Final Report 2000-06

Research

Automatic Passenger Counting in the HOV Lane

Technical Report Documentation Page

1. Report No.	2.	3. Recipient's Accession No.
MN/RC - 2000-06		

4. Title and Subtitle	5. Report Date
AUTOMATIC PASSENGER COUNTING IN THE HOV LANE	June 1999
	6.

7. Author(s)	8. Performing Organization Report No.
Ioannis Pavlidis, Bernard Fritz, Peter Symosek, Nikolaos P. Papanikolopoulos, Vassilios Morellas, Robert Sfarzo	

9. Performing Organization Name and Address	10. Project/Task/Work Unit No.
University of Minnesota Department of Computer Science 200 Union Street, S.E. Minneapolis, MN 55455	11. Contract (C) or Grant (G) No. c) 74708 wo) 74

12. Sponsoring Organization Name and Address	13. Type of Report and Period Covered
Minnesota Department of Transportation 395 John Ireland Boulevard St. Paul Minnesota, 55155	Final Report 1999
	14. Sponsoring Agency Code

15. Supplementary Notes

The first four authors listed are employed by Honeywell Technology Center, Minneapolis, MN

16. Abstract (Limit: 200 words)

This research applied wave band and computer vision methods to automatically count vehicle occupants in the High Occupancy Vehicle (HOV) lane at a high level of accuracy.

The research showed that use of near-infrared bandwidth offers potential as a method for developing an automatic vehicle occupant counting system. Near-infrared only can produce images when looking through glass, but not metal or heavy clothes, which limits its accuracy in counting children or occupants resting in vehicles. The mid-infrared camera did not produce clear images at highway speeds.

The next step involves additional research into a working device that can count vehicle occupants reliably, including analysis of device performance with more types of vehicles, passengers in the back seats, children in car seats, and passengers lying down.

17. Document Analysis/Descriptors		18. Availability Statement
HOV Automatic Passenger Counting	Near-infrared bandwidth	No restrictions. Document available from: National Technical Information Services, Springfield, Virginia 22161

19. Security Class (this report)	20. Security Class (this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	60	

AUTOMATIC PASSENGER COUNTING IN THE HOV LANE

Final Report

Prepared by:

Ioannis Pavlidis
Peter Symosek
Vassilios Morellas
Bernard Fritz
Nikolaos P. Papanikolopoulos
Robert Sfarzo

Department of Computer Science
University of Minnesota
Minneapolis, MN 55455

June 1999

Published by:

Minnesota Department of Transportation
Office of Research Services
Mail Stop 330
395 John Ireland Boulevard
St. Paul, MN 55155

Document Revision Control Log

Project Name	HOVL
Contract Number	F10041 (HTC)
Document Title	Automatic Passenger Counting in the HOV Lane
Revision	1.0
Revision Description	First Edition
Date	06/03/99

Prepared for

Minnesota Department of Transportation (Mn/DOT)

Office of Research Services

First Floor

395 John Ireland Boulevard, MS 330

St. Paul, MN 55155

Prepared by

Honeywell Technology Center (HTC)

3660 Technology Drive

Minneapolis, MN 55418

&

University of Minnesota (U of MN)

Minneapolis MN 55455

ACKNOWLEDGEMENTS

We would like to extend our deep appreciation to Kevin Schwartz, the Minnesota Department of Transportation (Mn/DOT) HOV program manager for his generous help and support. Many thanks to Ben Worel and Jack Herndon for accommodating our needs in the Minnesota Road Research Project (Mn/ROAD) facility. We would also like to thank Joe Keller for his help during the road tests. Finally, we would like to thank Scott Nelson for a valuable discussion regarding some technical issues in this project. This work was supported by the Minnesota Department of Transportation under contract $\#F10041$. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the funding agency.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

Background

There are compelling reasons for the development of an automatic vehicle occupant counting system. Currently, vehicle occupant counting is collected by human counters for documenting vehicle occupancy on the Twin Cities metro area freeways. Human counters are expensive and can disrupt traffic if they are visible to motorists, which in turn limits the amount of data collection that can be gathered. Also, human counters are not very accurate during poor weather or at nighttime and are usually available only during the longest daylight months.

