Appearance-Based Facial Recognition Using Visible and Thermal Imagery: A Comparative Study *

Andrea Selinger† Diego A. Socolinsky‡

†Equinox Corporation
9 West 57th Street
New York, NY 10019

‡Equinox Corporation
207 East Redwood Street
Baltimore, MD 21202

{andrea,diego}@equinoxsensors.com

Abstract

We present a comprehensive performance analysis of multiple appearance-based face recognition methodologies, on visible and thermal infrared imagery. We compare algorithms within and between modalities in terms of recognition performance, false alarm rates and requirements to achieve specified performance levels. The effect of illumination conditions on recognition performance is emphasized, as it underlines the relative advantage of radiometrically calibrated thermal imagery for face recognition.

1 Introduction

Face recognition in the thermal infrared domain has received relatively little attention in the literature in comparison with recognition in visible-spectrum imagery. Original tentative analyses have focused mostly on validating thermal imagery of faces as a valid biometric [1, 2]. The lower interest level in infrared imagery has been based in part on the following factors: much higher cost of thermal sensors versus visible video equipment, lower image resolution, higher image noise, and lack of widely available data sets. These historical objections are becoming less relevant as infrared imaging technology advances, making it attractive to consider thermal sensors in the context of face recognition. In the current study, we focus our attention on longwave infrared (LWIR) imagery, in the spectral range of 8μ–12μ. Other regions of the infrared spectrum also hold promise, and will be considered in upcoming work.

The influence of varying ambient illumination on systems using visible imagery is well-known to be one of the major limiting factors for recognition performance [2, 3]. A variety of methods for compensat-

*This research was supported by the DARPA Human Identification at a Distance (HID) program, contract # DARPA/AFOSR F49620-01-C-0008.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2006		**00-00-2006 to 00-00-2006**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Appearance-Based Facial Recognition Using Visible and Thermal Imagery: A Comparative Study	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Equinox Corporation,9 West 57th Street,New York,NY,10019**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**28**	

ing for variation in illumination have been studied in order to boost recognition performance, including histogram equalization, laplacian transforms, gabor transforms, logarithmic transforms, and 3-D shape-based methods. These techniques aim at reducing the within-class variability introduced by changes in illumination, which has been shown to be often larger than the between-class variability in the data, thus severely affecting classification performance. Detailed surveys of face recognition algorithms can be found in [4] and [5].

Thermal infrared imagery of faces is nearly invariant to changes in ambient illumination [6]. Consequently, no compensation is necessary, and within-class variability is significantly lower than that observed in visible imagery. As a matter of fact, it is well-known that under the assumption of Lambertian reflection, the set images of a given face acquired under all possible illumination conditions is a subspace of the vector space of images of fixed dimensions [7]. In sharp contrast to this, the set of LWIR images of a face under all possible imaging conditions is contained in a bounded set. It follows that under general conditions we can expect lower within-class variation for LWIR images of faces than their visible counterpart. It remains to demonstrate that there is sufficient between-class variability to ensure high discrimination.

Previous work by the authors provides a starting point for the current analysis. In [8], the authors perform a comparison of recognition performance between visible and longwave infrared imagery, based on two standard appearance-based algorithms: Eigenfaces and ARENA. The preliminary nature of that study limited the performance analysis to top-match recognition rates on various scenarios obtained by varying the training and testing sets, in a fashion reminiscent of n-fold cross-validation. No mention is made of false-alarm rates, receiver-operating-characteristic (ROC) curves or peformance-versus-rank curves.

The current work builds on our previous research and expands to cover those areas not touched-upon therein. In addition, we provide a much broader comparison including several other appearance-based face recognition algorithms based on more sophisticated representations, better approximating the state-of-the-art in the field. While still within the limitations imposed by existing data sets, we feel that the analysis below provides a firm basis for evaluation of thermal imagery as a valid biometric identification tool.

2 Data Collection and Calibration

All data used to obtain the results below was acquired with a newly developed sensor capable of capturing simultaneous coregistered video sequences with a visible CCD array and LWIR microbolometer. This is of particular significance for our tests, since it allows performance comparison on precisely the same imagery, much like using the red and blue channels of a color image.

We collected the data during a two-day period at the National Institute of Standards and Technology (NIST). The format consists of 240x320 pixel image pairs, co-registered to within $1/3$ pixel, where the visible image has 8 bits of grayscale resolution and the LWIR has 12 bits.

2.1 Calibration Procedure

In order to perform proper invariance analysis, it is necessary that thermal IR imagery be radiometrically calibrated. Radiometric calibration achieves a direct relationship between the grayvalue response at a

pixel and the absolute amount of thermal emission from the corresponding scene element. This relationship is called responsivity. Thermal emission is measured as flux in units of power such as W/cm^2. The grayvalue response of thermal IR pixels for LWIR cameras is linear with respect to the amount of incident thermal radiation. The slope of this responsivity line is called the *gain* and the y-intercept is the *offset*. The gain and offset for each pixel on a thermal IR focal plane array is significantly variable across the array. That is, the linear relationship can be, and usually is, significantly different from pixel to pixel. This is illustrated in Figure 1 where both calibrated and uncalibrated images are shown of the same subject.

Figure 1: Calibrated (right) and uncalibrated LWIR images. There is significant pixel-wise difference in responsivity which is removed by the calibration process.

While radiometric calibration provides non-uniformity correction, the relationship back to a physical property of the imaged object (its emmisivity) provides the further advantage of data where environmental factors contribute to a much lesser degree to within-class variability.

An added bonus of radiometric calibration for thermal IR is that it simplifies the problem of skin detection in cluttered scenes. The range of human body temperature is quite small, varying from $96°$F to $100°$F. We have found that skin temperature at $70°$F ambient room temperature to also have a small variable range from about $79°$F to $83°$F. Radiometric calibration makes it possible to perform an initial segmentation of skin pixels in the correct temperature range.

Since the responsivity of LWIR sensors is very linear, the pixelwise relation between grayvalues and radiant power can be computed by a process of two-point calibration. Images of a black-body radiator covering the entire field of view are taken at two known temperatures, and thus the gains and offsets are computed using the radiant power for a black-body at a given temperature.

Note that this is only possible if the emissivity curve of a black-body as a function of temperature is known. This is given by Planck's Law, which states that the flux emmited at the wavelength λ by a blackbody at a given temperature T in $W/(cm^2\mu m)$ is given by

$$W(\lambda, T) = \frac{2\pi hc^2}{\lambda^5 \left(e^{\frac{hc}{\lambda kT}} - 1\right)} \tag{1}$$

where h is Planck's constant, k is Boltzman's constant, and c is the speed of light in a vacuum. To relate this to the flux observed by the sensor, the responsivity, $R(\lambda)$ of the sensor must be taken into account. This allows the flux observed by a specific sensor from a black-body at a given temperature to be determined:

$$W(T) = \int W(\lambda, T) R(\lambda) d\lambda \ . \tag{2}$$

For our sensor, the responsivity is very flat between 8 and 12 microns, so we can simply integrate Equation (1) for λ between 8 and 12. The Planck curve and the integration process are illustrated in Figure 2.

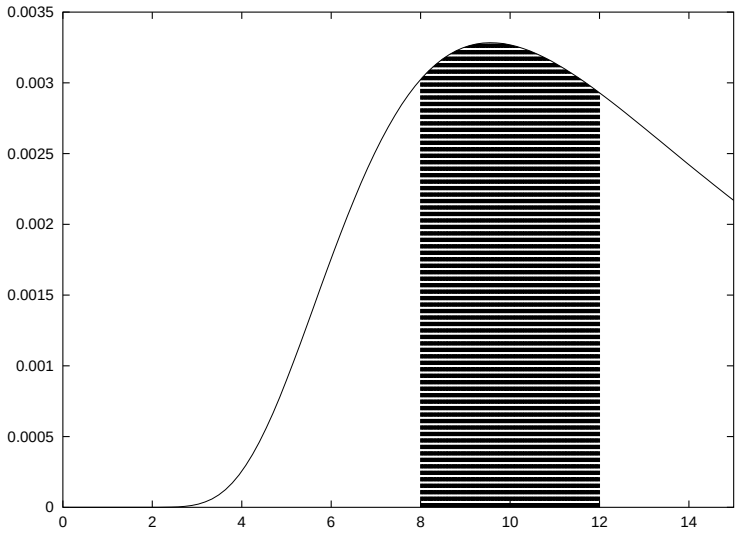

Figure 2: The Planck curve for a black-body at 303K (roughly skin temperature), with the area to be integrated for an 8-12μm sensor shaded.

One can achieve marginally higher precision by taking measurements at multiple temperatures and obtaining the gains and offsets by least squares regression. For the case of thermal images of human faces, we take each of the two fixed temperatures to be below and above skin temperature, to obtain the highest quality calibration for skin levels of IR emmision.

It should be noted that a calibration has a limited life span. If a LWIR camera is radiometrically calibrated indoors, taking it outdoors where there is a significant ambient temperature difference will cause the gain and offset of linear responsivity of the focal plane array pixels to change. Therefore, radiometric calibration must be performed again. This effect is mostly due to the optics and FPA heating up, and causing the sensor to "see" more energy as a result. Also, suppose two separate data collections are taken with two separate LWIR cameras but with the exact same model number, identical camera settings and under the exact same environmental conditions. Nonetheless, no two thermal IR focal plane arrays are ever identical and the gain and offset of corresponding pixels between these separate cameras will be different. Yet another example; suppose two data collections are taken one year apart, with the same thermal IR camera. It is very likely that gain and offset characteristics will have changed. Radiometric calibration standardizes all thermal IR data collections, whether they are taken under different environmental conditions or with different cameras or at different times.

The grayvalue for any thermal IR image is directly physically related to thermal emission flux which is a universal standard. This provides a standardized thermal IR biometric signature for humans. The images that face recognition algorithms can most benefit from in the thermal IR are not arrays of gray values, but rather arrays of corresponding thermal emission values. If there is no way to relate grayvalues to thermal emission values then it is not possible to do this.

2.2 The Collection Setup

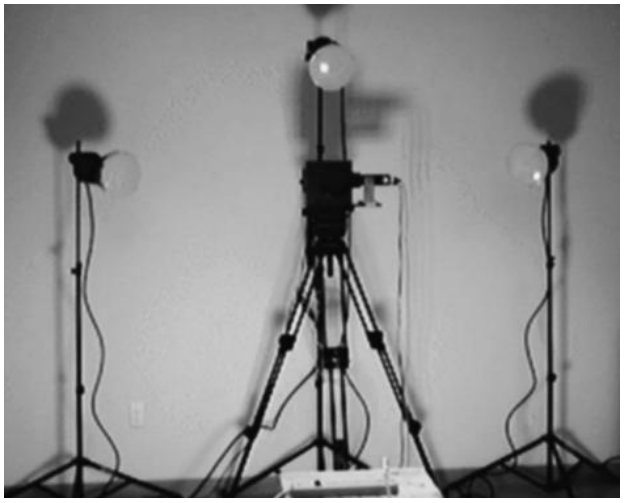

Figure 3: Camera and lighting setup for data collection.

For the collection of our images, we used the FBI mugshot standard light arrangement [9], shown in Figure 3. Image sequences were acquired with three illumination conditions: frontal, left lateral and right lateral. For each subject and illumination condition, a 40 frame, four second, image sequence was recorded while the subject pronounced the vowels looking towards the camera. After the initial 40 frames, three static shots were taken while the subject was asked to act out the expressions 'smile', 'frown', and 'surprise'. In addition, for those subjects who wore glasses, the entire process was done with and without glasses. Figure 4 shows a sampling of the data in both modalities.

