Pattern Recognition 45 (2012) 2445–2459

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

A comparative study of thermal face recognition methods in unconstrained environments

Gabriel Hermosilla [a,b], Javier Ruiz-del-Solar [a,b,*], Rodrigo Verschae [b], Mauricio Correa [a,b]

[a] Department of Electrical Engineering, Universidad de Chile, Av. Tupper 2007, 837-0451 Santiago, Chile
[b] Advanced Mining Technology Center, Universidad de Chile, Av. Tupper 2007, 837-0451 Santiago, Chile

ARTICLE INFO

Article history:
Received 8 August 2011
Received in revised form
22 December 2011
Accepted 1 January 2012
Available online 12 January 2012

Keywords:
Face recognition
Thermal face recognition
Unconstrained environments

ABSTRACT

The recognition of faces in unconstrained environments is a challenging problem. The aim of this work is to carry out a comparative study of face recognition methods working in the thermal spectrum (8–12 μm) that are suitable for working properly in these environments. The analyzed methods were selected by considering their performance in former comparative studies, in addition to being real-time, to requiring just one image per person, and to being fully online (no requirements of offline enrollment). Thus, in this study three local-matching methods based on histograms of Local Binary Pattern (LBP) features, on histograms of Weber Linear Descriptors (WLD), and on Gabor Jet Descriptors (GJD), as well as two global image-matching method based on Scale-Invariant Feature Transform (SIFT) Descriptors, and Speeded Up Robust Features (SURF) Descriptors, are analyzed. The methods are compared using the Equinox and UCHThermalFace databases. The use of these databases allows evaluating the methods in real-world conditions that include *natural* variations in illumination, indoor/outdoor setup, facial expression, pose, accessories, occlusions, and background. The UCHThermalFace database is described for the first time in this article and WLD is used for the first time in face recognition. The results of this comparative study are intended to be a guide for developers of face recognition systems. The main conclusions of this study are: (i) all analyzed methods perform very well under the conditions in which they were evaluated, except for the case of GJD that has low performance in outdoor setups; (ii) the best tradeoff between high recognition rate and fast processing speed is obtained by WLD-based methods, although the highest recognition rate in all cases is obtained by SIFT-based methods; and (iii) in experiments where the test images are acquired in an outdoor setup and the gallery images are acquired in an indoor setup, or vice versa, the performance of all evaluated methods is very low. As part of the future work, the use of normalization algorithms and calibration procedures in order to tackle this last issue will be analyzed.

1. Introduction

The recognition of human faces in unconstrained environments has attracted increasing interest in the research community in recent years. Several studies have shown that the use of thermal images can solve limitations of visible-spectrum based face recognition, such as invariance to variations in illumination and robustness to variations in pose [37,38], which are two of the major factors affecting the performance of face recognition systems in unconstrained environments [36]. This is possible, thanks to the physical properties of thermal technology (long-wave infrared spectrum, 8–12 μm), and the anatomic characteristics of the human body:

- Thermal sensors collect the energy emitted by a body instead of the reflected light, and the emissivity of human skin is between 8–12 μm,
- thermal sensors are invariant to changes in illumination; they can even work in complete darkness, and
- the anatomic and vascular information that can be extracted from thermal images is unique to each individual [13].

In addition, in recent years, the price of thermal cameras has decreased significantly, and their technology has improved, obtaining better resolution and quality, and the fixed pattern noise that was produced by old thermal cameras has been eliminated using non-uniformity correction techniques (NUC) [28,29]. Thus, the interest in the use of thermal technology in

* Corresponding author at: Advanced Mining Technology Center, Universidad de Chile, Av. Tupper 2007, 837-0451 Santiago, Chile.
E-mail address: jruizd@ing.uchile.cl (J. Ruiz-del-Solar).

0031-3203/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2012.01.001

face recognition applications has increased in recent years. Nevertheless, thermal face images still have undesirable variations due to (i) changes in ambient temperature, (ii) modifications of the metabolic processes of the subjects, (iii) camera susceptibility on extrinsic factors such as wind, and (iv) variable sensor response overtime when the camera is working for long periods of times [10,13,39].

In this general context, the aim of this article is to carry out a comparative study of thermal face-recognition methods in unconstrained environments. The main motivation is the lack of direct[1] and detailed comparisons of these kinds of method working under the same conditions. The results of this comparative study are intended to be a guide for developers of face recognition systems. This study concentrates on methods that fulfill the following requirements: (i) *Full online operation*: No offline enrollment stages. All processes must run online. The system has to be able to build the face database incrementally from scratch; (ii) *Real-time operation*: The recognition process should be fast enough to allow real-time interaction in case of HRI (Human-Robot Interaction) applications, and to search large databases in a reasonable time (a few milliseconds or some seconds depending on the application and the size of the database); (iii) *Single image per person problem*: One thermal face image of an individual should be enough for his/her later identification. Databases containing just one face image per person should be considered. The main reasons for this are savings in storage and computational costs, and the impossibility of obtaining more than one face image from a given individual in certain situations; and (iv) *Unconstrained environments*: No restrictions on environmental conditions such as illumination, indoor/outdoor setup, facial expression, scale, pose, resolution, accessories, occlusions, and background are imposed.

Thus, in this study three local-matching and two global image-matching method are selected by considering their fulfillment of the previously mentioned requirements, and their good performance in former comparative studies of face recognition methods [35,36,41,46]. Two local-matching methods, namely, histograms of LBP (Local Binary Pattern) features [3] and Gabor Jet Descriptors with Borda count classifiers [46] are selected based on their performance in the studies reported in [36,46]. The third local-matching method, histograms of WLD (Weber Linear Descriptor) features, recently proposed in [12], has shown very good performance in face detection applications, and is used here for the first time in face recognition. The SIFT (Scale-Invariant Feature Transform) image-matching method [26] is included following its good performance in a former face recognition study [36]. Finally, the SURF (Speeded Up Robust Features) image-matching method [5], which is inspired by the SIFT method, is included because in many applications in which the SIFT method is used, SURF obtains a similar performance and a higher speed.

The comparative study is carried out using the Equinox and UCHThermalFace databases. The Equinox database [16] was selected because it is one of the most frequently employed thermal face databases, and therefore it allows comparing the obtained results with former studies. The UCHThermalFace database was specially designed to study the problem of unconstrained face recognition in the thermal domain. The database incorporates thermal images acquired in indoor and outdoor setups, with *natural* variations in illumination, facial expression, pose, accessories, occlusions, and background. This database will be made public for future comparative studies, which is also a contribution of this paper.

This comparative study intends to be a complement to the recently published comparative study on visible-spectrum face recognition methods in unconstrained environments [36].

The paper is structured as follows: Related works are outlined in Section 2. The methods under analysis are described in Section 3. In Sections 4 and 5 the comparative analyses of these methods in the Equinox and UCHThermalFace databases are presented. Finally, in Section 6 results are discussed, and conclusions are given.

2. Related work

Several comparative studies of thermal face recognition approaches have been developed in recent years [37,38,40]. Most of the developed approaches make use of appearance-based methods, such as PCA (Principal Component Analysis), LDA (Linear Discriminant Analysis), and ICA (Independent Component Analysis), which project face images into a subspace where the recognition is carried out. These methods achieve a ~95% recognition rate in experiments that do not consider real-world conditions (unconstrained environmental conditions), as in [37,38,40], or when using the Equinox thermal face database [16]. The Equinox database is *de facto* standard database in thermal face recognition. It consists of indoor images of 91 individuals, captured with 3 different expressions and 3 different illumination conditions (see details in Section 4).

Other reported thermal face recognition approaches are based on the use of local-matching: Local Binary Pattern (LBP) [27] and Gabor Jet Descriptors (GJD) [2,20]. In the Equinox thermal database, a recognition rate of ~97% for the LBP approach [27] and ~80% for the GJD approach [2,20] has been reported. Methodologies based on global matching, such as Scale Invariant Features Transform (SIFT), have also been used for thermal face recognition [19,20]. These approaches are based on the use of local feature descriptors that are invariant to rotation, translation and scale changes. These local descriptors are used to match pairs of images by considering geometrical and probabilistic restrictions. In [19] the SIFT methodology is used to obtain the descriptors directly in the thermal face images, while in [20] they are computed in the vascular images generated by processing the thermal images. These approaches obtained a recognition rate that depended strongly on the database used; ~80% when using Equinox [20] and ~95% when using a non-public database [19].

Recent work uses vascular information of the face in order to develop thermal face recognition systems. This is accomplished by detecting thermal minutia points, and then matching them using a similar approach to the one used for fingerprint identity verification [8–11]. This kind of methodology achieves a ~80% recognition rate in a non-public database. In [4], an efficient approach for the extraction of physiological features from thermal face images is presented. The features represent the network of blood vessels under the skin of the face. This network is unique to each individual, and can be used to develop thermal face recognition systems. In [14] a similar approach based on thermal *faceprints* is presented. This approach uses new feature sets to represent the thermal face: the bifurcation points of the thermal pattern and the center of gravity of the thermal face region. The current study does not consider methods that are based on vascular information, because they still need to be improved in order to achieve the same performance as local-matching methods. They will be considered in a future version of this study.

In addition, in [6,15,31] methodologies based on the fusion of visible and thermal spectrum images are proposed. In [6,15]

[1] We mean a comparison carried out by a single research group, that makes an objective comparison of the methods, without considering particular details or tricks used in the different implementations.

standard appearance-based methods are used together with genetic algorithms for the analysis and fusion of visible and thermal data. The method achieved a recognition rate of ~96% in the Equinox database. In [1] two schemes of fusion, data fusion, and decision fusion, are applied. The algorithm is designed to detect and replace eyeglasses with an eye template in the case of thermal images. Commercial face recognition software, *FaceIt*, is used in the evaluation of the fusion algorithm. In [22], the fusion of visual and thermal images using the Discrete Wavelet Transform (DWT) domain is described. The results of the experiments demonstrate that the fusion method is effective in terms of visual quality compared to conventional fusion approaches. In [2], Gabor filters are used to extract facial features in the thermal and visual domains. In [7] different levels of fusion between visual and thermal data are analyzed. In [31] the advantages of combining thermal and visible face recognition are analyzed, and the recognition is achieved using a *k*-nearest neighbor classifier. The current study focuses on pure thermal-based methods. However, its results could be used in order to select the best thermal method to be used with methods that use visible images.

Although, it is out of the scope of this study, it is worth mentioning that thermal face images have also been used for facial expression recognition [21,42–44]. Some of the face recognition methodologies under analysis could be adapted for facial expression recognition. In [45] features extracted from thermal images, in addition to visible images, are used as inputs to a neural network that is in charge of recognizing facial expressions. In [44] the local temperature-difference caused by the rearrangement of facial muscle and inner temperature change is used as input data for a neural network used for face expression recognition. In [42] an unsupervised local and global feature extraction to solve the problem of facial expression is proposed, and the classification is made using Support Vector Machines. In [21] the facial skin temperature is used for recognizing and classifying positive and negative expressions. The classification results suggest that facial skin temperature can be used to help distinguish between positive and negative facial expressions and can assist in interpreting affective states. In [43] a wavelet transform is used to analyze multi-scale, multi-direction changes of thermal texture regarded as texture features of images.

In [20] the authors presented a preliminary comparative study of thermal face recognition methods that did not consider real-world conditions (it did not use the UCHThermalFace database; it only used the Equinox database), nor the use of the recently proposed WLD descriptors. The current work is an extension of that work that overcomes all these limitations.

3. Methods under comparison

As mentioned above, the methods under comparison were selected considering their fulfillment of the defined requirements (real-time, fully online, just one image per person), and their performance in former comparative studies of face-recognition methods [35,36,41,46] and in face detection applications [12].

