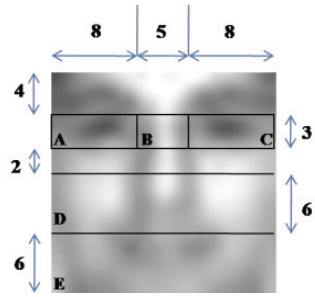

Fig. 1. Labeled regions in a 21 × 21 window.

3. Linear image transform

An unsupervised approach is proposed for face detection preprocessing. In the proposed method, pixel intensities in the input image will be transformed into *L* intensity levels linearly:

$$LIT(x,y) = \left\lfloor \frac{I(x,y)}{1/L} \right\rfloor \tag{1}$$

where, the values of input image, $I(x,y)$, are real numbers between $0-1$ and $\lfloor z \rfloor$ is the operator to return the largest integer not bigger than z. The values of $LIT(x,y)$ are integers within $0 \sim L-1$. The result is that those pixels with close intensities values in the input image will form a flat region in the *LIT* image, if they are adjacent to each other. As we see in figure 1, a face region has the non-homogenous vertical lines. The reason is that an arbitrary vertical line in a face window partially lies on at least one of the face features, e.g. eyebrows, eyes, nose, mouth and edges of face. Therefore, a flat line refers to a non-face window.

By considering 21×21 reference model in figure 1, let $l = \{l_4, l_5, \dots, l_{17}, l_{18}\}$ be a set of vertical lines in the reference model, where l_i denotes the *i-th* vertical line in this reference and $l_i = \{p_{i1}, p_{i2}, \dots, p_{i20}, p_{i21}\}$, where p_{ij} denotes the *j-th* pixel of *i-th* vertical line in the reference. Also, vertical lines 1 to 3 and 19 to 21 ignored in order to prevention from possible background effect on the face windows. A window is excluded if equation 1 satisfies for at least one vertical line:

$$\exists\, i \in \{4,5,\dots,17,18\} \mid variance(l_i) = 0 \tag{2}$$

where, variance is computed by using two integral images. First one is the integral image of the linearly transformed image, and second one is the integral image of squared intensities of linearly transformed image. The following equation had been used for computing the variance. Where σ and μ are the standard deviation and mean and x_i are the pixels in the image.

$$\sigma^2 = \left| \mu^2 - \frac{1}{N}\sum_{i=1}^{N} x_i^2 \right| \tag{3}$$

To speed-up the rejection process, two strategies used. First, if l_i causes the rejection of current window, then $(i-3)$ windows in the right side (according to search process) of this window also are ignored. Second, if a window not excluded by checking its $\{l_4, l_5, \dots, l_{17}, l_{18}\}$ vertical lines, it is efficient to checking only *18-th* line of next window.

Conversion to LIT form is straightforward. In the experiments, in order to achieving optimum detection rate, the following conditions used in formation of LIT:

Number of levels (L): 10 Level 1: pixels lower than 1.1×min(input pixels) Level 10: pixels higher than 0.8×max(input pixels)

4. Learning process

Proposed criteria are based on LIT, RM and ADT, where RM and ADT work on the binary images. Therefore for learning RM and ADT criteria, a thresholding process shown in figure 2, instead of using adaboost algorithm, is used. We first used adaboost for learning, but it produces many features for rejecting about 90% of non-face windows. Instead, by using a set of predefined features and a thresholding process, approximately the same rejection rate obtained while the number of features used are small. The rules obtained in sections 5 and 6 for RM and ADT criteria show that haar-like features used in learning process are different from traditional haar-like features [5]. This type of haar features are manually chosen due to the reference models shown in Fig. 5 at the end of section 6.2.

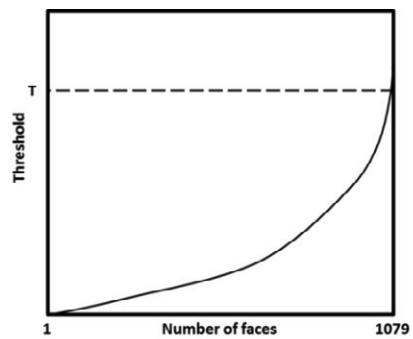

Fig. 2. The threshold learning process used in our experiments.