A total of 115 subjects were imaged during a two-day period. After removing corrupted imagery from 24 subjects, our test database consists of over $25,000$ frames from 91 distinct subjects. Much of the data is highly correlated, so only specific portions of the database can be used for training and testing purposes without creating unrealistically simple recognition scenarios. This is explained in Section 3. The entire image collection used for the experiments below is available at the authors' website[1].

3 Testing Methodology

Following the approach in [8], we selected subsets of our face database to be used as testing and training sets. In n-fold cross-validation experiments, one repeatedly selects a random subset of the available data

[1] http://www.equinoxsensors.com/hid

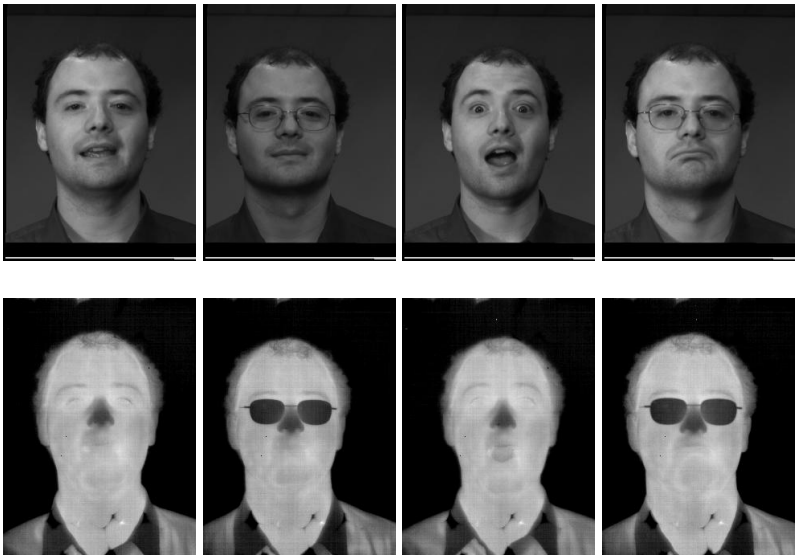

Figure 4: Sample imagery from our data collection.

as a training set, and testing is performed on the remaining data. Repeating this process multiple times and reporting mean performance yields statistically significant results. We are particularly interested in exposing the relation between illumination, as well as facial expression variation and recognition peformance. Therefore, we chose our training/testing pairs in a biased fashion rather than randomly, in order to elicit the desired information. Note that based on the choices below, our testing methodology is stricter, and should produce lower average results than random cross-validation. Additionally, since much of our data is highly correlated due to the acquisition procedure, the biased choices below help decorrelate testing and training sets.

We construct multiple query sets for testing and training. Frames 0, 3 and 9 from a given image sequence are referred to as vowel frames. Frames corresponding to 'smile', 'frown' and 'surprise' are referred to as expression frames. Our query criteria are as follows:

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 wearing glasses, all illuminations.
EG: Expression frames, subjects wearing glasses, all illuminations.
RR: 500 random frames, arbitrary illumination.

The same queries were used to construct sets for visible and LWIR imagery, and all LWIR images were radiometrically calibrated. Locations of the eyes and the frenulum were semi-automatically located in all visible images, which also provided the corresponding locations in the co-registered LWIR frames. Using these feature locations, all images were geometrically transformed to a common standard, and

cropped to eliminate all but the inner face.

When evaluating algorithms, three subsets of images are of interest [10]:

- The training set T
- The gallery set G
- The probe set P

The usage of T and P is essentially identical to that of training and test data in machine learning. However, the gallery contains exemplars for the people to be recognized. It has no direct correlate in the common machine learning nomenclature. In some cases T and G may be the same set. In other words, an algorithm may be trained on a set of images and then subsequently the trained algorithm may use the same set as exemplars against which to match probe images. However, this need not be the case. Query set RR is used as a training set in computing all relevant subspaces and basis sets for the algorithms below, unless otherwise noted. Additionally, some probe/gallery combinations are omitted from the tables due to inclusion relations.

The relation between vowel frames and expression frames is comparable to that between **fa** and **fb** sets in the FERET database, although our expression frames are often more different from vowel frames than in the FERET case. Frontal and lateral illumination frames are comparable to **fa** versus **fc** sets in FERET. We should also note that queries VG, EG and RR were only used as testing sets and not as training sets. The maximum possible correct classification performance achievable for those combinations is lower than 100%, and therefore those combinations were ignored to simplify the analysis.

Tabular performance results reported below are for the top match. We also report, in graphical form, recognition performance as a function of rank. In this case, for a fixed rank $k \geq 1$, a probe is considered correctly classified if any of the top k matches are correct. Note that this is not the same as a k-nearest-neighbor classifier.

When reviewing rank-ordered match results, in addition to the rate of correct recognition, we must also consider the false-alarm rate incurred by relaxing our correctness criterion. Let $\mathcal{T}$ be a training set and $\mathcal{P}$ a set of probes. For $p \in \mathcal{P}$, let M_p^k be the distance from p to the k^{th} closest training observation, and $H_p^k = \{t \in \mathcal{T} \mid \text{dist}(p,t) \leq M_p^k\}$. Define α_p to be 1 if any member of H_p^k belongs to the same class as p, and zero otherwise. Further define $\|H_p^k\|$ to be the number of distinct class labels among elements of H_p^k and $\|\mathcal{P}\|$ the number of probes in $\mathcal{P}$. With this notation, the correct classification rate and false alarm rate are respectively given by

$$\xi = \frac{1}{\|\mathcal{P}\|} \sum_{p \in \mathcal{P}} \alpha_p \ , \qquad \phi = \frac{1}{\|\mathcal{P}\|} \sum_{p \in \mathcal{P}} \frac{\|H_p^k\| - \alpha_p}{\|H_p^k\|}.$$

4 Algorithms Tested

The testing methodology outlined above was applied to several appearance-based algorithms. We should point out that the restriction to appearance-based techniques was motivated by the fact that geometry-based methods depend only on the ability to accurately locate facial landmarks in the image. While such landmarks may be more easily located in one modality over the other, the effect of the imaging modality

on the final recognition outcome is indirect, and thus an analysis of that effect would be less revealing. In addition, appearance-based methods have generally shown higher performance than those based on facial geometry alone.

All algorithms tested consist of a projection to a subspace of the image space followed by 1-nearest neighbor classification. We also experimented with k-nearest neighbor algorithms for different values of $k > 1$, and found that $k = 1$ achieves the best results. The different subspace constructions are outlined below. Digital images are converted into vectors by scanning in raster order.

4.1 Eigenfaces (PCA)

This is perhaps the most popular algorithm in the field [11], [12], [13] and it is a technique commonly used in dimensionality reduction in computer vision and particularly in face recognition. PCA techniques, also known as Karhunen-Loeve methods, choose a linear projection that reduces the dimensionality while maximizing the scatter of all projected samples.

Using the notation in [14], PCA can be described formally as follows. Given a set of N sample images $\{x_1, x_2, ..., x_N\}$ taking values in an n-dimensional image space, PCA finds a linear transformation W^T mapping the original n-dimensional image space into an m-dimensional feature space, where $m < n$. The new feature vectors $y_k \in \mathbb{R}^m$ are defined by the following linear transformation:

$$y_k = W^T x_k, k = 1, 2, \ldots, N \tag{3}$$

where $W \in \mathbb{R}$ is a matrix with orthonormal columns.

We define the total scatter matrix S_T as:

$$S_T = \sum_{k=1}^{N} (x_k - \mu)(x_k - \mu)^T \tag{4}$$

where n is the number of sample images, and $\mu \in \mathbb{R}^n$ is the mean image of all samples. After applying the linear transformation W^T, the scatter of the transformed feature vectors $\{y_1, y_2, \ldots, y_N\}$ is $W^T S_T W$. In PCA the projection W_{opt} is chosen to maximize the determinant of the total scatter matrix of the projected samples, i.e.,

$$W_{opt} = \arg\max_W |W^T S_T W| = [w_1 w_2 \ldots w_m] \tag{5}$$

where $\{w_i | i = 1, 2, \ldots, m\}$ is the set of n-dimensional eigenvectors of S_T corresponding to the m largest eigenvalues. Since these eigenvectors have the same dimension as the original images, they are referred to as eigenfaces. They are also referred to as principal components.

The *face space* is computed by taking a (usually separate) set of training observations, and finding the eigenfaces of this set. While each face from the training set can be represented as a linear combination of the eigenfaces, they can also be approximated by a linear combination of the "best" eigenfaces, those that have the largest eigenvalues. In fact, it is well-known that, for a fixed choice of n, the subspace spanned by the first n eigenvectors is the one with lowest L^2 reconstruction error for any vector in the training set used to create the face space. Under the assumption that the training set is representative of all face images, the face space is taken to be a good low-dimensional approximation to the set of all possible face images under varying conditions.

4.2 Linear Discriminant Analysis (LDA)

It is a classical result that while the feature subspace used by Eigenfaces, obtained through principal component analysis, is optimal in terms of L^2 reconstruction error, it has no optimality properties in terms of class discriminability. In fact, class membership is not taken into account in the construction of the face space.

Class specific methods, such as the one based on Fisher's Linear Discriminant (FLD) [15], [16] may get better recognition rates than the Eigenface method. This method selects the subspace W in such a way that the ratio of the between-class scatter and the within-class scatter is maximized [14]. The between-class scatter matrix is defined as

$$S_B = \sum_{i=1}^{c} N_i(\mu_i - \mu)(\mu_i - \mu)^T \tag{6}$$

and the within-class scatter matrix is

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in X_l} (x_k - \mu_i)(x_k - \mu_i)^T \tag{7}$$

where μ_i is the mean image of class X_i, and N_i is the number of sample in class X_i. If S_W is nonsingular, the optimal projection W_{opt} is chosen as the matrix with orthonormal columns that maximizes the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the within-class scatter matrix of the projected samples:

$$W_{opt} = argmax \frac{|W^T S_B W|}{|W^T S_W W|} = [w_1 w_2 \dots w_m] \tag{8}$$

where $\{w_i | i = 1, 2, \dots, w_m\}$ is the set of generalized eigenvectors of S_B and S_W corresponding to the m largest generalized eigenvalues $\{\lambda_i | i = 1, 2, \dots, m\}$, i.e.

$$S_B w_i = \lambda_i S_W w_i, i = 1, 2, \dots, m \tag{9}$$

There are at most $c - 1$ nonzero generalized eigenvalues, where c is the number of classes.

Since the rank of the within-class scatter matrix S_W is at most $N - c$, and often the number of images in the learning set N is much smaller than the number of pixels in each image n, S_W is often singular. To avoid this problem, the method known as *Fisherfaces* was proposed in [14]. Before computing the Fisher linear discriminant, the image set is projected to a lower dimensional space obtained using the PCA algorithm.

We have found that reducing the dimensionality of the space improves recognition results even if S_W is not singular.

In our experiments we considered two slight variants of the Fisherfaces algorithm: the classic approach (referred to as LDAg) where the projection matrix W was computed from the face gallery, and another approach (referred to as LDAt) that used a separate training set for computing the projection matrix. This training set contained random images of the people used during testing.