An automatic vehicle occupant counting system could replace human counters and facilitate the gathering of statistical data for traffic operations management, transportation planning, and construction programming. Also, if laws were changed to allow for such activity, it could give law enforcement a technical means to perform the High Occupancy Vehicle (HOV) lane monitoring task more effectively, as well as facilitate enforcement of a system to allow single-occupant vehicles to use the HOV lane for a fee.

The Honeywell Technology Center (HTC) in cooperation with the University of Minnesota (U of MN) carried out a feasibility study regarding the automatic counting of vehicle occupants in the HOV lane. Technology does not currently exist to automatically count vehicle occupants at a high level of accuracy. The purpose of the study was to determine if there was an appropriate wave band and computer vision method for reliable automatic counting of vehicle occupants. The study was funded

by the Minnesota Department of Transportation (Mn/DOT) and the performance period was from April 1998 to April 1999.

Key Tasks

The 3 key tasks for this project were: 1. Data Collection and Fusion 2. Algorithm Development 3. System Demonstration.

In order to develop a reliable vehicle occupant counting system, the most important step is to produce a clear imaging signal through the sensors with as distinct a signature for the vehicle occupant as possible. The Data Collection and Fusion task concentrated on the selection of the sensors and accessories, the sensor arrangement, and a fusion scheme to facilitate the vehicle occupant detection operation with the least amount of noise in the imaging signal. This task was the most critical task of the project, because if the imaging signal were noisy, then even the most powerful pattern recognition algorithms would not function accurately. Much of the project was spent developing this signal.

The study first focused on gathering images using a mid-infrared camera. The mid-infrared camera was chosen because it does not need infrared illumination, which makes it a simpler operation than using a bandwidth that would require infrared illumination.

After it was determined that the mid-infrared camera would not work well at high speeds, a near-infrared camera was tested using two filters set to different near-infrared bandwidths. These two near-infrared images were fused together to form signatures or silhouettes of passengers faces to distinguish people from inanimate objects. This test was conducted in both daylight and nighttime conditions. Near-infrared illumination, which is both invisible and safe to the human eye, was used to enhance the scene in the nighttime condition. The test only involved looking for passengers in the front seat who were sitting upright.

Once a clear silhouette of the front passengers faces was obtained from the near-infrared approach, the algorithm portion of this study was conducted to determine an automatic method for counting the number of silhouettes, or people, in the front seat of the vehicle. Two algorithm approaches were attempted: 1) a primary fuzzy neural network approach, and 2) an alternative approach.

Key Findings

The mid-infrared camera was not able to produce clear images at highway speeds. Until the technology improves and the cost is reduced, sensors in the mid-infrared range will not be able to detect passengers clearly at highway speeds. Also, the mid-infrared camera could only produce images through the side window. It could not produce images through the front windshield, which is made of a material that blocks some radiation.

The near-infrared approach produced good images and these cameras were not affected by tinted windshield or side glass. With specially modified cameras, the near-infrared cameras can work at highway speeds. With proper fusion, a clear signature or silhouette of the front passengers' faces was distinguishable from other inanimate objects in the vehicle including a dummy head. These silhouettes were obtained in both daylight and nighttime conditions. Due to the scope of this study, persons in the back seat or persons lying down were not tested.

The primary fuzzy neural network algorithm approach scored perfectly in classifying the 90-sample image set at car speeds of 0 - 40 mph. The alternative approach wasn't perfect in its classification, but it did well with the limited sample size.

Conclusions and Future Work

This research study shows that there is potential for developing an automatic vehicle occupant counting system using the near-infrared bandwidth. However, near-infrared

cameras can only produce images when looking through glass, not metal or heavy clothes. So there may be a limit to the level of accuracy that can be obtained with this technology if an automatic vehicle occupant counting system would be required to count children in car seats or persons who are lying down in an automobile.

