MNK Publication

International Journal of Latest Research in Science and Technology **ISSN (Online):2278-5299**
Volume 4, Issue 3: Page No.8-18, May-June 2015
https://www.mnkpublication.com/journal/ijlrst/index.php

ENSEMBLE OF FACE/EYE DETECTORS FOR ACCURATE AUTOMATIC FACE DETECTION

[1]Loris Nanni, [2]Alessandra Lumini, [3]Sheryl Brahnam
[1] Department of Information Engineering at the University of Padua, Padua, Italy
[2] DISI, University of Bologna, Cesena, Italy
[3] Computer Information Systems, Missouri State University, USA

Abstract*-In this work we propose a simple yet effective face detector that combines several face/eye detectors that possess different characteristics. Specifically, we report an extensive study for combining face/eye detectors that results in a final system we call FED that combines three face detectors that extract regions of candidate faces from an image with two approaches for eye detection: the enhanced Pictorial Structure (PS) model for coarse eye localization and a new approach proposed here (called PEC) that provides precise eye localization. PEC is an ensemble that utilizes three texture descriptors: multi-resolution local ternary patterns, local phase quantization descriptors, and patterns of oriented edge magnitudes. The extracted features are coupled with support vector machines trained on eye and non-eye samples to perform classification. The proposed framework for face detection could be considered an ad hoc integration of existing methods (the three face detectors and the PS coarse eye detector) that is combined with the proposed novel ensemble for precise eye localization (PEC). The aim of this approach is to maximize performance (not computation time). The quality of the proposed system is validated on three datasets (the well-known BioID and FERET datasets as well as a self-collected dataset). To the best of our knowledge, our system is one of the first fully automatic face detection approaches to obtain an accuracy of almost 100% on the BioID dataset (the most important benchmark dataset for frontal face detection) and 99.1% using the same dataset with only 12 false positives. A MATLAB version of our complete system for face detection can be downloaded from https://www.dei.unipd.it/node/2357.

Keywords - Eye detection; face detection; texture descriptors; local phase quantization; feature combination; support vector machine

I. INTRODUCTION

Over the last few years, face detection has become an essential task in the modern world [35][37][39]. The increasing importance of face detection is due to the widespread development of surveillance and security systems and to a wide range of new applications, such as face tagging, behavioral analysis, human-computer interaction, content-based image and video indexing, and many others [1]. Moreover, detecting faces in images is a necessary first step for facial analysis algorithms, including, for example, face alignment, face recognition/verification, head tracking, and facial expression recognition.

The face detection problem can be described as follows: given an arbitrary image, determine the position of all existing faces in an image by inscribing an ellipsoid (or a bounding box) around each face. From a machine learning point of view, the problem can be formulated as a two-class pattern recognition problem where each subwindow of a given image is classified as either containing or not containing a face [2].

Although great strides have been made in the field of face detection and several approaches are now available that produce accurate detection under variable conditions [3], it is still challenging at the present time to obtain reliable face estimation in unconstrained images, where face detection is arguably more difficult due to the existence of extreme facial poses. In order to improve face detection, additional adjunctive modules have been developed that aim at detecting specific facial features [4]: the nose, eyes, and mouth. Of course, facial feature detection is useful in a number of other applications. Mouth detection, for example, allows applications to capture lip movements [1] for lip reading and video sound synchronization, and eye detection is also important in eye tracking, facial expression recognition, and face modeling. Eye detection is especially useful in face recognition, with several studies reporting a close relation between accuracy in face recognition and eye localization [25][5][34][36][38]. Indeed, high eye detection accuracy is considered crucial in face recognition for avoiding the so-called "curse of misalignment" (i.e., the abrupt degradation of face recognition performance due to possible inaccuracy in the automatic localization of eyes) [6].

Automatic eye detection involves two tasks: (i) assessing the presence of eyes in an image and (ii) accurately detecting the position of the iris center. Although there have been significant advances in the development of precise eye localization methods, this remains an open problem. This problem is challenging because the appearance of eyes is highly variable, both in terms of the intrinsic dynamic features of eyes and ambient environmental changes, a challenge that is intensified because of the increasing demands for speed and accuracy with a margin of error of a few pixels in eye localization if it is to fit the needs of real-world applications.

Research in automatic eye detection can be divided into three major categories [19]: template-based methods, appearance-based methods, and feature-based methods. Template-based methods, as in [30], employ a generic eye model to search for eyes in images. Appearance-based

Publication History
Manuscript Received : 16 April 2015
Manuscript Accepted : 26 May 2015
Revision Received : 25 June 2015
Manuscript Published : 30 June 2015

methods locate eyes based on their photometric appearance using various global representations and statistical classification techniques (e.g., modular eigenspaces and HMM-based algorithms) [3]. Feature-based methods include approaches that extract discrete local features and that perform eye localization by means of standard pattern recognition techniques [5]; in other words, a classifier is trained using these features. A recent feature-based method of note is [29], which presents a new set of features called Feature Local Binary Pattern (FLBP) that combines the information of local textures and of local features (edges, Gabor wavelet features, color features, etc.) using the Local Binary Pattern (LBP) encoding scheme of the grey-level image and the binary representation of the selected feature.

Most early approaches developed for automatic eye detection focused on indoor environments. Because these approaches were unable to handle the extreme changes over time in light intensity levels and the direction of lighting (changes that are typically unavoidable in outdoor environments), they failed when transposed to outdoor scenes. To better address changing illumination conditions, recent studies have focused on illumination-invariant eye detection. Tan [26], for example, proposed a method based on an enhanced pictorial structure model that uses an Support Vector Machine (SVM) classifier to optimize the energy function of the model, and Jung et al. [27][28] adopted a method for illumination normalization based on retinex theory.

The aim of this paper is to produce an efficient and accurate face detection system that combines eye detection algorithms with state-of-the-art face detection methods. Specifically, we improve our previous work [20] by proposing a two-step system that combines three face detectors and two approaches for eye detection: a coarse method that is then followed by a novel precise method for eye localization that is proposed in this paper. The three face detectors are combined to extract regions of candidate faces from an image and are based on a split up sparse network of winnows (SN) [10], Viola-Jones (VJ) [22], and Eye Mouth Localization (EML) [5]. The two eye detectors are based on the enhanced Pictorial Structure (PS) model (a robust approach for eye localization in unconstrained conditions) for coarse localization and on our proposed ensemble of classifiers for precise eye detection. The proposed precise eye detector is a feature-based method that locally extracts several texture descriptors (Local Ternary Patterns, Local Phase Quantization, and Patterns of Oriented Edge Magnitudes [21]) from four subwindows of the eye region. The extracted features are then classified by SVMs trained using eye and non-eye samples.