3.1. LBP histograms

Face recognition using histograms of LBP (Local Binary Pattern) features was originally proposed in [3], and has been used by many groups since then. In the original approach, three different levels of locality are defined: pixel level, regional level, and holistic level. The first two levels of locality are achieved by dividing the face image into small regions from which LBP features are extracted and histograms are used for efficient texture information representation. The holistic level of locality, i.e. the global description of the face, is obtained by concatenating the regional LBP extracted features. The recognition is performed using a nearest neighbor classifier in the computed feature space, using one of the three following similarity measures: histogram intersection, log-likelihood statistic, or Chi square. We implemented this recognition system, without considering preprocessing (cropping using an elliptical mask and histogram equalization are used in [3]), and by choosing the following parameters: (i) images divided in 10 (2×5), 40 (4×10), or 80 (4×20) regions, instead of using the original divisions which range from 16 (4×4) to 256 (16×16), and (ii) using the mean square error as the similarity measure, instead of the log-likelihood statistic, in addition to histogram intersection and Chi square.

3.2. Gabor jet descriptors

Local-matching approaches for face recognition in the visible spectrum are compared in [46]. The study analyzes several local feature representations, classification methods, and combinations of classifier alternatives. Taking into account the results of their study, the authors implemented a system that integrates the best possible choice at each step. That system uses Gabor Jet Descriptors as local features, which are uniformly distributed over the images, one wave-length apart. In each grid position of the test and gallery images, and at each scale (multiscale analysis), the Gabor jets are compared using normalized inner products, and these results are combined using the Borda-Count method. In the Gabor feature representation, only Gabor magnitudes are used, and 5 scales and 8 orientations of the Gabor filters are adopted. We implemented this system using all parameters described in [46] (i.e. filter frequencies and orientations, grid positions, face image size).

3.3. SIFT descriptors

Wide-baseline matching approaches based on local interest points and descriptors have become increasingly popular and have experienced impressive development in recent years. Typically, local interest points are extracted independently from both a test and a reference image, and then characterized by invariant descriptors, and finally the descriptors are matched until a given transformation between the two images is obtained. Lowe's system [26], using SIFT descriptors and a probabilistic hypothesis rejection-stage, is a popular choice for implementing object-recognition systems, given its recognition capabilities, and near real-time operation. However, the main drawback of Lowe's system is the large number of false positive detections. This drawback can be overcome using several hypothesis rejection stages, as for example in the L&R system [33]. This system has already been used in the construction of robust fingerprint verification systems [33], for off-line signature verification [32] and for face recognition [36]. In this work, we use Lowe's system and the L&R system to build a face-recognition system. In the SIFT-Lowe case, we use only the number of matched key points to evaluate the performance of recognition (*matches* flavor). In the SIFT-L&R case, we consider 2 variants: the number of matched key points (*matches* flavor) and the number of matched key points after several probabilistic hypothesis rejection stages (*simple* flavor). See description in [33].

3.4. SURF descriptors

The SURF (Speeded Up Robust Features) method [5] proposes procedures to compute local interest points and descriptors at a higher speed than the SIFT approach. This is achieved using: (i) the so-called Fast-Hessian detector, which approximates

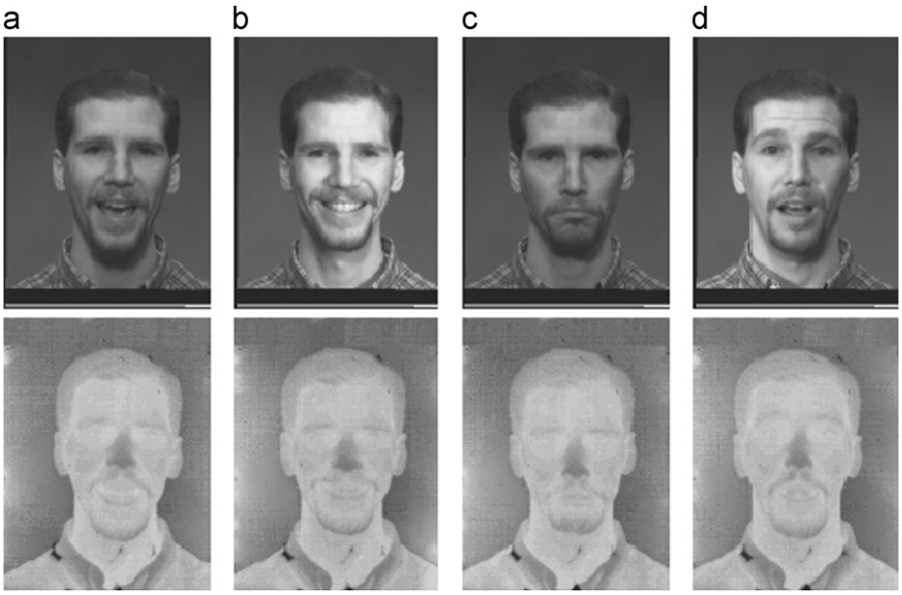

Fig. 1. Sample images from the Equinox database (taken from [17]). Top row: visible; bottom row: thermal. (a) speaking (frontal illumination), (b) smile (right illumination), (c) frown (frontal illumination), (d) surprise (left illumination).

Gaussians filters by box-filters and convolutions by the use of integral images, and (ii) SURF descriptors that are simpler to compute than SIFT descriptors, and correspond to the distribution of Haar-wavelets responses within the interest-point neighborhood. In addition, SURF descriptors have a lower number of components than SIFT descriptors. In this work we use the OpenSURF implementation provided by [30]. As in the SIFT-Lowe case, we use only the number of matched descriptors as a similarity measure in order to evaluate the performance of recognition (*matches* flavor).

3.5. WLD histograms

The WLD (Weber Linear Descriptor) descriptor [12] is inspired by the Weber's Law, and computes a two dimensional histogram of local intensity (differential excitations) and orientations. First, for a given pixel, the differential excitation component is computed as the ratio between the relative intensity differences of a current pixel against its neighbors, and the intensity of the current pixel descriptor. The orientation component is the gradient orientation of the current pixel. Afterwards, the 2D histogram of differential excitations and orientations is built. Thus, like LBP, WLD is a dense descriptor computed for every pixel, which makes it different from SIFT which corresponds to a sparse descriptor. In addition, WLD has much smaller pixel-granularity than SIFT, given that it is computed in a smaller region. In this work, the recognition is performed using a nearest neighbor classifier in the computed feature space using one of the following similarity measures between histograms: histogram intersection, Euclidean distance, or Chi square. As in the case of LBP, images are divided into a variable number of regions (10, 40, or 80), and the obtained histograms for each region are concatenated to obtain the descriptor. The 2D histograms are quantized to 8 orientations and 64 differential excitation values.

3.6. Notation: Methods and variants

We use the following notation to refer to the methods and their variations: A–B–C. (i) A describes the name of the face-recognition algorithm: LBP—Histogram of LBP features, WLD—Histogram of WLD features, GJD—Gabor Jet Descriptors, SIFT-Lowe—Lowe's original system with SIFT descriptors, SIFT-L&R—L&R system with SIFT descriptors, and SURF; (ii) B denotes the similarity measure or classification approach: HI—Histogram Intersection, XS—Chi square, EU—Euclidian Distance, BC—Borda Count, except for the case of SIFT and SURF, methods which do not use any explicit distance measure; (iii) C describes additional parameters: number of divisions in the case of the LBP-based and WLD-based methods, and matching method in the case of SIFT and SURF methods (*M*: number of matches; *S*: simple).

4. Comparative study using the equinox database

The performance of the methods is analyzed using the Equinox database [16], which has already been used to validate several face recognition methods for thermal images. This allows the direct comparison of the selected methods with face recognition methods implemented in previous works: in [37,38] methods based on LDA (Linear Discriminant Analysis), LFA (Local Feature Analysis), and ICA (Independent Component Analysis) have been compared. PCA (Principal Component Analysis) has been analyzed in [6,37,40]. Other methods that have been analyzed using this database are KPCA (Kernel PCA) and KFLD (Kernel Fisher Linear Discriminant) in [15], LBP (Local Binary Patterns) in [27], and Wavelets in [6].

The Equinox database consists of 18,629 thermal images (8–12 μm) and 18,629 visible images of 240×320 pixel size. All thermal images were radiometrically calibrated (see [40]). The database contains images of 91 individuals obtained under 3 different illuminations (frontal, left lateral, and right lateral), 3 different expressions (smile, frown, surprise), and also 40-frame sequences (4 s) in which the subject is speaking (vowels). People using eyeglasses were captured twice, with and without glasses. Fig. 1 shows some examples of IR and visible images obtained under different conditions. For 3 of the 91 individuals, there are only images where the subjects wear glasses. Therefore, it was not possible to perform the complete evaluation for these subjects. To have a fair evaluation, we removed these 3 individuals from the database, which finally contains 88 individuals.

The Equinox evaluation methodology names the thermal face images as *frames*. Following the Equinox terminology, frames 0, 3,

Table 1
Equinox frames (face images) description (see main text for details).

VA	Vowel frames	All subjects	All illuminations
EA	Expression frames	All subjects	All illuminations
VF	Vowel frames	All subjects	Frontal illumination
EF	Expression frames	All subjects	Frontal illumination
VL	Vowel frames	All subjects	Lateral illumination
EL	Expression frames	All subjects	Lateral illumination
VG	Vowel frames	Subjects using glasses	All illumination
EG	Expression frames	Subjects using glasses	All illuminations
RR	500 frames	Chosen at random	All illuminations

9 (from the 40 frames where the subject was speaking) will be called *vowel frames*, and frames "smile", "frown", "surprise" will be called *expression frames*. Multiple sets, as shown in Table 1 are built and used as test and gallery sets. For more details see [40]. Thus, each of these sets (VA, EA, VF, EF, VL, EL, VG, EG, and RR) may be used in the different experiments, and they can be part of a gallery set or a test set. The recognition experiments are carried out following the Equinox methodology [38,40] and for each of the 54 gallery-test set experiments, the top-1 recognition rate is computed.

Mean and standard deviation values of the top-1 recognition rates for the 54 gallery-test set experiments, as well as results for the most difficult test/gallery pairs (VL/EF, EF/VL, VF/VG, EF/EG) are presented in Table 2. These results were obtained using the same experimental conditions reported in previous works [37,38,40]. This experiment consists of a full gallery experiment that includes a gallery with between 3 and 9 images of each individual for each case. We use the evaluation methodology proposed by Equinox, which consists of evaluating the cross validation using all subsets. However we do not include the subsets VA or EA when they are subsets of the gallery or the test set, as in the results reported in [37,38,40]. Some results presented in these studies include those subsets in the evaluation, but the correct procedure would be to not consider them in order to avoid over-fitting. The obtained results can be summarized as:

- WLD-X-X variants show the best performance, even in difficult cases (e.g. glass). WLD-HI-80 obtains the best results.
- LBP-X-X variants obtain the second best results.
- SIFT-based, SURF-based, and Gabor-based methods do not obtain good results compared to WLD-X-X and LBP-X-X. It seems that these methods are affected by the quality of the face images, in this case the low quality of Equinox images. When these methods are used in other databases, such as those in the results presented later in Section 5, these methods show higher performance.
- Appearance-based methods also do not obtain good results compared to WLD-X-X and LBP-X-X. As it can be observed, appearance-based methods cannot properly handle variations produced by facial expressions or artifacts (e.g. eyeglasses).

