Comparison of Visible and Infra-Red Imagery for Face Recognition

Joseph Wilder[1], *P. Jonathon Phillips*[2], *Cunhong Jiang*[1], *Stephen Wiener*[1]
1. CAIP Center, Rutgers University, Piscataway, New Jersey 08855-1390
E-mail: wilder@caip.rutgers.edu
2. US Army Research Laboratory, Target Recognition Branch, Adelphi, MD 20783
E-mail: jonathon@arl.mil

Abstract

This paper presents initial results in a study comparing the effectiveness of visible and infra-red (IR) imagery for detecting and recognizing faces in areas where personnel identification is critical, (e.g., airports and secure buildings). We compare the effectiveness of visible versus IR imagery by running three face recognition algorithms on a database of images collected for this study. There are both IR and visible images for each person in the database collected using the same scenarios. We used three very different feature-extraction and decision-making algorithms for our study to insure that the comparisons would not depend on a particular processing technique. We also present recognition results when visible and infra-red decision metrics are fused. The recognition results show that both visible and IR imagery perform similarly across algorithms and that fusion of IR and visible imagery is a viable means of enhancing performance beyond that of either acting alone. We examine the relative importance of different regions of the face for recognition. We also discuss practical issues of implementation, along with plans for the next phase of the study, face detection in an uncontrolled environment. Preliminary face detection results are presented.

1 Introduction

Security concerns have stimulated a great deal of interest in automatic face detection and face recognition. One such concern involves the necessity for spotting people in public places, comparing their face images with a database of mug shots and presenting human decision makers with a likely set of possible matches. Another concern is ensuring and/or denying access to secure areas or computers. In this case, face verification can be used to corroborate other biometric (e.g., voice print, finger print, retinal scan) or documentary identifying information. Until recently, almost all research connected with this technology has dealt with imagery in the visible spectrum [1, 2]. However, in most practical implementations of the technology, performance is sensitive to variations in illumination and interactions between changes in pose and illumination. Consequently, infra-red (IR) imagery in the 8-12 micron wavelength region has been suggested as an alternative source of information for detection and recognition of faces. However, we are not aware of any studies that support or contradict this claim in the academic literature and to our best knowledge this is the first paper to address the issue. Since sensors in this wave band respond to the thermal radiation emitted by the face, it is impervious to variations in illumination. It is, however, subject to variations in temperature in the surrounding environment, and to variations in the heat patterns of the face when an individual moves from one temperature environment to another.

This paper reports on initial efforts in determining whether IR imagery represents a viable alternative to visible in the search for a robust, practical identification system. The results reported here deal only with the recognition aspects of the problem; we will deal with detection of faces in a complex background in the next phase of our project. An important part of the evaluation is the extraction of image data in a realistic setting (i.e., one that approximates the actual conditions in a public space where people are walking and talking). Section 2 outlines the experimental conditions under which we collected a database of visible and infra-red images. Section 3 describes the pre-processing, feature-extraction and recognition techniques that we employed in the visible/IR comparison. We used three very different feature-extraction and recognition algorithms to render the comparison independent of a particular processing technique. (We should emphasized that the three techniques were not,

0-8186-7713-9/96 $5.00 © 1996 IEEE

necessarily, implemented optimally, and that no conclusions as to their relative merits should be drawn.) We present recognition results using the three techniques in section 4, along with results for fusion of visible and IR imagery. Section 5 interprets these results in the light of the unexpected good performance obtained on the IR database. We also discuss practical considerations in implementing both visible and IR recognition systems along with our next steps in the detection of faces in complex backgrounds. Preliminary results on face detection are presented.