4.3 Local Feature Analysis (LFA)

Another subspace representation for facial data based on second order statistics results by enforcing topographic indexing of the basis vectors, and minimizing their correlation. Local Feature Analysis [17] achieves this by constructing a family of feature detectors based on a PCA decomposition, which are locally correlated. A selection, or sparsification, step is then used to produce a minimally correlated subset of features, which define the subspace of interest. While the original method is geared at optimal reconstruction, sparsification techniques consistent with the requirements of a recognition system are also possible. We use two subselection methods, one following [18] and the other explained in detail below, referred to as LFAb and LFAe, respectively.

Given the eigenvectors $\Psi_r(x)$ with eigenvalues λ_r, the following two functions are constructed:

$$K(x,y) = \sum_{r=1}^{N} \Psi_r(x) \frac{1}{\sqrt{\lambda_r}} \Psi_r(y) \tag{10}$$

$$P(x,y) = \sum_{r=1}^{N} \Psi_r(x) \Psi_r(y) \tag{11}$$

$K(x,y)$ is the kernel of the representation, and $P(x,y)$ is the residual correlation of the outputs. The rows of K contain kernels with spatially local properties. The kernel matrix K transforms the face set X into the LFA output O: $O = KX^T$. The original images can be reconstructed from O by $X^T = K^{-1}O$. LFA produces an n dimensional representation, where n is the number of pixels in the images. Since we have n outputs described by $p << n$ linearly independent variables, there are residual correlations in the output. Penev and Atick [17] proposed an algorithm for reducing the dimensionality of the representation by choosing a subset $\mathcal{M}$ of outputs that were as decorrelated as possible. Since their method was designed for image representation, not for recognition, the algorithm selected a different set of points for each image, which is problematic for recognition.

In order to make the representation suitable for recognition, we used two different sparsification methods, one described by Bartlett in [18], to which we will refer to as LFAb, and one devised by us, to which we will refer to as LFAe.

Similarly to Penev and Atick's method, LFAb is an iterative sparsification algorithm based on multiple linear regression. At each step, the point with the largest mean reconstruction error *across all of the images* was added to $\mathcal{M}$.

At each step, the point added to $\mathcal{M}$ is chosen as the kernel that maximizes the image reconstruction error:

$$\arg\max(||O - O^{rec}||^2) \tag{12}$$

where O^{rec} is a reconstruction of the complete output, O, using a linear predictor on the subset $\mathcal{M}$ of the outputs O. The linear predictor is of the form:

$$\mathcal{Y} = \beta \mathcal{X} \tag{13}$$

where $\mathcal{Y} = O^{rec}$, β is the vector of the regression parameters, and $\mathcal{X} = O(\mathcal{M}, N)$. $O(\mathcal{M}, N)$ denotes the subset of O corresponding to the points in $\mathcal{M}$ for all N images. β is calculated from

$$\beta = \frac{\mathcal{Y}\mathcal{X}}{(\mathcal{X}^T\mathcal{X})} = \frac{(O^{rec})^T O(\mathcal{M}, N)}{O(\mathcal{M}, N)^T O(\mathcal{M}, N)} \tag{14}$$

This equation can also be expressed in terms of the correlation matrix of the outputs, $C = O^T O$:

$$\beta = C(\mathcal{M}, N)C(\mathcal{M}, \mathcal{M})^{-1} \tag{15}$$

The termination condition can be either $|\mathcal{M}| = N$ or $(||O - O^{rec}||^2) \leq \epsilon$.

The LFAe algorithm sparsifies the LFA representation by choosing kernels that are as uncorrelated as possible and cover a large area of the image. At each step, the algorithm chooses the kernel that maximizes the square of the total residual correlation of $\mathcal{M}$:

$$\arg\max(||P^2 - P_{total}^2||) \tag{16}$$

where P_{total} is the sum of the residual correlations of the kernels already chosen.

4.4 Independent Component Analysis (ICA)

Principal component analysis seeks an orthonormal basis for the data space with respect to which the marginal training distributions are uncorrelated. Independent component analysis goes farther by requiring a basis (not orthogonal) such that the corresponding marginals are statistically independent. Note that these conditions are equivalent if the data is globally Gaussian, but that is hardly ever the case in practice. Original motivation for this decomposition came from the need to separate audio streams into independent sources without prior knowledge of the mixing process [19], [20]. Computation of the independent components cannot be done by solving an algebraic system of equations, and rather must be done by numerically minimizing a criterion function. Different criterion functions exist, based on kurtosis or other higher order moments, mutual information between marginals or entropy criteria, all yielding comparable results for our application. We used the FastICA package available as public-domain in MATLAB at
http://www.cis.hut.fi/projects/ica/fastica. This package implements the fast fixed-point algorithm for independent component analysis described in [21]. Similarly to the previously described algorithms, this algorithm computes a matrix W that is used to project the images onto a subspace where nearest-neighbor classification is performed.

If we denote by $x = (x_1, x_2, \ldots, x_m)^T$ a zero-mean m-dimensional random variable that can be observed, and by $s = (s_1, s_2, \ldots, s_n)^T$ its n-dimensional transform, then the ICA problem is the determine a constant (weight) matrix W so that the linear transformation of the observed variables

$$s = Wx \tag{17}$$

yields components s_i that are statistically as independent from each other as possible. ICA is only possible if every independent component has non-gaussian distribution.

The one-unit FastICA algorithm finds a direction, i.e. a unit vector w such that the projection $w^T x$ maximizes nongaussianity. A fundamental result of information theory is that *a gaussian variable has the largest entropy among all random variables of equal variance*. This means that entropy can be used to measure nongaussianity. To obtain a measure of nongaussianity that is zero for a gaussian variable

and always nonnegative, a slightly modified version of the differential entropy is used, called negentropy. Negentropy is defined as follows:

$$J(y) = H(y_{gauss}) - H(y) \tag{18}$$

where y_{gauss} is a Gaussian random variable of the same covariance matrix as y. J can be obtained using the following approximation:

$$J(y) \propto [E\{G(y)\} - E\{G(\nu)\}]^2 \tag{19}$$

for practically any non-quadratic function G. Useful choices for G are:

$$G_1(u) = \frac{1}{a_1} \log \cosh a_1 u, \qquad G_2(u) = -\exp(-u^2/2) \tag{20}$$

To estimate several independent components, the one-unit FastICA algorithm needs to be run on several units with weight vecors $w_1, \ldots, w_n$. To prevent different vectors from converging to the same maxima, the outputs $w_1^T x, \ldots w_n^T x$ have to be decorrelated after every iteration.

In order to control the dimensionality of the target subspace, that is the number of independent components, a PCA preprocessing step is applied before computation of the independent basis. Our subspace dimension is fixed at 100, just like for PCA and LFA. Comparisons of ICA-based face recognition with other methods can be found in [22], [18] and others.

5 Experimental Results and Discussion

Images were subsampled by a factor of 10 in each dimension prior to experimentation. Then they were masked using the binary mask in Figure 5 to remove any background clutter.

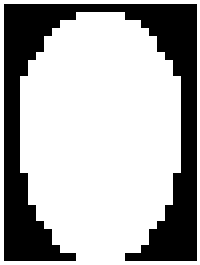

Figure 5: Binary mask used to remove background clutter from face images

After masking, visible images were zero-meaned and normalized to unit norm in order to provide some measure of illumination compensation. This process is comparable to the histogram equalization step in [23]. Thermal IR images were processed via two-point radiometric calibration.

With the exception of algorithm LDAg, we used query RR described in section 3 as a training set for each algorithm and we varied the gallery and probe sets. For each valid pair of training and testing sets, we computed the top-match recognition performance and receiver-operating-characteristic (ROC) curves for each algorithm and both modalities.

5.1 PCA

The subspace for PCA was chosen to be 100-dimensional and was computed from the RR set. Figure 6 shows the first 10 eigenfaces in the visible, as well as in the infrared domain.

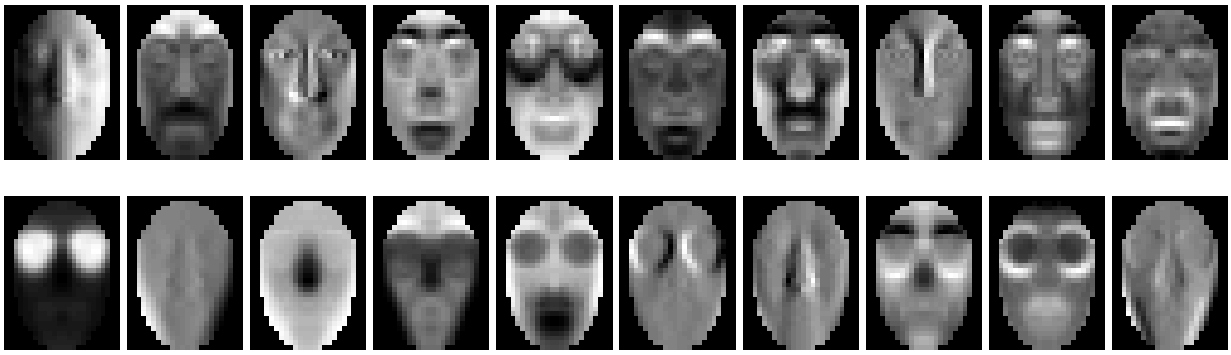

Figure 6: First 10 eigenfaces in the visible (top row) and LWIR (bottom row) domains

Table 1 shows the top-match performance of the PCA algorithm in visible and in LWIR. Visible results are reported above the corresponding LWIR results. While the algorithm performs reasonably well in the visible domain in some cases, the performance is extremely low if the illumination conditions change from the gallery set to the probe set, reaching a minimum of 35%. The LWIR domain deals much better with changes in visible illumination, and the overall performance increases from 72% in visible to 95% in LWIR (see also Table 7). Regardless of the variation in illumination conditions, the algorithm performs consistently better in the LWIR domain.

	VA	EA	VF	EF	VL	EL
VA		$\binom{0.857}{0.979}$	$\binom{0.749}{0.973}$	$\binom{0.511}{0.928}$	$\binom{0.887}{0.990}$	$\binom{0.741}{0.965}$
EA	$\binom{0.795}{0.949}$		$\binom{0.503}{0.899}$	$\binom{0.699}{0.951}$	$\binom{0.669}{0.934}$	$\binom{0.855}{0.983}$
VF		$\binom{0.794}{0.984}$		$\binom{0.792}{0.952}$	$\binom{0.664}{0.970}$	$\binom{0.449}{0.945}$
EF	$\binom{0.722}{0.939}$		$\binom{0.719}{0.914}$		$\binom{0.351}{0.909}$	$\binom{0.576}{0.951}$
VL		$\binom{0.890}{0.976}$	$\binom{0.622}{0.960}$	$\binom{0.369}{0.916}$		$\binom{0.890}{0.975}$
EL	$\binom{0.832}{0.954}$		$\binom{0.391}{0.891}$	$\binom{0.544}{0.925}$	$\binom{0.832}{0.947}$	
VG		$\binom{0.921}{0.970}$	$\binom{0.841}{0.993}$	$\binom{0.657}{0.947}$	$\binom{0.935}{1.000}$	$\binom{0.811}{0.963}$
EG	$\binom{0.867}{0.961}$		$\binom{0.620}{0.932}$	$\binom{0.786}{0.973}$	$\binom{0.754}{0.954}$	$\binom{0.899}{0.990}$
RR	$\binom{0.968}{0.994}$	$\binom{0.888}{0.984}$	$\binom{0.706}{0.968}$	$\binom{0.578}{0.948}$	$\binom{0.840}{0.974}$	$\binom{0.772}{0.964}$

Table 1: PCA performance in visible (top) and in LWIR (bottom)

Figure 7 shows the ROC curve and the performance versus rank curve for one of the best PCA performances, obtained with VA as gallery and RR as the set of probes. Although the performance is already high in the visible range at 96.8%, performance in LWIR is even better at 99.4% (note the smaller range on the vertical axis). The improvement in LWIR over the visible range is even more visible in cases where the algorithm performs poorly in the visible range, as in the case where VF was used as gallery,

and EL was the probe set. In this case both the illumination conditions and the facial expression were different in the gallery and probe images. Performance in the visible domain was 39.1%. In LWIR it increased to 89.1%. The ROC and performance versus rank curves can be seen in Figure 8.

