Distinctiveness of faces: a computational approach

MANUELE BICEGO
Deir - University of Sassari (Italy)
GAVIN BRELSTAFF
CRS4, Polaris, Pula (Italy)
LINDA BRODO
Dsl, University of Sassari (Italy)
ENRICO GROSSO
Deir - University of Sassari (Italy)
ANDREA LAGORIO
Deir - University of Sassari (Italy)
MASSIMO TISTARELLI
Dap - University of Sassari (Italy)

This paper develops and demonstrates an original approach to face-image analysis based on identifying distinctive areas of each individual's face by its comparison to others in the population. The method differs from most others—that we refer as *unary* — where salient regions are defined by analyzing only images of the same individual. We extract a set of multi-scale patches from each face image before projecting them into a common feature space. The degree of "distinctiveness" of any patch depends on its distance in feature space from patches mapped from other individuals. First a pair-wise analysis is developed and then a simple generalization to the multiple-face case is proposed. A perceptual experiment, involving 45 observers, indicates the method to be fairly compatible with how humans mark faces as distinct. A quantitative example of face authentication is also performed in order to show the essential role played by the distinctive information. A comparative analysis shows that performance of our n-ary approach is as good as several contemporary unary, or binary, methods - whilst tapping a complementary source of information. Furthermore we show it can also provide a useful degree of illumination invariance.

Categories and Subject Descriptors: I.4 [**Image Processing and Computer Vision**]: ; I.5 [**Pattern Recognition**]:

General Terms: Algorithms; Design; Human factors; Security; Perceptual models; Face recognition

1. INTRODUCTION

Automatic visual face analysis is an active research area in which interest has grown over recent years, for both scientific and industrial reasons. Identifying the most

Corresponding author: Manuele Bicego - Deir University of Sassari, via Torre Tonda 34 - 07100 Sassari (Italy) Tel: +39 79 2017321 - Fax: +39 079 2017312 - e-mail: bicego@uniss.it.

distinctive areas of a face [Bruce et al. 1994] ought to assist the performance in such analyses. Yet the methodologies proposed to date do not attempt to directly identify, or make use of, the areas of an individual's face that make it distinct from the rest of the population. For example, typical feature-based methods [Zhao et al. 2003] make the convenience assumption that anatomical features, such as the eyes nose and mouth, are, a-priori, the most distinctive areas [Campadelli and Lanzarotti 2004; Ming-Hsuan et al. 2002; Senior 1999]. Yet, the gaps around the eyes, nose and mouth could be as, or even more, characteristic of a given face - especially if they contain distinctive scars, spots and lines. Note that face recognition by human observers requires a series of eye movements (ocular saccades), that locate and process the distinctive areas within a face [Goren et al. 1975; Yarbus 1967; Nahm et al. 1997; Haith et al. 1979; Klin 2001]—which can not stereotypically be reduced to the eyes, mouth and nose.

Holistic methods, adopt the opposite approach by taking the neutral stance that all areas of a face are equally important—e.g Eigenfaces [Turk and Pentland 1991; Kirby and Sirovich 1990] and Fisher Faces [Belhumeur et al. 1997]. Similarly, Principal Component Analysis (PCA) and methods based on space decomposition have been applied to image windows but they also rely on global features of the face space—not distinctive areas.

An alternative is explored here: we compute those areas of an individual's face that appear distinct when compared to other faces selected from the population. This is performed before incorporating these areas in a subsequent face analysis, as detailed in later sections. Since we tap information from multiple individuals, this approach is conceptually different from most of the existing feature extraction methods that rely on the detection and analysis of specific face areas for authentication or recognition purposes—e.g. the Elastic Bunch Graph Matching technique [Wiskott et al. 1997]. It differs also from more elaborate techniques that identify the most "salient" parts within the face according to a pre-specified criterion. Among these [Tsotsos et al. 1995; Lindeberg 1993; Koch and Ullman 1985; Salah et al. 2002], the system described by [González-Jiménez and Alba-Castro 2005] that detects "key points" from a set of lines extracted from the face image and that in [Lowe 2004] which selects "characteristic points" in a generic image by means of a local optimization process applied to the difference of Gaussians image, filtered at different scales and orientations. Though they all vary in implementation, robustness, computational requirements and accuracy, each of the above approaches is essentially a *unary* technique: salient regions are defined by analyzing *only one* instance of the face class, namely only images of the *same* individual. On the contrary, we identify local patches within an individual's face that are *different* from other individuals by performing a pair-wise, or binary, analysis. This avoids issues that may arise when invoking a single average face, or canonical model, against which each face would then be distinguished. In particular, we find differences between faces by directly extracting from one individual's face image the most distinguishing or dissimilar patches with respect to another's. The details of how this is achieved are described in the next section. Here, we simply note that image patches from the same individual tend to cluster together when projected in a multi-dimensional space and the distance, in that space, of that patch from clusters formed by other

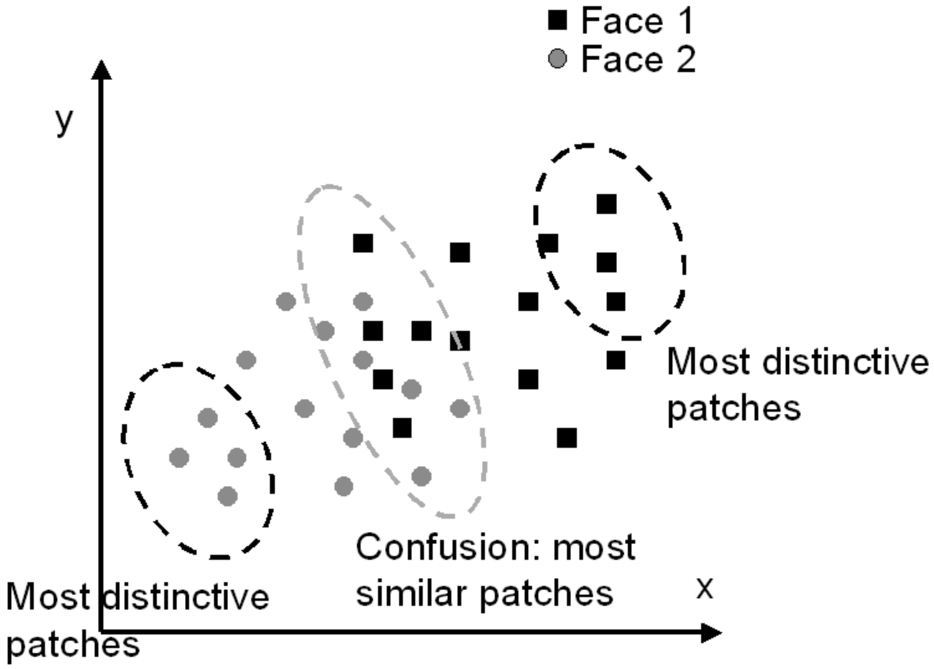


Fig. 1. Idea of the pair-wise differences algorithm.

faces can be used as a measure of "distinctiveness"—as sketched, in just 2-D, in Fig. 1.

Note that the concept of comparative face analysis is also inherent in the recent work by Penev and Atik [Penev and Atick 1996] (Local Feature Analysis), as well as by Li *et al.* [Li et al. 2001] (Local Nonnegative Matrix Factorization), and by Kim *et al.* [Kim et al. 2005] (Locally Salient Independent Component Analysis). These are locally salient versions of dimensionality reduction techniques, applied to a database of images so to obtain a local representation (as a set of basis) of the training set. Even if not explicitly developed to extract salient parts of a face, all these techniques find utility in characterizing a face by performing a comparative local analysis.