In a second set of experiments the methods are compared with each other by considering more realistic conditions: just 1 image per individual in the gallery and with images of size 150×81 pixels. The same evaluation methodology described above is used, but in this case the comparison is restricted to the five methods under comparison since appearance-based methods are not able to work properly when just 1 image per individual in the gallery is used (They need at least 2–3 images in order to build a proper face model). From Table 3 it can be observed that:

- Best results are obtained by WLD-HI-80. All WLD-X-X variants have good recognition performance in this case.

Table 2
Equinox database. Full gallery experiments (the galleries include between 3 and 9 images of each individual). The diagonals (where gallery and test sets correspond to the same set) and the cases containing test sets that are subsets of gallery sets are not considered. Mean/SDV: Mean and standard deviation values of top-1 recognition rates of 54 gallery/test set pairs (see main text for details).

Methods	Mean (%)	SDV	VL/EF (%)	EF/VL (%)	VF/VG (%)	EF/EG (%)
PCA [40]	96.0	3.2	86.8	91.2	99.7	98.3
PCA [38]	95.0	2.8	90.5	91.3	99.3	97.3
LDAg [38]	97.0	–	–	–	–	–
LDAt [38]	98.0	1.67	95.6	95.8	99.3	99.7
LFAe [38]	93.0	–	–	–	–	–
LFAb [38]	93.0	4.6	83.5	82.7	97.7	95.9
ICA [38]	94.0	3.9	87.9	85.7	99.3	96.8
PCA [37]	95.0	2.6	90.9	91.6	99.3	97.3
LDAt [37]	97.0	2.2	91.4	92.9	99.0	98..7
LDAg [37]	98.0	1.8	95.1	95.6	99.3	99.7
LFAb [37]	91.0	4.8	83.5	82.9	97.0	96.1
LFAe [37]	92.0	4.9	84.2	82.9	97.7	95.1
ICA [37]	94.0	3.9	86.8	86.2	99.3	96.6
Wavelets [6][b]	93.5	4.4	91.2	90.4	–	–
PCA [6][b]	92.9	4.3	91.2	90.4	–	–
KPCA [15][c]	82.7	–	–	–	–	–
KFLD [15][c]	96.3	–	–	–	–	–
LBP [27][d]	97.3	1.8	99.2	96.6	–	–
LBP [27] with NUC[d]	93.3	4.3	93.6	90.2	–	–
LBP-EU-10	95.3	2.9	88.9	90.4	97.4	93.0
LBP-HI-10	98.4	1.0	96.2	97.4	99.3	97.1
LBP-XS-10	98.8	0.8	97.6	98.0	99.5	97.1
LBP-EU-40	97.2	1.8	92.7	95.7	98.8	94.7
LBP-HI-40	98.5	1.0	96.5	98.1	99.5	96.2
LBP-XS-40	98.3	1.2	95.5	97.8	99.8	95.2
LBP-EU-80	97.9	1.3	94.8	97.4	99.0	95.4
LBP-HI-80	98.7	0.9	96.7	98.5	99.5	96.6
LBP-XS-80	98.1	1.2	94.6	98.1	98.6	96.4
GJD-BC[a]	91.2	4.8	85.3	82.4	97.1	91.1
WLD-EU-10	98.1	1.4	96.5	96.8	100	96.2
WLD-HI-10	99.0	0.8	98.1	98.9	100	97.8
WLD-XS-10	98.9	0.8	98.3	98.6	100	96.9
WLD-EU-40	98.5	1.1	96.5	98.2	99.3	96.4
WLD-HI-40	98.9	0.7	97.9	99.1	99.5	97.8
WLD-XS-40	98.7	1.0	97.2	98.1	99.8	96.4
WLD-EU-80	98.9	0.7	97.4	98.7	99.3	97.6
WLD-HI-80	99.1	0.7	98.1	99.3	99.8	97.8
WLD-XS-80	98.9	0.8	97.4	98.7	99.8	96.6
SIFT-L&R-M[a]	87.8	3.3	87.2	85.3	91.8	86.3
SIFT-L&R-S[a]	93.9	2.6	93.1	91.2	96.9	92.8
SIFT-Lowe-M[a]	91.9	3.3	84.9	92.0	96.2	92.8
SURF-M[a]	80.5	6.7	75.2	69.9	88.2	68.8

NUC: Non uniformity correction.

[a] Histogram stretching, range 0–255.
[b] The used subsets do not include glasses neither some cross relationship.
[c] It used eyes glasses and expression test subsets.
[d] The used subsets do not include glasses.

- The LBP-X-40 and LBP-X-80 obtain good results, followed by the other LBP and SIFT variants.
- GJD-BC and SURF-M present the worst results.

In summary, this first comparison includes aspects such as variable illumination, facial expression variations, facial variations observed when speaking, and the use of eyeglasses, aspects that directly influence face recognition accuracy. The best results are obtained using WLD-X-X and LBP-X-X. When compared to their behavior in other databases, GJD-BC, SIFT-X-X, and SURF-M obtain a lower performance. The main reason seems to be the low variability in the pixel values of the Equinox thermal images (the pixel values are highly quantized). Even when histogram stretching is applied, GJD-BC, SIFT-X-X, and SURF-M do not improve their performance. Therefore, the five methods under comparison are

Table 3
Equinox database. The galleries include just 1 image of each individual. All galleries, except VG, EG, and RR are used. Mean/SDV. Mean and standard deviation values of top-1 recognition rates of 36 gallery/test set pairs (see main text for details). The best variant of each method under analysis is presented in bold type.

Methods	Mean (%)	SDV	VL/EF (%)	EF/VL (%)	EL/VF (%)	VA/EF (%)
LBP-EU-10	77.1	6.0	75.1	66.7	76.4	75.1
LBP-HI-10	88.4	3.5	86.7	83.0	85.4	87.7
LBP-XS-10	91.4	3.7	92.6	85.9	87.2	91.9
LBP-EU-40	84.6	5.2	84.2	76.2	81.9	83.9
LBP-HI-40	91.0	3.6	90.9	85.6	86.8	90.9
LBP-XS-40	91.0	3.7	91.2	85.9	87.8	89.8
LBP-EU-80	88.1	4.2	83.0	86.3	82.1	85.5
LBP-HI-80	**92.5**	**3.0**	**88.7**	**90.9**	**89.7**	**90.0**
LBP-XS-80	89.8	3.6	85.6	88.4	85.0	86.9
GJD-BC[a]	**71.7**	**7.4**	**63.9**	**63.4**	**66.3**	**70.9**
WLD-EU-10	88.1	4.1	88.4	83.5	83.7	86.7
WLD-HI-10	92.2	2.6	92.3	89.4	88.2	91.6
WLD-XS-10	92.8	3.2	94.0	88.9	87.8	93.3
WLD-EU-40	91.5	4.1	91.6	87.2	87.2	93.3
WLD-HI-40	93.5	3.3	94.0	90.3	88.9	94.0
WLD-XS-40	93.2	3.6	94.4	89.4	89.2	94.0
WLD-EU-80	93.9	3.2	91.0	95.4	90.3	92.2
WLD-HI-80	**95.0**	**2.9**	**91.7**	**95.8**	**92.9**	**92.9**
WLD-XS-80	94.4	3.2	90.5	94.4	91.2	92.1
SIFT-L&R-M[a]	**77.4**	**9.9**	**75.8**	**75.0**	**74.3**	**79.6**
SIFT-L&R-S[a]	76.6	6.2	72.3	74.1	67.7	76.1
SIFT-Lowe-M[a]	**82.2**	**5.2**	**70.9**	**82.1**	**80.9**	**79.3**
SURF-M[a]	**63.3**	**8.7**	**55.4**	**54.8**	**55.2**	**59.3**

[a] Histogram stretching, range 0–255.

further analyzed in the following section using the UCHThermalFace database.

5. Comparative study using the UCHThermalFace database

The methods under study are analyzed considering real-world conditions that include indoor/outdoor setups and natural variations on facial expression, pose, accessories, occlusions, and background.

5.1. Database description

The UCHThermalFace database is divided into three sets, *Rotation*, *Speech*, and *Expressions*. The Rotation and Speech sets consist of indoor and outdoor thermal face images of 53 subjects obtained under different yaw and pitch angles, as well as a set of images captured while the subjects were speaking. The Expressions set consists of thermal images of 102 subjects captured in an indoor setup. The thermal images were acquired using a FLIR 320 TAU Thermal Camera[2] , with sensitivity in the range 7.5–13.5 μm, and a resolution of 324×256 pixels.

The Rotation set contains 22 images per subject, each one corresponding to a different rotation angle acquired in indoor and outdoor sessions (see experimental setup in Fig. 2). In both cases, indoor and outdoor, the distance from the subject to the thermal camera was fixed at 120 cm, and the thermal camera was situated at position P6 (see Fig. 2). The face images were acquired while subjects were observing positions 1–11 (see Fig. 2), which correspond to the following rotation angles: R1: (yaw$=-15^\circ$, pitch$=15^\circ$), R2: (yaw$=0^\circ$, pitch$=15^\circ$), R3: (yaw$=15^\circ$, pitch$=15^\circ$), R4: (yaw$=-30^\circ$, pitch$=0^\circ$), R5: (yaw$=-15^\circ$, pitch$=0^\circ$), R6: (yaw$=0^\circ$, pitch$=0^\circ$), R7: (yaw$=15^\circ$, pitch$=0^\circ$), R8: (yaw$=30^\circ$, pitch$=0^\circ$), R9: (yaw$=-15^\circ$, pitch$=-15^\circ$), R10: (yaw$=0^\circ$, pitch$=-15^\circ$), R11: (yaw$=15^\circ$, pitch$=-15^\circ$). Fig. 3 shows an example of 11 thermal face images corresponding to one individual of the database, acquired under different yaw and pitch angles in an indoor session. In addition to the rotation set, a video sequence was captured while each subject was observing point P6 and speaking the word, "Pa-ra-le-le-pi-pe-do", in the indoor and in the outdoor session. Later on, three frames were randomly selected from the video sequence of each individual in each session (indoor and outdoor). These images form the Speech set, which essentially contains images with different facial expressions.

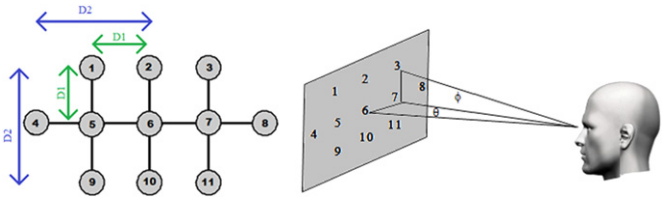

Fig. 2. Experimental setup for image acquisition at different yaw (θ) and pitch (ϕ) angles. The distance between the individual and the observed point *P*6 is 120 cm. *D*1 is 32.15 cm and *D*2 is 69.28 cm. See main text for details.

The Expression is captured in a different setup. In this setup subjects observe frontally the camera at a fixed distance of 150 cm. First, images were acquired while subjects were expressing three different expressions "Happy", "Sad", and "Angry". In addition a video sequence was captured while each subject was speaking different vowels. Later on, three frames were randomly selected from the video sequence of each individual.

In summary, for the Rotation and Speech sets, 14 indoor and 14 outdoor subsets are defined in order to carry out face recognition experiments. For the indoor session and the outdoor session, 11 subsets correspond to the different yaw-pitch combinations of the Rotation set (subsets R1 to R11), and 3 to the different images captured in the Speech set (subsets S1 to S3). In the Expressions sets, 3 expressions ("Happy", "Sad", and "Angry"; E1 to E3) and 3 vowels (V1 to V3) subsets are defined. The experiments reported in the next section make use of the 34 defined subsets. In each experiment a given subset is used as a test set, and a second one as a gallery set.