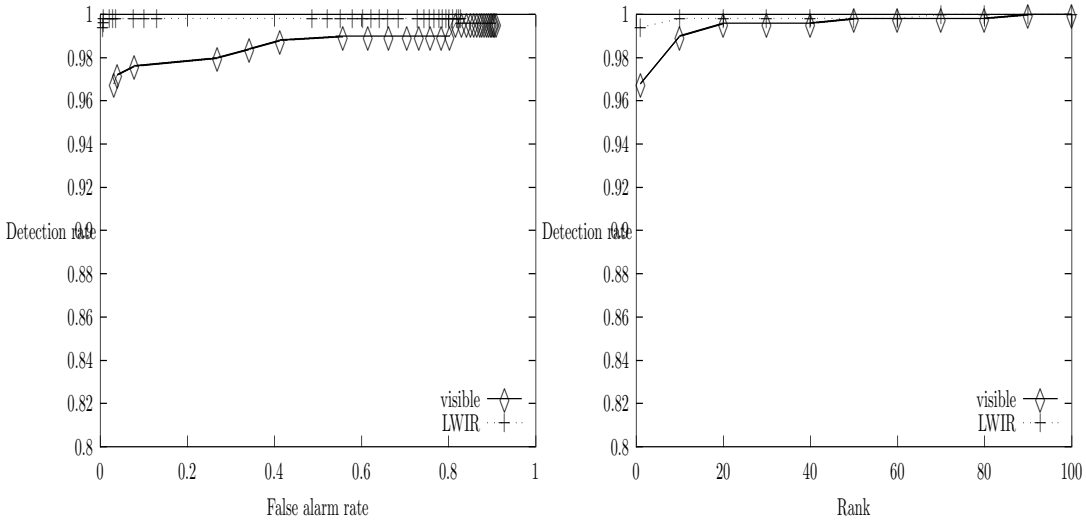

Figure 7: ROC and performance versus rank curves for PCA with gallery VA and probe set RR.

5.2 LDA

For the LDA algorithm we first reduced the dimensionality of the image space to 200 by performing PCA. We then reduced the dimensionality further by running the linear discriminant algorithm. LDA subspaces have as many dimensions as classes in the training set, minus one. In our case this amounted to 90.

The basis vectors of the Fisherface space have high frequency and at a high level do not seem to resemble the images they were derived from (see Figure 9).

We tested two versions of the Fisherfaces algorithm, one trained on query RR (referred to as LDAt), and the other trained on the image gallery (LDAg). The performance of the two versions was very similar. One possible reason is that query RR contains images of all the people that appear in the other queries, so all the classes are represented.

LDA (both versions that we tested) performs much better than PCA in the visible domain and deals well even with changes in illumination. The LWIR domain brings further improvements. In fact, LDA is by far our best performing algorithm in both domains. The top-match performance of LDAt and LDAg can be seen in tables 5.2 and 3 respectively.

For the LDAt version, the best performance in visible (99.4%) is obtained when VL is used as the gallery, and RR is used as the probe set. This doesn't leave much room for improvement, but in LWIR, the performance reaches 99.8%. The ROC and performance versus rank curves can be seen in Figure 10.

The ROC and performance versus rank curve for the worst performance (83.6% in visible and 93.7% in LWIR for gallery EL and probe set VF) can be seen in Figure 11. Again, use of the LWIR domain

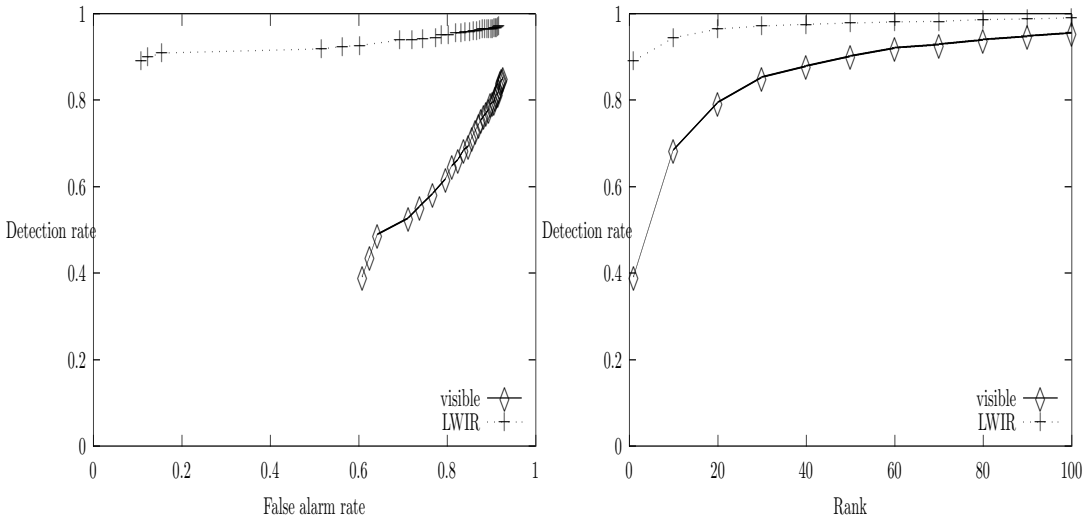

Figure 8: ROC and performance versus rank curves for PCA with gallery VF and probe set EL.

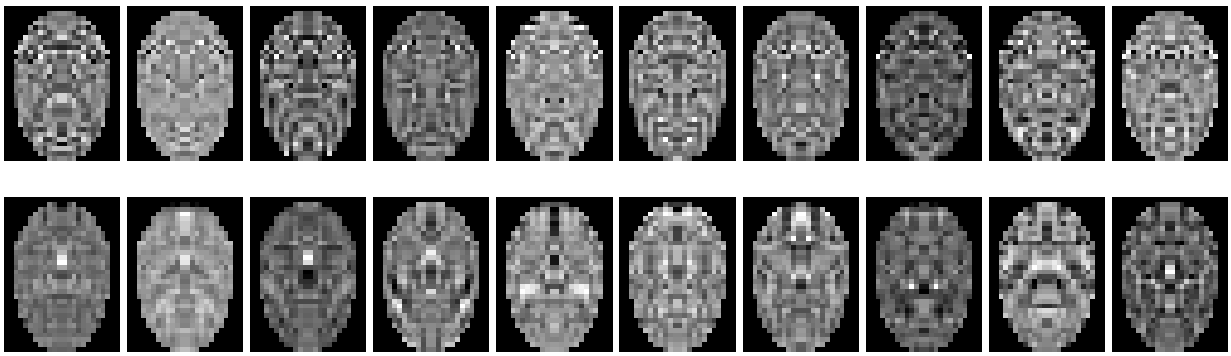

Figure 9: 10 fisherfaces in the visible (top row) and LWIR (bottom row) domains

	VA	EA	VF	EF	VL	EL
VA		$\binom{0.936}{0.984}$	$\binom{0.982}{0.985}$	$\binom{0.848}{0.935}$	$\binom{0.987}{0.991}$	$\binom{0.929}{0.973}$
EA	$\binom{0.897}{0.961}$		$\binom{0.852}{0.936}$	$\binom{0.952}{0.968}$	$\binom{0.877}{0.940}$	$\binom{0.981}{0.987}$
VF		$\binom{0.952}{0.979}$		$\binom{0.910}{0.947}$	$\binom{0.963}{0.974}$	$\binom{0.936}{0.952}$
EF	$\binom{0.891}{0.942}$		$\binom{0.884}{0.932}$		$\binom{0.840}{0.914}$	$\binom{0.946}{0.962}$
VL		$\binom{0.928}{0.987}$	$\binom{0.973}{0.978}$	$\binom{0.817}{0.929}$		$\binom{0.925}{0.984}$
EL	$\binom{0.900}{0.971}$		$\binom{0.836}{0.937}$	$\binom{0.928}{0.952}$	$\binom{0.896}{0.953}$	
VG		$\binom{0.928}{0.986}$	$\binom{0.986}{0.990}$	$\binom{0.862}{0.967}$	$\binom{0.993}{0.997}$	$\binom{0.928}{0.986}$
EG	$\binom{0.913}{0.971}$		$\binom{0.867}{0.959}$	$\binom{0.971}{0.987}$	$\binom{0.891}{0.959}$	$\binom{0.987}{0.992}$
RR	$\binom{0.996}{1.000}$	$\binom{0.960}{0.994}$	$\binom{0.978}{0.992}$	$\binom{0.886}{0.964}$	$\binom{0.994}{0.998}$	$\binom{0.952}{0.986}$

Table 2: LDAt performance in the visible (top) and LWIR (bottom) domains

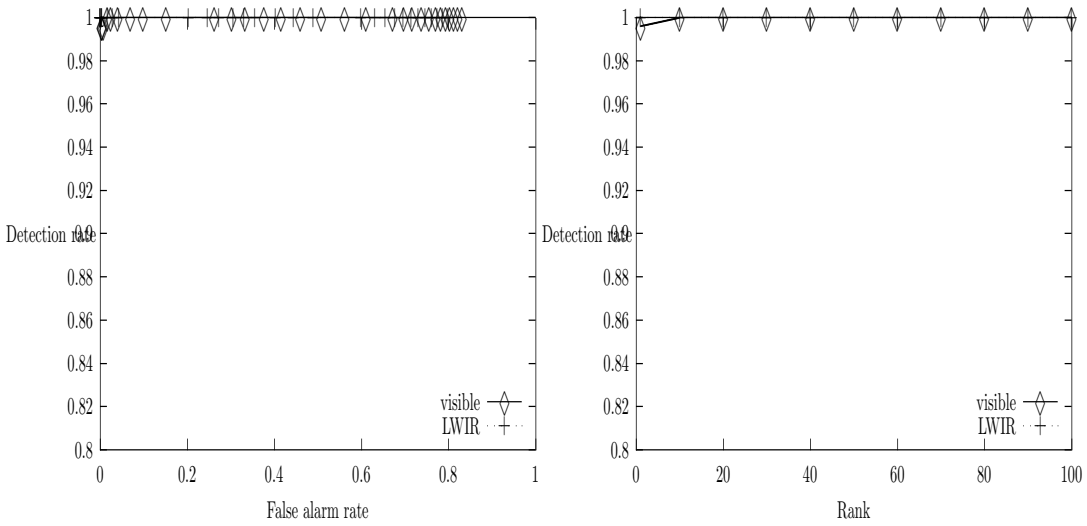

Figure 10: ROC and performance versus rank curves for LDAt with gallery VL and probe set RR.

improved performance significantly.

