# A Face Detection Method Based on Multi-Band Feature Extraction in the Near-IR Spectrum


Jonathan Dowdall
Honeywell Laboratories
3660 Technology Drive
Minneapolis MN 55418
jonathan.dowdall@honeywell.com

Ioannis Pavlidis[1]
Honeywell Laboratories
3660 Technology Drive
Minneapolis, MN 55418
ioannis.pavlidis@honeywell.com

George Bebis
Department of Computer Science
University of Nevada, Reno
Reno, NV 89557
bebis@cs.unr.edu



## Abstract

*Face detection is an important prerequisite step for successful face recognition. The performance of previous face detection systems reported in the literature seems to deteriorate ungracefully in outdoor environments where lighting conditions cannot be controlled. We propose a method in par, in terms of performance, with state-of-the-art face detection methods in indoor environments. In addition to performing indoors, our method also performs well in outdoor environments with variable lighting conditions. The approach capitalizes upon our near-IR skin detection method reported elsewhere [11][12]. It determines the orientation and extent of the face within the detected skin region by extracting features from multiple near-IR bands. Then, a generalized Hough transform, modeled after the expected appearance of the human eye in the feature image, is applied at each point. The strongest response to the generalized template yields an eye blob image from which we eventually locate the centers of the subject's eyes. The validity of the method is substantiated by extensive experimental results.*


## 1 Introduction

Face detection and recognition have been active research areas for more than thirty years. Face detection is an important preprocessing stage of an overall face recognition system. Although, it may appear rudimentary to a layman, face detection is a challenging machine vision operation, particularly in outdoor environments where illumination varies greatly. This is one of the primary reasons that face recognition is currently constrained to access control applications in indoor settings.

There is a pressing need for expanding the application of face recognition technologies to surveillance and monitoring scenarios. Such systems would be most advantageous in the context of protecting high value assets (e.g. perimeter of government buildings) from asymmetric (terrorist) threats. A major technical challenge that needs to be addressed in this direction is the low performance of face detectors in unconstrained environments. Visible-band face detectors, as those reported in the literature, opt for pure algorithmic solutions into inherent phenomenology problems. Human facial signatures vary significantly across races in the visible band. This variability coupled with dynamic lighting conditions present a formidable problem. Reducing light variability through the use of an artificial illuminator is rather awkward in the visible band because it may be distracting to the eyes of the people in the scene and "advertises" the existence of the surveillance system.

In the current paper we present a novel face detection system based on near-IR phenomenology, multi-band feature extraction, and the use of generalized Hough transforms. Facial signatures are less variable in near-IR aiding significantly the detection work. Illumination in the scene can be maintained at an optimal level through a feedback control loop that includes a near-IR illuminator. Since, near-IR light is invisible to the human eye the system can remain unobtrusive and covert. The above advantages in combination with the unique reflectance characteristics of the human skin in the near-IR spectrum allow for simple algorithmic-based face detection methods to perform extremely well.

The rest of the paper is organized as follows: In Section 2 we give an overview of previous work done in the area of face detection. In Section 3 we give a top-level description of the hardware and software architecture of our face detector. In Section 4 we provide a brief description of our skin detection method. In Section 5 we elaborate on our face detection method, which builds upon our skin detection method. In Section 6 we present and discuss the experimental results. Finally, in Section 7 we conclude the paper and mention our plans for future work.

[1] To whom all correspondence should be addressed.

## 2 Previous Work

In recent years a sizable body of research in the area of face detection has been amassed. The methodologies vary, but the research mainly centers around three different approaches: artificial neural networks, feature extraction, and wavelet analysis. Each of these approaches has its respective strengths and weaknesses when applied to face detection, but none has yet been able to attain results rivaling human perception.

The majority of face detection research has been focused around various types of feature extraction. Feature extraction methods utilize various properties of the face and skin to isolate and extract desired data. Popular methods include skin color segmentation [1][2], principal component analysis [3], Eigenspace modeling [4], histogram analysis [5], texture analysis [6], and frequency domain features [7].

Face detection research based on artificial neural networks has received a smaller share of the attention. One of the problems with this approach is finding a representative data set. This difficulty is compounded by the fact that a strong counter example set must also be compiled to train the individual networks. Despite these obstacles many of the most promising results have been reported form research involving artificial neural networks. In his work Rowley *et al.* [8] used an arbitration method among several networks to improve performance. His system produced some impressive results for forward facing subjects.