2 Experimental Conditions

The goals of the experiment were to collect comparable visible and IR image pairs, and to compute recognition results for various combinations of "train on one and test on the other" in a manner somewhat similar to experiments on the FERET database [2]. However, unlike the FERET experiments, where the subjects were always seated in a stable illumination environment, our subjects walked towards a camera while we captured a sequence of images. We captured eight sequences for each subject, four visible and four IR. A set of four included two images directly towards the camera and two at +/-22.5°. The subject was talking during one of the two direct sequences (saying "how are you" to guarantee significant mouth movement). The subject was silent during the other three sequences. The images used in the experiments reported here were extracted from the silent and talking direct paths. Each sequence consisted of 12 images captured while the subject was walking a distance of approximately 1 meter toward the camera. For the purposes of the experiments reported here, the fifth and eleventh image of each sequence were selected to ensure that the subject's mouth would be in motion during the talking sequence. Visible and cameras were positioned side-by-side at a distance of approximately 3.5 meters from the center of the 1-meter walking range, and angled very slightly to converge in the middle of that range.

The illumination for the visible camera consisted of an overhead ceiling fluorescent fixture, combined with diffuse, indirect lighting at about waist level directed from either side of a panel below the cameras. We erected uniform white backdrop approximately 2 meters behind the walking range.

The temperature of the room was allowed to range over the several degrees of a normally air-conditioned laboratory during the course of a day. Consequently, all subjects were not scanned at precisely the same temperature. Furthermore, due to the difficulty in getting all of our subjects to return for a second scanning, all sequences from each subject were acquired during one visit to the laboratory.

The visible camera was a CIDTEC 2250 CID camera, and the infra-red camera was an uncooled sensor on temporary loan from DARPA and the US Army Night Vision & Electronic Sensors Directorate. The IR camera is a Texas Instruments SMRTII, which supplies a standard RS170 video output signal readily digitized by our Datacube MV200/Sun Sparc5 imaging system. The lens and aperture stop of the visible camera were selected to match the magnification, center of focus and depth of field of the lens system of the IR camera, which was not selectable. The match that was achieved was quite close. We stored the raw images from both cameras as 512(H)x450(V) pixel images.

The database includes image sequences from 101 subjects without glasses (we deal with the issue of glasses in the Discussion section). We conducted four recognition experiments on this database. They were:

- train on visible silent image #5, test on visible talking image #11 -(VS5>VT11)
- train on visible talking image #5, test on visible silent image #11 -(VT5>VS11)
- train on IR silent image #5, test on IR talking image #11 -(IRS5>IRT11)
- train on IR talking image #5, test on IR silent image #11 -(IRT5>IRS11)

Examples of images from the database are shown in figure 1.

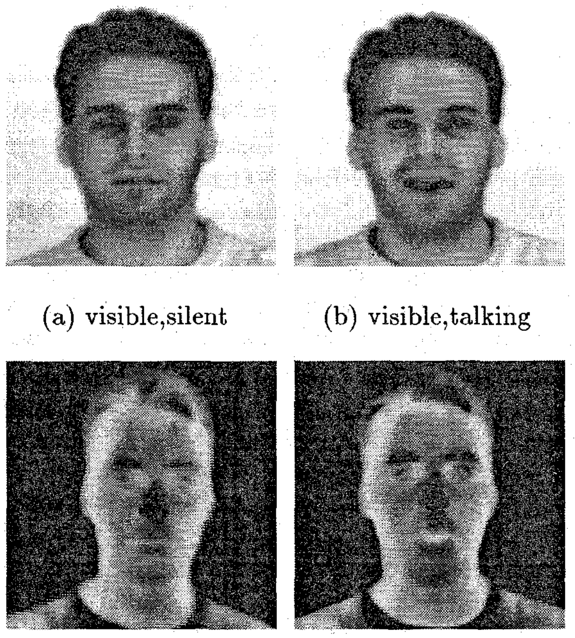
The preparation of these images for feature extraction and recognition is described in the following section.

3 Pre-processing, Feature Extraction and Recognition Algorithms

Since we were concerned with recognition in this phase of the investigation, we located faces manually by clicking with a mouse on four points on each face: the center of each eye, the tip of the nose and the center of the mouth. We then smoothed the images with a gaussian filter (s.d. = 3 pixels) to remove artifacts of the scanning process, and rotated and scaled the images so that the eyes of all subjects were in fixed columns of the same row. We then cropped the images and masked them to the individual requirements of three feature extraction and recognition algorithms. We describe these algorithms in the following paragraphs.