An interesting approach more related to our work extracts most salient patches (there denoted *fragments*) of a set of images [Ullman et al. 2002]. There a sufficient coverage of patches are extracted from a set of "client" images, before each patch is weighted in terms of its mutual information with respect to a selected set of classes. However, the optimality criterion there used to select the most relevant patches differs from ours. We use a *deterministic* criterion computing the distance from the "impostor" set, while they adopt a *probabilistic* criterion based on empirical estimation of probability function. In order to obtain a reliable estimate, their approach thus requires a considerably large training set—as illustrated in later sections.

Our computational analysis is in line with some psychophysical experimental results [Gauthier et al. 1999]. These indicate that the human visual system adopts a model formation process for objects, including faces, by making several continuous comparisons with other objects, or faces, i.e. by performing repeated comparative analysis.

Achieving biometric authentication generally involves the active cooperation of the individual. In the case of faces, client and impostor alike need to face towards the camera while adequately illuminated, else access will be denied. Although, three-quarters profiles may be more informative, [Bruce et al. 1987] passport-style frontal poses are the norm: they tend to reduce the preprocessing of the image to a simple affine transform—so to obtain a face representation invariant to 2-D translation, rotation and scaling. This is also in line with recent databases collected for testing face analysis systems—like Banca [Bailly-Baillire et al. 2003] or Face Recognition Grand Challenge (FRGC) [Phillips et al. 2005]. It is within such a context that our work has been developed: we presents results from a face authentication test, performed on the BANCA database [Bailly-Baillire et al. 2003] with encouraging results, adopting an impairment test. Of course, we should prefer to improve this authentication experiment, superseding the impairment test with one that actively uses the ranking of the face distinctive patches. To operate in more challenging scenarios, e.g. surveillance of video streams, invariance to 3-D pose and illumination would need reinforcing. We illustrate how we might achieve a greater degree of illumination invariance in the final section.

The remainder of the paper is organized as follows: in section 2 the pair-wise approach for computing differences between faces is described together with some qualitative visual analysis. Section 3 contains a perceptual study that asks human observers where in face-images do they see distinctive differences, and compares their responses to those of our algorithm. Section 4 of the paper presents results from the face authentication test. It ends with an illustration of illumination invariance. In section 5 conclusions are sought.

2. DISTINGUISHING FACES

Our method finds and applies the distinctive patterns in faces as detailed below in three parts: First we describe how candidate areas of each face-image are extracted and encoded as “feature vectors”. Second, we indicate how, in the resultant feature-space, face-pairs are analyzed. Finally, we elucidate some examples that motivate refinements of the process.

2.1 Multi-scale patches extraction

From each face-image, candidate patches are extracted as areas of gray-level image. These patches should be spatially distributed such that most of the face area is sampled—in a way similar to that adopted in patch-based image classification [Agarwal and Roth 2002; Fergus et al. 2003; Dorko and Schmid 2003; Csurka et al. 2004] and image characterization [Jojic et al. 2003]. Since face recognition may involve information apparent at a variety of spatial resolutions, there may be an advantage in extracting candidate patches at multiple scales. In particular, we adopt a variant multi-scale approach designed to avoid two notable pitfalls: (a) blind analysis - whereby information revealed at one scale is not usefully available at other scales, and (b) repeated image processing - which would add to the over-

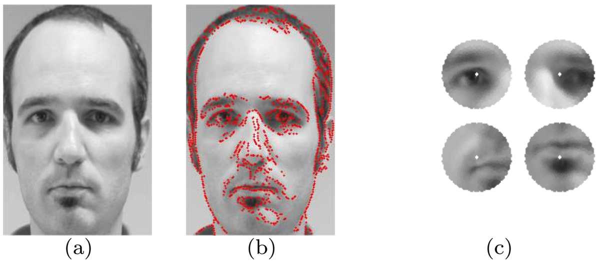


Fig. 2. Log polar sampling: (a) original image (b) all fixations (c) some reconstructed log-polar patches.

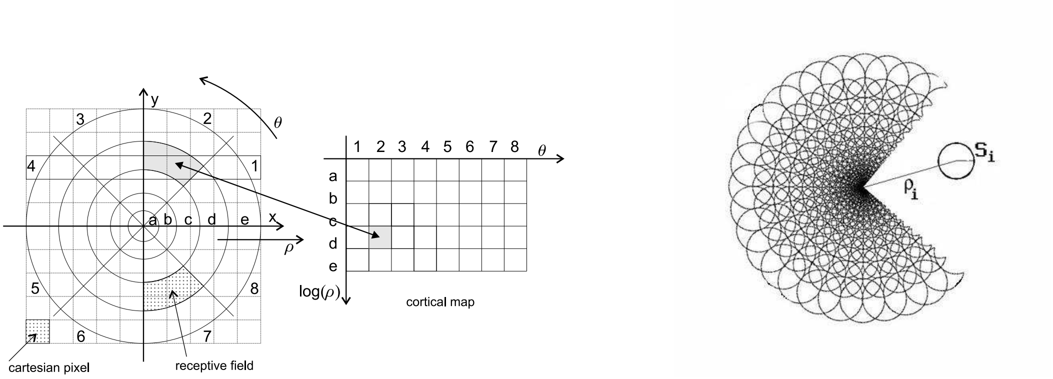
all computational expense. Thus we sample each face-image using patches derived from a log-polar mapping [Grosso and Tistarelli 2000], considering the resulting sampled vectors as our features. This mapping has been motivated by its resemblance to the distribution of the receptive fields in the human retina, where the sampling resolution is higher at the central fovea and decreases toward the periphery. The resultant sampling process ensures that each patch contains both low scale (fine resolution) and contextual (low resolution) information.

To ensure translation-independence and to side-step coarse registration issues the patches are extracted at a loci centered on the edges detected within the image - e.g. using zero crossing of Laplacian of Gaussian (LoG). This approach proved more practical than the alternatives of selecting points at random [Bicego et al. 2005] or by applying morphological operators. Thus we maintain translation-independence, while reducing the number of points required. Note, the use of edge points does not imply that the analysis occurs only at edges: the points actually represent the center (the *fovea*) of the log-polar mapping, which extends the processing also to the neighborhood. As an example, Figure 2(b) shows the sampling points (corresponding to fovea fixations) of one face.

In particular, the face-image is re-sampled at each point following a log-polar scheme so that the resulting set of patches represents a local space-variant remapping of the original image, centered at that point. Analytically, the log-polar scheme describes the mapping postulated to occur between the retina (retinal plane (r, q)) and the visual cortex log-polar or cortical plane (x, h). As illustrated in Fig. 3(a) the size of the "receptive fields" follows a linear increment moving from the central region (fovea) outwards into the periphery.

The log-polar transformation applied here is that described in [Grosso and Tistarelli 2000]—which differs from the models proposed in [Braccini et al. 1982; Tistarelli and Sandini 1993]. The parameters required to define the log-polar sampling are: the number of step in eccentricity (N_r), the number of receptive fields per eccentricity (N_a) and the radial and angular overlap of neighboring receptive fields (O_r and O_a).

For each receptive field, located at eccentricity ρ_i and with radius S_i, the angular overlap factor is defined by $K_0 = \frac{S_i}{\rho_i}$. The amount of overlapping is strictly related to the number of receptive fields per step in eccentricity N_a. In particular if $K_0 =$



(a) (b)

Fig. 3. (a) Log-polar sampling strategy and (b) the adopted log-polar model.