The combination of these five systems (the three face detection methods and the two eye detectors) into our proposed face detection system is performed in two serial steps. In the first step, two of the face detectors, EML and SN, are combined for fast extraction of candidate faces from the image. In the second step, only the challenging cases are handled using VJ and SN. The PS model [19] is used for coarse eye localization, and the proposed feature-based eye recognition method is used for precise eye localization and for filtering out false positives.

The main contributions of our system are the following: we produce a new, highly accurate ad hoc system for automatic face detection, which we call FED, and we introduce a novel precise eye localization system, which we call PEC, that is based on an ensemble of descriptors. The quality of our proposed face detection system is validated on three different datasets: the well-known BioID[1] and FERET[2] datasets as well as a self-collected dataset. Using a fully automatic face detection system, our ensemble obtains a detection rate in the BioID dataset of almost 100%, the highest we have seen reported for a fully automated system. This result suggests that combining approaches that have different characteristics is an excellent way to improve face detection performance, even in unconstrained conditions.

The remainder of this paper is organized as follows. In section II the eye localization systems (including PEC, the new system proposed here) are explained in detail. In section III the entire FED face/eye detection system is described, and in section IV the experimental results are presented that validate both PEC as a precise eye detector and FED (which utilizes PEC as a module) as one of the most accurate, fully automatic face detection systems developed to date. Finally, in section V some conclusions are drawn.

II. EYE LOCALIZATION

In our complete approach for face detection, two eye localization methods are employed to obtain a precise localization of eyes inside candidate face images: (i) coarse eye localization using the PS model, and (ii) our novel Precise Eye Classification (PEC) system. Both of these methods are detailed in this section.

A. Coarse Eye Localization (PS)

The enhanced PS model proposed in [19] is used for coarse eye localization in our proposed face detection system. PS [31] is a computationally efficient framework for part-based face modeling where appearance and structural information are combined into a unified framework. Using this approach, a face is first decomposed into parts, and then the best part candidates are searched, subject to some spatial constraints. A PS face model is expressed in terms of an undirected graph G = (V, E), where the vertices V correspond to its parts (the facial parts of two eyes, one nose, and one mouth) and where the edge set E characterizes the local pairwise spatial relationship between the different parts.

The PS approach proposed in [19] enhances the traditional PS model to handle the complicated appearance and structural changes of eyes under uncontrolled conditions. Specifically, it introduces some global constraints to improve translation, rotation, and scale invariance. Moreover, the enhance PS approach adopts a heuristic prediction method to deal with partial occlusion.

The specifics of this approach utilized in our proposed system can be described as follows. Each candidate image is resized to 100×100 pixels. For each input image, 1000 subwindows with the highest similarity to the eye class are paired and extracted according to [19]. All pairs whose

[1] http://www.bioid.com/download-center/software/bioid-face-database.html.
[2] http://www.itl.nist.gov/iad/humanid/feret/.

position is closer than 4 pixels are then merged together. All candidate windows smaller than 50×50 are discarded because we are not interested in localizing very small faces. Moreover, since we are searching for eyes, 50% of the pixels in the bottom and 20% in the central region of each candidate face image are discarded. Another constraint involves the need to search for eye pairs such that one eye is on the left-hand side while the other is on the right-hand side. Pairs found by PS that have a very low score are also discarded. In our experiments, two different threshold values are evaluated: a very low threshold (*th1*= -4) for retaining a larger number of candidate eyes and a higher threshold (*th2*=50) for discarding pairs with a low score (the rationale for using these two thresholds is provided in section III).

B. Precise Eye Classification (PEC)

In this work we propose a precise eye classification system (PEC) to enhance face detection that employs an ensemble of classifiers for precise localization. Given a candidate eye whose position is *B*, the search for precise localization is performed by considering a region of $d \times \frac{1}{2}d$ pixels moving around *B* by steps of {5, 10, 15} pixels. In order to scale to different eye sizes, three values for *d* are used during the search: 30, 35, and 40. Each candidate region is then classified as either "eye" or "non-eye" using a classification module based on the following texture descriptors:

- Local Ternary Pattern (LTP) [8]: this is a variant of the original Local Binary Pattern (LBP) texture descriptor. In LTP a ternary rather than a binary encoding scheme is used to represent pixel variations. The final pattern is then split into two binary patterns by considering its positive and negative components according to some threshold τ (τ=3 in our experiments). In order to achieve scale invariance, we consider in this work the concatenation of two different LTP descriptors obtained by varying the parameter settings (i.e., the number of pixels in the neighborhood *P* and the value of the search radius *R*). Our final descriptor is the concatenation of LTP(*R*=1; *P*=8)+LTP(*R*=2; *P*=16);
- Local Phase Quantization (LPQ) [9]: this is a local approach for texture analysis based on the quantized phase of the discrete Fourier transform that is computed in a local subregion of the image. In this work, local frequency estimation is performed by means of Gaussian derivative quadrature filters [9]. The final descriptor is obtained by the concatenation of two LPQ descriptors extracted at different values for the radius *R*: LPQ(*R*=3)+LPQ(*R*=5);
- Patterns of Oriented Edge Magnitudes (POEM) [21]: this descriptor is based on characterizing edge directions using the distribution of local intensity gradients. POEM measures the edge/local shape information and the relation between the information in neighboring cells. Extracting POEM descriptors is a three step process consisting of (i) gradient computation and orientation quantization, (ii) the calculation of accumulated magnitudes as local histograms of orientations, and (iii) the calculation of self-similarity using the LBP operator to encode accumulated magnitudes across different directions. In the present work, the default parameters proposed in [21] have been used for the POEM descriptors.

The similarity score to the eye class of a pair of eyes found by PEC is given by the sum of the scores of the two eyes. If this score is higher than the threshold τ (τ = -4 in our system), the pair of candidate eyes is classified as eye.