Further research and testing is needed in order to obtain a working device that can count vehicle occupants reliably. The sensor placement needs to be optimized, data collection techniques need to be automated, and the algorithm needs further development. Additional experimentation should analyze device performance with more types of vehicles, passengers in the back seats, children in car seats, and passengers lying down. Also, more rigorous testing in a variety of weather and lighting conditions is needed.

Chapter 1

DATA COLLECTION AND FUSION

1.1 Relevant Work

Our effort during the first half of the project's period was primarily concentrated in the task of *Data Collection and Fusion*. The objectives in this task were:

1. Select the appropriate sensors and accessories.

2. Design an appropriate sensor arrangement.

3. Devise a fusion scheme that will facilitate the passenger detection operation.

This was the most important task of the project since its successful conclusion was critical to the success of the entire project. The reason for this was simple: If we managed to acquire a clear imaging signal through the sensors, then even moderately powerful pattern recognition algorithms would succeed in the passenger detection task. If, however, the imaging signal were noisy, then even the most powerful pattern recognition algorithms would break.

Upon embarking on our effort in late April 1998 we were aware of one other similar study worldwide led by the Georgia Tech Research Institute (GTRI). The effort involved the use of a near-infrared camera ($0.55-0.90 \quad \mu m$) and a near-infrared illumination source in the same range. One reason for using near-infrared sensing was the ability to use non- distracting illumination source during the night. Illumination during night time certainly enhances the quality of the image.

In more general terms, Intelligent Transportation Systems (ITS) projects that involve imaging usually adopt the use of visible spectrum cameras. The strong point of the visible spectrum approach is that the relevant imaging sensors are the most advanced and at the same time the cheapest across the Electro-Magnetic (EM) spectrum. Visible spectrum cameras have a particular advantage in terms of speed which is an important consideration in the HOV Lane where vehicles are moving at rates of speed of 65 *mph*. They can also have very high resolution which accounts for very clear images under certain conditions. Unfortunately, there are also a lot of serious problems with the visible spectrum approach. To mention a few: Some vehicles have heavily tinted window glass to reduce glare from solar illumination; this glass is almost opaque to visible spectrum cameras. Also, visible spectrum cameras don't have operational capability during night time.

1.2 Methodology

One factor that is distinctly absent in the aforementioned research efforts, and unfortunately, in many other computer vision projects is a vigorous sensor phenomenology study. Most researchers without a second thought adopt the visible spectrum as the spectrum of choice, or, in rare cases, some other EM spectrum based primarily on intuition. The result is that they usually end up with a non-discriminating signal that makes the problem appear harder than it actually is. Then, they try to address the difficulty by devising powerful computer vision algorithms but often to no avail. The loss of information because of poor sensor choice and arrangement are usually irrevocable.

Our first consideration, was to consider what nature had to offer across the EM spectrum (see Fig. 1.1). The lower portion of the EM spectrum consists of the gamma rays, the x-rays, and the ultra-violet range. They are all considered harmful and they are used in a controlled manner in medical applications only. Then, the

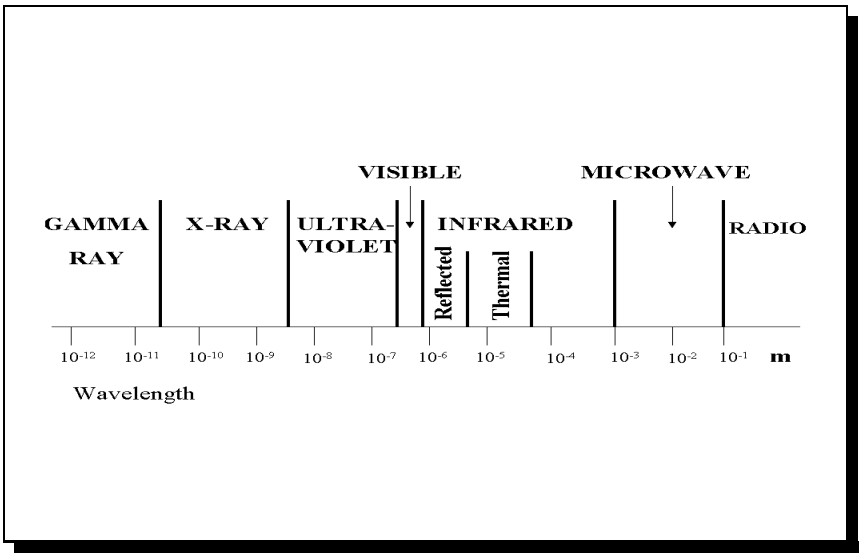


Figure 1.1: Electro-Magnetic (EM) spectrum.