5.2. Description of experiments

In order to evaluate the face recognition methods under analysis, six kinds of experiments were carried out: (i) variable window size, (ii) partial face occlusions, (iii) eye detection accuracy, (iv) indoor versus outdoor galleries, (v) facial expressions, and (vi) variable distance. In all experiments face images are aligned using the annotated eye position; faces are aligned by centering the eyes in the same relative positions, at a fixed distance of 42 pixels between the eyes, except in the case of the variable distance experiments where the distanced between the eyes is decreased accordingly with the reduction in the image resolution. The experiments are:

Variable Window Size. The effect of using different window sizes in the performance of the methods is analyzed. Increasing the size of the windows corresponds to adding or removing different amounts of background to the region being analyzed, given that we are not decreasing the scale of the faces (there is no change in image resolution). Thus, experiments were performed including window sizes of 81×150 pixels, 100×185 pixels, and 125×225 pixels (see Fig. 4 for examples).

Partial Face Occlusions. In order to analyze the behavior of the different methods in response to partial occlusions of the face area, images were divided into 10 different regions (2 columns

[2] ⟨http://www.flir.com/cvs/cores/uncooled/products/tau/⟩

Fig. 3. Example of the 11 thermal face images of the rotation set of one individual of the UCHThermalFace database, captured at the indoor session.

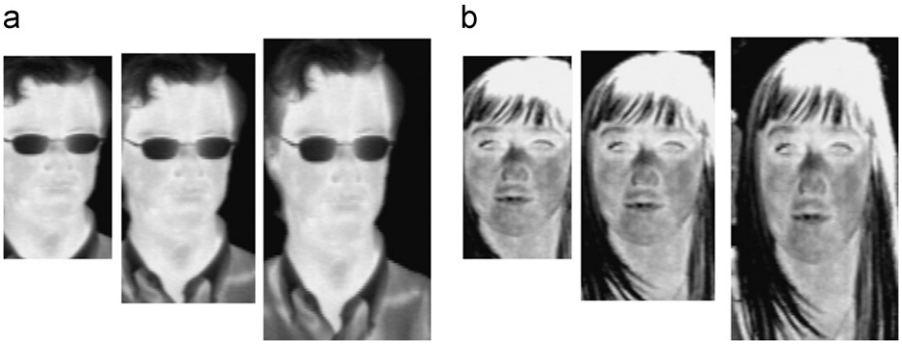

Fig. 4. Examples of faces with different cropping sizes (UCHThermalFace database). (a) Indoor session. Window size (in pixels): 81×150, 100×185, 125×225. (b) Outdoor session. Window size (in pixels): 81×150, 100×185, 125×225.

Fig. 5. Example of images with partial occlusion (UCHThermalFace database). (Left) Indoor session. Window size 125×225. (Right) Outdoor session. Window size 125×225.

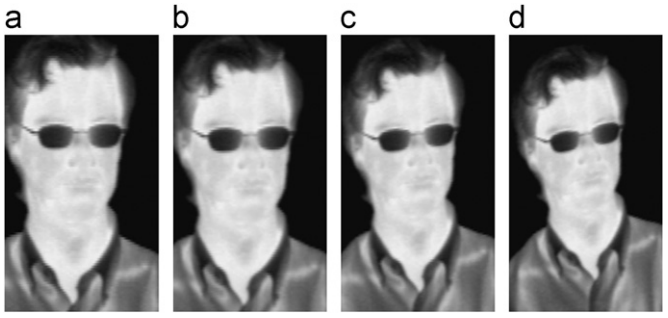

Fig. 6. Examples of aligned images after adding noise in the eyes' position (UCHThermalFace database). Indoor session. Window size 125×225 pixels. (a) No noise; (b) 2.5% noise; (c) 5% noise; (d) 10% noise.

and 5 rows), and one of the regions was randomly selected and its pixels set to 0 (See Fig. 5 for some examples). This experiment analyzes the behavior of the different methods in response to these partial occlusions.

Eye Detection Accuracy. Most face-recognition methods are very sensitive to face alignment, which depends directly on the accuracy of the eye detection process; eye position is usually the primary, and sometimes the only, source of information for face alignment. In order to analyze the sensitivity of the different methods on eye position accuracy, we added white noise to the position of the annotated eyes in the test images. The noise was added independently to the x and y eye positions of each eye, using the procedure described in [36]. In the different experiments, the noise can take up to 2.5%, 5%, or 10% of the distance between the eyes. Examples are shown in Fig. 6. Note that this affects both the position and the scale of the face in the image.

Indoor versus Outdoor Galleries. The performance of face recognition methods depends largely on environmental conditions, particularly under the indoor or outdoor conditions. In these experiments the test and gallery images correspond to images taken in an indoor session or in an outdoor session. When the test images are indoor images, then the gallery images are outdoor images, and vice versa. The outdoor images were

captured in summer (with high temperatures up to 30 °C), and at times the faces, as well as the camera, were receiving direct sunlight.

Facial Expressions. The effect of having different facial expressions in the subjects is analyzed. The experiments consider subjects with different facial expressions, as well as subjects speaking different vowels.

Variable Distance. The sensitivity of the methods to the distance between the subject and the camera is analyzed. The effect of having variable distances between the subject and the camera is simulated by decreasing the resolution of the thermal images.

The Expressions set is used (E1–E3, V1–V3) in the facial expressions experiments, while in all other experiments the Rotation and Speech datasets are used (R1–R11, S1–S3).

5.3. Recognition results

The performance of the different methods is evaluated using the top-1 recognition rate. In all experiments the rotation subset R6, without any occlusion and without noise in the eye position, is selected as a gallery set because it contains clean frontal views of the faces. Naturally, in the indoor experiments the indoor R6 subset is used, while in the outdoor experiments the outdoor R6 subset is employed.

In Tables 4–9, the top-1 recognition rate is given separately for the 2 different sets, Rotation and Speech. For each category a mean value is calculated: the mean recognition rate over all rotation subsets for the rotation set, and the mean recognition rate over all speech subsets. In addition, the average between these two results is given. Figs. 8 and 9 include just this average value. Table 10 reports average results obtained using the Expressions set.

Variable Window Size. Table 4 shows the performance of the different methods when different window sizes are used in the test and gallery sets. In these experiments all gallery and test sets correspond to indoor images.

It can be concluded from the experiments that the best window size depends on the method. For SIFT-X-X, SURF-M, and GJD-BC the best size is 125×225, while for WLD-X-X and LBP-X-X the best size is 81×150. These results are consistent with the ones obtained in [36] for GJD-BC and LBP-X-X methods. As noted in that study, GJD-BC works better with windows that contain large portions of background. The reason is three-fold: (i) the Gabor-filters encode information about the contour of the face, (ii) large regions allow the use of large filters, which encode large-scale information, and (iii) the Borda Count classifier may reduce the effect of regions (or jets) that are not relevant for the recognition. For the LBP-X-X variants, adding some background does not help, and in most cases reduces the performance. We can observe that WLD variants have similar behavior compared to that of LBP variants in terms of optimal window size. This seems to be due to the similar kind of analysis of the face information that both methods carry out, which basically computes histograms of local features over non-overlapping regions of the face area. In the cases where the histograms include background information (as in larger croppings of the image), the discriminability of the methodology decreases. Regarding SIFT-X-X

Table 4
UCHThermalFace database. Experiment using different windows sizes. Indoor session. Top-1 recognition rate. Rotation and speech test sets. (See main text for details). The best variant of each method under analysis is presented in bold type.

Methods	Rotation												Speech mean (%)	Average (%)
	R1 (%)	R2 (%)	R3 (%)	R4 (%)	R5 (%)	R6 (%)	R7 (%)	R8 (%)	R9 (%)	R10 (%)	R11 (%)	Mean (%)		
LBP-EU-80 81 × 150	54.72	77.36	54.72	50.94	81.13	100	90.57	39.62	73.58	92.45	84.91	72.73	81.76	77.24
LBP-HI-80 81 × 150	**73.58**	**88.68**	**73.58**	**64.15**	**96.23**	**100**	**96.23**	**50.94**	**90.57**	**100**	**96.23**	**84.56**	**92.45**	**88.51**
LBP-XS-80 81 × 150	64.15	90.57	73.58	62.26	88.68	100	92.45	41.51	86.79	100	81.13	80.10	89.94	85.02
GJD-BC 81 × 150	66.04	96.23	75.47	50.94	92.45	100	88.68	33.96	81.13	98.11	69.81	77.53	98.11	87.82
WLD-EU-80 81 × 150	**77.36**	**96.23**	**90.57**	**64.15**	**96.23**	**100**	**92.45**	**71.70**	**90.57**	**98.11**	**90.57**	**87.99**	**95.60**	**91.80**
WLD-HI-80 81 × 150	**73.58**	**96.23**	**88.68**	**64.15**	**94.34**	**100**	**96.23**	**66.04**	**90.57**	**100**	**92.45**	**87.48**	**95.60**	**91.54**
WLD-XS-80 81 × 150	73.58	94.34	90.57	64.15	88.68	100	92.45	62.26	92.45	96.23	84.91	85.42	94.97	90.20
SIFT-L&R-M 81 × 150	86.79	96.23	84.91	73.58	96.23	100	92.45	66.04	92.45	98.11	83.02	88.16	94.97	91.57
SIFT-L&R-S 81 × 150	86.79	100	81.13	56.60	98.11	100	96.23	50.94	86.79	98.11	81.13	85.08	98.11	91.59
SIFT-Lowe-M 81 × 150	92.45	100	84.91	73.58	98.11	100	96.23	66.04	94.34	100	90.57	90.57	99.37	94.97
SURF-M 81 × 150	69.81	94.34	67.92	50.94	86.79	100	71.70	30.19	83.02	92.45	71.70	74.74	89.94	82.19
LBP-EU-80 100 × 185	45.28	75.47	41.51	45.28	81.13	100	69.81	30.19	77.36	90.57	66.04	65.69	81.76	73.73
LBP-HI-80 100 × 185	67.92	86.79	67.92	54.72	98.11	100	92.45	39.62	94.34	100	84.91	80.62	93.08	86.85
LBP-XS-80 100 × 185	71.70	90.57	62.26	56.60	96.23	100	94.34	33.96	90.57	100	81.13	79.76	91.82	85.79
GJD-BC 100 × 185	83.02	98.11	84.91	50.94	94.34	100	96.23	37.74	79.25	98.11	79.25	81.99	100	91.00
WLD-EU-80 100 × 185	79.25	90.57	79.25	54.72	88.68	100	94.34	47.17	86.79	100	84.91	82.33	91.82	87.08
WLD-HI-80 100 × 185	69.81	88.68	81.13	56.60	92.45	100	96.23	45.28	90.57	100	86.79	82.50	92.45	87.48
WLD-XS-80 100 × 185	81.13	92.45	84.91	60.38	96.23	100	92.45	49.06	90.57	98.11	94.34	85.42	92.45	88.94
SIFT-L&R-M 100 × 185	94.34	98.11	83.02	77.36	98.11	100	96.23	71.70	92.45	100	81.13	90.22	94.34	92.28
SIFT-L&R-S 100 × 185	86.79	98.11	86.79	60.38	100	100	94.34	47.17	90.57	100	77.36	85.59	97.48	91.54
SIFT-Lowe-M 100 × 185	86.79	100	96.23	79.25	98.11	100	96.23	75.47	94.34	100	94.34	92.80	100	96.40
SURF-M 100 × 185	79.25	98.11	71.70	64.15	96.23	100	88.68	49.06	84.91	100	83.02	83.19	97.48	90.34
LBP-EU-80 125 × 225	39.62	73.58	37.74	24.53	66.04	100	62.26	24.53	58.49	84.91	49.06	56.43	81.76	69.10
LBP-HI-80 125 × 225	58.49	84.91	58.49	37.74	86.79	100	83.02	39.62	79.25	96.23	75.47	72.73	86.16	79.44
LBP-XS-80 125 × 225	60.38	90.57	58.49	39.62	90.57	100	90.57	39.62	84.91	98.11	79.25	75.64	90.57	83.11
GJD-BC 125 × 225	**84.91**	**98.11**	**84.91**	**49.06**	**96.23**	**100**	**92.45**	**41.51**	**86.79**	**100**	**83.02**	**83.36**	**99.37**	**91.37**
WLD-EU-80 125 × 225	62.26	92.45	69.81	49.06	88.68	100	84.91	33.96	79.25	96.23	81.13	76.16	89.94	83.05
WLD-HI-80 125 × 225	58.49	88.68	71.70	47.17	90.57	100	92.45	41.51	88.68	100	86.79	78.73	93.08	85.91
WLD-XS-80 125 × 225	67.92	94.34	73.58	50.94	92.45	100	94.34	47.17	88.68	98.11	86.79	81.30	95.60	88.45
SIFT-L&R-M 125 × 225	**86.79**	**100**	**94.34**	**73.58**	**100**	**100**	**94.34**	**73.58**	**96.23**	**100**	**94.34**	**92.11**	**98.11**	**95.11**
SIFT-L&R-S 125 × 225	92.45	100	90.57	67.92	100	100	94.34	56.60	98.11	100	92.45	90.22	99.37	94.80
SIFT-Lowe-M 125 × 225	**90.57**	**98.11**	**96.23**	**88.68**	**100**	**100**	**98.11**	**83.02**	**98.11**	**100**	**94.34**	**95.20**	**100**	**97.60**
SURF-M 125 × 225	**83.02**	**100**	**84.91**	**67.92**	**100**	**100**	**96.23**	**56.60**	**98.11**	**100**	**88.68**	**88.68**	**99.37**	**94.03**