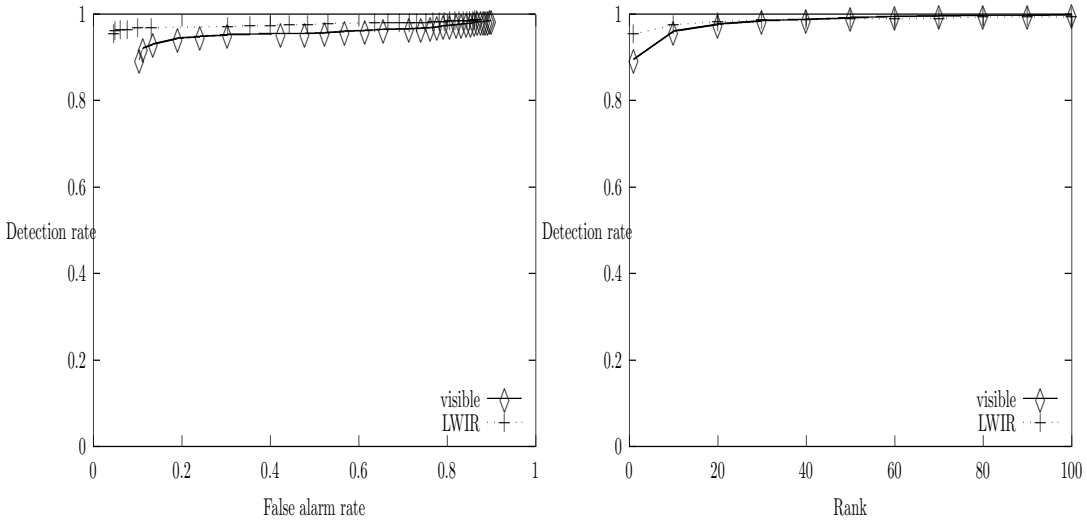

Figure 11: ROC and performance versus rank curves for LDAt with gallery EL and probe set VF.

The performance of the LDAg version of the algorithm is very close to that of LDAt. Performance reaches 100% in one case in the visible domain, and in several cases in LWIR. One of the best performances is obtained by gallery VL and probe set RR, while one of the worst is obtained for gallery VF and probe set EL. Figures 12 and 13 show the ROC and performance versus rank curves for these cases.

	VA	EA	VF	EF	VL	EL
VA		$\binom{0.976}{0.997}$	$\binom{0.949}{0.986}$	$\binom{0.832}{0.958}$	$\binom{0.995}{0.996}$	$\binom{0.960}{0.992}$
EA	$\binom{0.939}{0.977}$		$\binom{0.800}{0.936}$	$\binom{0.888}{0.979}$	$\binom{0.917}{0.966}$	$\binom{0.985}{0.994}$
VF		$\binom{0.986}{1.000}$		$\binom{0.956}{0.963}$	$\binom{0.986}{0.988}$	$\binom{0.942}{0.984}$
EF	$\binom{0.928}{0.979}$		$\binom{0.925}{0.944}$		$\binom{0.884}{0.951}$	$\binom{0.956}{0.983}$
VL		$\binom{0.971}{0.996}$	$\binom{0.923}{0.979}$	$\binom{0.769}{0.956}$		$\binom{0.969}{0.996}$
EL	$\binom{0.945}{0.977}$		$\binom{0.735}{0.931}$	$\binom{0.830}{0.968}$	$\binom{0.934}{0.974}$	
VG		$\binom{0.965}{0.993}$	$\binom{0.954}{0.993}$	$\binom{0.862}{0.970}$	$\binom{1.000}{1.000}$	$\binom{0.960}{0.990}$
EG	$\binom{0.971}{0.985}$		$\binom{0.848}{0.975}$	$\binom{0.918}{0.997}$	$\binom{0.956}{0.985}$	$\binom{0.990}{1.000}$
RR	$\binom{0.996}{1.000}$	$\binom{0.970}{1.000}$	$\binom{0.926}{0.984}$	$\binom{0.824}{0.998}$	$\binom{0.992}{0.998}$	$\binom{0.954}{0.998}$

Table 3: LDAg performance in the visible (top) and LWIR (bottom) domains

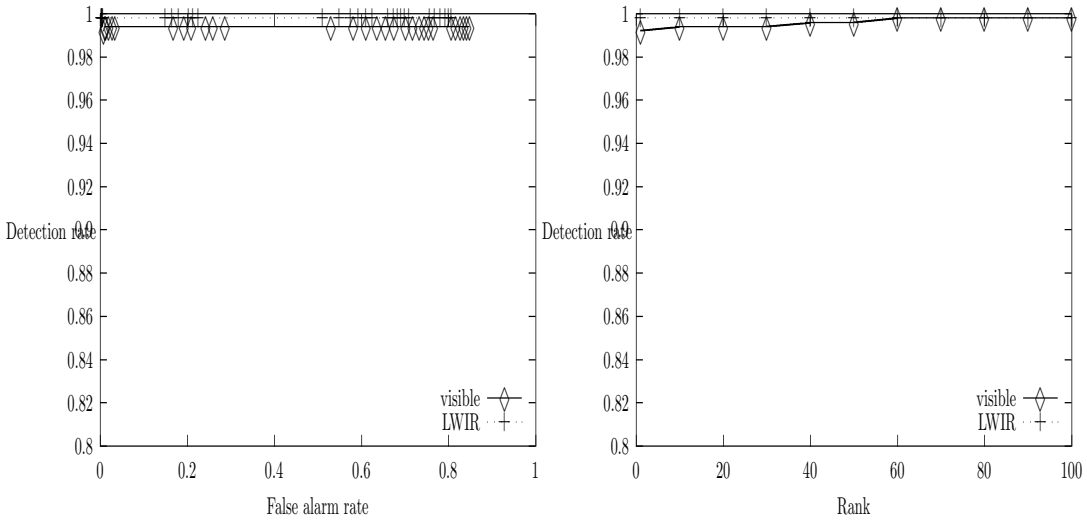

Figure 12: ROC and performance versus rank curves for LDAg with gallery VL and probe set RR.

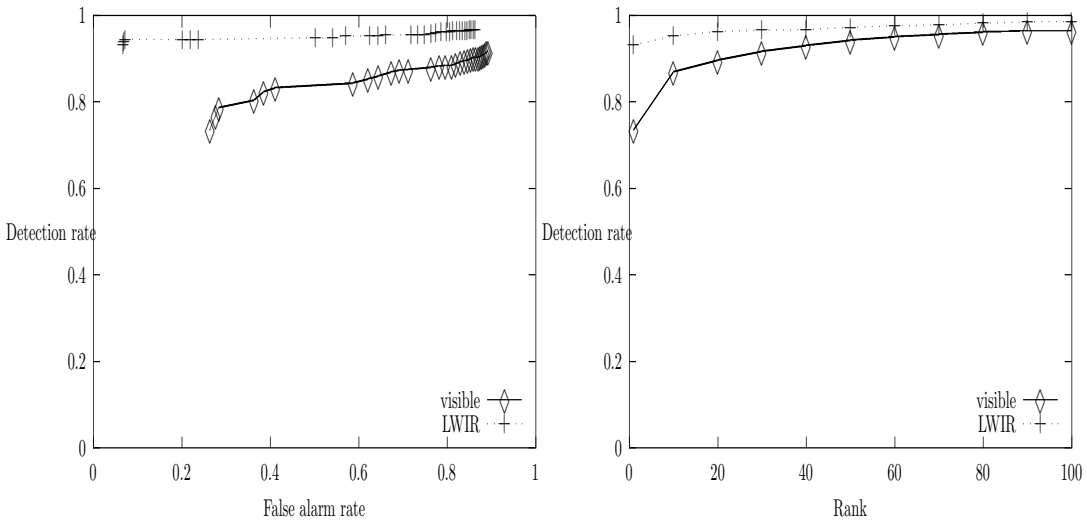

Figure 13: ROC and performance versus rank curves for LDAg with gallery VF and probe set EL.

5.3 LFA

After we computed the LFA representation of the images in query RR, we used two different methods for sparsifying the local feature set and thus reducing the dimensionality of the space. The two methods yielded very different results, as it can be seen from the first 10 features selected by each (Figures 14 and 15 respectively).

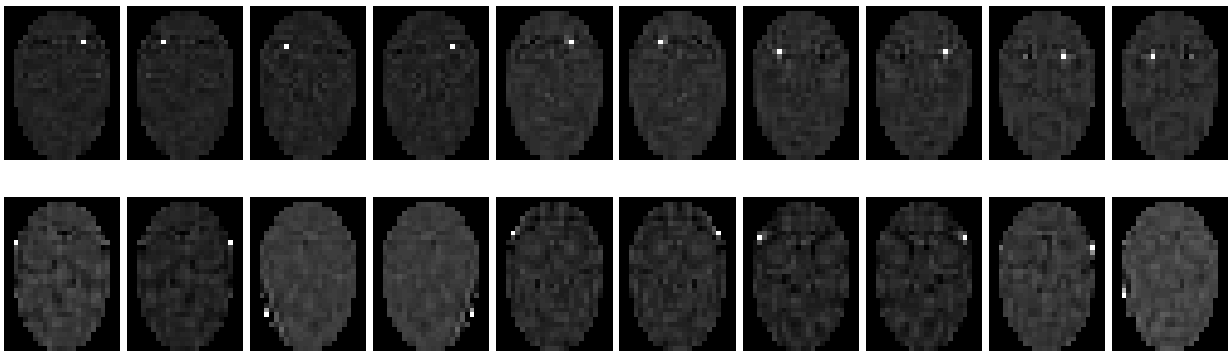

Figure 14: First 10 LFA components using Bartlett's sparsification method in the visible (top row) and LWIR (bottom row) domains

Although the features selected by our sparsification algorithm (LFAe) span a larger area of the face, Bartlett's sparsification method (LFAb) performs much better in the visible domain, obtaining an overall recognition rate of 83%, as opposed to 74% (see Tables 4, 5, and 7). Similarly to the other algorithms, performance in LWIR is significantly better than performance in the visible domain. But in this case the difference between the performances of the two sparsification methods is insignificant.