$\frac{\pi}{N_a}$ all receptive fields are disjoint. The radial overlap is determined by:

$$K_1 = \frac{S_i}{S_{i-1}} = \frac{\rho_i}{\rho_{i-1}}.$$

The two overlap parameters K_0 and K_1 are not independent, in fact:

$$K_1 = \frac{\rho_i}{\rho_{i-1}} = \frac{1+K_0}{1-K_0}.$$

As for the angular overlap, the radial overlap is not null only if:

$$K_1 < \frac{1+K_0}{1-K_0}.$$

Given the log-polar parameters N_r, N_a, O_r, O_a, K_0 and K_1 are computed as:

$$K_0 = \pi\frac{O_a}{N_a}, \qquad K_1 = \frac{O_r + K_0}{O_r - K_0}.$$

The image resolution determines the physical limit in the size of the smallest receptive fields in the modeled fovea. This, in turn, determines the smallest eccentricity value:

$$\rho_0 = \frac{S_0}{K_0}$$

The set of log-polar image patches, sampled from each face-image, are vectorized, e.g. a 20 × 20 pixel patch is transformed in a vector of 400 raw gray-level values—that represent the face in feature space.

2.2 Finding differences between face-pairs

Without loss of generality, we start by considering the two-face case, i.e. when client set and impostor set contain only one face each. Later we examine how this process can be expanded to the multi-face case.

The main idea is that the patches from one face-image will tend to form their own cluster in the feature space, while those of the other face-image ought to form a different cluster—e.g. see Fig. 1. The "distinctiveness" of each patch can be related to its locus in feature space with respect to other faces. Any patches of the first face, found near loci of a second face can be considered less distinctive since they may easily be confused with the patches of that second face, and thus may lead to algorithmic misclassification. Conversely, a patch lying on the limb of its own cluster, that is most distant from any other cluster, should turn out to be usefully representative, and may thus be profitably employed by a classifier.

2.2.1 *Patch weighting* . We formalize the degree of distinctiveness of each face patch by weighting it according to its distance from the projection of the other data-cluster. Patches with the highest weights are then interpreted as encoding the most important differences between the two face-images.

More formally, let S_1, S_2 be the set of patches of face 1 and 2, respectively. The weight of distinctiveness ω of a patch $p_1(x, y)$, centered at the position (x, y) in the face 1 is computed as:

$$\omega(p_1(x, y)) = d(p_1(x, y), S_2) \tag{1}$$

where

$$d(p_1(x, y), S_2) = \min_{(x', y')} d_E(p_1(x, y), p_2(x', y')) \tag{2}$$

where d_E is some distance metric between features vectors. Here, for clarity, we adopt an Euclidean metric. It might be worthwhile investigating other metrics, such as those due to transforming feature space via say a Principal Component Analysis or Linear Discriminant Analysis.

Another possibility for the extraction of weights is to train a classifier able to separate set S_1 and S_2 and compute a distinctiveness measure based on the distance from the separating surface. We investigated this in [Bicego et al. 2005], using a Support Vector Machine— with mixed results. However this is computationally demanding and much affected by the choice of the parameters. Here we focus instead on a more basic approach that is easier to control.

2.3 Details and qualitative examples

This section presents details of our pair-wise face differences algorithm: in particular, the parameters used throughout the experimental session are given; moreover, some visual examples are presented, in order to assess the capability of the proposed approach in finding differences between pairs of faces, showing different extra-class results. Images used were all gray-level, of dimensions 320 × 200 pixels, and cropped in order to reduce the influence of the background. Fixations, or centers of the patch sampling process (edge-points), were computed using zero-crossings of a LoG filter. After a preliminary evaluation, log-polar patch resolution was set to 15 eccentricity steps (N_r), at each of which there were 35 receptive fields (N_a), with a 70% overlap along the two directions (O_r and O_a). This represents a reasonable compromise between fovea resolution and peripheral context. Some examples of log-polar patches, rebuilt from the log-polar representations, are shown on Fig. 2(c).

Fig. 4 represents the comparison between different individuals. In particular two pairs of images are displayed in two columnal groups. In the first row are the original images, in the second the fixations used for the log-polar analysis, and in

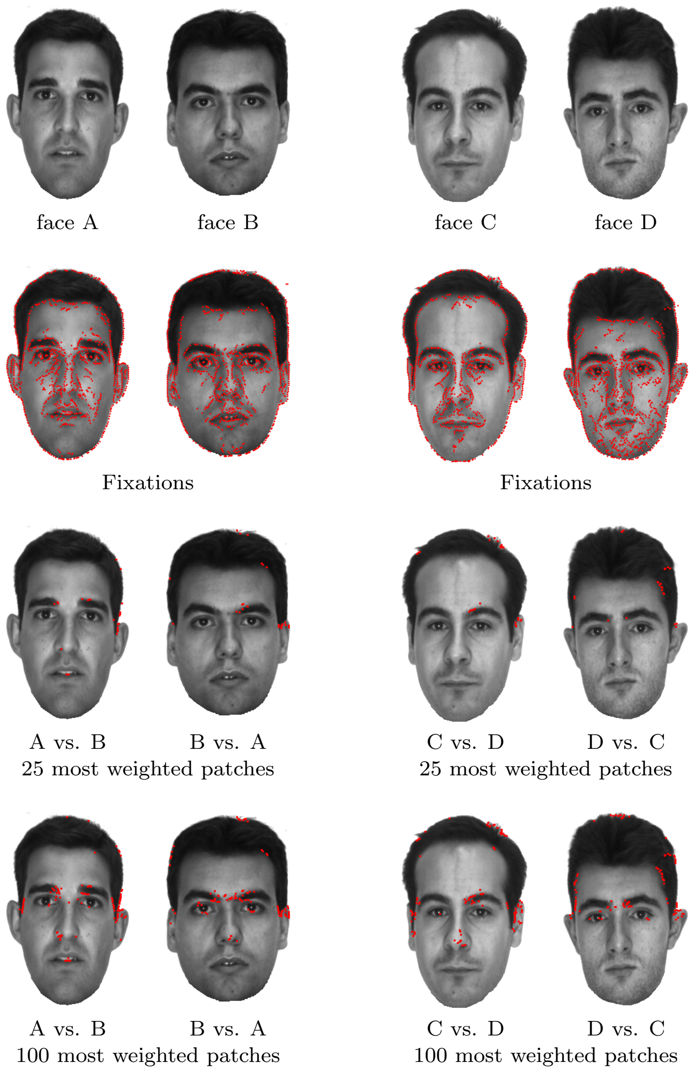


Fig. 4. Two examples of differences extracted from pairs of images of different persons: (A,B) and (C,D).

the third and fourth are, respectively, the plots of the fixations that result in the 25 and the 100 most weighted patches.

The first two columns (subjects A and B) reveal that the main differences are in the ears and in the eyebrows: this is clearly evidenced in row 3 that shows that the first 25 patches are located on the ear in the right part of the face and on the eyebrows. This result is re-enforced when adding patches (last row): note how the left ear is now highlighted.

As a general consideration, it seems that the method is able to capture the essential differences between these face pairs. A similar reasoning applies to faces C and D (column 3 and 4).