Wavelet analysis is the newest of the face detection approaches under discussion. The general aim of the wavelet approach is maximum class discrimination and signal dimensionality reduction [9]. Due to the reduced dimensionality, wavelet-based methods are computationally efficient.

All of the above approaches are associated with visible spectrum imagery. Therefore, they are susceptible to light changes [10] and the variability of human facial appearance in the visible band. A distinct line of research pursued by our group proposed the fusion of two near-IR bands for the detection of face and other exposed skin areas of the body [11][12]. The method capitalizes upon some unique properties of the human skin in the near-IR spectrum. Our dual-band system maintains an optimal illumination in the scene through the liberal use of artificial non-distracting near-IR lights. As a result, the system performs superb skin detection both in indoor and outdoor settings. In the present paper, we report further algorithmic work that accurately locates the face within the detected skin region.

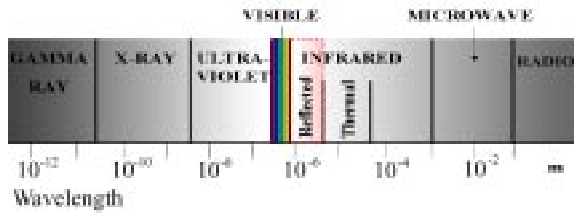



**Figure 1:** The EM spectrum.

## 3 System Overview

### 3.1 Hardware Architecture

The system uses two cameras as the input medium. The cameras are sensitive to the so called near-IR spectrum in the range 0.9-1.7 $\mu$m. This range clearly falls within the reflected portion of the infrared spectrum and has no association with thermal emissions (see Figure 1). The two cameras are set at perpendicular angles (see Figure 2) and a beam splitter is used to allow both cameras to view the scene from the same vantage point, yet in different sub-bands. The splitter divides the light reflected from the scene into the lower-band beam (0.9-1.4 $\mu$m) and the upper-band beam (0.4-1.7 $\mu$m). The two beams are funneled to the Focal Plane Arrays (FPA) of the corresponding cameras. Each camera is connected to a frame grabber, which digitizes the incoming video.

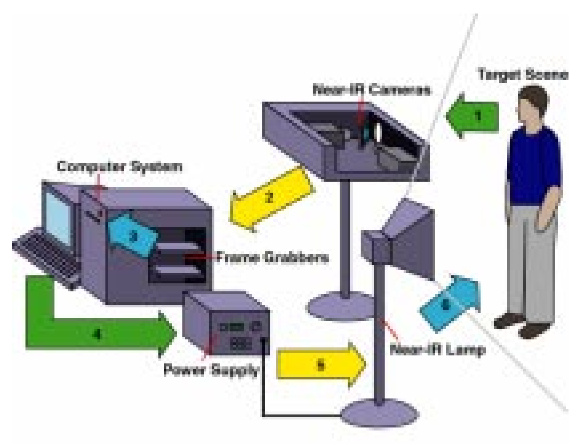



**Figure 2:** Hardware diagram.

A component of the software running on the computer with the frame grabbers analyzes the luminance in the incoming frames. The system then appropriately adjusts the output voltage on the programmable power supply unit connected to the computer via the serial port. The power supply provides power for the near-IR lamp that illuminates the scene. Through this feedback the system is able to keep the scene at a constant luminance regardless of external conditions.

One of the main benefits of using the near-IR spectrum is that subjects in the scene are unaware that they are being illuminated by the system. This is especially beneficial for covert operation in surveillance applications. One consideration, however, that must be made for the near-IR lamp is that like any intense light source it can be harmful to human eyes if direct exposure occurs for a prolonged period [13]. One possible method for damage avoidance is to strobe the lamp when a subject gazes at the system unknowingly for too long.

### 3.2 Software Architecture

The system's software consists of four units (see Figure 3):