Figures 16, 17, 18 and 19 show ROC curves and performance versus rank curves for the best and

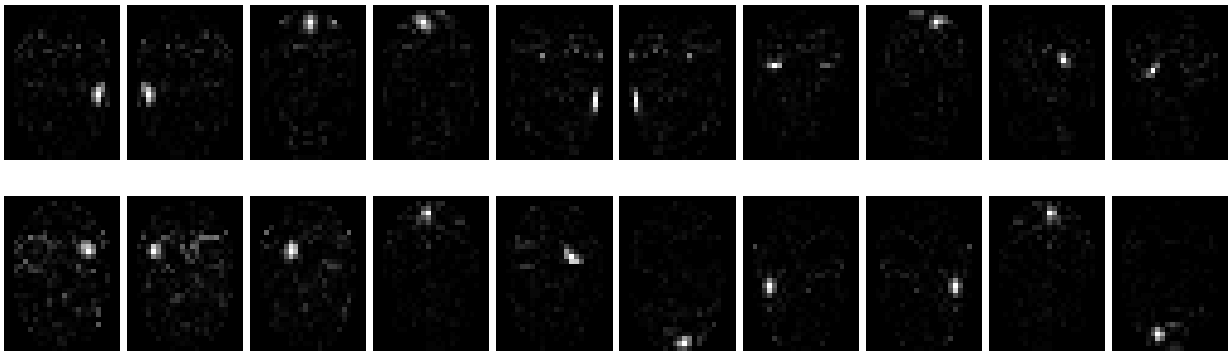

Figure 15: First 10 LFA components using our sparsification algorithm in the visible (top row) and LWIR (bottom row) domains

	VA	EA	VF	EF	VL	EL
VA		$\binom{0.880}{0.933}$	$\binom{0.930}{0.956}$	$\binom{0.724}{0.845}$	$\binom{0.973}{0.993}$	$\binom{0.846}{0.911}$
EA	$\binom{0.795}{0.897}$		$\binom{0.666}{0.832}$	$\binom{0.903}{0.929}$	$\binom{0.749}{0.861}$	$\binom{0.966}{0.974}$
VF		$\binom{0.883}{0.940}$		$\binom{0.833}{0.876}$	$\binom{0.920}{0.979}$	$\binom{0.789}{0.890}$
EF	$\binom{0.780}{0.881}$		$\binom{0.747}{0.840}$		$\binom{0.668}{0.835}$	$\binom{0.900}{0.923}$
VL		$\binom{0.878}{0.930}$	$\binom{0.894}{0.934}$	$\binom{0.668}{0.829}$		$\binom{0.875}{0.922}$
EL	$\binom{0.804}{0.905}$		$\binom{0.624}{0.827}$	$\binom{0.853}{0.892}$	$\binom{0.790}{0.874}$	
VG		$\binom{0.891}{0.942}$	$\binom{0.912}{0.970}$	$\binom{0.712}{0.889}$	$\binom{0.972}{0.995}$	$\binom{0.839}{0.935}$
EG	$\binom{0.822}{0.942}$		$\binom{0.687}{0.872}$	$\binom{0.879}{0.961}$	$\binom{0.774}{0.913}$	$\binom{0.961}{0.985}$
RR	$\binom{0.960}{0.986}$	$\binom{0.900}{0.950}$	$\binom{0.888}{0.926}$	$\binom{0.760}{0.874}$	$\binom{0.926}{0.962}$	$\binom{0.888}{0.926}$

Table 4: LFAb performance in the visible (top) and the LWIR(bottom) domains

worst performing gallery-probe combinations. As we observed with other algorithms, pairs that perform poorly in the visible domain improve considerably in LWIR, but even good performances are improved by switching modalities.

LFA performs better than PCA in the visible domain, worse in LWIR, and worse than LDA in both modalities. Similarly to PCA, worst performance is achieved when both the illumination and facial expression are varied between the gallery and probe images.

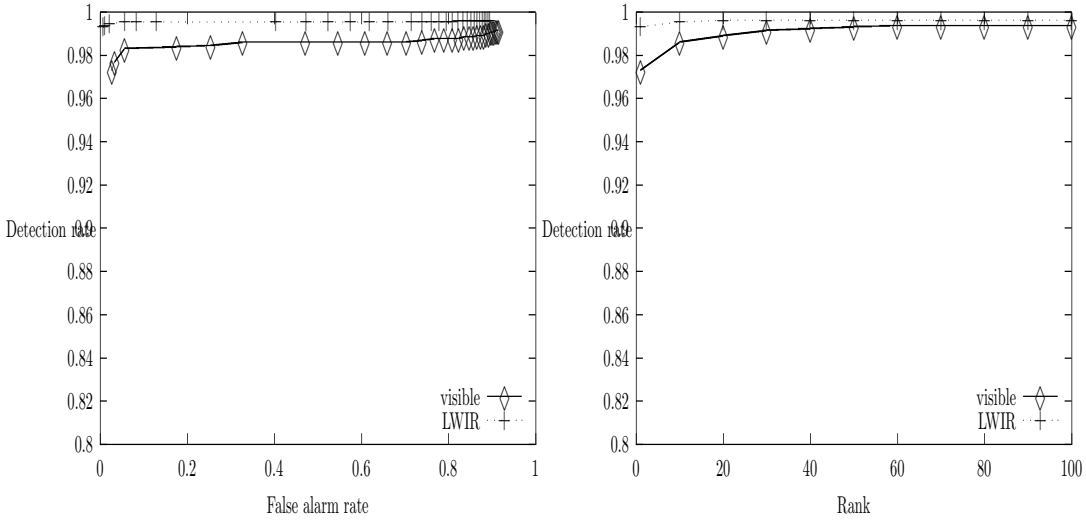

Figure 16: ROC and performance versus rank curves for LFAb with gallery VL and probe VA.

	VA	EA	VF	EF	VL	EL
VA		$\binom{0.794}{0.936}$	$\binom{0.886}{0.966}$	$\binom{0.567}{0.838}$	$\binom{0.958}{0.988}$	$\binom{0.761}{0.928}$
EA	$\binom{0.689}{0.905}$		$\binom{0.531}{0.833}$	$\binom{0.829}{0.934}$	$\binom{0.641}{0.880}$	$\binom{0.935}{0.976}$
VF		$\binom{0.762}{0.933}$		$\binom{0.682}{0.856}$	$\binom{0.876}{0.965}$	$\binom{0.680}{0.913}$
EF	$\binom{0.638}{0.879}$		$\binom{0.608}{0.835}$		$\binom{0.527}{0.842}$	$\binom{0.810}{0.930}$
VL		$\binom{0.811}{0.937}$	$\binom{0.828}{0.949}$	$\binom{0.509}{0.829}$		$\binom{0.803}{0.936}$
EL	$\binom{0.715}{0.918}$		$\binom{0.491}{0.832}$	$\binom{0.740}{0.900}$	$\binom{0.700}{0.899}$	
VG		$\binom{0.816}{0.954}$	$\binom{0.832}{0.977}$	$\binom{0.581}{0.882}$	$\binom{0.926}{0.993}$	$\binom{0.777}{0.949}$
EG	$\binom{0.697}{0.925}$		$\binom{0.512}{0.867}$	$\binom{0.795}{0.951}$	$\binom{0.656}{0.906}$	$\binom{0.913}{0.985}$
RR	$\binom{0.884}{0.986}$	$\binom{0.782}{0.952}$	$\binom{0.764}{0.944}$	$\binom{0.586}{0.852}$	$\binom{0.844}{0.970}$	$\binom{0.752}{0.934}$

Table 5: LFAe performance in the visible (top) and LWIR (bottom) domains

5.4 ICA

The first 10 independent components in the visible and LWIR domains can be seen in Figure 20.

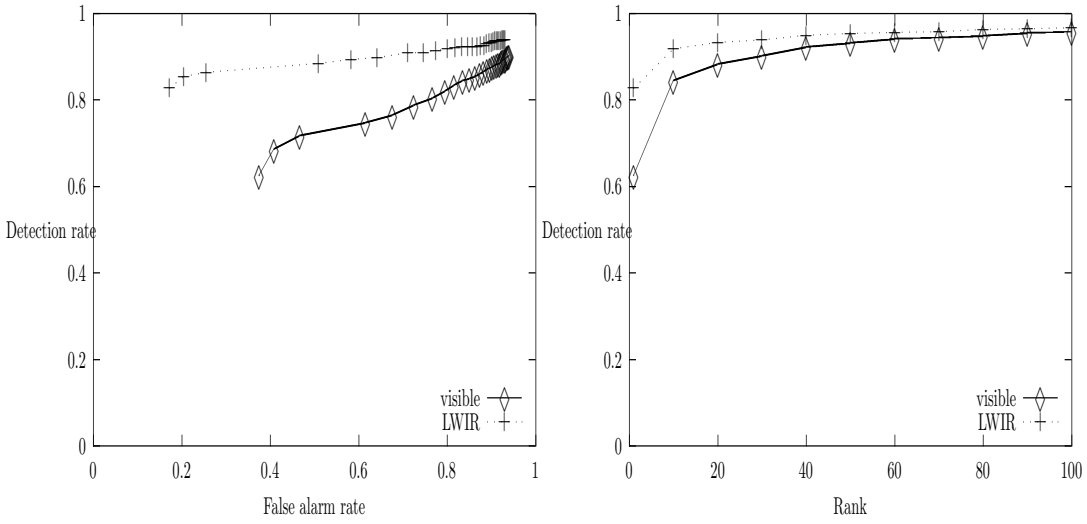

Figure 17: ROC and performance versus rank curves for LFAb with gallery VF and probe set EL.

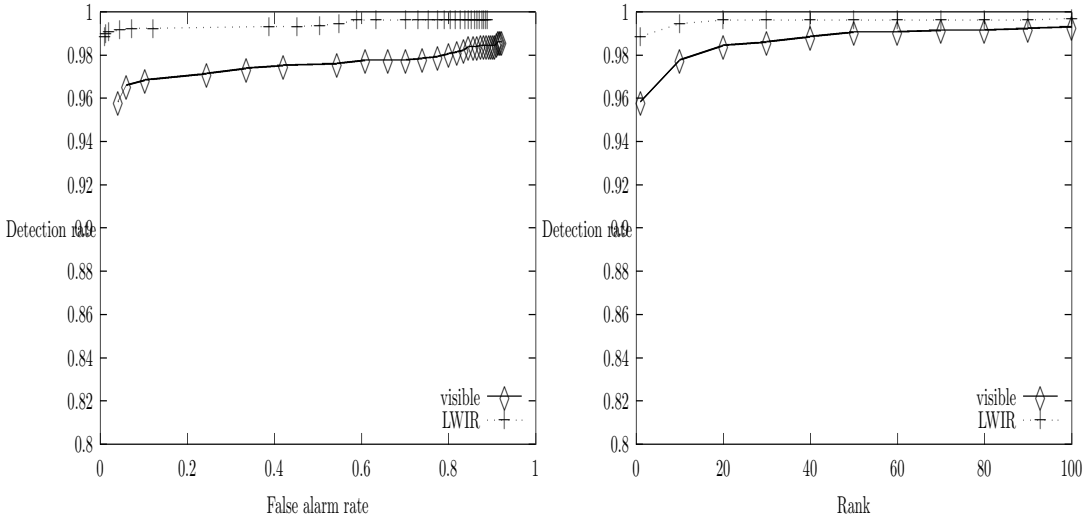

Figure 18: ROC and performance versus rank curves for LFAe with gallery VL and probe VA.

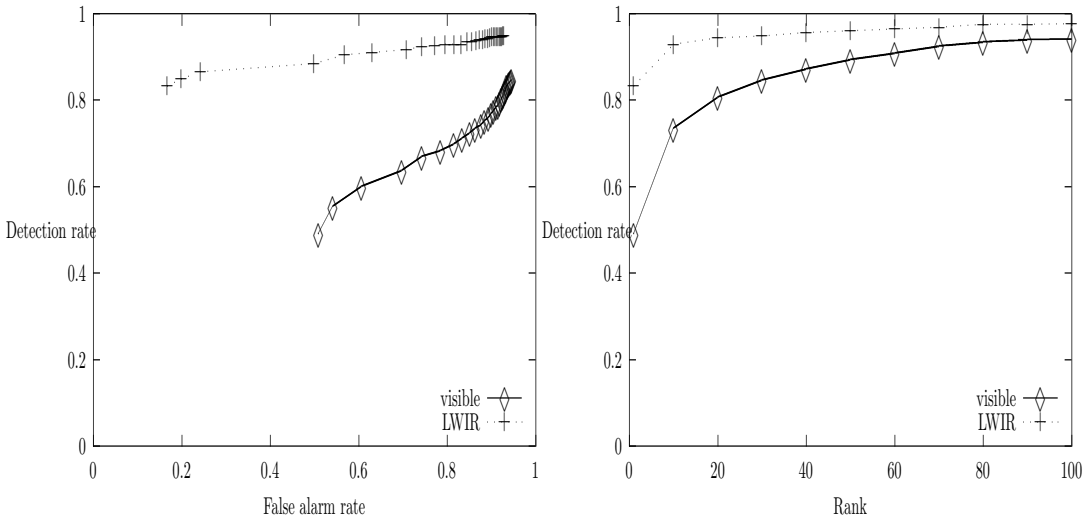

Figure 19: ROC and performance versus rank curves for LFAe with gallery VF and probe set EL.