3. PERCEPTUAL COMPARISON

3.1 Motivation

Do human beings observe differences between faces around the areas designated as "distinctive" by our algorithm? We carried out a perceptual experiment that asked observers to click on images where they perceived important differences and then compared the results to those produced by our algorithm. We preferred this direct approach over others such that require the interpretation of eye-movements. The degree of correspondence may indicate whether the algorithm can replicate, or complement, the work of human observers. As such this work follows the paradigm adopted in other human face recognition studies [O'Toole 2004; O'Toole et al. 2006; O'Toole et al. 2005]. We selected image-pairs from the BANCA database each showing faces of two different individuals—such that the observer is confronted with a range of difficulties when asked to distinguish between the pairs. By an informal survey of the data-set we were able to select one pair that even expert observers might confuse: i.e. where the image of face A appeared very like that of face B. Three other pairs of decreasing level of difficulty were also selected so as to cover a small but representative spectrum. All four pairs are shown in the left two columns of Fig. 5.

3.2 Method

The image-pairs were included as stimuli in a perceptual experiment (cf. [Collishaw and Hole 2000]) carried out on an LCD screen of a PC, implemented in Matlab. In all, 45 university students (13 male, 32 female), all with normal, or corrected vision, were asked to perform one trial per selected image-pair. Each trial consisted of displaying the two gray-scale face-images side-by-side on a mid-gray background for 20s while the subject was free to scrutinize the two faces and mark, by a mouse-click, the points in the right image that indicate differences. During the 5s between-trial interval each face image was replaced by a Gaussian noise gray-level image—to minimize any influence of after-images. No observer saw the same image twice. On each click a confirmatory audio high-pitch beep was given. Clicking outside of the left-hand image resulted in a lower-pitch beep—as a warning.

Viewing parameters were fixed as follows: viewing distance: 50 cm; image height: 6.5 cm (11 deg, 220 pixels); image width: 6 cm (7 deg, 200 pixels); image-pair separation 1 cm; full contrast LCD screen viewed under indoor illumination.

3.3 Results

A preliminary analysis [Brelstaff et al. 2006] of a larger set of 12 face-pair stimuli permitted a selection of data suitable for a direct comparison with our algorithm's output. Since the algorithm models a low-level visual process our results plot only those loci marked by observers on their first click, and within the following time slice of one second. Fig. 5 shows the resultant loci in the third column for all 45

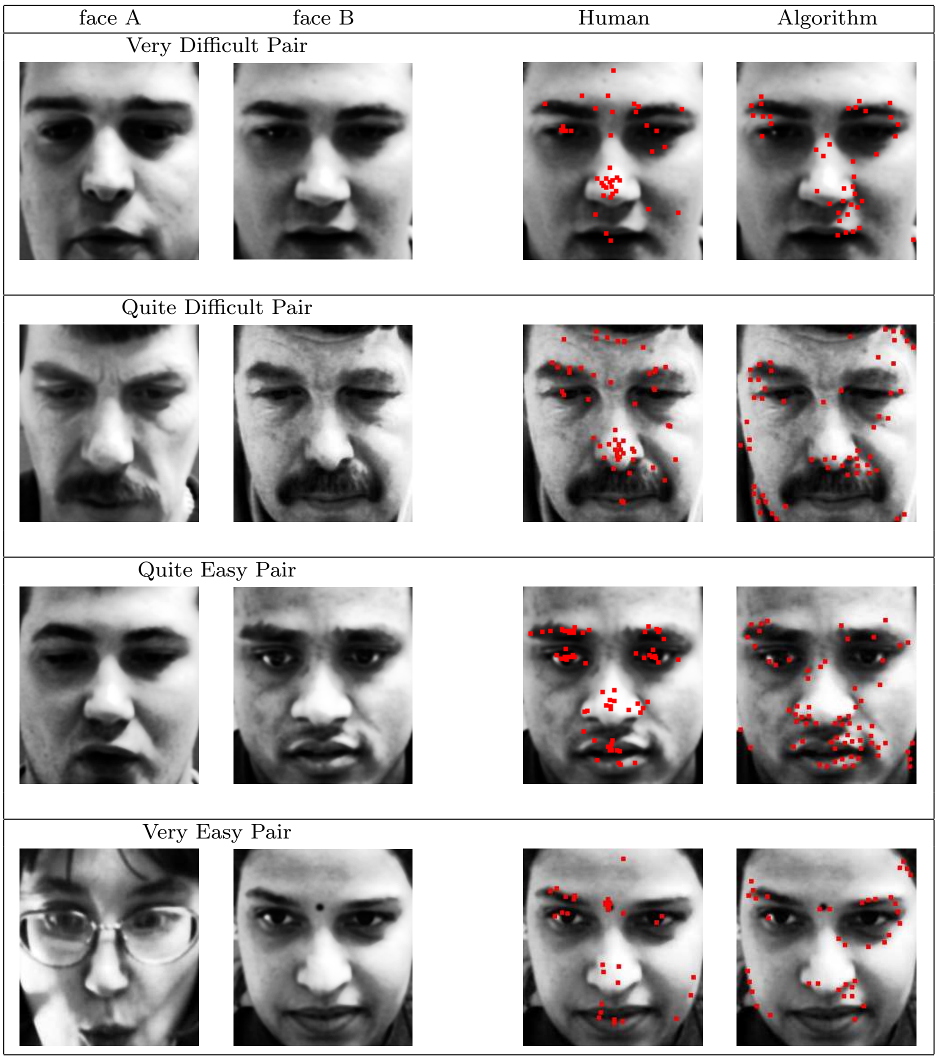


Fig. 5. Results of perceptual experiment: original face-pairs (each of 2 different individuals) are shown side-by-side in the first two columns, in the third the points clicked by human observers, and in the fourth the result of our pair-wise algorithm.

human observers, and in the fourth column for the algorithm.

A degree of correspondence is apparent even when examining by eye the third and fourth columns in the figure. Several areas not usually included as distinct features are marked both by the human observer and by our algorithm: the bridge of the nose in the top-most face-pair; the right jowl zone in the second one down; the crease on the right cheek of the face in the third row; and the central spot between the eyebrows in the fourth row. However, some traditional features are marked by observers and not by the algorithm, and vice-versa. To make a quantitative assessment it is important to account for the fact that observers were free to click

Pair	Cov. Image	Match	Non Match	Cov. Image	Match	Non Match
Very difficult	44.09%	75.00%	25.00%	20.10%	7.50%	92.50%
Quite difficult	58.10%	69.64%	30.36%	54.70%	62.50%	37.50%
Quite easy	64.63%	81.16%	18.84%	52.87%	57.97%	42.03%
Very easy	48.18%	71.79%	28.21%	26.47%	10.26%	89.74%
	(a)			(b)		

Table I. Quantitative comparison: percentage of points of the perceptual map matched in the algorithmic map: (a) our algorithm; (b) Ullman *et al.* 2002.

at any point within the image, while the algorithm is shown only to mark points near edges—i.e only the centers of each spatially extended patch. Such patches can often overlap the points marked by the observers, e.g. those marked by the algorithm as the base and flanks of the nose often correspond to those marked by observers in the central nose-zone. Thus we quantified the degree of correspondence as follows: the percentage of points in the perceptual map that "matched" the algorithm map was computed (cf. [Brelstaff et al. 1990]). A match was deemed to occur where a click lay inside the analysis area, analogous to a receptive field, of one fixation, i.e. within a circular zone of radius $k \cdot \rho_{max}$, where ρ_{max} is the largest radius of the log-polar patch (see Section 2.2.1), where $0 < k \leq 1$ is a parameter determining what fraction of the image gets covered. We used $k = 0.8$.

In Table I(a) the results are presented. The table indicates a fair correspondence between results obtained by our algorithm and those of the perceptual test. The measures are slightly lower when the algorithm erroneously latches on to the face border at the bottom left corner of the image, in face-pair 2 and 4.