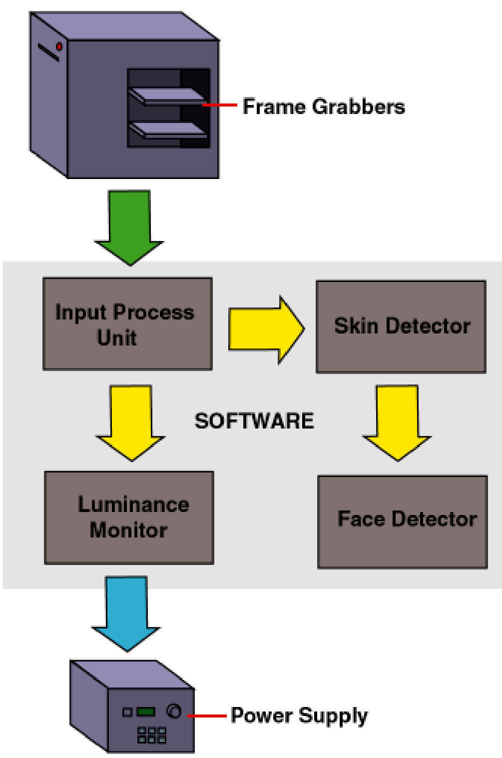



**Figure 3:** Software diagram.

1. **Input Process Unit:** The initial step for the system is to get the input images for both bands from the frame grabbers. The images are then aligned and sent to the luminance monitor and the skin detector module.
2. **Luminance Monitor:** The luminance monitor evaluates the current luminance level in the scene and dynamically adjusts the power output on the power supply. A simple mapping between the output voltage and the corresponding luminance (see Figure 4) allows the system to accurately achieve the desired light level.
3. **Skin Detector:** Upon receiving the dual input images the skin detector performs a series of operations to isolate the skin in the image. The output of the face detection module is a binary image where all skin appears black against a white background. This image is then passed to the final unit of the current software system, the face detector.
4. **Face Detector:** The face detector uses a series of generalized Hough transforms on a feature image extracted from the two near-IR input images and the skin image. After the transforms are complete a good approximation of the location of the eyes can be made. Based on the distance between the eyes we can determine heuristically the 2D orientation and extent of the face.

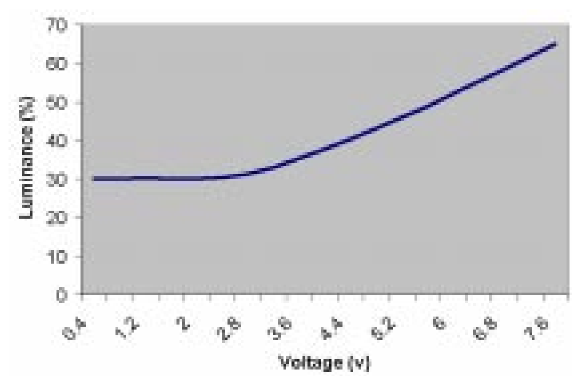



**Figure 4**. Voltage versus luminance diagram for the near-IR lamp.

## 4 Skin Detection in the Near-IR Spectrum

The near-IR spectrum is particularly beneficial for skin detection purposes [11][12]. Human skin exhibits an abrupt change in reflectance around 1.4 $\mu$m. This phenomenology allows for a highly accurate skin mapping by taking a weighted difference of the lower band near-IR image and the upper band near-IR image. A consequence of the phenomenological basis of our skin detection method is that artificial human heads cannot fool the system (see Figure 5).

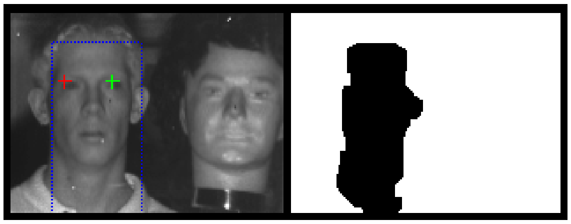


**Figure 5**. Example of successful discrimination between a real and artificial human head. The binary image to the right is the output of the skin detection unit.

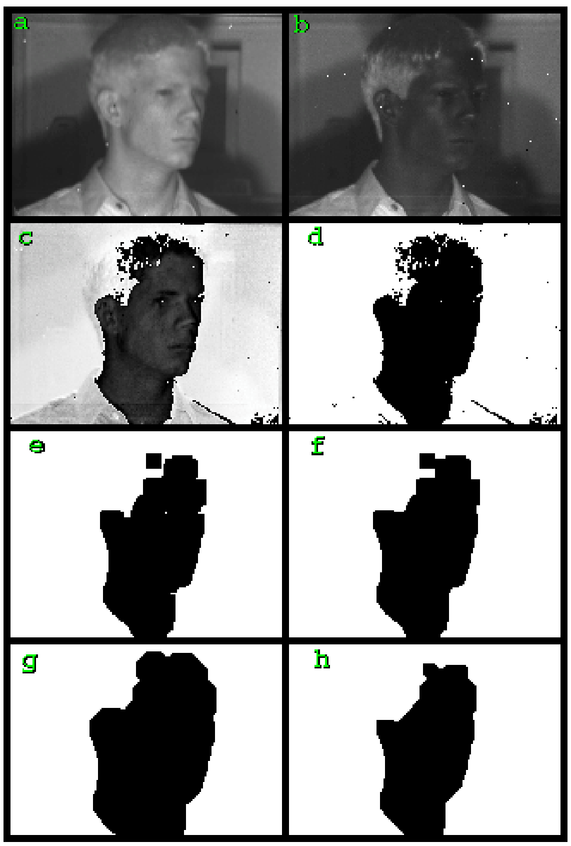



**Figure 6:** The skin detection process: (a) The lower near-IR band image (b) The upper near-IR band image (c) The weighted subtraction image (d) The thresholded image. (e) The opened image. (f) The closed image. (g) The dilated image. (h) The eroded image.