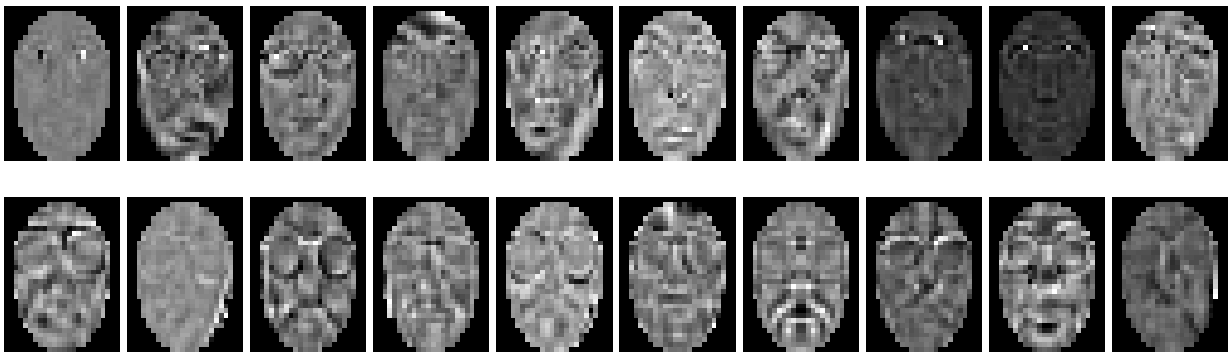

Figure 20: First 10 independent components in the visible (top row) and LWIR (bottom row) domains

	VA	EA	VF	EF	VL	EL
VA		$\binom{0.902}{0.949}$	$\binom{0.955}{0.977}$	$\binom{0.769}{0.870}$	$\binom{0.983}{0.992}$	$\binom{0.881}{0.938}$
EA	$\binom{0.843}{0.933}$		$\binom{0.734}{0.876}$	$\binom{0.928}{0.948}$	$\binom{0.819}{0.907}$	$\binom{0.977}{0.983}$
VF		$\binom{0.906}{0.945}$		$\binom{0.858}{0.888}$	$\binom{0.952}{0.977}$	$\binom{0.853}{0.915}$
EF	$\binom{0.817}{0.916}$		$\binom{0.798}{0.868}$		$\binom{0.750}{0.868}$	$\binom{0.935}{0.951}$
VL		$\binom{0.900}{0.951}$	$\binom{0.932}{0.966}$	$\binom{0.724}{0.862}$		$\binom{0.895}{0.950}$
EL	$\binom{0.857}{0.941}$		$\binom{0.701}{0.880}$	$\binom{0.891}{0.922}$	$\binom{0.855}{0.928}$	
VG		$\binom{0.919}{0.965}$	$\binom{0.958}{0.993}$	$\binom{0.813}{0.919}$	$\binom{0.990}{0.997}$	$\binom{0.889}{0.958}$
EG	$\binom{0.891}{0.951}$		$\binom{0.790}{0.908}$	$\binom{0.937}{0.966}$	$\binom{0.850}{0.930}$	$\binom{0.987}{0.997}$
RR	$\binom{0.978}{0.994}$	$\binom{0.918}{0.964}$	$\binom{0.910}{0.962}$	$\binom{0.798}{0.896}$	$\binom{0.956}{0.980}$	$\binom{0.906}{0.952}$

Table 6: ICA performance in the visible (top) and LWIR (bottom) domains

The top-match performance of the ICA algorithm in the visible and LWIR domains can be seen in Table 6. Again, performance in the LWIR is always better than in the visible.

ICA performs better than LFA with the two sparsification algorithms, but still worse than the Fisherfaces algorithm. As in the case of PCA and LFA, worst performance is achieved when both the illumination and facial expression differ in the gallery and the probe set. LWIR brings a significant improvement in this case, but it improves even the best visible performance. ROC and performance versus rank curves for the best and worst performing gallery-probe set pairs can be seen in Figures 21 and 22.

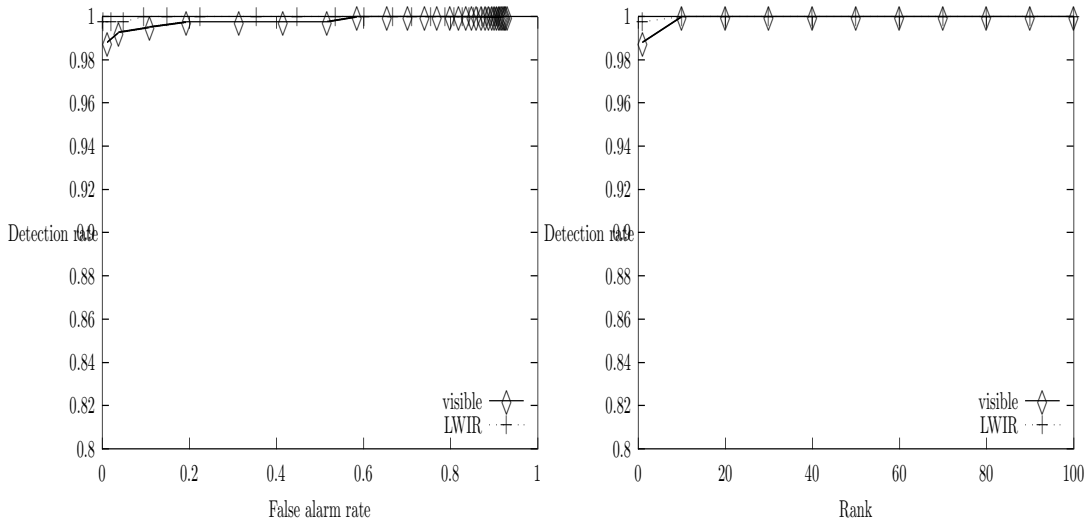

Figure 21: ROC and performance versus rank curves for ICA with gallery EL and probe set EG.

5.5 Discussion

Over all experiments performed, results on visible imagery are always inferior to those on LWIR imagery. This is not only the case for testing/training pairs where the illumination conditions are different, but indeed holds even for those pairs where we have no intuitive reason to expect performance on LWIR to be superior.

Recognition performance on visible imagery, regardless of algorithm, is worst for pairs where both illumination and facial expression vary between the training and testing sets, followed by pairs where either illumination or expression differ. Note that due to the reflective nature of visible light imaging, a change in facial expression implies a change in shading (even in uniform areas of the face) as a result of varying surface normals. Worst performance for LWIR recognition occurs for similar condition pairs. We should briefly mention that the best improvement between algorithms on visible imagery occurs also for these challenging pairs, indicating that more powerful representational methods are better able to reject features with poor classification potential.

Table 7 shows mean, minimum and maximum performances for each algorithm over the multiple experiments described above. Mean results are weighted according to the number of images in each testing set. The most notable property of these results is that recognition performance is always better with LWIR over visible imagery. Average error is reduced anywhere from 47% to 84%, depending on

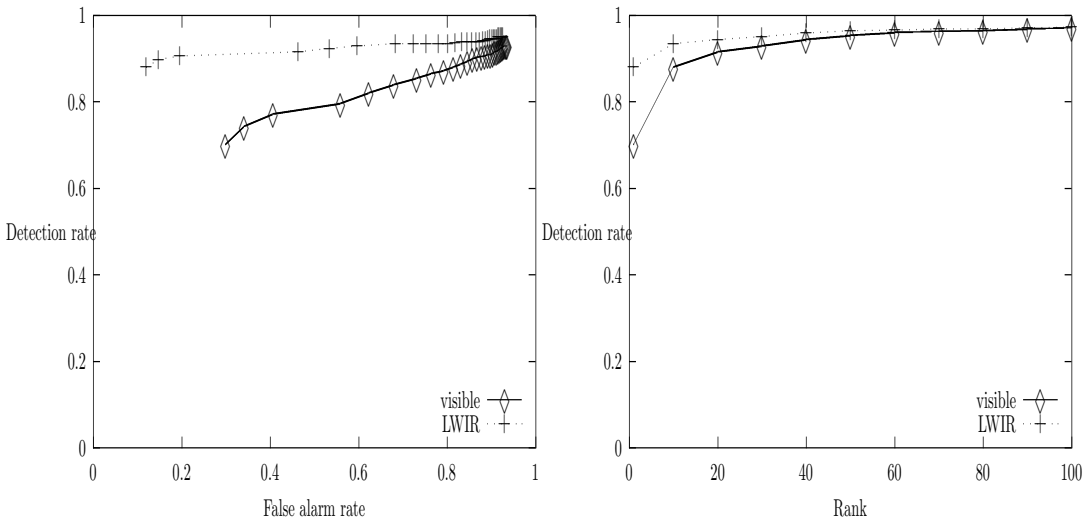

Figure 22: ROC and performance versus rank curves for ICA with gallery VF and probe set EL.

the algorithm. Similar improvement is seen for the worst- and best-case results. An additional measure of relative accuracy and stability of recognition results in the visible versus LWIR is given by the average ratio of worst to mean performance. For visible imagery we have a ratio of 0.729, while for LWIR we have 0.925, which indicates that LWIR recognition is both more accurate and more stable.

	Visible	LWIR	Error Reduction %
PCA	0.72 / 0.35 / 0.97	0.95 / 0.89 / 1.00	84/83/100
LDAt	0.92 / 0.82 / 0.99	0.97 / 0.91 / 1.00	58/53/100
LDAg	0.92 / 0.73 / 1.00	0.98 / 0.93 / 1.00	74/74/0
LFAb	0.83 / 0.62 / 0.97	0.91 / 0.83 / 0.99	47/54/82
LFAe	0.74 / 0.49 / 0.96	0.92 / 0.83 / 0.99	68/66/83
ICA	0.88 / 0.70 / 0.99	0.94 / 0.86 / 0.99	49/54/75

Table 7: Weighted mean, minimum and maximum performance on each modality, plus percentual reduction of error from visible to LWIR.

Representative receiver-operating-characteristic curves for each algorithm and both modalities show that LWIR imagery is superior not only in terms of correct classification, but also in terms of lower false alarm rates. In fact, in order to obtain recognition performance with visible imagery comparable to top-match performance in LWIR, one must be willing to accept untenable false-alarm levels. Meanwhile representative plots of performance as a function of rank-ordered result show that top-match performance in the LWIR is comparable to that obtained with visible imagery when considering the top 10-50 matches.

Our experiments allow us not only to compare the performance in the visible domain versus the performance in the LWIR domain, but also to compare average performances of different algorithms in the same domain.

