A Comparative Analysis of Face Recognition Performance with Visible and Thermal Infrared Imagery*

Diego A. Socolinsky† Andrea Selinger‡

†Equinox Corporation
207 East Redwood Street
Baltimore, MD 21202

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

{diego,andrea}@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 compensating 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.

Thermal infrared imagery of faces is nearly invariant to changes in ambient illumination [4]. 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. 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 [5], 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

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

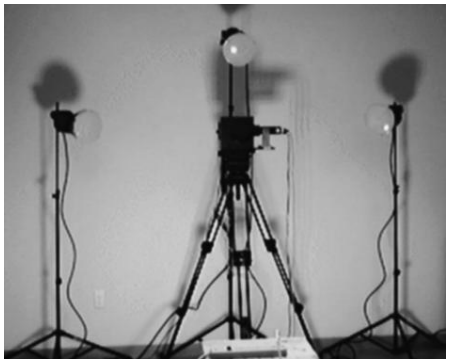

Figure 1: Camera and lighting setup for data collection.

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.

All of the LWIR imagery was radiometrically calibrated. 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. A complete explanation of the process can be found in [5], but we should note here that radiometic calibration of thermal images removes extrinsic variations due to sensor and environmental factors, yielding a physically meaningful measurement of the scene's radiance.

2.1 The Collection Setup

For the collection of our images, we used the FBI mugshot standard light arrangement, shown in Figure 1. 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 2 shows a sampling of the data in both modalities.

Figure 2: Sample imagery from our data collection.

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 [5], 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 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,

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

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 coregistered LWIR frames. Using these feature locations, all images were geometrically transformed to a common standard, and cropped to eliminate all but the inner face. Query set RR is used to compute all relevant subspaces and basis sets for the algorithms below, unless otherwise noted. Additionally, some testing/training combinations are omitted from the tables due to inclusion relations.

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. The different subspace constructions are briefly outlined below. For complete details see [6]. 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 [7]. The *face space* is computed by taking a (usually separate) set of training observations, and finding the unique ordered orthonormal basis of the data space that diagonalizes the covariance matrix of those observations, ordered by the variances along the corresponding one-dimensional subspaces. These vectors are known as principal components, or *eigenfaces*. It is well-known that, for a fixed choice of n, the subspace spanned by the first n basis vectors 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. Under the assumption of homoscedastic gaussianly distributed classes and linear separability, one can show that the optimal subspace in which to perform classification is spanned by the solution vectors w of the following generalized eigenvalue problem $S_b\, w = \lambda S_w\, w$, where S_w and S_b are the within-class and between-class scatter matrices, respectively. This gives rise

to the algorithm popularized as Fisherfaces [8]. We consider two slight variants, referred to below as LDAg and LDAt, details on the differences may be found in [6].

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 [9] 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 [10] and the other explained in detail in [6], referred to below as LFAb and LFAe, respectively.

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. 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 algorithm described in [11].

5 Experimental Results and Discussion

Images were subsampled by a factor of 10 in each dimension prior to experimentation. Visible images were demeaned and normalized to unit norm in order to provide some measure of illumination compensation. Thermal images were processed via two-point radiometric calibration. Subspaces for PCA, LFA and ICA were chosen to be 100-dimensional, and the LDA subspaces have as many dimensions as classes in the training set, minus one.

For each valid pair of training and testing sets, we computed the top-match recognition performance, and reported it below in Tables 1, 2, 3, and 4. Each column in a given table corresponds to a training set, and each row to a testing set. Visible results are reported above the corresponding LWIR results. Note that, 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 5 shows mean, minimum and maximum performances for each algorithm over the multiple experiments in Tables 1, 2, 3, and 4. 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 83%, depending on 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 0.719, while for LWIR we have 0.936, which indicates that LWIR recognition is both more accurate and more stable.

Figure 3 shows representative receiver-operating-characteristic curves for each algorithm and both modalities. We can see 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. Figure 4 shows representative plots of performance as a function of rank-ordered result. Once again, we see that top-match performance in the LWIR is comparable to that obtained with visible imagery when considering the top 10-50 matches. A more thorough analysis of these phenomena can be found in [6].