3.1 Transform Coding of Grey Scale Projections

Grey scale projections, (that is grey scale sums along one or more directions through a face image) converts a



(a) visible,silent (b) visible,talking

(a) IR,silent (b) IR,talking

Figure 1. **Examples of Visible and IR face images.**

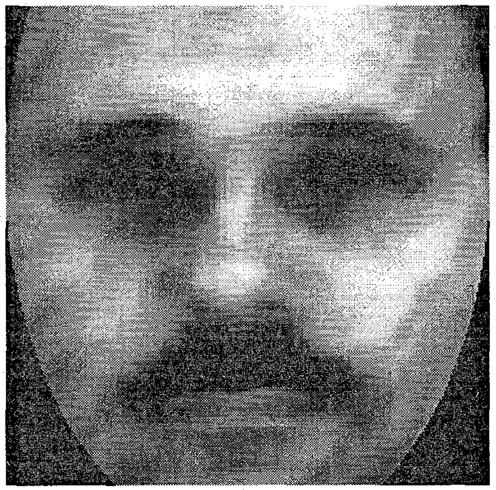


Figure 2. **Face image masked with truncated ellipse.**

two-dimensional image to one or more one-dimensional signatures. Unlike tomographic applications, where image reconstruction is the goal, a small number of projections can provide sufficient information for classification [3], [4]. Grey scale projections are readily extracted with low-cost, high-speed video hardware. In this instance, four projections are used: horizontal, vertical and +/-45°. The horizontal projections, which carry the most useful information for recognition, are also relatively insensitive to rotations of the head about the vertical axis [4]. We obtained the projections. after masking with a truncated ellipse, as shown in figure 2. We carried out transform coding of these four projections to provide greater data reduction; to decorrelate samples from the input waveforms; and, as a practical matter, to distribute local distortions due to changes in expression across all output features. We used the Discrete Cosine Transform; the feature vectors were comprised of 18 low-pass components, excluding the first two (i.e. the DC and next higher component were excluded to provide immunity to absolute light level and gradual shading). Recognition was based on a weighted sum of the distance metrics $M_h, M_v, M_{+45}, M_{-45}$ for each projection. The metric, M_j, is given by $Max(M_i) = Max(1 - D_i/D_{max})$, where D_i is the L1-norm distance between the test vector and the i^{th} training vector, and D_{max} is the L1-norm distance between the test vector and the most distant training vector. In the results reported here, $M = 0.375M_h + 0.125M_v + 0.25M_{+45} + 0.25M_{-45}$. These results are consistent with a number of experiments that show horizontal projections providing the greatest discrimination and vertical projections (at least on the full face) providing the least discrimination.

3.2 Eigenface Algorithm

The eigenface approach to face recognition was developed and reported in Turk and Pentland [7]. The basic approach of this technique is to develop a multidimensional feature space, based on a holistic analysis of the faces in the database, and then use the projections of the test face onto the feature space axes for recognition. This work followed the development of a technique for efficiently storing face images as a collection of weights, which represents the projections of the face onto the various directions of the feature space, Kirby and Sirovich [8]. The feature space is generated by determining the eigenvectors of the covariance matrix of the set of faces in the database (principle component analysis), so each face in the training set can be represented by a summation of the contribution of each eigenvector to that face. To recognize an unknown test face, the contribution of each eigenvector to the test face is calculated, and the resulting set of weights is then compared to the weights of each training set face.

We implemented the algorithm as described in Turk and Pentland [7]. For each of the four recognition experiments, we calculated a separate set of eigenvectors. We computed each set of eigenvectors for all 101 known images, and used the normalized and masked images. For recognition of visible images, each face was represented by its projection onto the first 70 eigenvectors, and for IR, each face was represented by its projections on eigenvectors 5 through 100.