Table 5
UCHThermalFace database. Best results windows size. Outdoor session. Top-1 recognition rate. Rotation and speech test sets. (See main text for details.)

Methods	Rotation												Speech mean (%)	Average (%)
	R1 (%)	R2 (%)	R3 (%)	R4 (%)	R5 (%)	R6 (%)	R7 (%)	R8 (%)	R9 (%)	R10 (%)	R11 (%)	Mean (%)		
LBP-HI-80 81×150	86.79	86.79	83.02	60.38	98.11	100	96.23	73.58	88.68	98.11	90.57	87.48	90.57	89.03
GJD-BC 125×225	62.26	98.11	71.70	39.62	88.68	100	94.34	35.85	60.38	96.23	67.92	74.10	86.79	80.45
WLD-EU-80 81×150	62.26	73.58	79.25	47.17	84.91	100	84.91	58.49	77.36	84.91	64.15	74.27	72.96	73.62
WLD-HI-80 81×150	69.81	86.79	92.45	60.38	96.23	100	94.34	66.04	84.91	96.23	79.25	84.22	96.23	90.23
SIFT-L&R-M 125×225	94.34	100	92.45	81.13	100	100	100	75.47	98.11	98.11	88.68	93.48	98.11	95.80
SIFT-Lowe-M 125×225	92.45	100	96.23	84.91	100	100	100	81.13	94.34	100	90.57	94.51	98.11	96.31
SURF-M 125×225	94.34	100	96.23	86.79	96.23	100	98.11	75.47	94.34	100	88.68	93.65	97.48	95.57

Table 6
UCHThermalFace database. Partial occlusion. Indoor session. Top-1 recognition rate. Rotation and speech test sets. (See main text for details.)

Methods	Rotation												Speech mean (%)	Average (%)
	R1 (%)	R2 (%)	R3 (%)	R4 (%)	R5 (%)	R6 (%)	R7 (%)	R8 (%)	R9 (%)	R10 (%)	R11 (%)	Mean (%)		
LBP-HI-80 81×150	60.38	77.36	62.26	58.49	96.23	100	90.57	47.17	83.02	100	81.13	77.87	86.79	82.33
GJD-BC 125×225	71.70	86.79	66.04	35.85	71.70	100	83.02	26.42	66.04	92.45	64.15	69.47	93.71	81.59
WLD-EU-80 81×150	75.47	96.23	86.79	62.26	96.23	100	92.45	71.70	84.91	98.11	90.57	86.79	93.08	89.94
WLD-HI-80 81×150	75.47	94.34	86.79	62.26	94.34	100	94.34	71.70	90.57	100	90.57	87.31	96.23	91.77
SIFT-L&R-M 125×225	75.47	100	86.79	64.15	98.11	100	92.45	66.04	90.57	98.11	88.68	87.31	93.87	90.59
SIFT-Lowe-M 125×225	86.79	98.11	96.23	73.58	92.45	100	94.34	67.92	94.34	100	88.68	90.22	98.78	94.50
SURF-M 125×225	71.70	98.11	73.58	49.06	94.34	100	88.68	43.40	84.91	96.23	73.58	79.42	95.60	87.51

methods, in [36] smaller window sizes obtain better performance in the case of visible images. The reason is that the background disturbs the recognition process in the case of visible images because incorrect matchings between local interest points corresponding to the face area and the background can be produced. However, this is not the case for thermal images, in which the background is usually uniform around the face area, because the temperature in this region is lower than that the one in the face and head of humans. This is an additional advantage of using thermal images.

In terms of the best overall performance, SIFT-X-X variants are the ones with the highest top-1 recognition rate, followed by the SURF-M method. The third best performance is obtained by WLD-X-X methods. In fourth place, GJD-BC and LBP-X-X obtain similar performance.

It can also be observed in Table 4 that SIFT-X-X variants are the most robust against rotations. The main reason seems to be that the SIFT-matching paradigm is robust against out-of-plane face rotations of up to $\sim 40^{\circ}$. For moderate yaw and pitch rotations (R2, R5, R7, and R10 test sets) most of the other methods behave well, and in most of the cases, when the appropriate window size is selected, top-1 recognition rates larger than 90% are obtained.

The SIFT-X-X and SURF-M methods are the most robust ones in terms of facial expression variations (Speech sets); in many cases these methods obtain a 100% top-1 recognition rate in the experiments reported in Table 4. The main reason is the way in which the SIFT/SURF matching paradigm works, by basically matching pairs of interesting points, and finding a coherent transformation between a subset of the matches (see example in Fig. 7). GJD-BC variants are also robust to facial expressions. One of the variants (GJD-BC, 100×185 pixels) obtains a 100% top-1 recognition rate in the experiments reported in Table 4. This robustness is achieved thanks to the use of an ensemble of classifiers, which is implemented by the Borda Count voting of the Gabor-jets. WLD-X-X and LBP-X-X variants also show some invariance to facial expression variations. When a window size of 81×150 pixels is used, they achieve a top-1 recognition rate of 95.6% and 92.5%.

For each method, the best results in top-1 recognition rate are obtained by the following variants (considering different parameters and window sizes):

- SIFT-Lowe-M, 125×225 pixels: 97.6%
- SIFT-L&R-M, 125×225 pixels: 95.1%
- SURF-M, 125×225 pixels: 94.0%
- WLD-EU-80, 81×150 pixels: 91.8%
- WLD-HI-80, 81×150 pixels: 91.5%
- GJD-BC, 125×225 pixels: 91.4%
- LBP-HI-80, 81×150 pixels: 88.5%.

In the case of the WLD and SIFT methods two variants are selected, either because the performance of the two variants is very similar (as in the WLD case), or they correspond to a different approach for implementing the method (as in the SIFT case). For the next experiments, results for only these seven variants are presented and analyzed.

Table 5 shows the performance of the seven best variants of the methods for the defined Rotation and Speech sets in the outdoor case. Best results are obtained by SIFT-X-X and SURF-M variants, followed by WLD-HI-80. Fourth best performance is obtained by LBP-EU-80, and the fifth place is taken by GJD-BC. The SIFT-Lowe-M method shows a high top-1 recognition rate (96.3%), which is very similar to the one obtained in the indoor case (97.6%). The SIFT-L&R-M and SURF-M methods show a similar behavior. In the case of WLD-HI-80 and LBP-HI-80, the performance is also similar in the indoor and outdoor cases. Interestingly, in the case of WLD-EU-80, a variant that uses the Euclidian distance, the performance decreases greatly in the outdoor case. The same happened for GJD-BC, whose performance decreased about 11 percentage points in the outdoors setting. These results indicate that GJD-BC does not behave appropriately in outdoor conditions.

Eye Detection Accuracy. Fig. 8 shows the method's sensitivity to eye detection accuracy in the indoor case, while Fig. 9 shows the method's sensitivity in the outdoor case.

Table 7
UCHThermalFace database. Partial occlusion. Outdoor session. Top-1 recognition rate. Rotation and speech test sets. (See main text for details.)

Methods	Rotation												Speech mean (%)	Average (%)
	R1 (%)	R2 (%)	R3 (%)	R4 (%)	R5 (%)	R6 (%)	R7 (%)	R8 (%)	R9 (%)	R10 (%)	R11 (%)	Mean (%)		
LBP-HI-80 81 × 150	73.58	73.58	75.47	52.83	90.57	100	79.25	52.83	81.13	92.45	67.92	76.33	81.13	78.73
GJD-BC 125 × 225	49.06	79.25	52.83	24.53	81.13	100	77.36	24.53	56.60	94.34	49.06	62.61	83.02	72.82
WLD-EU-80 81 × 150	52.83	71.70	71.70	39.62	81.13	100	81.13	47.17	64.15	81.13	50.94	67.41	71.70	69.56
WLD-HI-80 81 × 150	73.58	88.68	90.57	60.38	94.34	100	90.57	66.04	86.79	92.45	79.25	83.88	87.42	85.65
SIFT-L&R-M 125 × 225	84.91	98.11	86.79	77.36	100	100	98.11	69.81	94.34	98.11	88.68	90.57	94.34	92.46
SIFT-Lowe-M 125 × 225	81.13	98.11	92.45	77.36	98.11	100	100	75.47	94.34	100	84.91	91.08	97.48	94.28
SURF-M 125 × 225	79.25	98.11	92.45	73.58	92.45	100	92.45	54.72	94.34	94.34	73.58	85.93	97.48	91.71

Table 8
UCHThermalFace database. Different gallery sets: Indoor gallery set, Outdoor test sets. Top-1 recognition rate. Rotation and speech test sets. (See main text for details.)

Methods	Rotation												Speech mean (%)	Average (%)
	R1 (%)	R2 (%)	R3 (%)	R4 (%)	R5 (%)	R6 (%)	R7 (%)	R8 (%)	R9 (%)	R10 (%)	R11 (%)	Mean (%)		
LBP-HI-80 81 × 150	24.53	30.19	16.98	20.75	32.08	32.08	16.98	9.43	22.64	20.75	11.32	21.61	22.01	21.81
GJD-BC 125 × 225	30.19	39.62	24.53	18.87	35.85	54.72	30.19	7.55	28.30	33.96	16.98	29.16	42.77	35.97
WLD-EU-80 81 × 150	22.64	28.30	16.98	9.43	15.09	20.75	11.32	13.21	16.98	16.98	11.32	16.64	22.01	19.33
WLD-HI-80 81 × 150	28.30	35.85	28.30	13.21	33.96	33.96	18.87	11.32	26.42	35.85	18.87	25.9	32.08	28.99
SIFT-L&R-M 125 × 225	9.43	7.55	15.09	9.43	16.98	13.21	9.43	5.66	7.55	11.32	5.66	10.12	11.32	10.72
SIFT-Lowe-M 125 × 225	13.21	13.21	9.43	0.00	15.09	16.98	11.32	1.89	5.66	11.32	3.77	9.26	14.47	11.87
SURF-M 125 × 225	24.53	30.19	22.64	11.32	24.53	35.85	13.21	7.55	22.64	26.42	5.66	20.41	22.01	21.21

Table 9
UCHThermalFace database. Different gallery sets: Outdoor gallery set, Indoor test sets. Top-1 recognition rate. Rotation and speech test sets. (See main text for details.)