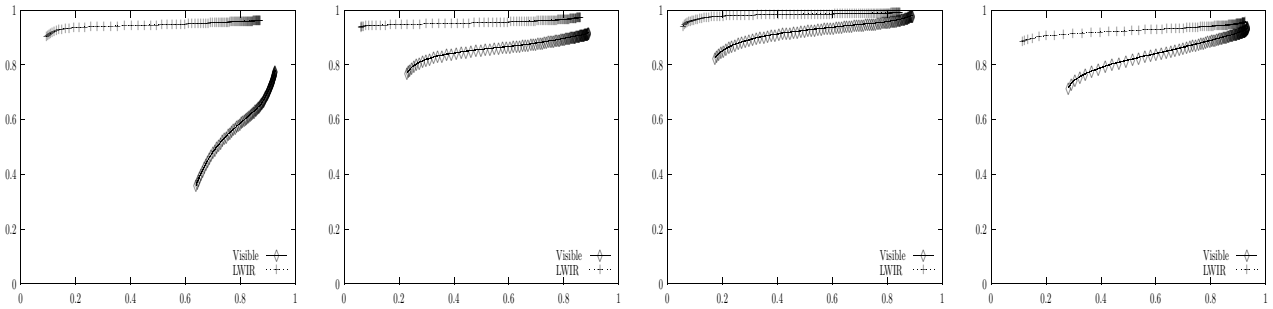

Figure 3: ROC curves for Eigenfaces, LDA, LFAb and ICA, respectively.

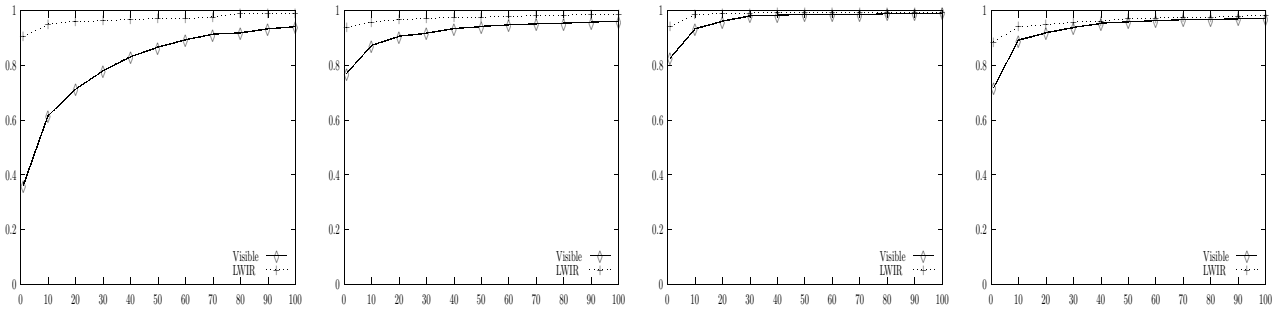

Figure 4: Performance-vs-rank curves for Eigenfaces, LDA, LFAb and ICA, respectively.

	VA	EA	VF	EF	VL	EL
VA		$\binom{0.867}{0.979}$	$\binom{0.731}{0.971}$	$\binom{0.509}{0.927}$	$\binom{0.877}{0.990}$	$\binom{0.747}{0.967}$
EA	$\binom{0.806}{0.945}$		$\binom{0.498}{0.894}$	$\binom{0.694}{0.950}$	$\binom{0.682}{0.930}$	$\binom{0.856}{0.983}$
VF		$\binom{0.805}{0.981}$		$\binom{0.805}{0.954}$	$\binom{0.636}{0.970}$	$\binom{0.447}{0.947}$
EF	$\binom{0.722}{0.935}$		$\binom{0.722}{0.905}$		$\binom{0.361}{0.905}$	$\binom{0.578}{0.951}$
VL		$\binom{0.899}{0.978}$	$\binom{0.596}{0.957}$	$\binom{0.359}{0.913}$		$\binom{0.899}{0.978}$
EL	$\binom{0.849}{0.951}$		$\binom{0.383}{0.888}$	$\binom{0.536}{0.924}$	$\binom{0.848}{0.943}$	
VG		$\binom{0.928}{0.970}$	$\binom{0.820}{0.993}$	$\binom{0.652}{0.949}$	$\binom{0.926}{1.000}$	$\binom{0.811}{0.967}$
EG	$\binom{0.887}{0.963}$		$\binom{0.620}{0.932}$	$\binom{0.781}{0.973}$	$\binom{0.766}{0.959}$	$\binom{0.894}{0.990}$
RR	$\binom{0.968}{0.994}$	$\binom{0.898}{0.982}$	$\binom{0.688}{0.964}$	$\binom{0.564}{0.942}$	$\binom{0.838}{0.978}$	$\binom{0.788}{0.964}$