3.3 Matching Pursuit Filters

A matching pursuit filter is a self-organizing technique for creating efficient and compact models from data. The design of a matching pursuit filter is based on an adapted wavelet expansion, where the expansion is adapted to both the data and the pattern recognition problem being addressed. This contrasts with most adaptation schemes, where the representation is a function of the data, but not of the problem to be solved. This approach does not decompose the images in the training set individually; rather, it determines the expansion by simultaneously decomposing all the images. By using two-dimensional wavelets as the building blocks for the decomposition, the representation is explicitly two-dimensional and is composed of local information. Matching pursuit filters are applied to face recognition by encoding a small set of facial features. Because the features are restricted to the nose and eye regions of the face, the algorithm is robust to variations in facial expression, hair style and the surrounding environment. The algorithm uses coarse-to-fine processing to estimate the location of a small set of key facial features. Based on the hypothesized locations of the facial features, the identification module searches the database for the identity of the unknown face. For further details see Phillips and Vardi [5] and Phillips [6].

Matching pursuit filters are designed from a training set of images; in this study, the training set consisted of images from a different data set than those used here. This avoids adapting the filters to a particular subset of images in the database (matching pursuit filters are usually designed from a subset of images in the database). There are two sets of filters: one for the visible images and one for the IR images. In each set of filters, there is a filter for each feature. For the visible images the features used were the left and right eyes, tip and bridge of the nose, and the face and in the IR the features are the left and right eyes, and tip and bridge of the nose. A feature corresponds to a region of the face (i.e., the left eye is the region surrounding the left eye and the face feature is the face after it has been cropped). Both the visible and IR filters were trained on images not used in the experiment. The visible filters are trained on images from the DARPA/ARL FERET database, which consist of images taken with a 35mm camera and then digitized [5, 6]. The size of the training set was 58 images. We designed the IR filters from 30 IR images acquired with an earlier version of the SMRTII [6].

4 Experimental Results

The recognition results that we obtained with the three algorithms described in the previous section are presented in the following table. We should emphasize that the algorithms we used were not optimized, and no conclusions about their relative performance should be drawn. Rather, the results are intended to show that, under the experimental conditions imposed, the relative performance of visible and IR imagery for face recognition is roughly consistent, with slight favoritism shown for one type of imagery over another, across algorithms.

Train > Test	GS Proj.		Eigen Face		Match Purs.	
	Top Match	Top 2 Matches	Top Match	Top 2 Matches	Top Match	Top 2 Matches
IRS5 > IRT11	92	93	85	91	93	95
IRT5 > IRS 11	94	95	85	91	95	98
VS5 > VT11	89	96	86	94	95	97
VT5 > VS11	93	96	88	93	97	99

Table 1. Number of correctly identified (top match) and correct within top 2 matches out of 101 subjects. Two visible and two IR experiments using three feature extraction and recognition algorithms.

In a second set of experiments, we investigated the importance for recognition of the overall structure of the face, which was accomplished with two runs of the matching pursuit filter recognition algorithm. The matching pursuit filter results in table 1 use different sets of features for the visible and IR. In this experiment we ran the matching pursuit filter algorithm on both visible and IR using three sets of features. In the first run, all the features were used (the left and right eyes, the bridge and tip of the nose, and the face). In the second run the effect of removing the face feature was investigated (the features are the left and right eyes, and the bridge and tip of the nose). In the third run the reverse effect was studied (the only feature was the face). This experiment shows that the overall shape of the face contributes to recognition for the visible imagery, whereas IR performance increases if the overall shape of the face is not included (table 2). The third run demonstrates that performance based on overall shape of the face is better in visible than IR.

	IRS5> IRT11	IRT5> IRS11	VS5> VT11	VT5> VS11
5 features	90	95	95	97
Eyes & nose	93	95	89	95
Face only	40	54	87	73

Table 2. Performance for the matching pursuit filter face recognition algorithm as the number of features is varied. The number reported is the number correctly identified (top match) out of 101 subjects.

In addition to comparing visible and infra-red imagery, it was interesting to compute the performance when both visible and infra-red information is fused. We computed a straightforward equal-weight combination of the normalized distance metrics for the grey scale projection algorithm. The results are presented in table 3.