The PEC ensemble is obtained (i) by dividing the candidate image into four equal subwindows, (ii) by extracting three texture descriptors from each subwindow, and (iii) by classifying each with an ensemble of SVMs combined by sum rule (where the scores of the classifiers are summed). In other words, each SVM is trained using a given descriptor extracted from a given subwindow of the image (e.g., given 4 subwindows and 3 descriptors, we train 12 SVMs).

The training set is obtained by selecting positive samples from the FRGC Eye_Centered dataset and 25000 negative samples taken from the Yale-B dataset (see section IV for details). The SVM parameters are grid-searched using a 10-fold cross validation on the training data. A complete schema of the PEC eye classification approach is presented in Figure (1).

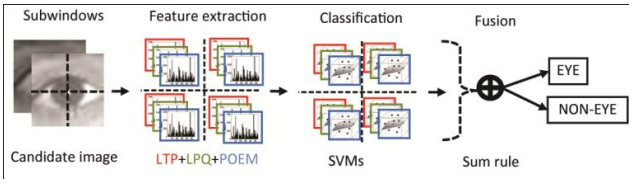

Fig. 1 Schema of PEC, the Proposed Eye Classification System

III. ENSEMBLE OF FACE/EYE DETECTORS (FED)

In this section our novel Face/Eye Detection system is described. FED is inspired by our previous work [20] and by the system proposed by Kroon et al. [7], which is based on the combination of the well-known OpenCV face detector and an eye localizer to refine the eye position.

FED is composed of two steps (see Figure 2). In brief, the first step combines two face detectors: Sparse Network of Winnows (SN) [12] and EML [5]. The second step is aimed at handling more difficult cases, and is composed of the VJ face detector [22] combined with two eye localization modules (i.e., the two eye detectors described in section II). The rationale behind this sequential structure is to begin with a light approach (the first step)–that works almost in real time to detect faces–that is then combined with a heavier approach (the second step) that is applied to difficult cases only. The three face detectors (SN, EML, and VJ) are described in section A, and the two sequential steps are presented more fully in section B.

A. FED Face Detectors

FED combines three face detectors: SN[3], EML, and VJ. The face detector SN [12] is based on local Successive Mean Quantization Transform (SMQT) features applied to a Split up sparse Network of Winnows (SN) classifier. According to the experiments presented in [12], SN performance can be adjusted by a sensitivity parameter σ, which is a threshold of the maximum similarity to the face class allowed for classifying an image. By varying σ in a fixed range [σ_{min}, σ_{max}] (in the original implementation σ_{min}=1 and σ_{max}=10), it is possible to tune the system between a low sensitivity value that determines the presence of some false positives and a high sensitivity value that retains very few false positives but at the cost of missing some faces. In order to take advantage of both results, our system combines the performance obtained by fixing σ to the two extreme values of σ_{min} and σ_{max}.

The face detector EML[4] is a simplified version of the face detection approach described in [5]. The output of this system is the eye position (detected with a very high accuracy when the face is found but with a lower face detection rate than SN).

The face detector VJ [22] is a widely used method for real-time face detection that is characterized by slow training but very fast classification. This approach involves a very simple image representation that is based on (i) Haar wavelets, (ii) an integral image for rapid feature detection, (iii) the AdaBoost machine-learning method for selecting a small number of important features, and (iv) a cascade combination of weak learners for classification.

B. Serial Combination of FED Face Detectors

In FED the output of the two face detectors (EML and SN with $\sigma=\sigma_{max}$) is considered as a first step in our approach to face detection. The component EML outputs both the bounding box of the candidate face (B_{EML}) and the assumed eye position (E_{EML}), while the component SN gives only the bounding box (B_{SNmax}) for the face; therefore, a fixed eye position (E_{SN}) is assumed. The combination of all the resulting bounding boxes discovered by the two face detectors, B_{EML} and B_{SNmax}, is performed according to the following rule: if the Euclidean distance between the positions of E_{EML} and E_{SNmax} is lower than a fixed threshold υ, we accept the eye position returned by EML (E_{EML}). The rationale of this choice is that we use the SN approach to validate the output of EML, which is very accurate when it does find a face.

When SN fails to find a face, the system moves on to carry out a second step that considers the outputs of three face detectors (VJ, EML, and SN with $\sigma=\sigma_{min}$). In this step the lowest value of the parameter σ is used (σ_{min}) in order to maximize the number of true positive faces, disregarding the false positives that will be rejected by the eye detectors. All the resulting candidate faces obtained as the output of VJ, EML, and SN_{min} are processed for eye detection by means of the coarse and the precise eye localization modules described in section II. Precise localization is performed by considering all the candidate eyes resulting from coarse localization (PS) and classifying them using precise eye classification (PEC), the system detailed in section II.B that produces the E_{PEC} eye position. Finally, all the output images that are within a distance of $md \leq 30$ pixels are merged together. For handling non-upright frontals images, we rotate the candidate face image by 25° and -25° before the eye classification step.

In Figure (2) we present a schematic of the two steps involved in our complete system for face detection.

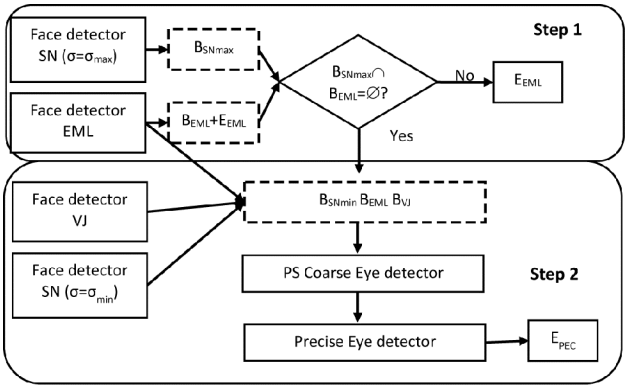

Fig. 2 Outline of the Complete Face/Eye Detection (FED) System

IV. EXPERIMENTAL RESULTS

In this section we present the experimental results validating both the PEC system (the novel eye localization system proposed here and described in section II.B) and the complete face detection system FED (described in section III) that utilizes PEC as a module.

A. Eye Classification with PEC

The intention of the first PEC experiment is to validate the selection of texture descriptors used to build the ensemble (see section II.B). The compared methods in this study are based on an ensemble of SVMs, each trained on different descriptors extracted from four subwindows of the image.