Methods	Rotation												Speech mean (%)	Average (%)
	R1 (%)	R2 (%)	R3 (%)	R4 (%)	R5 (%)	R6 (%)	R7 (%)	R8 (%)	R9 (%)	R10 (%)	R11 (%)	Mean (%)		
LBP-HI-80 81 × 150	13.21	15.09	3.77	11.32	24.53	24.53	13.21	7.55	24.53	24.53	15.09	16.12	20.13	18.13
GJD-BC 125 × 225	22.64	28.30	13.21	5.66	22.64	35.85	24.53	5.66	24.53	33.96	22.64	21.78	35.85	28.82
WLD-EU-80 81 × 150	15.09	20.75	13.21	15.09	30.19	24.53	20.75	11.32	22.64	22.64	15.09	19.21	25.79	22.50
WLD-HI-80 81 × 150	15.09	16.98	11.32	9.43	18.87	18.87	20.75	15.09	16.98	22.64	18.87	16.81	20.13	18.47
SIFT-L&R-M 125 × 225	18.87	20.75	18.87	7.55	11.32	20.75	9.43	5.66	9.43	16.98	9.43	13.55	16.04	14.80
SIFT-Lowe-M 125 × 225	9.43	7.55	11.32	3.77	9.43	15.09	11.32	7.55	13.21	9.43	7.55	9.61	16.35	12.98
SURF-M 125 × 225	7.55	18.87	7.55	3.77	13.21	24.53	15.09	3.77	15.09	24.53	11.32	13.21	27.04	20.13

Table 10
UCHThermalFace database. Gallery/test expressions sets: E1–E3, V1–V3. Average top-1 recognition rate of 6 experiments, in which 1 set is chosen as gallery and 1 as test. (See main text for details.)

Methods	Average (%)
LBP-HI-80	94.9
GJD-BC	94.4
WLD-EU-80	93.9
WLD-HI-80	94.1
SIFT-L&R-M	98.9
SIFT-Lowe-M	99.6
SURF-M	87.2

It can be observed that the highest top-1 recognition rate is obtained by SIFT-Lowe-M, and that this method is very robust to variations in the eye position, both in the indoor and outdoor case. Top-1 recognition rates vary between 97.6% and 96.8% when 10% noise is added in the eye position in the indoor experiments, and between 96.3% and 96.6% under the same level of noise in the outdoor experiments. The SIFT-L&R-M and SURF-M methods show a similar behavior. The main reason for this high robustness is the way in which the SIFT/SURF matching paradigm works. (see explanation in previous paragraphs and examples in Fig. 7).

In the indoor experiments, the performance of LBP-HI-80, WLD-X-80, and GJD-BC methods decrease uniformly with the level of noise in the eyes' position. When 10% noise is added in the eye position, the top-1 recognition rate decreases by ~12% in all cases. LBP-HI-80 and GJD-BC have the same behavior in the outdoor experiments (maximal decrease of ~12% top-1 recognition rate). However, outdoor conditions affect the WLD-X-80 variants a little more, when 10% noise is added in the eyes position the top-1 recognition rate decreases by ~15%. Nevertheless, in most cases LBP-HI-80 and WLD-HI-80 obtain the third best performance after the SIFT-X-X methods.

In addition, in Figs. 8 and 9 it can be observed that in some few cases the performance of the methods increased slightly with the noise. These small variations have no statistical significance and they are produced by the statistical nature of the methods, which includes the computation of histograms in the case of LBP-X-X and WLD-X-X variants, and the use of hypothesis rejection stages in the case SIFT-X-X variants, and by the fact that a small variation in the eyes position information can eventually improve the face alignment.

Partial Face Occlusions. Table 6 shows the method's sensitivity to partial occlusions of the face area for the indoor case, while Table 7 shows the method's sensitivity for the outdoor case.

In the indoor and outdoor case we observe that the best performance is achieved by SIFT-X-X variants, and the second best performance by WLD-X-80. In both cases, the top-1 recognition rate decreases by about 2–5% with occlusions of 10% of the face area, which is considered very good behavior. SURF-M,

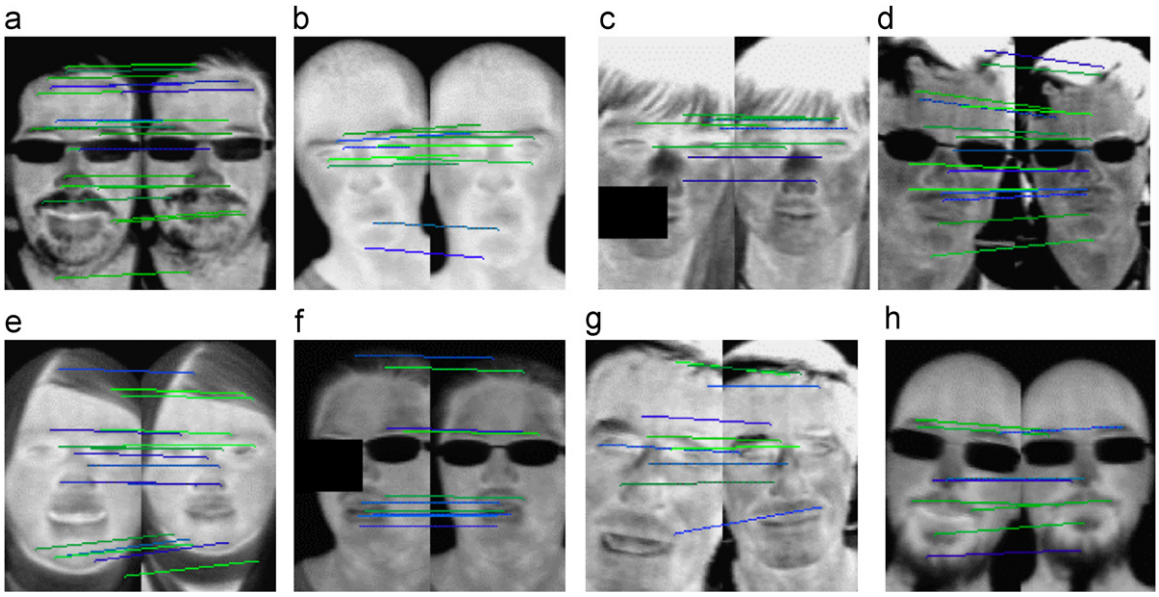

Fig. 7. Examples of correct matches of some database subjects. In all cases the right image corresponds to the gallery image (R6), and the left image to the test image. (a) Speech Set S1. Outdoor session. (b). Rotation set R11. 10% noise in eye position. Indoor session (c). Rotation set R9. Occlusion. Outdoor session. (d) Rotation set R5. 10% noise in eye position. Outdoor session. (e) Speech set S3. Indoor session. (f) Rotation set R7. Occlusion. Indoor session. (g) Speech set S3. 10% noise in eye position. Outdoor session. (h) Speech set S1. 10% noise in eye position. Indoor session.

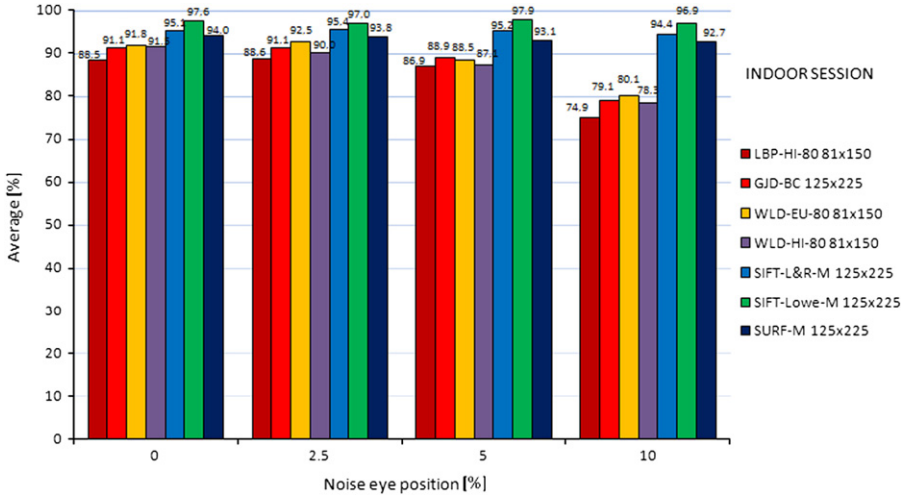

Fig. 8. UCHThermalFace database. 0%, 2.5%, 5%, and 10% noise in eye position. Indoor session. Average top-1 recognition rate. Rotation and speech test sets. (See main text for details.)

LBP-HI-80, and GJD-BC show lower performance, and they are more affected by occlusions (the top-1 recognition rate decreases by 6–11%).

Indoor versus Outdoor Galleries. In Table 8 face recognition experiments that use an indoor gallery set together with outdoor test sets are reported. Conversely, in Table 9 experiments that use an outdoor gallery set together with indoor test sets are reported. It can be clearly observed that in these cases all methods under comparison decrease their performance dramatically compared with previous experiments. In all cases the top-1 recognition rate is very low. Interestingly, the best performance is achieved by GJD-BC. Another interesting observation is that SURF-M outperforms SIFT-X-X variants, showing a higher robustness to indoor–outdoor variations.

Facial Expressions. In Table 10 face recognition experiments that use the Expressions sets are reported. The table shows the average top-1 recognition rate of 6 experiments. In each experiment one set is chosen as gallery and one as test. The sets are E1–E3 and V1–V3. It can be observed that best results are obtained by SIFT-X-X variants, with a very high top-1 recognition rate ($\sim$99.6%). Second best results are obtained by LBP-X-X, GJD-BC, and WLD-X-X, and third best results by SURF-M. These results are consistent with the ones obtained with the Speech dataset (see Table 4), except for the case of SURF-M. The reason seems to be a decrease of the performance of the SURF-M's matching process produced by large variations in the face expressions.

Variable Distance. In Table 11 face recognition experiments that correspond to the following subject–camera distances 1.2, 1.69, 2.4, 3.39, and 4.8 m are shown. In these experiments the image resolution decreases with the distance in a factor of $\sqrt{2}$. For instance, in the case of GJD-BC, where the distance is 1.69 m, the corresponding resolution is 88×159 pixels. From Table 11 it can be observed that all methods, except SURF-M, are robust to variations in the subject–camera distance. Most robust methods are SIFT-X-X and LBP-X-X. It seems that SURF-M has a minimum resolution for working appropriately.