	Top Match	Top 2 Matches
IRS5 > IRT11 / VS5 > VT11	99	100
IRT5 > IRS11 / VT5 >VS11	98	99

Table 3. Number of correct recognitions out of 101 for multi-sensor fusion of infra-red and visible distance metrics (using the grey scale projection technique).

5 Discussion

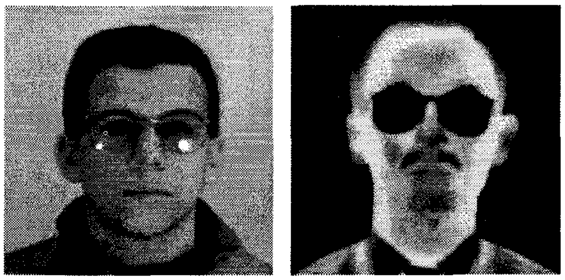
In this paper we compared recognition performance of visible versus IR imagery. We collected the images in a manner that allowed some variation in geometry and pose of the face, and illumination (as the subject walked through the image sequence space). Temperatures remained constant. To obtain a robust measure of performance, we ran the images through three face recognition algorithms. One of the algorithms performs better on IR images (GS), and two of the algorithms perform better on visible images (eigenface and matching pursuit filters). None of the algorithms showed performance significantly better for one modality than another.

We did not address a number of important issues that will concern any real-world application in this study. The main issues that we have not addressed are: significant variations in illumination for visible images, and changes in temperature for IR images. The relative importance of each of these issues will depend on the face recognition scenario under scrutiny. For verification scenarios with cooperative subjects, illumination can be controlled, whereas variations in temperature cannot be controlled.

The key to answering these questions is to measure the effects of temperature versus illumination variations. Thermal changes can be due both to physiological (running versus resting) or environmental variations. To measure these effects will require the collection of a large IR database. For visible images, there is a marked decrease in performance when images are taken at different times. In the phase-two FERET test the average performance of the three algorithms tested is 92 percent when the gallery and probe images of each person were collected within 5 minutes. (The gallery is the set of people (images) of known identity, and the task is to identify the person in the probe image.) Performance decreases to 57 percent when the gallery and probe images are collected on different days [2].

One potential limitation of IR imagery is that IR is opaque to glass (figure 3), which makes it very simple for a person to block out a large portion of their face (for example by wearing eyeglasses). This is a serious consideration if most of the information for identification is in the eye and nose region. (Visible imagery can suffer from highlights on the glasses under certain illumination conditions (figure 3), but the problems are considerably less severe than with IR). Another limitation is the considerably greater cost of an IR sensor, compared with a visible sensor. Because of their relative costs, any system with an IR sensor would probably include a visible camera. Thus, a future line of research is to develop face recognition and detection algorithms that fuse information from the two sensors.

There are two key areas in automatic processing of faces: face recognition and detection. In this paper, we have reported on the first area. We have begun our comparison of visible and IR imagery for face detection. To pursue face detection, we have collected both indoor and outdoor scenes that contain faces (see figure 4 for example). For each scene, we have collected a visible and an IR image. Our preliminary experiment, using a grey scale projection-based face finder, showed approximately equivalent performance for face finding in visible and IR imagery (as indicated by the effect of threshold sensitivity on the ROC curve). This algorithm is illustrated in figures 4(a) and 4(b). Differences in performance between visible and IR may become evident when a more detailed analysis is carried out, using several detection algorithms.



(a) Visible (b) Infra-red

Figure 3. Images with glasses

Acknowledgements

This work is supported by DARPA's FERET program with ARL serving as technical agent. Rutgers is funded by ARL contract DAAL01-93-K-0119.

References

[1] R. Chellappa, C.L. Wilson, and S. Sirohey, "Human and Machine Recognition of Faces: A Survey," *Proc. IEEE*, Vol. 83, No. 5, pp. 705-740, May, 1995.