Table 1: Eigenfaces performance.

	VA	EA	VF	EF	VL	EL
VA		$\binom{0.888}{0.937}$	$\binom{0.937}{0.963}$	$\binom{0.723}{0.844}$	$\binom{0.977}{0.993}$	$\binom{0.860}{0.914}$
EA	$\binom{0.801}{0.911}$		$\binom{0.669}{0.852}$	$\binom{0.919}{0.929}$	$\binom{0.772}{0.875}$	$\binom{0.974}{0.977}$
VF		$\binom{0.872}{0.940}$		$\binom{0.817}{0.876}$	$\binom{0.933}{0.979}$	$\binom{0.799}{0.894}$
EF	$\binom{0.768}{0.898}$		$\binom{0.745}{0.870}$		$\binom{0.685}{0.835}$	$\binom{0.923}{0.935}$
VL		$\binom{0.896}{0.935}$	$\binom{0.905}{0.945}$	$\binom{0.675}{0.827}$		$\binom{0.892}{0.924}$
EL	$\binom{0.818}{0.918}$		$\binom{0.630}{0.843}$	$\binom{0.878}{0.892}$	$\binom{0.817}{0.896}$	
VG		$\binom{0.903}{0.947}$	$\binom{0.926}{0.977}$	$\binom{0.770}{0.901}$	$\binom{0.981}{0.995}$	$\binom{0.878}{0.935}$
EG	$\binom{0.826}{0.942}$		$\binom{0.709}{0.891}$	$\binom{0.918}{0.959}$	$\binom{0.786}{0.913}$	$\binom{0.983}{0.990}$
RR	$\binom{0.970}{0.986}$	$\binom{0.888}{0.950}$	$\binom{0.898}{0.936}$	$\binom{0.784}{0.866}$	$\binom{0.946}{0.980}$	$\binom{0.880}{0.928}$

Table 3: LFAb performance.

	VA	EA	VF	EF	VL	EL
VA		$\binom{0.974}{0.996}$	$\binom{0.952}{0.983}$	$\binom{0.824}{0.962}$	$\binom{0.995}{0.996}$	$\binom{0.960}{0.993}$
EA	$\binom{0.933}{0.974}$		$\binom{0.825}{0.940}$	$\binom{0.897}{0.981}$	$\binom{0.910}{0.968}$	$\binom{0.988}{0.992}$
VF		$\binom{0.984}{1.000}$		$\binom{0.949}{0.970}$	$\binom{0.986}{0.988}$	$\binom{0.947}{0.986}$
EF	$\binom{0.930}{0.972}$		$\binom{0.930}{0.946}$		$\binom{0.879}{0.956}$	$\binom{0.965}{0.979}$
VL		$\binom{0.969}{0.995}$	$\binom{0.928}{0.974}$	$\binom{0.761}{0.958}$		$\binom{0.967}{0.996}$
EL	$\binom{0.935}{0.974}$		$\binom{0.770}{0.937}$	$\binom{0.844}{0.971}$	$\binom{0.925}{0.974}$	
VG		$\binom{0.963}{0.993}$	$\binom{0.960}{0.993}$	$\binom{0.845}{0.981}$	$\binom{1.000}{1.000}$	$\binom{0.963}{0.993}$
EG	$\binom{0.961}{0.987}$		$\binom{0.865}{0.975}$	$\binom{0.925}{0.997}$	$\binom{0.947}{0.985}$	$\binom{0.995}{1.000}$
RR	$\binom{0.998}{1.000}$	$\binom{0.976}{1.000}$	$\binom{0.940}{0.984}$	$\binom{0.822}{0.978}$	$\binom{0.994}{0.998}$	$\binom{0.962}{0.994}$

Table 2: LDAt performance.