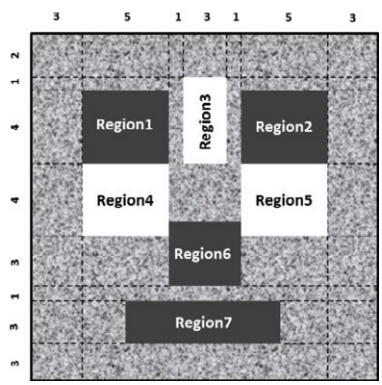

Fig. 3. A 21 × 21 reference model used in our experiments. Black regions are left and right eyes (regions 1, 2), nose (6) and mouth (7). White regions are bottom left eye (5), bottom right eye (6) and between eyes (3).

Furthermore in learning RM and ADT criteria, a 21 × 21 reference model which is shown in Fig. 3 is used. The images used for learning process are 1079 face images from AR dataset [14], of course resized to the size of 21 × 21. This reference model is compatible with the reference model shown in Fig. 1 which applied for

LIT criterion.

5. Regional extrema

Two classes of regional extrema are regional minima and regional maxima. Image minima and maxima are important morphological features [15]. Regional minima are flat zones, which are connected to the larger values while regional maxima are flat zones, which are connected to the smaller values. The regional extrema of an image are defined as union of its regional minima and maxima.

5.1. Regional minima

A regional minimum M of an image f at elevation t is a connected component of pixels with the value t whose external boundary pixels have a value strictly greater than t [15]:

$$\begin{cases} \forall p \in M, & f(p) = t, \\ \forall q \in \delta^{(1)}(M) \backslash M, & f(q) > t. \end{cases} \tag{4}$$

where, $\delta^{(1)}(M)$ denotes the elementary dilation of M and equation 5 defines the difference between two images:

$$\delta^{(1)}(M) \backslash M = \delta^{(1)}(M) \cap \bar{M} \tag{5}$$

Regional minima repeat in the face features such as eyes, mouth and nose. This pattern has been used for pre-processing. A threshold learning process described in section 4 is used. By considering the reference model in Fig. 3, the following rules obtained from threshold learning:

Rule 1: $Regions(1 + 2 + 6 + 7) > k \times W$, where $0 < k < 0.5$ and W denotes the pixels lying in the whole reference model shown in figure 2.
Rule 2: $Region3 >= \max(Region1, Region2)$.

6. Preprocessing by modified adaptive thresholding

A modified version of an adaptive thresholding (ADT) scheme introduced by Bradley and Roth [15] is proposed and used in this paper. We first introduce the proposed ADT method and then use it to preprocess input images in face detection application.

6.1. The proposed face representation method

The most thresholding methods choose a fixed threshold value and compare each pixel value with that. A fixed thresholding method is not robust under illumination changes and fails in such an environment conditions. ADT is a technique to handle illumination variations. Bradley & Roth [16] modified ADT method introduced by Wellner [17] by using the integral image. We improved this approach to obtain a representation of the images containing faces. For each pixel, we compute the average of SxS surrounding pixels of that pixel and compare value of that pixel with the product of the obtained average with a coefficient C. The proposed coefficient is computed with the following equation:

$$C = 0.95 + \frac{|\mu - \sigma|}{1000} \tag{6}$$

where, μ and σ are the average and standard deviation of intensities of input image. A pixel is set to 0 (dark) if the value of that pixel is smaller than the product of S mean values and C, otherwise pixel is set to 1 (white). Wellner [17] and Bradley et al [16] choose the value of the C as 0.85 and number of S as a $1/8$ of image length. Furthermore, Bradley & Roth suggested that for different applications one can use a different C. Number of surrounding pixels S must be chosen due to the fact that face features such as eyes, nose and mouth have no relation with background pixels. Therefore, we choose the number of surrounding pixels a small constant. Suppose that we have an image with high intensity values, ADT may fail because pixel value usually becomes more than surrounding pixels and maybe set to 1 erroneously. In images with low intensity values the same scenario repeated in setting pixels with 0 (dark). The proposed method interfere the mean and standard deviation in the coefficient to overcome this problem.