We, furthermore, made a comparison with the patches produced by Ullman *et al.* [Ullman et al. 2002] maximizing mutual information algorithm discussed earlier. For that purpose, images were sub-sampled with rectangular patches of nine different sizes, equally distributed in the image. In all, more than 6000 patches were collected for the client image. The corresponding quantitative comparison with our perceptual results as given in table I(b): it is evident that the method fails—as was to be expected since the key step (computation of mutual information) is based on calculation of probabilities empirically estimated from client set and impostor set. Yet here, the client set contains only one item (as does the impostor set), so the probability estimation is unreliable. Indeed, in the paper [Ullman et al. 2002], the best patches are estimated using a training set of 138 face images, which somewhat limits the application of that approach in this kind of scenario.

4. QUANTITATIVE ANALYSIS: AN EXAMPLE OF FACE AUTHENTICATION

Here we aim to assess the usefulness of the information extracted by our algorithm in a face authentication context, i.e. to answer the question: "Is the information extracted instrumental in face authentication?" Below we describe an authentication test carried out towards that end. Note, for pragmatic reasons, we adopt an assessment based on an impairment: by omitting those face parts marked by our algorithm we find that the results are impaired more than when other parts of equivalent area are omitted.

4.1 Subject-specific face authentication

Our algorithm, that extracts differences between face-pairs, might be naturally used for the so-called subject-specific face authentication. In this case the challenge is to develop a method able to reliably compute differences between a client and the "rest of the world" population, employing in the recognition process only those features really relevant for that individual (in principle different for each subject). This involves computing differences between multiple faces, not just face-pairs. In this case there are two sets of images, one related to client images (training data) and one related to impostor images (the rest of the world). Here we compute the differences between two distinct ensembles: the first consists of the patches extracted from the client faces, and the second from those of the impostors. This is achieved by projecting both ensembles into feature space. The advantage of this approach is that differences can be computed in the feature space irrespectively of the number of faces involved. The resultant "features" represent those parts of the client faces that differ from *all* the impostor faces. Clearly, since the analysis is image-based, the client images should be more or less registered, in order to aggregate the results.

As in the two-face case, the analysis assigns a weight to each patch sampled from the face. Here, the patches can also be ordered by the weights, so as to display the K most weighted patches thus visualizing the most distinctive parts of the client face.

4.2 Database description and authentication methodologies

Here we carry out face authentication experiments on the BANCA database [Bailly-Baillire et al. 2003]—a multimodal database, containing face (and voice) data. The part used for face authentication contains 52 subjects (26 female and 26 male). For each subject, 12 different sessions recorded under different conditions are available (4 controlled, 4 degraded, and 4 adverse). From each session we extracted five images for uses as training, client and impostor testing.

The BANCA protocol defines seven different experimental configurations of increasing difficulty. We adopt the Matched Controlled protocol, where the images gathered from the first session are used for training, while the testing images are taken from second, third, and fourth sessions. Here we selected landmarked images, that is images for which standard landmark features are pre-located: this permits straight-forward registration of the images allowing us to concentrate on authentication task. In particular, all the images were preprocessed using a simple geometric normalization, followed by standard histogram equalization [Gonzalez and Woods 2002]. Geometric normalization maps each face on to a 220 pixel high by 200 pixel wide output image, via an affine transform that makes use of the manually annotated eye positions so that the eyes always map to loci 50 pixels in from vertical borders, and 77 pixels from the top, of the output image.

The BANCA protocol splits the dataset in two groups –G1 and G2. These are alternatively used as validation (to set thresholds) and as test set (to estimate errors). Different Weighted Error rates (WER) are computed, for different cost ratios between false positives and false negatives. Since the goal is to demonstrate the discriminative power of the information extracted by the algorithm, we adopt a simple authentication technique based on the established Eigenfaces approach [Turk and Pentland 1991], founded on PCA. In particular, a PCA space is built using the world model of the BANCA database (separated from data used for training and testing), retaining a number of components such that 99% of the variance is preserved. Given a test-face and a claimed identity, that face is projected into the PCA space, and the minimum distance to the projected training images of the claimed subject is extracted. This distance is the matching score. Various distance metrics might be employed, such as Euclidean, Manhattan, angle-based or others. We adopted the Whitened Angle Distance since it has been shown to be very effective in PCA-based face recognition problems [Perlibakas 2004].

In order to understand the significance of the methodology, we masked some parts of the face, according to three impairment criteria, examining three different approaches. Note that the authentication technique (and the PCA space) remained unchanged throughout: the changes occurred in the images.

— *Deleting K most important points*: this approach masks out the K most important patches, as determined by our algorithm—as previously configured: i.e. with log-polar patch resolution set to 15 eccentricity steps and 35 receptive fields at each, with a 70% overlap along the two directions. Mask templates are obtained by extracting differences between the five training images and another five images randomly selected from the XM2VTS database [Messer et al. 1997]. The K most important patches ($K = 100$ in all our runs) are replaced in all the relevant images with random noise. Thus, here the template is represented by the set of training images plus a mask in which the most distinctive parts of the face have been removed. This mask is then applied to the test image and to the training images.

— *Deleting random points*: this approach masks a number of random points equivalent to the number of points marked by the algorithm.

— *All points* are retained and so no masking is performed;

Clearly we expect that removing points determined by our algorithm will be more deleterious for authentication than when deleting *the same quantity* of random points.

Two examples of the application of the three masking methods on faces of the BANCA database are presented in Fig. 6. Note that in the upper example, we compute the most distinguishing zones to be around the mouth and the nose—which may be reasonable since the database largely contains European people, and this is a African face with a very different mouth and nose. In the lower example, the zones around the eyes are compactly marked as most distinct.

4.3 Results

Since random variations are inherent in our masking procedure we repeated the tests 30 times, averaging results in order to obtain a reasonable statistical significance, as detailed below. Weighted Error Rates for the three different methodologies are presented in Table II, for three different values of the standard parameter R.

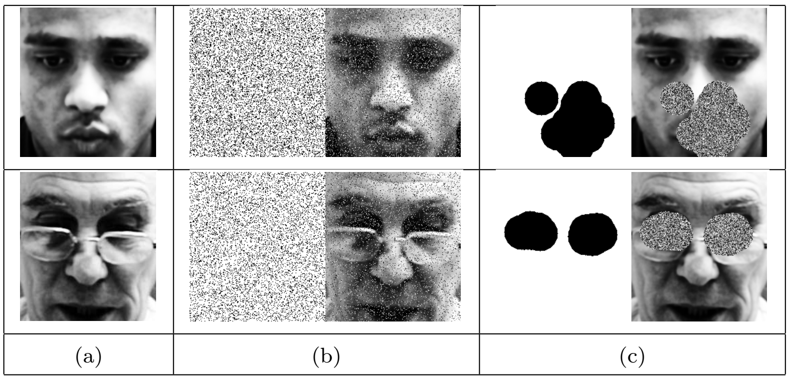


Fig. 6. Example of the masking process on two faces of the database BANCA: (a) original images, (b) random mask and application on the image (b) mask obtained by deleting the K most important patches ($K = 100$) and its application on the image.

Table II. Different WER for the three methodologies.

	WER (R = 0.1)		**WER (R = 1)**		**WER (R = 10)**		
Methodology	**G1**	**G2**	**G1**	**G2**	**G1**	**G2**	**Average**
All points	10.76%	4.85%	10.58%	8.21%	5.10%	1.97%	6.91%
Del. random pnt	9.55%	6.37%	15.71%	8.71%	7.64%	3.00%	8.50%
Del. most imp. pnt	11.62%	7.46%	19.58%	13.37%	9.11%	4.46%	10.93%

From this table the important role played by the information extracted by our algorithm is evident. When removed, the performance of the classifier deteriorates substantially and significantly. Note that much greater impairment is obtained by omitting those parts extracted by our algorithm than those omitted at random. The significance of this interpretation is supported by the fact that these random deletions have been repeated several times and results averaged.