	VA	EA	VF	EF	VL	EL
VA		$\binom{0.911}{0.956}$	$\binom{0.959}{0.979}$	$\binom{0.791}{0.870}$	$\binom{0.983}{0.993}$	$\binom{0.892}{0.946}$
EA	$\binom{0.850}{0.933}$		$\binom{0.751}{0.887}$	$\binom{0.932}{0.948}$	$\binom{0.821}{0.913}$	$\binom{0.973}{0.984}$
VF		$\binom{0.913}{0.958}$		$\binom{0.872}{0.894}$	$\binom{0.952}{0.981}$	$\binom{0.865}{0.924}$
EF	$\binom{0.824}{0.925}$		$\binom{0.817}{0.891}$		$\binom{0.743}{0.879}$	$\binom{0.921}{0.953}$
VL		$\binom{0.910}{0.956}$	$\binom{0.938}{0.968}$	$\binom{0.750}{0.857}$		$\binom{0.906}{0.957}$
EL	$\binom{0.863}{0.937}$		$\binom{0.718}{0.885}$	$\binom{0.897}{0.922}$	$\binom{0.862}{0.930}$	
VG		$\binom{0.926}{0.956}$	$\binom{0.949}{0.993}$	$\binom{0.841}{0.910}$	$\binom{0.988}{1.000}$	$\binom{0.898}{0.949}$
EG	$\binom{0.896}{0.961}$		$\binom{0.800}{0.918}$	$\binom{0.949}{0.968}$	$\binom{0.858}{0.944}$	$\binom{0.975}{0.997}$
RR	$\binom{0.978}{0.994}$	$\binom{0.930}{0.968}$	$\binom{0.922}{0.964}$	$\binom{0.824}{0.896}$	$\binom{0.954}{0.982}$	$\binom{0.912}{0.954}$

Table 4: ICA performance.

	Visible	LWIR	Error Reduction %
PCA	0.73 / 0.36 / 0.97	0.95 / 0.89 / 1.00	83/83/100
LDAg	0.93 / 0.85 / 0.99	0.97 / 0.92 / 1.00	57/47/100
LDAt	0.92 / 0.76 / 1.00	0.98 / 0.94 / 1.00	77/74/0
LFAe	0.82 / 0.62 / 0.97	0.93 / 0.84 / 0.99	61/59/92
LFAb	0.85 / 0.63 / 0.98	0.93 / 0.83 / 0.99	47/53/73
ICA	0.88 / 0.72 / 0.99	0.94 / 0.86 / 1.00	49/50/100

Table 5: Weighted mean, minimum and maximum performance on each modality, plus percentual reduction of error from visible to LWIR.

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 [5], 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. J. 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, 2000.

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

[3] Yael Adini, Yael Moses, and Shimon Ullman, "Face Recognition: The Problem of Compensating for Changes in Illumination Direction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 721–732, July 1997.

[4] L. Wolff, D. Socolinsky, and C. Eveland, "Quantitative Measurement of Illumination Invariance for Face Recognition Using Thermal Infrared Imagery," in *Proceedings CVBVS*, Kauai, Dec. 2001.

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

[6] A. Selinger and D. Socolinsky, "Appearance-Based Facial Recognition Using Visible and Thermal Imagery: A Comparative Study," Tech. Rep., Equinox Corporation, 2001, Available at www.equinoxsensors.com.

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

[8] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Transactions PAMI*, vol. 19, no. 7, pp. 711–720, July 1997.

[9] 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.

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

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