The pixel mapping for the fusion of the two near-IR images is as follows:

$$\mathbf{P(i,j)}_{\text{fused}} = \mathbf{P(i,j)}_{\text{lower}} - \boldsymbol{f}^{*}\mathbf{P(i,j)}_{\text{upper}} \qquad (4.1)$$

where, $\mathbf{P(i,j)}_x$ is the pixel value at position **(i,j)** in the respective image and $\boldsymbol{f}$ is the weight factor used. The constant $\boldsymbol{f}$ = 1.38 (determined through experimentation) is used. The weighted subtraction operation increases substantially the contrast between human skin and the background in the image. This prepares the ground for the successful application of a thresholding operation. Then, the resulting binary image undergoes a series of morphological operations (see Figure 6).

(a) **Opening and Closing:** The opening operation smoothes the contour of the skin region, breaks narrow isthmuses, and eliminates small islands and sharp peaks or capes. The closing operation fuses narrow breaks and long, thin gulfs; eliminates small holes; and fills gaps on the contours.

(b) **Dilation and Erosion:** The application of dilation and erosion results in the elimination of small image detail.

## 5 Face Detection in the Near-IR Spectrum

The main objective of the face detector is to determine the location and extent of the face. It achieves this objective by finding the location of the eyes. The detector needs to determine precisely the location of at least one eye in order to provide information of some use to a face recognizer. A major strength of our method is the exploitation of the phenomenology exhibited by the skin, eyes, and hair (explained below) in the near-IR band of the EM spectrum. The face detector uses a three-step approach to determine the location of the eyes. As input the detector uses the high and low band near-IR images (see Figure 6 (a) and (b)), and the output image from the skin detector module (see Figure 6 (h)).

**Step 1:** The face detector extracts regions in the upper and lower near-IR images that are likely to be the eyebrows (see Figure 7(a)) and eyes (see Figure 7(b)) respectively. This is accomplished by capitalizing upon the unique reflectance characteristics of human hair and skin in the two near-IR bands.

In the upper near-IR band eyebrow hair stands out comparatively to the extremely low reflectivity human skin. The threshold values found to be most suitable for the eyebrow extraction are as follows:

$$B(P_u(x,y)) = \begin{cases} 0 & ,65 < P_u(x,y) < 80 \\ 255, & 65 >= P_u(x,y) \text{ or } P_u(x,y) >= 80 \end{cases} \quad (5.1)$$

Where $B(\bullet)$ is the eyebrow threshold function, and $P_u(x, y)$ is the pixel value of the upper near-IR image at position $(x, y)$.

In the lower near-IR band the eyes stand out comparatively to the extremely high reflectivity human skin. The threshold values found to be most suitable for the eye extraction are as follows:

$$E(P_l(x,y)) = \begin{cases} 150, & 30 < P_l(x,y) < 90 \\ 255, & 30 >= P_l(x,y) \text{ or } P_l(x,y) >= 90 \end{cases} \quad (5.2)$$

Where $E(\bullet)$ is the eye threshold function, and $P_l(x, y)$ is the pixel value of the lower near-IR image at position $(x, y)$.

As with the skin detector, a dynamic adjustment algorithm will replace the constants used for the eye and eyebrow detection in the future. The eyebrow and eye feature images are then fused into a composite feature image (see Figure 5(c)). This is a tri-level image: the black areas denoting likely eyebrow regions, the gray areas likely eye regions, and the white areas all the rest.