The training set is obtained by selecting positive samples from the FRGC_Eye_Centered dataset[5], which includes eye images having a resolution higher than 15×30, and 25000 negative samples taken from the Yale-B dataset[6] and from other public images. The SVM parameters are grid-searched by 10-fold cross-validation on the training data.

The testing set is obtained from the "red-eyes" dataset [13], which is a set of 390 images containing 1049 images of red-eyes collected for the red-eye classification task. Since this dataset is freely available, we use a labeled subset, called the "candidate red-eyes" dataset (shared by the author of [13]), which includes 2513 images (848 eyes and 1665 non-eyes) at a resolution of 30×30 pixels extracted from the original photographs through an image filtering pipeline that performs the search for reddish zones in the HSL color space (see [13] for details). Unfortunately, this filtering step loses 201 true positive samples. It is important to note that the presence of the red color in the eyes has no effect on results,

[3] http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=13701&objectType=FILE.

[4] http://lipori.dsi.unimi.it/download.html.

[5] http://www.ecse.rpi.edu/~cvrl/database/ISL_IR_Eye_Database.htm.

[6] http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html.

since all our experiments are performed on grey-level images. Because the "candidate red-eyes" dataset includes eyes of different sizes, the eye classification results have been obtained using moving windows of different sizes and angles, each resized to 20×40 to perform the search. The scores obtained from each window are combined by max rule.

TABLE I TEXTURE DESCRIPTORS PERFORMANCE IN TERMS OF AUC FOR THE EYE CLASSIFICATION PROBLEM

	Systems	False Negative	False Positive
Commercial Systems	NikonView V6.2.7	259	35
	KodakEasyShare V6.4.0	293	25
	StopRedEye! V1.0	249	20
	HP RedBot	283	71
	Arcsoft PhotoPrinter V5	235	88
	Cyberlink MediaShow	256	59
Our best approach: PEC		264	50
Error! Reference source not found.		207	4

Table 1 presents the classification performance obtained considering the different texture descriptors. The performance indicator is the Area under ROC curve (AUC) [32]. The area under the ROC curve can be interpreted as the probability that the classifier will assign a higher score to a randomly picked positive sample than to a randomly picked negative sample.

The approaches we compare are the following:

- Whole Eye [20]: a global method, where the entire candidate image is used for extracting three features: LTP(R=1; P=8), LTP(R=2; P=16), and LPQ(R=3);
- "Four Subwindows" [20]: the final method proposed in [20], where the eye image is divided in four subwindows for extracting three features: LTP8, LTP16, and LPQ(R=3).
- LTP, LPQ, and POEM: comparison of the single descriptors LTP, LPQ, and POEM. Each descriptor is extracted from one of the four subwindows of the image;
- PEC: the ensemble described in section II.B based on the combination of subwindows and descriptors based on LTP, LPQ, and POEM.

Examining Table 1, it is clear that our proposed approach, PEC, outperforms our previous work [20]. Moreover, the POEM descriptor seems to be particularly effective for this classification problem.

TABLE 2 SYSTEM PERFORMANCE IN "RED-EYES" DATASET

The purpose of the second PEC experiment is to compare the performance of PEC (in terms of False Negative and False Positive) with other well-known approaches evaluated on the "red-eyes" dataset. This comparison can be read in Table 2 by assuming that our approach is combined with the filtering step in [13] using the "red-eyes" dataset.

	Descriptors					
	Whole Eye Error! Reference source not found.	Four Subwindows Error! Reference source not found.	LTP	LPQ	POEM	PEC
AUC	0.923	0.971	0.955	0.964	0.975	0.984

It should be noted that the best performance obtained by [13] has been achieved using a more advantageous testing protocol, one that involved training on the same dataset using the leave-one-out approach.

Examining the results reported in Table 2, it is clear that our system works similarly to many commercial products.

B. Face Detection with FED

The complete face/eye detection system proposed in this paper (see section III) is evaluated on the following three datasets, all of which contain frontal images:

- The BioID dataset: this dataset includes 1521 images of 23 different people acquired during several sessions;
- The FERET dataset: this dataset is a subset of 250 pictures in Dup2[7] ;
- SC: this dataset is a self-collected dataset containing 64 images of approximately 40 persons collected in unconstrained conditions. This is a challenging dataset, since faces have very different backgrounds.

To validate the performance of our complete face/eye detection system, we use a relative error measure based on the distances between the expected eye positions and the estimated eye positions in faces. Let d_l and d_r be the Euclidean distance between the manually extracted eye centers C_l and C_r and the detected eye centers C'_l and C'_r, with l and r denoting the left and right eye respectively. The relative error of detection is defined as $DER=\max(d_l, d_r)/d_{lr}$ where the normalization factor d_{lr} is the Euclidean distance of the expected eye centers used to make the measure independent of the scale of the face in the image and of the image size. In [15] a face is assumed to be correctly detected if $DER<0.25$ (i.e., if there is an error of less than half an eye width).

In this work the face detection performance is measured using the following indicators:

DR: detection rate (or recall), i.e., the percentage of images where a face is detected according to the above criterion ($DR_{0.25}$ for DER<0.25, $DR_{0.35}$ for DER<0.35);

ADER: the average relative error of detection (evaluated on the whole dataset);

NFP: the number of false positive faces retrieved with $DR_{0.25}$.

In Tables 3-5 we first report the performance of several stand-alone methods and their fusions on the three datasets. In the ensembles, all candidate images resulting from each

[7] For fair comparisons the list of the selected images in the FERET dataset will be available along with the source code used in this work.

face detector are considered (and merged together, in the case of overlapping).

To improve the performance on the SC dataset (Table 5), it is possible to apply a prefiltering step based on skin detection [23], which removes all non-skin patches from the candidate list. In our tests, this prefiltering step succeeded in removing some false positive images while retaining all the true positives; the number of false positives after application of skin filtering is denoted in Table 5 as *NFPsk*.