In order to investigate if the significance of this impairment is statistically relevant, we compute the statistical confidence of the results, using the method recently proposed in [Bengio and Mariétoz 2004]. This method, given the number of client tests NC and the number of the impostor tests NI, computes the confidence at a level of 95% starting from the FAR and FRR rates—as follows:

$$\sigma_{95} = 1.96 \times \sqrt{\frac{FAR(1-FAR)}{4NI} + \frac{FRR(1-FRR)}{4NC}}$$

The value σ_{95} indicates that the confidence is computed for a probability of 95%. We compute this confidence for all the results, for both groups G1 and G2, and the average error rates confidence is 2.32%: this indicates the statistical significance of the result.

4.4 Comparison with other methods

In this section we compare our algorithm to other established approaches for the extraction of salient parts of the face. In particular, we adapt four methods to our analysis:

— Lowe [Lowe 2004][1]: which represents the standard SIFT features, whereby the image is processed with multiscale filters, returning a set of key points (position, scale, orientation and key descriptor). Here, we mask the zones around the key point locations, using the scale parameter to decide the extent of removal.

— Salah *et al.* 2002 [Salah et al. 2002]: this method simulates the attentive visual selection mechanism of human beings. The starting point is a saliency map, which determines where the "eye" should observe. This map is computed using a bank of multi-frequency Gabor filters. Here we derive the masking map by considering the N most salient points in the map—where N matches the number of points that our algorithm used.

— Walther [Walther 2006][2]: in his PhD thesis, Walther extends previous work on saliency-based visual attention by Koch and Ullman [Koch and Ullman 1985] and Itti *et al.* [Itti et al. 1998], producing a bottom-up salient region selection scheme. As for Salah, we consider just the relevant saliency map—that is obtained by extracting maps derived from (color) luminance and orientation contrasts across different scales.

— Ullman *et al* [Ullman et al. 2002]: Here the parameters were set as in the perceptual comparison in section 3. Again we used the five training images allowed by the BANCA protocol together with five randomly selected images.

For all four methods we employed default parameter settings without attempting further optimization. Since the first three methods represent unary operators and deal only with client images we decided to apply each method to the training images and then combine saliency maps derived by summing the output from the different images.

Two examples of the application of each method are shown in Fig. 7—as mask and resultant image. Note the wide-spread spatial distribution of the zones considered of importance by the three unary methods. They tend to disperse attention between standard "a priori" zones, such eyes, nose and mouth, with little adjustment for the particular classification task. That contrasts markedly with the compacted focused zones indicated by our pair-wise approach—see Fig. 6(c). These zones are distinctive and salient with respect to other people in the population, e.g. the "African mouth," previously mentioned, from the first row of Fig. 6.

For completeness we computed the averaged WER, as described in previous section, to each method employed here. As before, the most important areas were masked with random noise, repeating the process 30 times and running the authentication impairment test. We report the final averaged errors in Table III along with the results obtained by the pair-wise algorithms.

[1] Code currently available at http://www.cs.ubc.ca/ lowe/keypoints

[2] Saliency Toolbox currently available at http://www.saliencytoolbox.net/

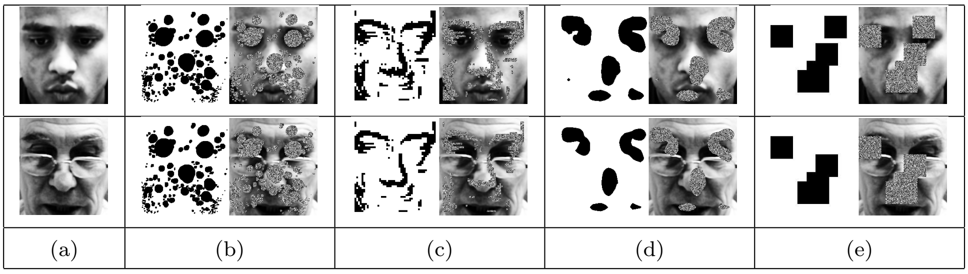


Fig. 7. Example of the masking process on two faces of the database BANCA: (a) original images, (b) Lowe 2004 (SIFT) (c) Salah *et al.* 2002; (d) Walther 2006; (e) Ullman *et al.* 2002

Method	Averaged WER
Lowe [Lowe 2004]	10.15%
Salah *et al.* [Salah et al. 2002]	10.19%
Walther [Walther 2006]	10.74%
Ullman *et al.* [Ullman et al. 2002]	11.87%
Our algorithm	10.93%

Table III. Averaged WER on the impairment authentication test, for each method.

The table shows that, in this test scenario, the pair-wise discriminant methods (lower two rows) perform marginally better than any of the unary methods (top three rows)—though differences in WER are not statistically significant. Note, how the Ullman *et al.* method improves when moving from one face-per-set to five faces-per-set: the pdf estimates become more reliable. From this performance base-line there is much scope to refining the way our algorithm in search of further improvement. A thorough comparison on a direct face recognition problem might then be better performed.

In order to deeply understand the differences between the methodologies, we computed, for each pair of them, the percentage of overlap in the errors occurred in the authentication test. Clearly, low percentages indicate low correlations between the occurred errors. Obtained results are displayed in Table IV, with also the averaged values: from this table it is evident that the different methodologies are quite uncorrelated, with the proposed approach being the most uncorrelated one. This confirms that the presented methodology is a valid and complementary alternative to standard techniques.

4.5 Discussion on illumination invariance

Illumination represent a intrinsic problem relevant to all face recognition algorithms [Adini et al. 1997]. As discussed in the *Introduction*, we here illustrate how our algorithm may embody a degree of invariance to illumination variations. To this end, we apply the algorithm—at first without any normalization stage—to four

	Prop. Approach	Lowe	Salah *et al.*	Walther	Ullman *et al.*
Prop. Approach	100%	65.43%	71.90%	67.91%	71.54%
Lowe	65.43%	100%	71.38%	74.42%	68.19%
Salah *et al.*	71.90%	71.38%	100%	74.03%	74.16%
Walther	67.91%	74.42%	74.03%	100%	69.94%
Ullman *et al.*	71.54%	68.19%	74.16%	69.94%	100%
Average	69.20%	69.86%	72.87%	71.57%	70.96%

Table IV. Error correlation matrix.

images of the same face lit from different directions: (1) normal to the face, (2) from on high, (3) from the left, and (4) from the right.

We extract the differences between each face-image and a corresponding impostor face, and then we visualize its 100 most distinctive points. Fig. 8 shows the results: the first two columns being, respectively, (a) the client and (b) imposter faces. Column (c) shows the results of the algorithm omitting the normalization/registration phase—they are not very robust to illumination changes. Failures occur at points bordering on the most illuminated zones. This is mainly due to the log-polar scheme using the average gray-level to assign a value to each of its cells—which obviously varies with illumination change.

Column (d) shows the results of restoring the histogram equalization preprocessing (the one used in the authentication test)—there is an immediate benefit. Column (e) are the results obtained using a more sophisticated illumination compensation scheme[3] [Gross and Brajovic 2003], where again increased invariance is achieved. Again, similar results—column (f)—were obtained after internally modifying our log-polar scheme—whereby its previously failing average-cell value was substituted by one computed by a LoG filter. Since the resultant values are now differential they are largely unaffected by variations in local illumination levels [Marr 1982]. The filter was implemented as difference of two Gaussians, with widths that were directly linked to the radius of the receptive field[4], thus realizing multiscale filtering.