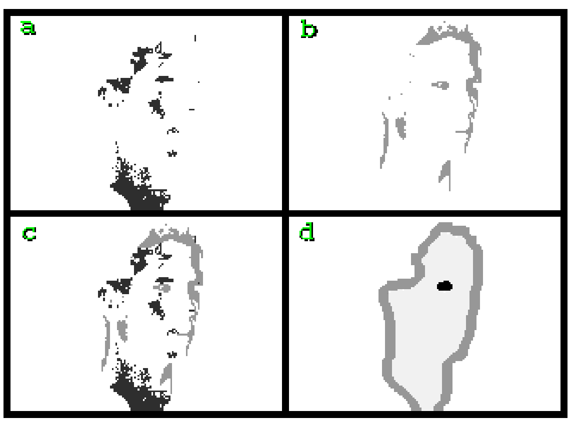



**Figure 7:** (a) Eyebrow feature image extracted from the upper near-IR band. (b) Eye feature image extracted from the lower near-IR band**.** (c) Composite eyebrow-eye feature image. (d) The result of the Hough Transform superimposed on the skin image.

**Step 2:** In the second step the face detector utilizes the Hough transform to find the eye regions on the composite eyebrow-eye feature image. We use a generalized Hough transform template [14] (see Figure 8) that is modeled after the expected appearance of an eye region in the composite feature image. This consists of a black region (modeling the eyebrow) over a gray region (modeling the eye). The template is rotated and sized at each point of implementation to account for the rotation and variation of individual faces. The result of this transformation is a tri-level image where the background shows as white, the skin region as gray and within the skin region the area(s) that exhibited the strongest response to our eye template as black (see Figure 7(d)).

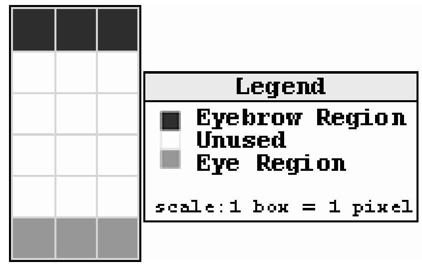



**Figure 8:** The Hough Transform template. The template models the appearance of an eye region in the composite feature image, given the constraints of human anatomy.

**Step 3:** In the final step the face detector estimates the center of the subject's eyes through blob analysis. Because of the variation in human faces many different patterns of 'eye' blobs can arise in the resulting Hough Transform image (see Figure 9). Specifically:

*Cas***e** *1***:** There is a single blob that spans the width of the face region. The blob is bisected in the middle and processed as two smaller blobs.

*Cas***e 2:** There are two blobs that are roughly equal size, which are higher than any other blobs.

*Cas***e 3:** There is a single small blob set apart and higher than any other blobs.

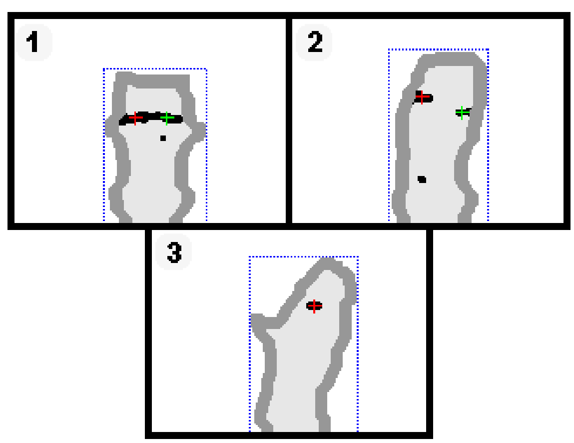



**Figure 9:** Eye region blob cases (1) A single blob covers both eyes (2) Each eye region appears as a distinct blob (3) Single eye region in a side view.

Ultimately, the face detector locates the center of the eyes as the centroids of the selected blobs. Once the eyes are located then the 2D orientation (the face in the sagittal plane) of the head can also be determined heuristically based on the observed distance between the eyes. After the orientation and location of the head are known then it is possible to ascertain if a good image of the face can be extracted for recognition purposes. It is also a viable option to extrapolate, given a partially obstructed view of the face, where necessary to create a frontal image of the face from any given rotation, provided that at least half of the face is visible.

## 6 Experimental Results

We tested the performance of the system on a stream of single facial images taken live through our dual-band system (see Figure 2). The images were taken inside our laboratory using the near-IR illuminator as the sole lighting source. Our experimental data set was composed of 474 images taken from 18 different subjects. we used a wide variety of people including both genders, subjects with glasses, and subjects with facial hair. Each subject performed a series of head movements to test the system's detecting power in different facial orientations and positions (see Figure 10).