TABLE 3 COMPARISON OF PERFORMANCE, DETECTION RATE (DR), AND NUMBER OF FALSE POSITIVES (NFP) IN THE BIOID DATASET

Face Detection System	BioID		
	$DR_{0.25}$	$DR_{0.35}$	*NFP*
SN_{min}	96.5	97.9	193
SN_{max}	98.4	99.9	1361
EML	95.9	98.4	220
VJ	77.5	90.1	2396
SN_{max}+EML	99.3	99.9	1581
SN_{max}+VJ	99.7	100	3757
SN_{max}+VJ+EML	99.9	100	3977

TABLE 4 COMPARISON OF PERFORMANCE, DETECTION RATE (DR), AND NUMBER OF FALSE POSITIVES (NFP) IN THE FERET DATASET

Face Detection System	FERET		
	$DR_{0.25}$	$DR_{0.35}$	*NFP*
SN_{min}	98.4	100	4
SN_{max}	98.4	100	13
EML	99.2	99.2	2
VJ	89.2	98.4	85
SN_{max}+EML	100	100	15
SN_{max}+VJ	99.6	100	98
SN_{max}+VJ+EML	100	100	100

TABLE 5 COMPARISON OF PERFORMANCE, DETECTION RATE (DR), AND NUMBER OF FALSE POSITIVES (NFP) IN THE SC DATASET

Face Detection System	SC			
	$DR_{0.25}$	$DR_{0.35}$	*NFP*	*NFPsk*
SN_{min}	87.5	89.1	24	24
SN_{max}	95.3	98.4	116	97
EML	93.8	93.8	19	17
VJ	62.5	75.0	520	343
SN_{max}+EML	98.4	98.4	135	115
SN_{max}+VJ	96.9	100	636	440
SN_{max}+VJ+EML	98.4	100	655	457

Recall that the aim of this work is to design a method that improves stand-alone face detection systems by combining approaches having different characteristics, for instance, by combining a method that finds many true positives, but gives an inaccurate localization (SN), with a more accurate approach resulting in a higher number of missed positives (EML). Notice in Tables 3-5 that by considering the bounding boxes found by the three tested face detectors a face is always discovered (even if not precisely detected, i.e., it is found with $DR_{0.35}$). The eye classification step is then able to refine the face position.

In Tables 6 and 7, we compare the face detection performance of FED with other methods in the literature. For a better comparison, some of the approaches evaluated above are also reported[8]:

- SN: the face detector SN [12] as described in section III with different values for the sensitivity parameter σ;
- EML: the face detector EML [5] as described in section III. Both results were obtained using the free code[4] published in [5];
- FED: our full system described in section III;

Some variant (partial) versions of FED are also evaluated:

- FED_{SE}: a version of FED using SN and EML, but not VJ, as the face detectors;
- FED_{SE*}: as above, but no rotation is performed in the image for finding rotated faces (we do this for a fair comparison with [20]);
- FED_{POEM}: as in FED but only the POEM descriptor is used for eye detection;
- FED_{SN}: a version of FED using SN and the eye detection methods (PS+PEC);
- FED_{EML}: a version of FED using EML and the eye detection methods (PS+PEC);
- FED_{PS}: a version of FED using only PS for eye detection;
- $FED_{SE*POEM}$: as in FED_{SE*} but using only the POEM descriptor for eye detection (this is the fastest ensemble approach tested) ;

We present as well the performance on BioID using some other well-known approaches: the first work on BioID [15], the work that inspired this paper [7], and some of the most recent state-of-the-art approaches [11] [15][16][17][18] [24] [33].

TABLE 6 DETECTION RATE (DR), AVERAGE RELATIVE ERROR OF DETECTION (ADER), AND NUMBER OF FALSE POSITIVES (NFP) IN FACE RECOGNITION PROBLEM. NFP IS OBTAINED WITH $DR_{0.25}$

Eye Detection Approaches		BioID		
		$DR_{0.25}$	*ADER*	*NFP*
SN	$\sigma=\sigma_{min}$	96.5%	0.12	193
	$\sigma=\sigma_{max}$	98.4%	0.13	1361
EML	Free code[4]	95.9%	0.05	220
	Error!	99.3%	-	-

[8] If not specified we use *th1* in the PS method.

	Reference source not found. (2009)			
FED		99.4%	0.08	339
FED_{SE}		99.3%	0.08	300
FED_{SE*}		99.1%	0.06	12
FED_{POEM}		98.4%	0.08	1095
FED_{SN}		98.0%	0.08	265
FED_{EML}		95.5%	0.06	87
FED_{PS}		87.6%	0.07	753
$FED_{SE*POEM}$		98.8%	0.06	43

TABLE 6 CONTINUED

Eye Detection Approaches		**FERET**		
		$DR_{0.25}$	***ADER***	***NFP***
SN	$\sigma=\sigma_{min}$	98.4%	0.13	4
	$\sigma=\sigma_{max}$	98.4%	0.13	13
EML	Free code[4]	99.2%	0.04	2
	Error! Reference source not found. (2009)	-	-	-
FED		100%	0.05	2
FED_{SE}		100%	0.05	0
FED_{SE*}		100%	0.05	0
FED_{POEM}		100%	0.05	28
FED_{SN}		100%	0.05	0
FED_{EML}		98.4%	0.05	0
FED_{PS}		97.6%	0.06	3
$FED_{SE*POEM}$		100%	0.05	1
Eye Detection Approaches		**SC**		
		$DR_{0.25}$	***ADER***	***NFP***
SN	$\sigma=\sigma_{min}$	87.5%	0.15	24
	$\sigma=\sigma_{max}$	95.3%	0.15	116
EML	Free code[4]	93.8%	0.05	19
	Error! Reference source not found. (2009)	-	-	-
FED		98.5%	0.08	29
FED_{SE}		98.5%	0.08	26
FED_{SE*}		97.0%	0.07	9
FED_{POEM}		98.5%	0.08	48
FED_{SN}		95.3%	0.09	22
FED_{EML}		87.5%	0.07	10
FED_{PS}		84.5%	0.07	459
$FED_{SE*POEM}$		93.7%	0.08	19