In the visible domain, PCA has the lowest performance, followed by LFA with the two sparsification methods, with Bartlett's sparsification achieving better results, then ICA, then the two versions of LDA, with very close performances. This can be expressed as:

$$PCA < LFAe < LFAb < ICA < (LDAg \lesssim LDA) \tag{21}$$

with the performance increasing from left to right. Interestingly, this order changes in the LWIR domain. Worst performance is achieved by LFA (the two sparsification methods obtain very similar results), followed by ICA, then PCA, and finally LDA is still the best performing algorithm. So the previous formula becomes:

$$(LFAb \lesssim LFAe) < ICA < PCA < (LDA \lesssim LDAg) \tag{22}$$

with the performances increasing from left to right.

The difference in algorithm performance can further be illustrated with ROC and performance versus rank curves for representative gallery/probe set pairs. We have chosen some of the more difficult cases: a change in illumination for the visible domain, with VF as gallery and VL as probe set (see Figure 23), and a change in facial expression regardless of illumination in LWIR, with EA as gallery and VA as probe set (see Figure 24.

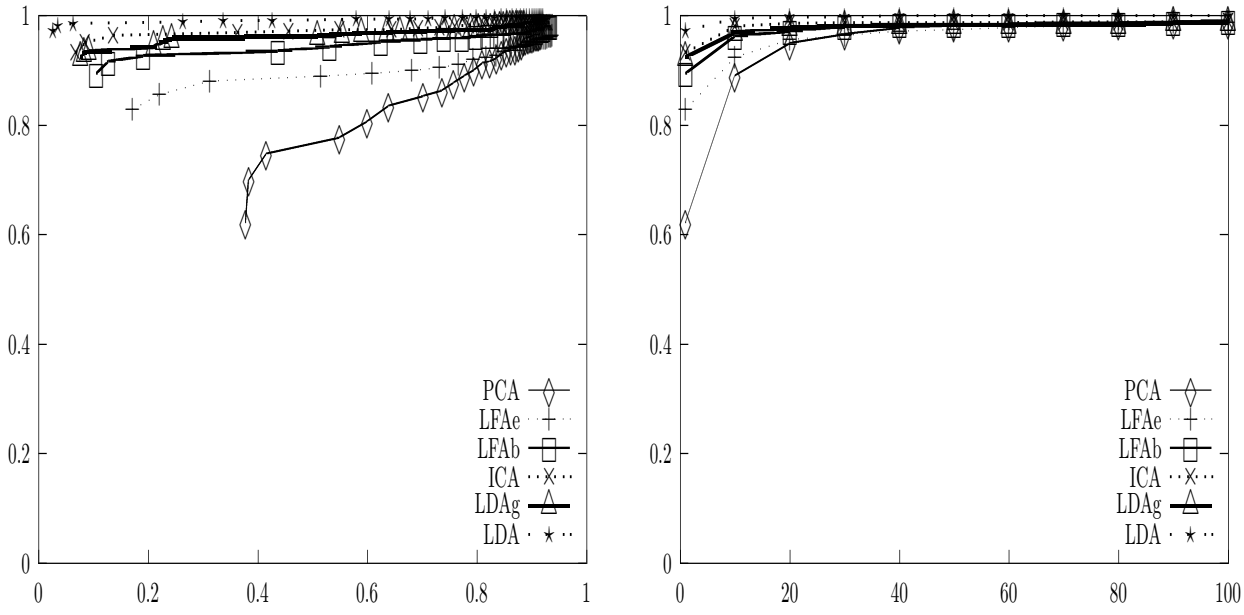

Figure 23: ROC and performance-vs-rank curves of different algorithms in the visible domain

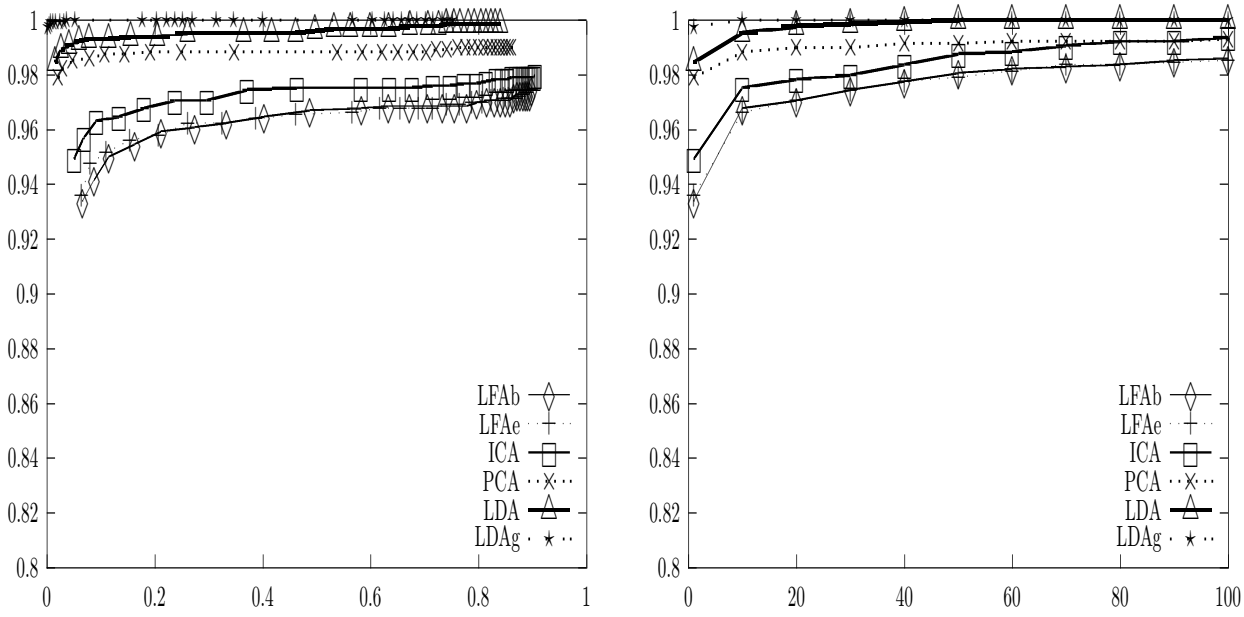

Figure 24: ROC and performance-vs-rank curves of different algorithms in the LWIR domain

6 Conclusions

We performed a comprehensive comparison of classical and state-of-the-art appearance-based face recognition algorithms applied to visible and LWIR imagery. Building on previous work, we emphasized the role of varying the training and testing sets, as a tool to uncover strengths and weaknesses of algorithms and imaging modalities. Confounding variation in imaging conditions were minimized by collecting data with an innovative sensor capable of simultaneous coregistered acquisition of both modalities.

It becomes clear from our analysis, that LWIR imagery of human faces is not only a valid biometric, but almost surely a superior one to comparable visible imagery. This conclusion must be tempered somewhat by the fact that while our data collection includes many challenging situations for visible recognition algorithms, it may not contain sufficiently challenging ones for LWIR recognition. Unfortunately, collecting such challenging imagery is costly and complicated, since we must introduce variation due to ambient temperature, wind, and metabolic processes in the subject. Nonetheless, such data collection is currently underway, and experimental results will be reported elsewhere. As noted in [8], while our current working database may not include the most challenging scenarios for LWIR face recognition, it is representative of uncontrolled indoor imagery, and thus our results are very encouraging in that context.

Ongoing and future work includes analysis on more challenging LWIR imagery, improved calibration methods to further reduce environmental distractors, and most importantly fusion of both modalities. Preliminary results on fusion of modalities are extremely promising, indicating that a further reduction of error of 50% over LWIR performance may be possible.

References

[1] F. Prokoski, "History, Current Status, and Future of Infrared Identification," in *Proceedings IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, (Hilton Head, SC), 2000.

[2] J. Wilder, P. Phillips, C. Jiang, and S. Wiener, "Comparison of Visible and Infra-Red Imagery for Face Recognition," in *Proceedings of 2nd International Conference on Automatic Face & Gesture Recognition*, (Killington, VT), pp. 182–187, 1996.

[3] Y. Adini, Y. Moses, and S. Ullman, "Face Recognition: The Problem of Compensating for Changes in Illumination Direction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 721–732, July 1997.

[4] R. Chellappa, C. Wilson, and S. Sirohey, "Human and Machine Recognition of Faces: a Survey," *Proc. IEEE*, vol. 83, no. 5, pp. 705–740, 1995.

[5] A. Samal and P. Iyengar, "Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey," *Pattern Recognition*, vol. 25, pp. 65–77, 1992.

[6] L. Wolff, D. Socolinsky, and C. Eveland, "Quantitative Measurement of Illumination Invariance for Face Recognition Using Thermal Infrared Imagery," in *Proceedings IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, (Kauai, HI), 2001.

[7] P. N. Belhumeur and D. J. Kriegman, "What is the Set of Images of an Object under all Possible Lighting Conditions?," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1996.

[8] D. Socolinsky, L. Wolff, J. Neuheisel, and C. Eveland, "Illumination Invariant Face Recognition Using Thermal Infrared Imagery," in *Proceedings CVPR*, (Kauai, HI), December 2001.

[9] R. M. McCabe, "Best Practice Recommendations for the Capture of Mugshots," tech. rep., NIST, September 1997. Available at http://www.itl.nist.gov/iaui/894.03/face/bprmug3.html.

[10] J. R. Beveridge, K. She, B. A. Draper, and G. H. Givens, "Parametric and Nonparametric Methods for the Statistical Evaluation of Human ID Algorithms," in *Third Workshop on Empirical Evaluation Methods in Computer Vision*, (Kauai, HI), December 2001.

[11] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience*, vol. 3, pp. 71–86, 1991.

[12] M. Turk and A. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1991.

[13] L. Sirovich and M. Kirby, "Low-Dimensional Procedure for the Characterization of Human Faces," *J. Optical Soc. of Am. A*, vol. 2, pp. 519–524, 1987.

[14] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Transactions PAMI*, vol. 19, pp. 711–720, July 1997.

[15] R. A. Fisher, “The Use of Multiple Measures in Taxonomic Problems,” *Ann. Eugenics*, vol. 7, pp. 179–188, 1936.

[16] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

[17] P. Penev and J. Attick, “Local Feature Analysis: A General Statistical Theory for Object Representation,” *Network: Computation in Neural Systems*, vol. 7, no. 3, pp. 477–500, 1996.

[18] M. Bartlett, *Face Image Analysis by Unsupervised Learning*, vol. 612 of *Kluwer International Series on Engineering and Computer Science*. Boston: Kluwer, 2001.

[19] P. Comon, “Independent component analysis: a new concept?,” *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[20] A. J. Bell and T. J. Sejnowski, “An Information-Maximization Approach to Blind Separation and Blind Deconvolution,” *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.

[21] A. Hyvarinen, “Fast and Robust Fixed-Point Algorithms for Independent Component Analysis,” *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[22] C. Liu and H. Wechsler, “Comparative Assesment of Independent Component Analysis (ICA) for Face Recognition,” in *Proceedings of the Second Int. Conf. on Audio- and Video-based Biometric Person Authentication*, (Washington, DC), March 1999.

[23] P. J. Phillips, H. Moon, S. A. Rizvi, and P. A. Rauss, “The FERET Evaluation Methodology for Face Recognition Algorithms,” Tech. Rep. NISTIR 6264, National Institute of Standards and Technology, January 1999.