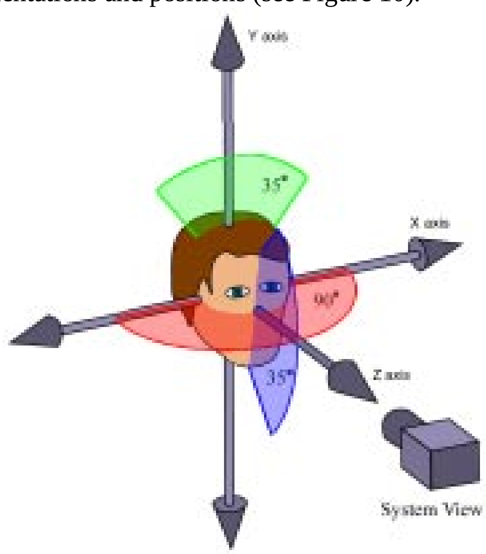



**Figure 10:** Subject head motion range.

For the purpose of quantifying results we defined a correct face detection by the system as the system correctly locating at least one eye (provided that at least one was visible) in a skin region. The system performed very well for subjects facing forward as well as most degrees of rotation (see Figure 11 and Figure 12). In Table 1, where we report the experimental performance of the system, we cluster under the "Frontal Face" category all the images featuring head orientations within the ($-30^0 \div +30^0$) range in the y direction. Under "Rotated Face" we cluster all the other images featuring heads at extreme orientations. The system also achieved lower frontal face detection performance in subjects with glasses than subjects without glasses (see Figure 13). Glasses interfered with the skin phenomenology in the near-IR and hindered an accurate feature mapping as described in Section 5. One very interesting aspect of the system's performance was that in subjects with glasses the detection rate was higher at extreme head orientations. At extreme orientations (e.g. $90^0$ with respect to y axis) the camera could view more of the naked skin around the eye while the specular reflection from the glasses was less of an issue (see the top right image in Figure 12).

The system seemed to encounter difficulties when examining an image where the subject was looking up or down (more then a $-30^0$ rotation in the x axis, see Figure 14). This obscured the eyes and therefore made it difficult to create a good composite feature image from which to find the eye regions.

Our method is also quite efficient operating at 5 frames per second in a 500 MHz Pentium II PC, a speed sufficient for most security applications.

| | **Subjects Without Glasses** | **Subjects With Glasses** |
|---|---|---|
| **Detection Rate Frontal Face** | 96.35% | 72.25% |
| **Detection Rate Rotated Face** | 66.88% | 67.40% |

**Table 1:** Face detection results from 474 images of 18 subjects.

**Figure 11**: Examples of the system's performance in frontal faces. The images belong to the lower near-IR stream and are superimposed by the skin box and the crosses indicating the locations of the eyes.

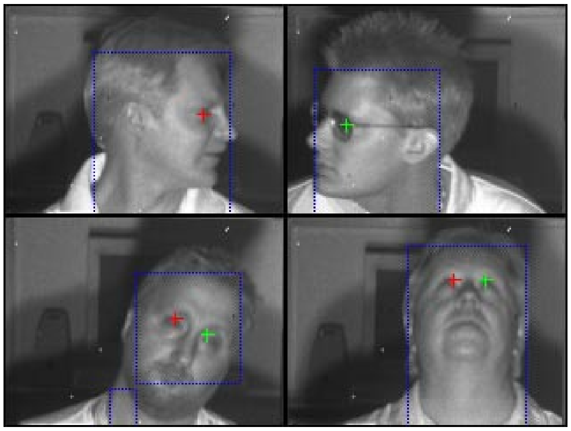


**Figure 12:** Examples of the system's performance in rotated faces. The images belong to the lower near-IR stream and are superimposed by the skin box and the crosses indicating the locations of the eyes.

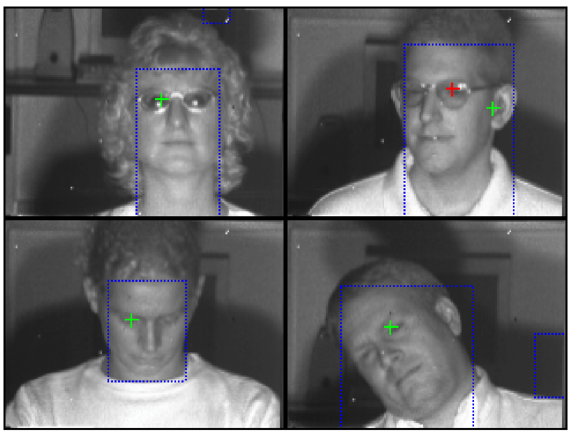


**Figure 13:** Examples of the system detecting the face and finding correctly at least one eye.

**Figure 14:** Examples of the system detecting the overall facial region, but not correctly finding both eyes.

### 6.1 Experimental Comparison

To fully understand the significance of the results from a system it is important to establish a basis for comparison using similar works. The area of face detection has received a lot of attention in recent years providing for a comprehensive body of literature. Table 2 lists the reported performance of several referenced systems as well as our own work.