TABLE 7 COMPARISON OF WITH STATE-OF-THE-ART

Eye Detection Approaches	**BioID**		
	$DR_{0.25}$	***ADER***	***NFP***
FED	99.4%	0.08	339
FED_{SE}	99.3%	0.08	300
Error! Reference source not found.	99.3%	0.06	26
Error! Reference source not found. (2012)			
SN	$\sigma=\sigma_{min}$	96.5%	0.12	193
	$\sigma=\sigma_{max}$	98.4%	0.13	1361
EML	Free code[4]	95.9%	0.05	220
	Error! Reference source not found. (2009)	99.3%	-	-
Error! Reference source not found. (2001)		91.8%	-	-
Error! Reference source not found. (2003)		94.5%	-	-
Error! Reference source not found. (2005)		98.1%	-	-
Error! Reference source not found. (2006)		98.5%	-	-
Error! Reference source not found. (2007)		98.8%	-	-
Error! Reference source not found. (2009)		96.0%[9]	0.0365	171
Error! Reference source not found. (2011)		99.1%[10]	-	-
Error! Reference source not found. (2011)		99.1%[11]	-	-

Eye Detection Approaches		**FERET**		
		$DR_{0.25}$	***ADER***	***NFP***
FED		100%	0.05	2
FED_{SE}		100%	0.05	0
Error! Reference source not found. (2012)		100%	0.07	2
SN	$\sigma=\sigma_{min}$	98.4%	0.13	4
	$\sigma=\sigma_{max}$	98.4%	0.13	13
EML	Free code[4]	99.2%	0.04	2
	Error! Reference source not found. (2009)	-	-	-
Error! Reference source not found.		-	-	-

[9] The best accuracy claimed in the paper (99.9%) is obtained on a subset of the BioID Database, considering only the 1457 images found by the OpenCV face detector.

[10] This accuracy has been obtained for $DR_{0.1}$ using a *semi-automated* face localization module (personal communication).

[11] Obtained with a different testing protocol: two-fold cross validation on BioID.

(2001)			
Error! Reference source not found. (2003)	-	-	-
Error! Reference source not found. (2005)	-	-	-
Error! Reference source not found. (2006)	-	-	-
Error! Reference source not found. (2007)	-	-	-
Error! Reference source not found. (2009)	-	-	-
Error! Reference source not found. (2011)	-	-	-
Error! Reference source not found. (2011)	-	-	-

Eye Detection Approaches		**SC**		
		DR$_{0.25}$	***ADER***	***NFP***
FED		98.5%	0.08	29
FED_{SE}		98.5%	0.08	26
Error! Reference source not found. (2012)		97.0%	0.09	28
SN	$\sigma=\sigma_{min}$	87.5%	0.15	24
	$\sigma=\sigma_{max}$	95.3%	0.15	116
EML	Free code[4]	93.8%	0.05	19
	Error! Reference source not found. (2009)	-	-	-
Error! Reference source not found. (2001)		-	-	-
Error! Reference source not found. (2003)		-	-	-
Error! Reference source not found. (2005)		-	-	-
Error! Reference source not found. (2006)		-	-	-
Error! Reference source not found. (2007)		-	-	-
Error! Reference source not found. (2009)		-	-	-
Error! Reference source not found. (2011)		-	-	-

Our proposed method FED is one of the few methods in the literature that obtains $DR_{0.25}$ (a correctly matched pair) higher than 99% on the BioID dataset, and FED used a fully automated face detector system and the standard testing protocol.

It is possible to improve the DR0.25 by using a lower threshold τ in PEC for classifying a given pair of candidate eyes as "eye," but this increases the number of false positives. For instance, if we allow the retention of 433 false positives, FED obtains a $DR_{0.25}$ of 99.6%. Another method for improving the detection rate of FED is to change the parameter *md* by merging two candidate pair of eyes when their distance is lower than 10 pixels (instead of 30). In this way FED obtains a $DR_{0.25}$ detection rate of 99.8% on the BioID, with only three faces missed, 481 false positives, and a high number of different true positives for the same face. Although this behavior is not a problem for face detection/recognition, it would need to be avoided for eye tracking applications.

Among the variant (partial) FED ensembles tested in this work (FED_{SE} through $FED_{SE*POEM}$), it is noteworthy that $FED_{SE*POEM}$ provides a good detection rate with a lower computation time and a lower number of false positives then the complete version of FED was able to obtain. However, the main drawback with $FED_{SE*POEM}$ is that it cannot handle rotated faces, as we can infer by examining its performance on SC, the dataset containing rotated faces.

FED also obtains an ADER higher than that obtained by the stand-alone EML in the BioID (but the EML detection rate is lower than that obtained by the fusion approaches). For face authentication, a system with a high detection rate might be preferable, especially when the matching algorithm makes use of additional poses (see [6]). In order to further validate the precision of our face detection system, we tested the recognition performance of the Eigenface method coupled with the additional poses [6] on the faces detected by our approach: the equal error rate obtained in the BioID dataset was the same (6%) gained when using a "perfect" manual detection.

A very valuable finding of this work is supporting evidence that it is possible to obtain very coarse face detection even on difficult datasets by combining different face detectors. Using PS, for instance, it is possible to extract different pairs of eyes within each image with at least one pair always obtaining DER<0.25.

In our previous work [20], we searched only upright frontal faces, so this should be compared with FED_{SE*}. The main drawbacks of [20] were the following:

- In the coarse eye detection step, some pairs of eyes were missed;
- The system in [20] assumed that SNσ*min* always discovers a true face (step 2 of [20], when the face is not detected in step 1); as a result the false positive found by SNσ_{min} could not be discarded.

In contrast, our new method FED always finds pairs of true eyes inside a given image (using coarse eye detection), and it relies on the performance of the texture-based eye detector. As a result, a perfect detection rate without any false positives is possible by improving the precise eye detection component.

From the perspective of efficiency, the focus of our research was not on developing real-time computation but rather on localization accuracy. The computation time required by the PEC module strictly depends on the number of pairs of eyes found by PS. This value is related to the "acceptance" threshold *th*, as explained in section II. In table 5 the average number of retained pairs of eyes for each image (#Eyes/image) and the detection rate of PS are reported as a function of *th*. Using a high value of *th* (*th*2), we have greatly reduced the candidate eyes returned by PS, and this fact improves the performance in terms of computation time.

In Table 8 we report the average number of the pairs of eyes (*#Eyes/Image*) found in a given image, and the $DR_{0.25}$ obtained using all pairs found by PS for the following methods:

FULL: all three face detectors are used, and the image is rotated for finding the rotated faces;

HALF: only SN and EML are used as face detectors, and the image is rotated for finding the rotated faces;

$HALF_{norot}$: only SN and EML are used as face detectors, and the image is not rotated for finding the rotated faces.