[2] P. Rauss, P. J. Phillips, A. T. DePersia, and M. Hamilton, "Face Recognition Technology Program Overview and Results," *Proc. SPIE Conf. on AIPR96* October, 1996.

[3] J. Wilder, "Face Recognition Using Transform Coding of Gray Scale Projections and the Neural Tree Network," *Artificial Neural Networks with Applications in Speech and Vision*, Chapman Hall, Neural Computing Series, 1994.

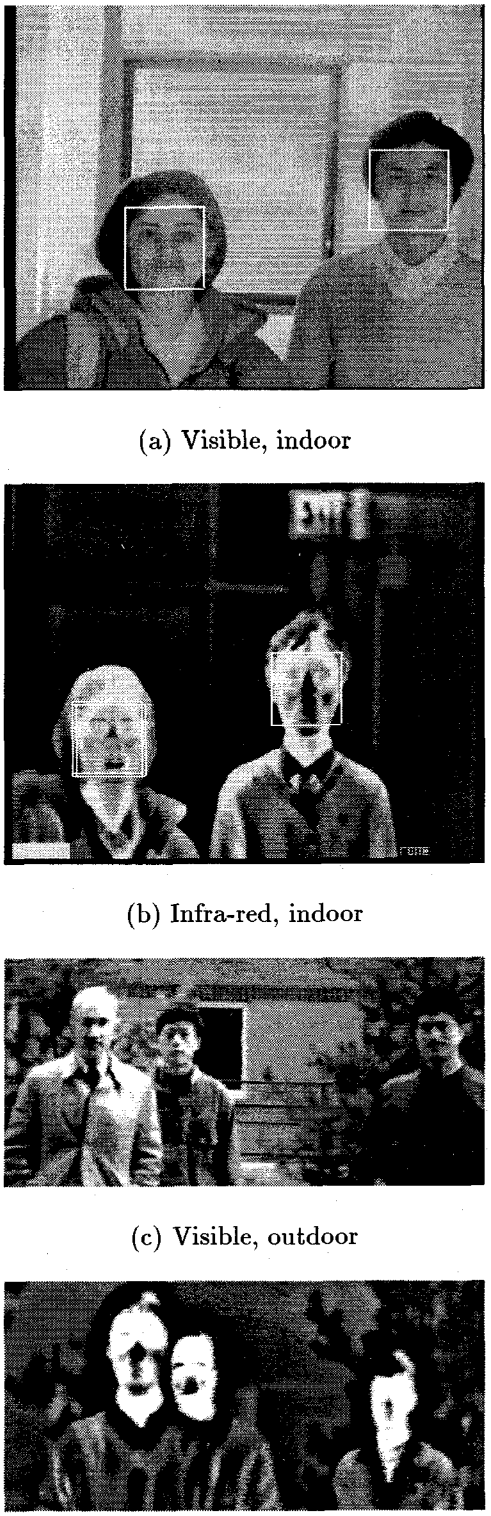
[4] J. Wilder, R. J. Mammone, P. Meer, A. Flanagan, X. Kai, A. Tsai, S. Weiner, and X. Zhang, "Projection-Based Face Recognition," *Proc. SPIE Conf. on Automatic Systems for the Identification and Inspection of Humans*, Vol. 2277, pp. 22-31, July, 1994.

[5] P. J. Phillips and Y. Vardi, "Data Driven Methods in Face Recognition," *International Workshop on Automatic Face and Gesture Recognition*, pp. 65-70, June, 1995.

[6] P. J. Phillips, *Representation and Registration in Face Recognition and Medical Imaging*, PhD thesis, RUTCOR, Rutgers University, 1996.

[7] M. Turk and A. Pentland, "Eigenfaces for Recognition", *J. Neuroscience*, Vol. 3, No. 1, 1991.

[8] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces," *IEEE trans. PAMI*, Vol. 12, No. 1, 1990.



(a) Visible, indoor

(b) Infra-red, indoor

(c) Visible, outdoor

(d) Infra-red, outdoor

Figure 4. Group images to be used for face detection