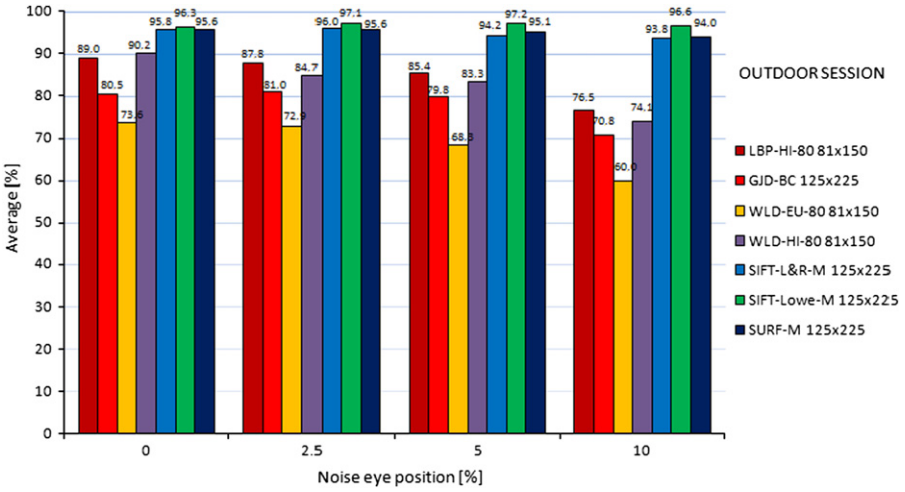

Fig. 9. UCHThermalFace database. 0%, 2.5%, 5%, and 10% noise in eye position. Outdoor session. Average top-1 recognition rate. Rotation and speech test sets. (See main text for details).

Table 11
UCHThermalFace database. Gallery set: R6. Test set: S1. Top-1 recognition rate. The image resolution decreases with the distance in a factor of $\sqrt{2}$. (See main text for details.)

Methods	1.2 m Mean (%)	1.69 m Mean (%)	2.4 m Mean (%)	3.39 m Mean (%)	4.8 m Mean (%)
LBP-HI-80. Initial resolution 81×150	92.5	94.3	94.3	92.5	92.5
GJD-BC. Initial resolution 125×225	100	100	100	94.3	88.7
WLD-EU-80. Initial resolution 81×150	96.2	98.1	92.5	86.8	66.0
WLD-HI-80. Initial resolution 81×150	96.2	96.2	94.3	94.3	67.9
SIFT-L&R-M. Initial resolution 125×225	100	98.1	98.1	96.2	86.8
SIFT-Lowe-M. Initial resolution 125×225	100	98.1	100	100	92.5
SURF-M. Initial resolution 125×225	100	92.5	79.2	41.50	7.5

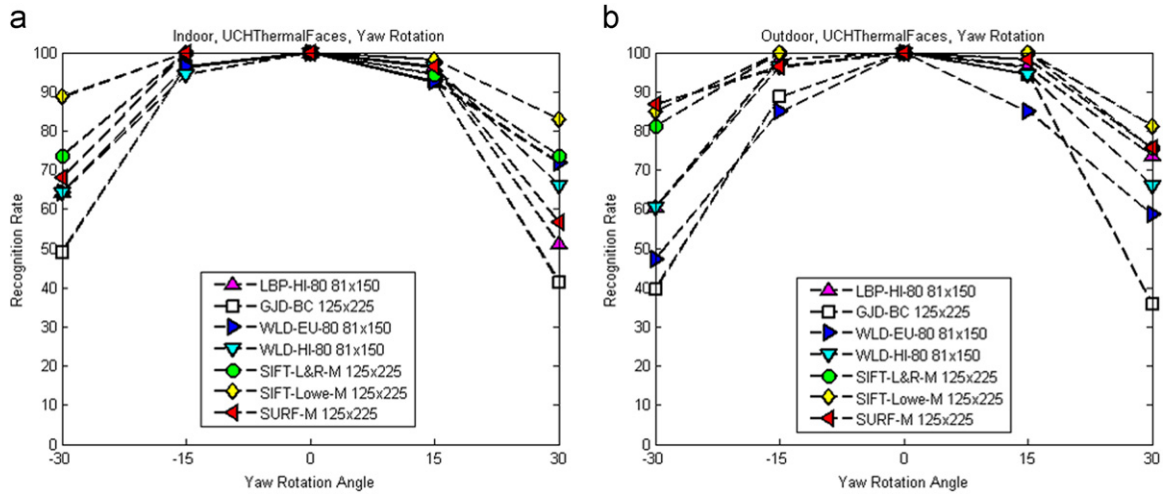

Fig. 10. This figure shows the rotation angle (yaw) versus recognition rate. The horizontal rotation angles (yaw) correspond to the position R4, R5, R6, R7, and R8 of Fig. 2. (a) Indoor; (b) Outdoor.

Fig. 10(a) and (b) show the yaw angle ($-30°$, $-15°$, $0°$, $15°$, $30°$) versus the recognition rate obtained for each selected methodology and for faces with a pitch rotation of $0°$ (a $0°$ yaw rotation corresponds to a frontal face). Angles ($-30°$, $-15°$, $0°$, $15°$, $30°$) correspond to the rotation sets R4–R8. These results summarize part of the results presented in Tables 4 and 5. Here it can be observed that for all methodologies the performance decreases as the yaw rotation increases. For low rotations ($\pm 15°$), in the indoor case (Fig. 10(a)) the performance of all methods is very similar, with SIFT variants performing best, while in the outdoor case GJD-BC and WLD-EU-80 clearly show a lower performance than other methods, methods which all have a similar recognition rate. For large rotations ($\pm 30°$), in the indoor case, only SIFT-Lowe has good performance, while in the outdoor case, both SIFT variants, SIFT-Lowe and SIFT-L&R, have a performance of approximately an 80% recognition rate. All other methods show worst performance for large rotations, with WLD-HI-80 working slightly better than other methods, both in the indoor and outdoor settings.

5.4. Computational performance

The speed of the recognition process is an important constrain in many face recognition applications (e.g. Human–Robot-Interaction or identity identification using large face databases). For this reason, in this section we present a comparative analysis of the selected methods in terms of processing time. In order to achieve this, we have evaluated the time required for feature extraction (FET—Feature Extraction Time), the time required for matching two feature vectors (MT—Matching Time), and the total processing time (PT—Processing time) required to recognize a face depending on the size of the database. Note that in the case of GJD, the total processing time is not linear but $n \log n$ on the size of the database. This is because of the way in which the Borda Count classifier works (After our experience the $\log n$ factor is relevant for large gallery databases, for example more than 1000 images). All other methods are linear on the size of the gallery. Note also that in this analysis we are only considering the time required during operation, and not the time required to create the database.

The experiments were carried out on a computer running Windows 7 Ultimate (64-bits) with an Intel Core 2 duo CPU T5870@2.00 GHz GHz (4GB RAM) processor. For all methods, with the exception of SIFT-Lowe and SURF, we used our own C/C++ implementations compiled as 32-bit applications. For SIFT-Lowe we used the original MATLAB implementation[3], run using MATLAB 7.10.0 (R2010a). According to our experience, a C/C++ implementation of SIFT-Lowe would have a similar processing time to SIFT-L&R. For the evaluation of SURF we used the OpenSURF[4] C implementation.

Table 12 shows the computed processing times of all methods under comparison in terms of feature extraction, and matching and processing times. In terms of feature extraction, LBP-X-80 is the fastest method, followed closely by WLD-X-80. The third fastest method is SURF-M. The fourth fastest method, GJD-BC, has a feature extraction two times shorter than SIFT-L&R-M, three times shorter than SIFT-Lowe-M, and more than one order of magnitude slower than LBP-HI-80 and WLD-X-80.

In terms of Matching Time MT (time for pairs of images), the fastest methods are WLD-X-80, GJD-BC, LBP-X-80, and SURF-M, all of them with MT lower than 1 ms. SIFT-Lowe-M is the slowest method, with a matching time more than 8 times slower than WLD-X-80, GJD-BC, and LBP-X-80, and 4 times slower than SIFT-L&R-M. When we consider the total processing time of the methods (PT), the method with the shortest processing time is LBP-X-80, independently of the size of the database. The second fastest methods are the WLD-X-80 variants. For large databases of large sizes (1000 images in gallery), WLD-EU-80 and WLD-HI-80 are almost two times slower than LBP-X-80. The slowest method is SIFT, with the SIFT-L&R variant being from 1.5 to 4 times faster than the SIFT-Lowe-M variant, depending on the size of the database. For small databases, SIFT-Lowe-M, the slowest method, is two orders of magnitude slower that LBP-X-80, while for large databases SIFT-Lowe-M becomes 70 times slower that LBP-X-80. GJP-BC is 2 to 10 times faster than SIFT variants depending on the size of the database, and 4 to 20 times slower than LBP-X-80. SURF-M is faster than SIFT-X variants for all database sizes. In summary, there is a clear distinction on the speed of the methods, with LBP and WLD being the fastest methods, GJD-BC standing in an intermediate position, and SIFT variants being the slowest methods.

[3] Available at ⟨http://www.cs.ubc.ca/~lowe/keypoints/⟩

[4] Available at ⟨http://www.chrisevansdev.com/computer-vision-opensurf.html⟩

Table 12
Processing time. Time measures are in milliseconds. The experiments are those we carried out on a computer running Windows 7 Ultimate with an Intel Core 2 duo CPU T5870 @2.00 GHz GHz (4 GB RAM) processor. FET/MT: Feature Extraction/Matching Time [ms]. PT: Processing Time [ms]. DB sizes (gallery) of 1, 10, 100, and 1000 faces are included.

Method	FET [ms]	MT [ms]	PT (FET+MT) [ms]			
			1	10	100	1000
LBP-X-80 81×150	2.6	<1	3.6	3.6	12.9	122.3
WLD-EU-80 81×150	3.9	<1	4.9	5.9	23.9	202.9
WLD-HI-80 81×150	3.9	<1	4.9	6.9	26.9	229.9
WLD-XS-80 81×150	3.9	<1	4.9	4.9	16.9	130.9
GJD-BC 125×225	95.94	<1	96.9	96.9	127.9	532.9
SIFT-L&R-M 125×225	189.01	2.2537	191.26	211.55	414.38	2442.7
SIFT-Lowe-M 125×225	290.22	8.4386	298.66	374.61	1134.1	8728.8
SURF-M 125×225	36.00	<1	37.0	46.0	136.0	1036.0

6. Discussion and conclusions

In this article, a comparative study of thermal-based face-recognition methods in unconstrained environments was presented. The analyzed methods were selected by considering their suitability for the defined requirements —real-time operation, just one image per person, fully online (no training), and robust behavior in unconstrained environments—and their performance in former studies. The comparative study was carried out using two databases: Equinox and UCHThermalFace. The well-known Equinox database was used as a baseline for comparison, and experiments were carried out and compared with results presented in previous work on images obtained under controlled conditions. The UCHThermalFace database includes aspects such as yaw and pitch rotations, environment condition variations (indoor/outdoor) and facial expressions. In addition, bad alignment of the images and occlusions were simulated. The methods under comparison are LBP histograms, Gabor Jet descriptors, SIFT descriptors, and WLD histograms. Comments on the main results of this study, and some conclusions drawn from this work, follow.

Comments on the Size of the Face Region. Unlike the results of [36] for visible images, here the dependence of the methods on the size of the images is not large, and for the best working methods, like SIFT and WLD, the effect is rather low. This seems to be due to the mostly uniform background observed for thermal images, which in addition allows the methods to use more information about the contour of the face.

Comments on Alignment, *Occlusions*, *and Expressions*. From our experiments we conclude that to a large degree only some of the analyzed methods are robust to inaccurate alignment, face occlusions, and variations in expressions. Accepting that these factors affect the face recognition process, their influence in the method's performance is much lower than outdoor conditions or pose variations.