In Table 8 a $DR_{0.25}$ of 99.8% with *#Eyes/Image* = 100 means that in 99.8% of the images at least one pair of eyes among the 100 found by PS is the true eye pair.

TABLE 8 NUMBER OF EYE PAIRS FOUND BY PS AND THE DETECTION RATE (DR) AS A FUNCTION OF THE PS THRESHOLD

Method		**BioID**		**FERET**		**SC**	
	th	***#Eyes/ image***	***$DR_{0.25}$***	***#Eyes/ image***	***$DR_{0.25}$***	***#Eyes/ image***	***$DR_{0.25}$***
FULL	*th1*	133.2	100%	40.0	100%	938.8	100%
	th2	66.7	99.8%	22.5	100%	293.0	100%
HALF	*th1*	51.2	99.9%	12.6	100%	187.5	98.4%
	th2	35.6	99.4%	11.5	100%	128.7	98.4%
HALF norot	*th1*	29.2	99.9%	12.6	100%	125.1	96.9%
	th2	24.3	99.2%	11.5	100%	93.7	96.9%

It is clear in examining Table 8 that when using only SN and EML as the face detectors in the system the computation time is greatly reduced; however, in these cases, the $DR_{0.25}$ detection rate also drops below 100%.

Finally, in Table 9 we report the results obtained by our complete eye detector system as a function of *th*. The computational time has been evaluated on a PC E5-2609 dual CPU, 2.4Ghz, with 32 GB Ram, using MATLAB code with the parallel toolbox. In Table 9 we report the average time in seconds spent for texture descriptor extraction on images in the BioID dataset. The variant system based only on the POEM descriptor (FED_{POEM}) can be performed in a feasible amount of time, but its performance is lower than that obtained using all three descriptors.

V. CONCLUSION

In this work we studied face detection for frontal faces using an ensemble of eye detectors coupled with an ensemble of face detectors. Our main goal was to improve the accuracy and reliability of the face detection system. Our proposed face detection system FED is composed of two steps. In the first step, two face detectors are combined that extract regions of candidate faces from an image. In the second step, only the challenging cases are handled. The PS model [19] is used for coarse eye localization, and a novel feature-based eye recognition method proposed here is used for precise eye localization and filtering out of false positives. In our proposed precise eye localization method (PEC) the eye descriptors are evaluated by dividing the candidate eye-region into four subwindows followed by extracting sets of features separately from inside each subwindow. Experimental results show that PEC, based on multi-resolution LTP, LPQ, and POEM descriptors, obtains a performance similar to commercial eye detection software.

Face detection experiments using FED, the ensemble of face/eye detectors, which uses PEC as one module, obtains a coarse face detection rate of 100% on the images tested in all three datasets used for validation: the well-know BioID, the FERET dataset, and a self-collected dataset. The coarse eye localization succeeds in extracting several pairs of eyes from each image, and among these pairs, DER<0.25 is always found, that is in all the three tested datasets and in 100% of the tested images.

TABLE 9 DETECTION RATE (DR), AVERAGE RELATIVE ERROR OF DETECTION (ADER), AND NUMBER OF FALSE POSITIVES (NFP) OF DIFFERENT FED METHODS AS A FUNCTION OF *TH*

Method	**th**	**BioID**			
		$DR_{0.25}$		***NFP***	**Average Time (Seconds)**
FED	*th1*	99.4%	*14.5*	339	*14.5*
	th2	*99.2%*	*7.2*	*509*	*7.2*
FED_{SE}	*th1*	99.3%	*5.3*	300	*5.3*
	th2	98.6%	*3.8*	10	*3.8*
FED_{SE*}	*th1*	99.1%	*3.1*	12	*3.1*
	th2	98.8%	*2.9*	190	*2.9*
FED_{POEM}	*th1*	98.4%	1.2	1095	1.2
	th2	98.0%	0.8	156	0.8
$FED_{SE*POEM}$	*th1*	98.8%	0.6	43	0.6
	th2	98.0%	0.5	25	0.5

Method	**th**	**FERET**		
		$DR_{0.25}$	***ADER***	***NFP***
FED	*th1*	100%	0.05	2
	th2	100%	0.05	0
FED_{SE}	*th1*	100%	0.05	0
	th2	100%	0.05	0
FED_{SE*}	*th1*	100%	0.05	0
	th2	100%	0.05	0
FED_{POEM}	*th1*	100%	0.05	28
	th2	100%	0.05	1
$FED_{SE*POEM}$	*th1*	100%	0.05	1
	th2	100%	0.05	1

Method	th	SC		
		$DR_{0.25}$	*ADER*	*NFP*
FED	*th1*	98.5%	0.08	29
	th2	100%	0.09	30
FED_{SE}	*th1*	98.5%	0.08	26
	th2	95.3%	0.07	4
FED_{SE*}	*th1*	97%	0.07	9
	th2	98.5%	0.09	14
FED_{POEM}	*th1*	98.5%	0.08	48
	th2	98.5%	0.08	38
$FED_{SE*POEM}$	*th1*	93.7%	0.08	19
	th2	93.7%	0.08	10

Future plans of investigation include studying and developing texture descriptors for precise eye localization that are higher performing yet reduce the computation time and the number of false positive (without reducing the detection rate). As noted in the experimental section, improving the precise eye module in FED will result in a more powerful face detector.

REFERENCES

[1] Zeng Z., Pantic M., Roisman G.I., and Huang T.S., "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, no. 1, pp. 39-58, Jan. 2009.

[2] H. L. Jin, Q. S. Liu, and H. Q. Lu, "Face detection using one-class based support vectors," in Proc. 6th IEEE Int. Conf. Autom. Face Gesture Recog., 2004, pp. 457–462.

[3] Zhang C. and Zhang Z., "A Survey of Recent Advances in Face Detection", Microsoft Research Technical Report, MSR-TR-2010-66, Jun. 2010.

[4] Hansen D. and Ji Q., "In the eye of the beholder: a survey of models for eyes and gaze", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 3, pp. 478–500, 2010.

[5] Campadelli P., Lanzarotti R., Lipori G., Precise Eye and Mouth Localization. IJPRAI 23(3): 359-377 (2009).