We realize that because of the varying methods and data sets the results reported in Table 2 should be interpreted with caution. In general, our data set comprises fewer subject faces but features the greatest percentage of real-time imagery and extreme facial rotation. Also, the data sets used by others are almost void of glass wearing subjects. To achieve a somewhat meaningful comparison we report in Table 2 the performance of our system in the subset featuring frontal faces ($-30^0 \div +30^0$ rotation in the y direction) and without glasses.

| | **Method** | **Detection Rate** |
|---|---|---|
| **Jeon [7]** | Clustering Algorithm | 91% *(z-axis 360$^0$)* |
| **w y** | Neural Network | ~96.0% *(frontal)* |
| **Z** | Wavelet Analysis | ~95% |
| **O** *r* **y***st* **m** | Feature Extraction in Near-IR | 96.35% *(frontal)* |

*a l* Reported performance results from various face detection systems.

Our system seemed to achieve superior detection performance from most other systems reported in the literature. Only the performance of Rowley's system [8] came close to ours for frontal facing subjects. Rowley **et** *al***.** used multiple data sets to test their system. The most comparable of those with our data set was the FERET database. The FERET data set contained a broad range of faces under good lighting with uncluttered backgrounds. One major disadvantage of Rowley's approach relative to ours is that it requires extensive training to perform satisfactorily.

Our approach, being the only one based on near-IR phenomenology, has the major advantage of performing robustly under different natural lighting conditions including total darkness. This resilient performance is attained through the liberal use of artificial near-IR illumination and the distinct reflectance characteristics of the human skin.

## 7 Conclusion and Future Work

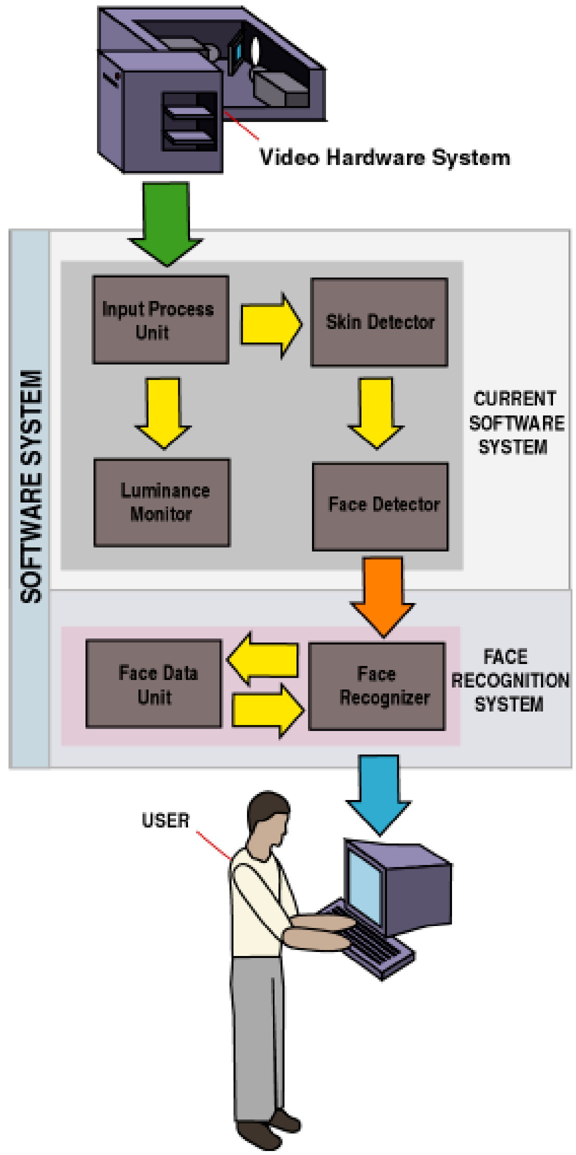
We have expanded the skin detection work reported earlier by our group [11][12] by developing a face detection method based on multi-band feature extraction in the near-IR spectrum. Within the detected skin area, an eye feature image is extracted from the lower near-IR band and an eyebrow feature image is extracted from the upper near-IR band. The two feature images are superimposed to create a composite feature image. Then a generalized Hough transform is rotated and sized at each point in the image. The generalized template is modeled after the expected appearance of an eye region in the composite feature image. The strongest response to the eye template produces an eye blob image and finally the center locations of the subject's eyes. The determination of the eye locations within the skin region allows the system to calculate heuristically the 2D orientation of the face and extent of the face based on the observed distance between the eyes.