Comments on the Indoor/Outdoor Conditions. Most of the methods behave very well in natural, indoor conditions, as well as in outdoor conditions, with the one exception of GJD-BC, whose performance decreases considerably under outdoor conditions. This aspect should be further analyzed with additional experiments. In experiments where the test images are acquired in an outdoor setup and the gallery images are acquired in an indoor setup, or vice versa, the performance of all methods is very low. The reason seems to be the very different range of pixel values of thermal images acquired in indoor setups compared to images acquired in outdoor settings. The saturation observed in images acquired under outdoor conditions, because of the heat of the environment, may be the main reason for this. This eventually

could be improved by a better calibration of the camera, or using normalization algorithms, as in the case of images of the visible spectrum [35]. This aspect needs to be further analyzed.

Conclusions about the Performance of Methods. The question of which method is the best is a very difficult one. However, we could say that WLD-based methods are an excellent choice if real-time operation is needed as well as good recognition rates. WLD is the second fastest method, and it has better performance than LBP, which is the fastest method. If the processing time is not an important consideration, it is clear that SIFT variants would be the method to use, given their robustness to alignment errors, to rotations and to facial expressions. In particular, SIFT-L&R seems to be a good choice for middle and large size databases, given that it has a shorter processing than SIFT-Lowe and reasonably good performance in these cases.

Demonstrating the fact that a SIFT-variant is a method of choice in some setups is one of the main results of this work, as SIFT-based methods are not widely used as yet for face recognition in visible images.

Future Work. We believe that there is still room for improvement in many aspects of recognition of faces in unconstrained environments. The main open questions in the case of thermal images are: (i) how to achieve invariance to environment conditions such as temperature and saturation of the images because of the heating of the camera produced by long periods of operation or environmental conditions (e.g. outdoor operation under direct sun exposure), (ii) how to combine the use of different methods in order to achieve, at the same time, high recognition rates and processing speed, (iii) what is the influence of face resolution in the recognition process, (iv) what can be learned through a deeper analysis of the facial expression effect in the recognition of faces, and (v) how to develop fast methods that are robust to rotations. In addition, we will extend our comparative study by incorporating methods that use vein information in the recognition process (e.g. [19]), apply new texture classification methods (e.g. [23,24]) in face recognition using visible and thermal images, as well as explore the use of information fusion approaches in multimodal recognition (e.g. [18]).

Acknowledgements

This research was partially funded by the FONDECYT-Chile Grant 1090250 and by the Advanced Mining Technology Center.

References

[1] B. Abidi, S. Huq, M. Abidi, Fusion of visual, thermal, and range as a solution to illumination and pose restrictions in face recognition, in: Proceedings of IEEE International Carnahan Conference on Security Technology, Albuquerque, NM, October 2004, pp. 325–330.
[2] J. Ahmad, U. Ali, R.J. Qureshi, Fusion of thermal and visual images for efficient face recognition using Gabor filter, In: Proceedings of the 4th ACS/IEEE International Conference on Computer Systems and Applications, March 8–11, 2006, Dubai/Sharjah, UAE, pp. 135–139.
[3] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition,, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (12) (2006) 2037–2041.
[4] M. Akhloufi, A. Bendada, Thermal faceprint: a new thermal face signature extraction for infrared face recognition, In CRV 2008, Fifth Canadian Conference on Computer and Robot Vision, Windsor, Ontario, 28–30 May 2008, pp. 269–272.
[5] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, SURF: Speeded Up Robust Features, Computer Vision and Image Understanding (CVIU) 110 (3) (2008) 346–359.
[6] G. Bebis, A. Gyaourova, S. Singh, I. Pavlidis, Face recognition by fusing thermal infrared and visible imagery, Image Vision Computing 24 (no. 7) (2006) 727–742.
[7] M.K. Bhowmik, D. Bhattacharjee, M. Nasipuri, D.K. Basu, M. Kundu, Optimum fusion of visual and thermal face images for recognition, in: Proceedings of the Sixth International Conference on Information Assurance and Security, IAS 2010, Atlanta, USA, IEEE Intelligent Transportation Systems Society, Aug 23–25, 2010, pp. 311–316.
[8] P. Buddharaju, I. Pavlidis, Multi-spectral face recognition—fusion of visual imagery with physiological information, in: R.I. Hammoud, B.R. Abidi, M.A. Abidi (Eds.), Face Biometrics for Personal Identification: Multi-Sensory Multi-Modal Systems, Springer, 2007, pp. 91–108.
[9] P. Buddharaju, I. Pavlidis, C. Manohar, Face recognition beyond the visible spectrum, advances in biometrics: sensors, Algorithms and Systems (2007) 157–180. (Springer).
[10] P. Buddharaju, I. Pavlidis, I. Kakadiaris, Physiology-based face recognition, in: Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, Lake Como, Italy, 2005, pp. 354–359.
[11] P. Buddharaju, I. Pavlidis, I. Kakadiaris, Pose-invariant physiological face recognition in the thermal infrared spectrum, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York, USA, 2006, pp. 53–60.
[12] J. Chen, S. Shan, Ch. He, G. Zhao, M. Pietikäinen, Ch.X. Chen, W. Gao, WLD: a robust local image descriptor, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (.9) (2010) 1705–1720.
[13] X. Chen, P. Flynn, K.W. Bowyer, PCA-based face recognition in infrared imagery: baseline and comparative studies, IEEE International Workshop on Analysis and Modeling of Faces and Gestures, Nice, France, 2003, pp. 127–134.
[14] S.Y. Cho, L. Wang, W.L. Ong, Thermal imprint feature analysis for face recognition, Industrial Electronics, 2009. ISIE 2009, IEEE International Symposium on, pp. 1875–1880.
[15] S. Desa, S. Hati, IR and visible face recognition using fusion of kernel based features, In: Proceedings of the 19th International Conference on Pattern Recognition ICPR, Tampa, Florida, USA, December 8–11 2008, pp. 1–4.
[16] Equinox, 2011. Equinox Database. Available on July 2011 in:〈 http://www.equinoxsensors.com/products/HID.html〉.
[17] R. Gross, Face databases, in: S. Li, A. Jain (Eds.), Handbook of Face Recognition, Springer-Verlag, 2005. (Chapter 13).
[18] M. He, S.-J. Horng, P. Fan, R.-S. Run, R.-J. Chen, J.-L. Lai, M. Khan, K. Sentosa, Performance evaluation of score level fusion in multimodal biometric systems, Pattern Recognition 43 (5) (2010) 1789–1800.
[19] G. Hermosilla, P. Loncomilla, J. Ruiz-del-Solar, Thermal Face Recognition using Local Interest Points and Descriptors for HRI Applications, Lecture Notes in Computer Science 6556, RoboCup Symposium, 2010, pp. 25–35.
[20] G. Hermosilla, J. Ruiz-del-Solar, R. Verschae, M. Correa, Face recognition using thermal infrared images for human–robot interaction applications: a comparative study, In: Proceedings of the 6th IEEE Latin American Robotics Symposium—LARS, Oct. 29–30 2009, Valparaíso, Chile, (CD Proceedings).
[21] M.M. Khan, R.D. Ward, M. Ingleby, Infrared thermal sensing of positive and negative affective states, in: Proceedings of the 2006 IEEE Conference on Robotics Automation and Mechatronics, 2006, pp. 1–6.
[22] O.K. Kwon, S.G. Kong, Multiscale fusion of visual and thermal images for robust face recognition, in: Proceedings of the IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety, FL, vol. IV, March 2005, pp. 112–116.
[23] W. Lan, Face Recognition System Based on Spatial Constellation Model and Support Vector Machine, Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Master's Thesis, 2010.
[24] H. Lategahn, S. Gross, T. Stehle, T. Aach, Texture classification by modeling joint distributions of local patterns with Gaussian mixtures, IEEE Transactions on Image Processing 19 (6) (2010) 1548–1557.
[26] D. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2) (2004) 91–110.
[27] H. Mendez, C. San Martín, J. Kittler, Y. Plasencia, E. García, Face recognition with LWIR imagery using local binary patterns, LNCS 5558 (2009) 327–336.
[28] P. Narendra, Reference-free nonuniformity compensation for IR imaging arrays, Proceedings of the SPIE 252 (1980) 10–17.
[29] P. Narendra, N. Foss, Shutterless fixed pattern noise correction for infrared imaging arrays, Proceedings of the SPIE 282 (1981) 44–51.
[30] OpenSURF library from Matlab. Available in October 2011 in 〈http://www.mathworks.com/matlabcentral/fileexchange/28300-opensurf-including-image-warp〉.
[31] F.M. Pop, M. Gordan, C. Florea, A. Vlaicu, Fusion based approach for thermal and visible face recognition under pose and expressivity variation, Roedunet International Conference (RoEduNet), 2010, pp. 61–66.
[32] J. Ruiz-del-Solar, Ch. Devia, P. Loncomilla, F. Concha, Offline Signature Verification using Local Interest Points and Descriptors, Lecture Notes in Computer Science 5197 (CIARP 2008), 2008, pp. 22–29.
[33] J. Ruiz-del-Solar, P. Loncomilla, Ch. Devia, A. New, Approach for Fingerprint Verification based on Wide Baseline Matching using Local Interest Points and Descriptors, Lecture Notes in Computer Science 4872 (PSIVT 2007), 2007, pp. 586–599.
[35] J. Ruiz-del-Solar, J. Quinteros, Illumination compensation and normalization in eigenspace-based face recognition: a comparative study of different pre-processing approaches, Pattern Recognition Letters 29 (14) (2008) 1966–1979.
[36] J. Ruiz-del-Solar, R. Verschae, M. Correa, Recognition of Faces in Unconstrained Environments: A Comparative Study, EURASIP Journal on Advances in Signal Processing, special issue, Recent Advances in Biometric Systems: A

Signal Processing Perspective, vol. 2009, Article ID 184617, p. 19, 2009. doi:10.1155/2009/184617.

[37] A. Selinger, D. Socolinsky, Appearance-Based Facial Recognition Using Visible and Thermal Imagery: A Comparative Study, Technical Report, Equinox Corporation, 2001.

[38] D. Socolinsky, A. Selinger, A comparative analysis of face recognition performance with visible and thermal infrared imagery, in: Proceedings of the International Conference on Pattern Recognition (ICPR), Quebec, Canada, 2002.

[39] D. Socolinsky, A. Selinger, Thermal Face Recognition Over Time, in: Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04), vol. 4, 2004, pp. 187–190.

[40] D. Socolinsky, L. Wolff, J. Neuheisel, C. Eveland, Illumination invariant face recognition using thermal infrared imagery, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2001.

[41] X. Tan, S. Chen, Z.H. Zhou, F. Zhang, Face recognition from a single image per person: a survey, Pattern Recognition 39 (2006) 1725–1745.

[42] L. Trujillo, G. Olague, R. Hammoud, B. Hernandez, Automatic feature localization in thermal images for facial expression recognition, in: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)—Workshops, June 20–26 2005, p. 14 .

[43] S. Wang, S. Lv, X. Wang, Infrared facial expression recognition using wavelet transform, in: Proceedings of the International Symposium on Computer Science and Computational Technology, vol. 2, 2008, pp. 327–330.

[44] Y. Yoshitomi, N. Miyawaki, S. Tomita, S. Kimura, Facial expression recognition using thermal image processing and neural network, in: Proceedings of the IEEE International Workshop Robot and Human Interactive Communication, 1997, pp. 380–385.

[45] Y. Yoshitomi, K. Sung-Ill, T. Kawano, T. Kilazoe, Effect of sensor fusion for recognition of emotional states using voice, face image and thermal image of face, in: Proceedings of the 9th IEEE International Workshop on Robot and Human Interactive Communication, 2000, pp. 178–183.

[46] J. Zou, Q. Ji, G. Nagy, A comparative study of local matching approach for face recognition, IEEE Transactions on Image Processing 16 (10) (2007) 2617–2628.