[6] Min J., Bowyer K., Flynn P., "Eye Perturbation Approach for Robust Recognition of Inaccurately Aligned Faces", Hilton Rye Town, NY, USA, July 20-22, 2005, 544-554.

[7] Kroon B., Maas S., Boughorbel S., Hanjalic A., Eye localization in low and standard definition content with application to face matching. Computer Vision and Image Understanding, 113(8):921-933, 2009.

[8] Tan X., and Triggs B., "Enhanced local texture feature sets for face recognition under difficult lighting conditions," Analysis and Modelling of Faces and Gestures, vol. LNCS 4778, pp. 168-182, 2007.

[9] Ojansivu V., and Heikkila J., "Blur insensitive texture classification using local phase quantization", in ICISP, 2008.

[10] R. Sun and Z. Ma, Robust and Efficient Eye Location and Its State Detection, Advances in computation and intelligence, Lecture Notes in Computer Science, 2009, Volume 5821/2009, pages 318-326.

[11] Wu J., Zhou Z.-H., Efficient face candidates selector for face detection, Pattern Recognition 36 (2003) 1175 – 1186.

[12] Nilsson M., Nordberg J., Claesson I., "Face detection using local SMQT features and split up SNOW classifier", in IEEE International conference on Acoustics, Speech, and signal processing (ICASSP), 2007, vol 2, pp. 589-592.

[13] Battiato S., Farinella G.M., Guarnera M., Messina G., Ravì D., "Red-Eyes Removal through Cluster Based Linear Discriminant Analysis", Proceedings of IEEE ICIP 2010 - International Conference on Image Processing, Hong Kong, September 2010.

[14] Safonov, I.V.: Automatic red-eye detection. In: GraphiCon., International conference on the Computer Graphics and Vision, Moscow, Russia (2007).

[15] Jesorsky O., Kirchberg K., Frischholz R., Robust face detection using the Hausdorff distance, in proc. Int. Conf. on Audio- and Video-Based Biometric Person Authentication, pp. 90-95, 2001.

[16] D. Cristinacce and T.F. Cootes. Feature detection and tracking with constrained local models. Proc. the British Machine Vision Conf., 3:929-938, 2006.

[17] S. Kim, S. Chung, S. Jung, D. Oh, J. Kim, and S. Cho. Multi-scale gabor feature based eye localization. Proc. of World Academy of Science, Engineering and Technology, 21:483-487, 2007.

[18] X. Tang, Z. Ou, T. Su, H. Sun and P. Zhao. Robust Precise Eye Location by AdaBoost and SVM Techniques. Proc. Int'l Symposium on Neural Networks, pages 93-98, 2005.

[19] Xiaoyang Tan; Fengyi Song; Zhi-Hua Zhou; Songcan Chen, "Enhanced Pictorial Structures for precise eye localization under incontrolled conditions," Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on , vol., no., pp.1621,1628, 20-25 June 2009

[20] Loris Nanni, Alessandra Lumini: Combining Face and Eye Detectors in a High- Performance Face-Detection System. IEEE MultiMedia 19(4): 20-27 (2012)

[21] Ngoc-Son Vu, Hannah M. Dee and Alice Caplier "Face Recognition using the POEM descriptor", Pattern Recognition Volume 45 Issue 7, July, 2012 Pages 2478-2488.

[22] Paul viola and Michael J. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", CVPR 2001.

[23] Loris Nanni, Alessandra Lumini, Mauro Migliardi: Learning based Skin Classification. International Journal of Automated Identification Technology, In Press.

[24] Dong Yi, Zhen Lei, Stan Z. Li. "A Robust Eye Localization Method for Low Quality Face Images", International Joint Conference on Biometrics, 2011.

[25] T. Riopka and T. Boult. The eyes have it. In Proceedings of ACM SIGMM Multimedia Biometrics Methods and Applications Workshop, pages 9–16, 2003.

[26] X. Tan, Enhanced pictorial structures for precise eye localization under uncontrolled conditions, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 1621–1628

[27] C. Jung, L.C. Jiao, T. Sun, Illumination invariant eye detection in facial Images based on the Retinex theory, in: Proceeding of IscIDE, 2011, pp. 175–183

[28] Cheolkon Jung, Tian Sun, Licheng Jiao, Eye detection under varying illumination using the retinex theory, Neurocomputing, Volume 113, 3 August 2013, Pages 130-137.

[29] Jiayu Gu, Chengjun Liu, Feature local binary patterns with application to eye detection, Neurocomputing, Volume 113, 3 August 2013, Pages 138-152

[30] Z. Zhou, X. Geng, Projection functions for eye detection, Pattern Recognition, 37 (5) (2004) 1049–1056

[31] F. Felzenszwalb and P. Huttenlocher. Pictorial structures for object recognition. IJCV, 61(1):1573–1405, 2005

[32] Fawcett, Tom (2004); ROC Graphs: Notes and Practical Considerations for Researchers, Pattern Recognition Letters, 27(8):882-891.

[33] F. Yang, J. Huang, P. Yang, and D. Metaxas. Eye localization through mul-tiscale sparse dictionaries. In Proceedings of IEEE Conference on Automatic Face and Gesture Recognition, pages 514–518, 2011.

[34] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, N. Kumar, "Localizing Parts of Faces Using a Consensus of Exemplars", Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

[35] S. Charles Brubaker, Jianxin Wu, Jie Sun, Matthew D. Mullin, James M. Rehg, On the Design of Cascades of Boosted Ensembles for Face Detection, International Journal of Computer Vision, May 2008, Volume 77, Issue 1-3, pp 65-86

[36] Fasel, I., Fortenberry, B., Movellan, J.: A generative framework for real time object detection and classification. Computer Vision Image Understand. 98, 182-210, 2005

[37] C. Huang, H. Ai, Y. Li, S. Lao, High-performance rotation invariant multiview face detection, Pattern Analysis and Machine Intelligence, IEEE Transactions on 29 (4), 671-686

[38] R. Verschae, J. Ruiz-del-Solar, M. Correa, A unified learning framework for object detection and classification using nested cascades of boosted classifiers, Machine Vision and Applications, Vol 19, pp 85-103, 2008

[39] C. Garcia, M. Delakis, "Convolutional Face Finder: A Neural Architecture for Fast and Robust Face Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 11, pp. 1408-1423, November, 2004.