In the experimental testing so far, our face detector exhibited state-of-the-art performance in real-time imagery. Our test set included a substantial percentage of faces wearing glasses and faces at extreme rotations.

Our ongoing work focuses on the exploitation of the face detection information for face recognition purposes. We are working towards incorporating the face recognition engine FaceIt® [15] by Visionics into our overall system. Since FaceIt® relies primarily on facial geometry for face recognition, it can be invariably applied to visible as well as near-IR imagery. By replacing the nominal face detector in the FaceIt® system with our face detector we will be able to readily extend the application scenarios to outdoor surveillance.

## Acknowledgements

We would like to thank Mr. Jeff Radke and Murray Cooper from the management team of Honeywell Laboratories for their financial support. We would also like to thank Mr. Pete Reutiman, Alan Greisbach, and Justin Droessler for their valuable technical contributions. Part of this project was supported by an NSF grant (NSF/CRCD 0088086) through the University of Nevada at Reno. The views expressed in this article reflect the opinions of the authors only and should not be linked in any way to the funding agencies.




*i r* **5** Diagram of the extended face detection/recognition system under development.

## References

[1] S. Kawato and J. Ohya, "Two-step Approach for Real-time Eye Tracking with a New Filtering Technique," in *Proceedings 2000 IEEE International Conference on Systems, Man, and Cybernetics*, 2000, Vol. 2, pp. 1366 –1371.

[2] S.H. Kim, and H.G. Kim, "Face Detection Using Multi-modal Information," in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000, pp. 14–19.

[3] C. Morimoto, and M. Flickner, "Real-Time Multiple Face Detection Using Active Illumination," in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000, pp. 8-13.

[4] Y. Li, S. Gong, S. Liddel, and H. Liddel, "Multi-view Face Detection Using Support Vector Machines and Eigenspace Modeling," in *Proceedings Fourth International Conference on Knowledge-Based Intelligent Engineering Systems & Allied Technologies*, 2000, Vol. 1, pp. 241-244.

[5] X. Lv, J. Zhou, and C. Zhang , "A Novel Algorithm for Rotated Human Face Detection," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, 2000, Vol. 1, pp. 760–765.

[6] W. Huang, and R. Mariani, "Face Detection and Precise Eyes Location," in *Proceedings 15th International Conference on Pattern Recognition*, 2000, Vol. 4, pp. 722–727.

[7] B.H. Jeon, S.U. Lee, and K.M. Lee, "Rotation Invariant Face Detection Using a Model-Based Clustering Algorithm," in *Proceedings 2000 IEEE International Conference on Multimedia and Expo, 2000*, Vol. 2, pp. 1149-1152.

[8] H.A. Rowley, S. Baluja, and T. Kanade, "Neural Network-Based Face Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 1, pp. 23–38, January 1998.

[9] Y. Zhu, S. Schwartz, and M. Orchard, "Fast Face Detection Using Subspace Discriminant Wavelet Features," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, 2000, Vol. 1, pp. 636–641.

[10] . Wilder, P. Phillips, C.Jiang, and S. Wiener, "Comparison of Visible and Infra-Red Imagery for Face Recognition," *Proceedings Second IEEE International Conference on Automatic Face and Gesture Recognition*, 1996, pp. 182–187.

[11] I. Pavlidis, and P. Symosek, "The Imaging Issue in an Automatic Face/Disguise Detection System," in *Proceedings IEEE Workshop on Computer Vision beyond the Visible Spectrum: Methods and Applications*, 2000, pp. 15–24.

[12] Pavlidis, V. Morellas, and N. Papanikolopoulos, "A Vehicle Occupant Counting System Based on Near-Infrared Phenomenology and Fuzzy Neural Classification," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 1, No. 2, pp. 72-85, June 2000.

[13] D. Sinley, "Laser and Led Eye Hazards: Safety Standards," *Optics and Photonics News*, pp. 32-37, September 1997.

[14] D.H. Ballard, "Generalizing the Hough Transform to Detect Arbitrary Shapes," *Pattern Recognition,* Vol. 13, No. 2, pp. 111-122, 1981.

[15] http://www.visionics.com/faceit/whatis.html