We compared our proposed representation of face images with previous works. It should be noticed that a fair comparison cannot be performed since these approaches proposed to handle the images containing character information or augmented reality. We just make an example to show that coefficient C which mentioned in equation 1 works well. In Fig. 4 four approaches compared. Proposed method scale binary images obtained from input image to find faces in images with different scales. A problem with Wellner approaches is that they did not highlight face features completely. Therefore, sometimes scaling images result in missing face features. Fig. 4 qualitatively shows that our proposed method makes a better highlight on the face features.

Fig. 4. An input image includes a face. a) Bradley and Roth representation [16], b) Wellner representation with a 4×4 Gaussian filter [17], c) Wellner representation with a 4×4 Median filter [17], d) Our representation. This figure shows advantage of the proposed face representation method (d) which is a modified version of Bradley and Roth method (a).

6.2. Preprocessing phase

For preprocessing, we convert arbitrary input images into binary images by using the proposed representation method. Then the reference model shown in figure 3 is used. Following rules obtained from the threshold learning process mentioned in section 4:

Rule 1: $(Region1 + Region2) > (Region3 + Region4)$.
Rule 2: $W > T1$ and $W < T2$, where W denotes the summation of pixels in whole window.
Rule 3: $(Region1 + Region2) > T3$.
Rule 4: $Regions(1 + 2 + 6 + 7) > k \times W$, where $0 < k < 0.5$.

Fig. 5 shows the distribution of pixels obtained from 1079 images in the AR dataset for ADT, RM and regional maxima.

a b c

Fig. 5. Normalized reference models obtained from learning process. (a) Normalized reference model obtained from ADT, (b) RM and (c) regional maxima. A brighter pixel shows the more frequent pattern of that pixel.

7. Experimental results

The proposed criteria are tested on two benchmark datasets CMU-MIT [18] and BioID [19]. First dataset contains 130 images with 507 labeled faces, and BioID contains 1521 images with 1521 faces. A total number of 53,395,311 windows had been checked in the CMU-MIT dataset. In this paper, we compute the average time for scanning $42K$ windows with the following equation:

$$Average\ time = (TT \times 42K)/TW \tag{7}$$

where, TT is total time elapsed for scanning all of the windows (TW) in each dataset. It's clear that using equation 7 to report the search time for an arbitrary approach is better than reporting for a single image. Intel Pentium 4, 2.0 GHz Celeron with 2GB of RAM used in our experiments. Average time includes input image(s) scaling, converting to the desired form, computing integral image(s) and passing rules. LIT criterion highly depends on the image complexity. We achieved 80 fps to convert images in the BioID dataset (286×384). ADT criterion has better performance among other criteria which evaluated in this paper. Furthermore, we achieved 56 fps and 30 fps to convert images in the BioID dataset into ADT and RM representation respectively. A good attribute of the RM criterion is its low variation in terms of rejection rate over images with different complexity. Regional minima is the slowest criterion according to its slow conversion process. Figure 6 provides the ROC curve for the proposed criteria separately and in a cascade.

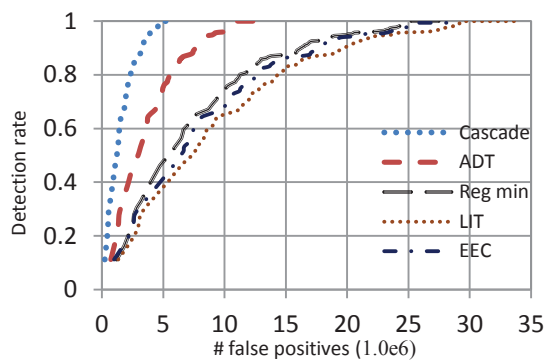

Fig. 6. ROC for the proposed criteria and the EEC on the CMU-MIT dataset. Cascade includes ADT, RM and LIT respectively.