For completeness, we quantify the degree of illumination invariance manifest by each of the four marked columns in Fig. 8. This is done by computing the average similarity of the points mapped on each column's four consituent images, in a pair-wise manner. For this purpose we adapt a method established for matching fingerprint minutiae [Maio et al. 2003]. Given two maps, pairs of spatially corresponding points are identified as follows: for each point in map-1 take the closest point in map-2 and measure the distance between them; if it is shorter than a threshold then the points are deemed to correspond. The similarity measure between the two maps (known as matching score in the fingerprint context) is then computed as the percentage of points in correspondence divided by the total number

[3] Currently available at `http://torch3vision.idiap.ch/downloads.php`.

[4] In particular, σ_i of the inhibitory Gaussian is the radius divided by 2.5, whereas for the other Gaussian it holds $\sigma_e = \sigma_i/1.6$. [Marr 1982]

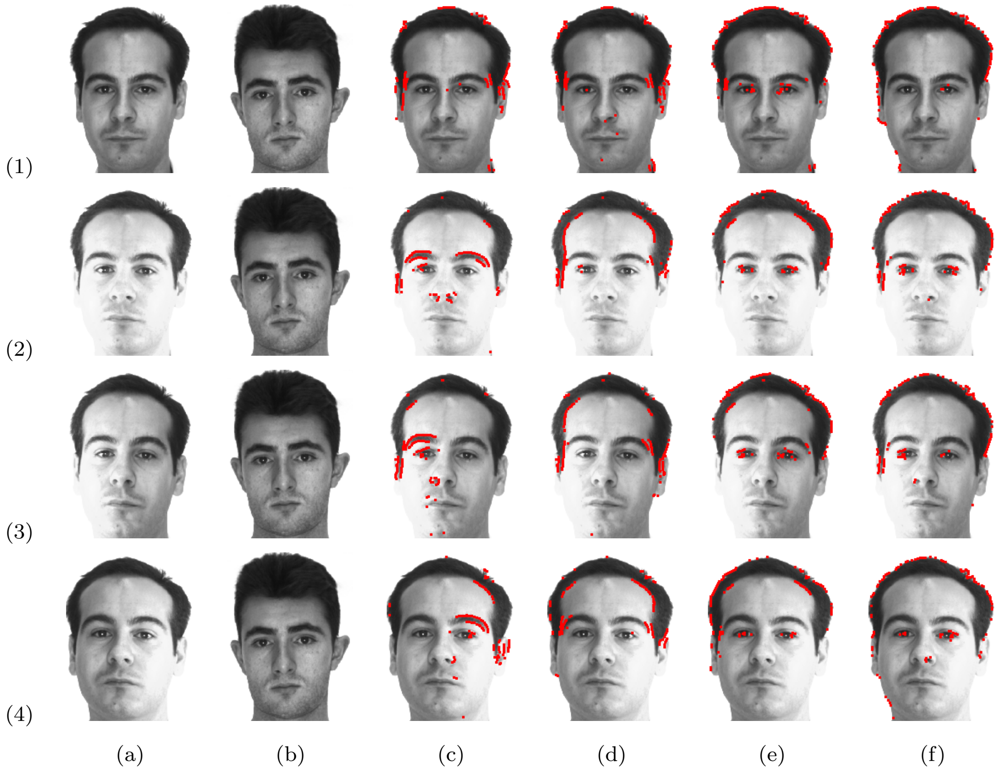


Fig. 8. Experiment on illumination changes: (a) client images; (b) impostor images; (c) Our algorithm on unnormalized image; (d) Our algorithm on histogram equalized images; (e) Our algorithm on images normalized using illumination compensation; (f) Our algorithm using LoG log-polar features.

of points. Such similarities are computed for all pairs of maps relative to different illumination conditions and then the averaged similarities are reported in Table V. It is evident, here, that all methods but the first alleviate problems deriving from

Method	Averaged similarity between maps
Log-polar	68.63%
Histogram equalization	92.25%
Illumination compensation	96.44%
LoG log-polar	94.56%

Table V. Quantifying illumination invariance via averaged similarities between maps generated with different illumination conditions, for the four different methods.

illumination changes—including the relatively efficient histogram equalization that we adopted in earlier sections.

5. CONCLUSIONS AND FUTURE WORK

In this article we developed and demonstrated the feasibility of an alternative approach to face analysis based on identifying distinctive areas of individual faces by comparing them with others in the population. Scope exists to improve several aspects of the method. Patches might be extracted using different techniques, e.g. DCT, or Gabor filters. More sophisticated distance metrics might be explored. Alternative generalizations from pair-wise to multiple-face scenarios might be usefully elaborated. Nevertheless, even before such investigations are attempted it is encouraging to observe a base-line performance already on a par with existing methods, and an apparent compatibility with how humans mark faces as distinct.

Although we illustrated that our method embodies some degree of invariance to illumination variations, more work is necessary if invariance to 3-D pose is to be provided. Furthermore, we should prefer to improve the authentication experiment, superseding the impairment test with one that actively uses the ranking of the face distinctive patches.

Acknowledgements

Authors would like to thank A.A. Salah for the code of the saliency map of [Salah et al. 2002]. Authors are also indebted to the anonymous reviewers for their valuable suggestions.

REFERENCES

Adini, Y., Moses, Y., and Ullman, S. 1997. Face recognition: The problem of compensating for changes in illumination direction. *IEEE Trans. on Pattern Analysis and Machine Intelligence 19,* 7, 721–732.

Agarwal, S. and Roth, D. 2002. Learning a sparse representation for object detection. In *Proc. European Conf. on Computer Vision*. Vol. 4. 113–130.

Bailly-Baillire, E., Bengio, S., Bimbot, F., Hamouz, M., Kittler, J., Marithoz, J., Matas, J., Messer, K., Popovici, V., Pore, F., Ruiz, B., and Thiran, J.-P. 2003. The BANCA database and evaluation protocol. In *Proc. Int. Conf. on Audio- and Video-Based Biometric Person Authentication (AVBPA03)*. Springer-Verlag, 625–638.

Belhumeur, P., Hespanha, J. P., and Kreigman, D. J. 1997. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. on Pattern Analysis and Machine Intelligence 19,* 7, 711–720.

Bengio, S. and Mariétoz, J. 2004. A statistical significance test for person authentication. In *Proc. of ODISSEY 2004, Speaker and Language Recognition Workshop*. 237–244.

Bicego, M., Grosso, E., and Tistarelli, M. 2005. On finding differences between faces. In

Audio- and Video-based Biometric Person Authentication, T. Kanade, A. Jain, and N. Ratha, Eds. Vol. LNCS 3546. Springer, 329–338.

Braccini, C., Gambardella, G., Sandini, G., and Tagliasco, V. 1982. A model of the early stages of the human visual system: Functional and topological transformation performed in the peripheral visual field. *Biol. Cybern. 44*, 47–58.

Brelstaff, G., Brodo, L., Bicego, M., and Grosso, E. 2006. Face-pair scrutiny - subject-type classification. *Perception 35 sup*, 209. ECVP06.

Brelstaff, G., Ibison, M., and Eliot, P. 1990. Edge-region integration for segmentation of mr images. In *Proc. British Machine Vision Conference*. 151–156.

Bruce, V., Burton, A., and Dench, N. 1994. What's distinctive about a distinctive face? *Quarterly Journal of Experimental Psychology 47A*, 119–141.