Two integral images used for computing the mean and variance in EEC. Also in our experiments the values for variances and k^* has been set with 64 and 0.9, respectively. These values determined by a thresholding process. By using these values, 3 faces have been lost in the CMU-MIT dataset. Also, RM and ADT criteria

lose one face, while LIT criterion lost another face. Totally, proposed criteria failed to detect two faces in CMU-MIT dataset. The proposed preprocessing criteria show significant advantage against EEC in terms of detection rate, rejection rate and speed. The first part of EEC is fast enough but the second part is not. Therefore, still variance classifier part of EEC can be a proper preprocessing step. Table 1 provides the performance of the proposed criteria and EEC.

Table 1. The Experimental result comparison between EEC [8,9,10] and the proposed criteria. Cascade consisted of ADT, LIT and RM critera, respectively.

Dataset	Rejection rate (%)					Average time (ms)				
CMU-MIT	EEC	LIT	ADT	RM	Cascade	EEC	LIT	ADT	RM	Cascade
	44.19	36.46	76.18	47.98	89.82	126	16	17	26	37
BioID	54.81	59.31	87.18	49.81	92.27	108	16	17	28	35

In order to using the proposed criteria in conjunction with a face detection system, we learned Viola-Jones face detector by using 6000 face images and 30000 non-face images for each stage. The non-face samples were chosen randomly from scene category dataset [20]. The cascade architecture consist of 14 stages, each rejecting at least 50% of non-face images, while totally 57 faces rejected during learning. We put the proposed pre-processing criteria at the first step of Viola-Jones detector. Table 2 shows the detection/rejection results with/without using the proposed pre-processing criteria in conjunction with Viola-Jones face detection system.

Table 2. The effect of using proposed preprocessing criteria in conjunction with Viola-Jones detector.

	Detection rate (%)		Rejection rate (%)	
Dataset	VJ	Pre + VJ	VJ	Pre + VJ
CMU-MIT	85.51	85.51	99.9991965	99.9992845
BioID	92.65	93.07	99.9999712	99.9999801

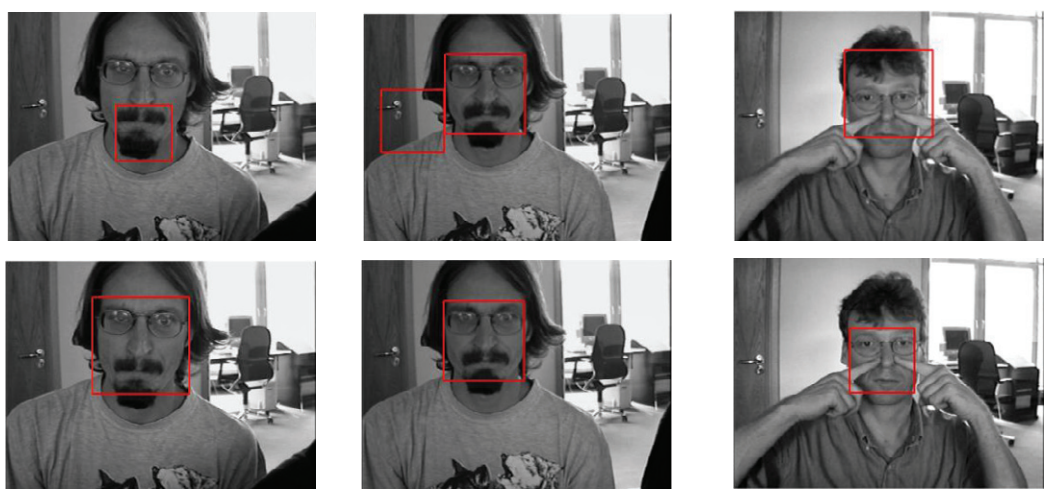

Fig. 7. Top images show the detection results of Viola-Jones detector, while bottom images show the effect of using pre-processing criteria in conjunction with the Viola-Jones detector. Images are from BioID dataset.