Bruce, V., Valentine, T., and Baddeley, A. 1987. The basis of the 3/4 view advantage in face recognition. *Applied Cognitive Psychology 1*, 109–120.

Campadelli, P. and Lanzarotti, R. 2004. Fiducial point localization in color images of face foregrounds. *Image and Vision Computing 22*, 863–872.

Collishaw, S. M. and Hole, G. J. 2000. Featural and configurational processes in the recognition of faces of different familiarity. *Perception 29*, 893–909.

Csurka, G., Dance, C., Bray, C., Fan, L., and Willamowski, J. 2004. Visual categorization with bags of keypoints. In *Proc. Workshop Pattern Recognition and Machine Learning in Computer Vision*.

Dorko, G. and Schmid, C. 2003. Selection of scale-invariant parts for object class recognition. In *Proc. Int. Conf. on Computer Vision*. Vol. 1. 634–640.

Fergus, R., Perona, P., and Zisserman, A. 2003. Object class recognition by unsupervised scale-invariant learning. In *Proc. Int. Conf. on Computer Vision and Pattern Recognition*. Vol. 2. 264.

Gauthier, I., Tarr, M., Anderson, A., Skudlarski, P., and Gore, J. 1999. Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. *Nature Neuroscience* 2, 568–573.

Gonzalez, R. and Woods, R. 2002. *Digital Image Processing*, 2 ed. Prentice Hall.

González-Jiménez, D. and Alba-Castro, J. 2005. Biometrics discriminative face recognition through gabor responses and sketch distortion. In *Pattern Recognition and Image Analysis: Second Iberian Conference*. Vol. LNCS 3523. Springer-Verlag, 513–520.

Goren, C., Sarty, M., and Wu, P. 1975. Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics 56*, 544–549.

Gross, R. and Brajovic, V. 2003. An image preprocessing algorithm for illumination invariant face recognition. In *Audio- and video-based biometric person authentification*, J. Kittler and M. Nixon, Eds. Vol. LNCS 2688. 10–18.

Grosso, E. and Tistarelli, M. 2000. Log-polar stereo for anthropomorphic robots. In *Proc. European Conference on Computer Vision*. Vol. 1. Springer-Verlag, 299–313.

Haith, M., Bergman, T., and Moore, M. 1979. Eye contact and face scanning in early infancy. *Science 198*, 853–854.

Itti, L., Koch, C., and Niebur, E. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence 20,* 11, 1254–1259.

Jojic, N., Frey, B., and A.Kannan. 2003. Epitomic analysis of appearance and shape. In *Proc. Int. Conf. on Computer Vision*. Vol. 1. 34–41.

Kim, J., Choi, J., Yi, J., and Turk, M. 2005. Effective representation using ica for face recognition robust to local distortion and partial occlusion. *IEEE Trans. on Pattern Analysis and Machine Intelligence 27,* 12, 1977–1981.

Kirby, M. and Sirovich, L. 1990. Application of the karhunen-loeve procedure for the characterization of human faces. *IEEE Trans. on Pattern Analysis and Machine Intelligence 12,* 1, 103–108.

Klin, A. 2001. Eye-tracking of social stimuli in adults with autism. Paper presented at the meeting of the NICHD Collaborative Program of Excellence in Autism. Yale University, New Haven, CT.

Koch, C. and Ullman, S. 1985. Shifts in selective visual-attention towards the underlying neural circuitry. *Human Neurobiology 4*, 219–227.

Li, S., Hou, X., and Zhang, H. 2001. Learning spatially localized, parts-based representation. *Computer Vision and Image Understanding 1*, 207–212.

Lindeberg, T. 1993. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. *Int. Journal of Computer Vision 11,* 3, 283–318.

Lowe, D. 2004. Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision 60,* 2, 91–110.

Maio, D., Maltoni, D., Jain, A. K., and Prabhakar, S. 2003. *Handbook of Fingerprint Recognition*. Springer Verlag.

Marr, D. 1982. *Vision*. Freeman Publishers.

Messer, K., Matas, J., Kittler, J., Luettin, J., and Maitre, G. 1997. XM2VTSDB: The extended M2VTS database. In *Proc. Int. Conf. on Audio and Video-based biometric person authentication.*

Ming-Hsuan, Y., Kriegman, D., and Ahuja, N. 2002. Detecting faces in images: a survey. *IEEE Trans. on Pattern Analysis and Machine Intelligence 24*, 3458.

Nahm, F., Perret, A., Amaral, D., and Albright, T. 1997. How do monkeys look at faces? *Journal of Cognitive Neuroscience 9*, 611–623.

O'Toole, A. 2004. Psychological and neural perspectives on human face recognition. In *Handbook of Face Recognition*, S. Li and A. Jain, Eds. Springer-Verlag. in press.

O'Toole, A., Jiang, F., Abdi, H., and Haxby, J. 2005. Partially distributed representations of objects and faces in ventral temporal cortex. *Journal of Cognitive Neuroscience 17,* 4, 580–590.

O'Toole, A., Jiang, F., Roark, D., and Abdi, H. 2006. Predicting human performance for face recognition. In *Face Processing: Advanced models and methods*, R. Chellappa and W. Zhao, Eds. Academic Press. in press.

Penev, P. and Atick, J. 1996. Local feature analysis: a general statistical theory for object representation. *Network: computation in Neural Systems 7,* 3, 477–500.

Perlibakas, V. 2004. Distance measures for PCA-based face recognition. *Pattern Recognition Letters 25,* 6, 711–724.

Phillips, P., Scruggs, P. F. T., Bowyer, K., Chang, J., Hoffman, K., an J. Min, J. M., and Worek, W. 2005. Overview of the face recognition grand challenge. In *Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition.* Vol. 1. 947–954.

Salah, A., Alpaydin, E., and Akarun, L. 2002. A selective attention-based method for visual pattern recognition with application to handwritten digit recognition and face recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence 24,* 3, 420–425.

Senior, A. 1999. Face and feature finding for a face recognition system. In *Proc. of Audio- and Video-based Biometric Person Authentication.* 154–159.

Tistarelli, M. and Sandini, G. 1993. On the advantages of polar and log-polar mapping for direct estimation of time-to-impact from optical flow. *IEEE Trans. on Pattern Analysis and Machine Intellingence 15,* 4, 401–410.

Tsotsos, J., Culhane, S., Wai, W., Lai, Y., Davis, N., and Nuflo, F. 1995. Modelling visual attention via selective tuning. *Artificial Intelligence 78*, 507–545.

Turk, M. and Pentland, A. 1991. Eigenfaces for recognition. *Journal of Cognitive Neuroscience 3,* 1, 71–86.

Ullman, S., Vidal-Naquet, M., and Sali, E. 2002. Visual features of intermediate complexity and their use in classification. *Nature Neuroscience 5*, 682–687.

Walther, D. 2006. Interactions of visual attention and object recognition: computational modeling, algorithms, and psychophysics. Ph.D. thesis, Pasadena - CA.

Wiskott, L., Fellous, J.-M., and der Malsburg, C. V. 1997. Face recognition by elastic bunch graph matching. *IEEE Trans. on Pattern Analysis and Machine Intelligence 19*, 775–779.

Yarbus, A. 1967. *Eye movements and vision.* Plenum Press, New York.

Zhao, W., Chellappa, R., Phillips, P., and Rosenfeld, A. 2003. Face recognition: A literature survey. *ACM Computing Surveys 35*, 399 – 458.

Received Month Year; revised Month Year; accepted Month Year