We noticed that Viola-Jones detector itself is faster without using the proposed criteria. However, it receives better false alarm rate and even better detection rate. While the proposed pre-processing steps reject about 90%

of non-face windows, results show that proposed steps can improve the performance of other detectors. Figure 7 shows the effect of using proposed criteria in conjunction with a face detector.

8. Conclusion

In this research, we proposed a set of criteria to be used with any face detection system. Linear image transform (LIT) was able to reject acceptable non-face windows in an image, especially when the input image has low complexity. A method based on regional minima have been proposed to reject non-face windows in images, which shown an acceptable prune over non-face windows independent of image complexity. We also proposed a new representation of face images and use it as a preprocessing step in face detection task. A comparison has been made with the early exclusion criterion (EEC). The proposed criteria not only can speed up slow face detection systems, but also have enough potential to reduce the number of false positives in any face detection system.

References

[1] Yang MH, Kriegman D, Ahuja N. Detecting faces in images:A survey. IEEE Trans. Pattern Analysis and Machine Intelligence; vol. 24, no. 1, 2002, p. 34–58.
[2] Zhang C, Zhang Z. Boosting-Based Face Detection and Adaptation. Synthesis Lectures on Computer Vision, Morgan & Claypool Press; 2010.
[3] Saberian MJ, Vasconcelos N. Boosting Classifier Cascades. Neural Information Processing Systems; 2010
[4] Masnadi-Shirazi H, Vasconcelos N. Cost-Sensitive Boosting. IEEE Trans. Pattern Analysis and Machine Intelligence; vol. 33, no. 2, 2011, p. 294–309.
[5] Viola P, Jones M. Robust Real-Time Face Detection. International Journal of Computer Vision; 57(2), 2004 p. 137–154.
[6] Rowley H, Baluja S, Kanade T. Neural Network-Based Face Detection. IEEE Trans. Pattern Analysis and Machine Intelligence; vol. 20, no.1, 1998, p. 23–38.
[7] Scheiderman H, Kanade T. A statistical method for 3D object detection applied to faces and cars. in International Conference on Computer Vision; 2000, p. 746–751.
[8] Liu, C. A Bayesian discriminating features method for face detection. IEEE Trans. Pattern Analysis and Machine Intelligence; vol. 25, no. 6, 2003, p. 725–740.
[9] Shih P, Liu C. Face detection using discriminating feature analysis and Support Vector Machine. Pattern Recognition; vol. 39, 2006, p. 260–276.
[10] Tsao WK, Lee AJT, Liu YH, Chang TW, Lin HH. A data mining approach to face detection. Pattern Recognition; vol. 43, 2010, p. 1039–1049.
[11] Elad M, Hel-Or Y, Keshet R. Rejection based classifier for face detection. Pattern Recognition Lett; vol. 23, 2002, p. 1459–1471.
[12] Sahbi H, Boujemaa N. Coarse-To-Fine Support Vector Classifiers for Face Detection. in International Conference on Pattern Recognition; 2002.
[13] Fleuret F, Geman D. Coarse-to-Fine Face Detection. International Journal of Computer Vision. vol. 41, 2001, p. 85–107.
[14] Martínez AM, Benavente R. The AR Face Database. CVC Technical Report; no.24, 1998.
[15] Soille P. Morphological Image Analysis, Principles and Applications. Springer press; 2007
[16] Bradley D, Roth G. Adaptive thresholding using the integral image. Journal of graphic tools; vol. 12, no. 2, 2007, p. 13–21.
[17] Wellner PD. Adaptive thresholding for the digitaldesk. Tech. Rep. EPC–110, 1993.
[18] http://vasc.ri.cmu.edu/idb/html/face/frontal_images
[19] http://www.bioid.com
[20] Lazebnik S, Schmid C, Ponce J. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. in Computer Vision and Pattern Recognition; 2006, p. 2169–2178.