visible spectrum, is the range we are mostly acquainted with since it is used by the human eye and the vast majority of cameras. Visible spectrum cameras use mature technology and they feature the best quality to price ratio. They achieve very high resolution and speed, and with the recent introduction of digital technology negligible systemic noise levels. Unfortunately, their systemic noise levels increase during poor environmental conditions like bad weather, night time, and direct sunlight. Some of these weaknesses are incurable. Some others, like night time, can be overcome by using artificial lighting. Nevertheless, this is not an option in the case of transportation applications. The artificial light should match the spectrum of the camera (visible range) and consequently it will distract the drivers with perhaps fatal results. One other characteristic, that is very important in computer vision applications, is also absent in this range. The image understanding task becomes feasible or easier when the object of interest, the human face in our case, appears to have consistent qualities under a variety of conditions. In visible spectrum cameras, the passenger faces appear darker or lighter depending on the physical characteristics of the passenger,

the incident angle of illumination, and the illumination intensity.

At the far end of the EM spectrum there are the microwave and radio regions. This area was just started to be exploited for imaging purposes. Sensors operate in active or in passive mode. The major advantage is that the long wavelengths in these regions can penetrate clouds, fog, and rain producing weather independent imaging results. The technology is very new, and thus prohibitively expensive. Also the sensors are bulky, and feature very low resolution. Their application is currently constrained in the military and the remote sensing domain [13].

Between the low and the far end of the EM spectrum there is a middle region which is known as the infrared range ($0.7-100 \quad \mu m$). Within the infrared range two bands of particular interest are the reflected infrared ($0.7-3.0 \quad \mu m$) and the thermal infrared ($3.0-5.0 \quad \mu m, \quad 8.0-14.0 \quad \mu m$) bands. The reflected infrared band on one hand is associated with reflected solar radiation that contains no information about the thermal properties of materials. This radiation is for the most part invisible to the human eye. The thermal infrared band on the other hand is associated with the thermal properties of materials.

1.3 Mid-Infrared Approach

Upon embarking in our study we concentrated on the thermal infrared region for the following reasons.

1. The human body maintains a relatively constant temperature (about 36.6^o Celsius) irrespectively of physical characteristics or illumination conditions. This would translate into a consistent light color pattern for the faces of the vehicle passengers in infrared imaging. This consistency is lacking in the visible spectrum and would greatly facilitate the image understanding task. Incidentally, the thermal property can serve as a differentiator between humans and dummies.

2. A thermal infrared sensor is operational day and night without any need for an external illumination source.

Our only concern, was the attenuation introduced by the presence of the vehicle glass. Glass disrupts severely the transmission of infrared illumination beyond 2.8 μm. This is the range where thermal energy is just beginning to appear. If we were to capture anything at all we needed an extremely sensitive mid-infrared camera in the range $2.5 - 3.5$ μm. The Mitsubishi Thermal Imager $IR - 700$ proved to be the appropriate sensor for the task. An additional consideration was the composition of the glass in vehicle windows. Vehicle windows are not made from common glass for a variety of reasons (safety, energy efficiency, visibility). Notably, the composition of the front windshield differs substantially from the composition of the side windows. Spectral measurements for a typical vehicle windshield (see Fig. 1.2) compared with spectral measurements for typical side window glass (see Figs. 1.3 and 1.4), revealed that it would be beneficial to place the infrared camera by the side of the road. Initial experiments with the Mitsubishi camera confirmed these theoretical predictions. We could not see anything inside the vehicles when we were shooting in frontal view. In contrast we were getting very clear images when we were shooting from the side.

These initial experiments were performed in a parking lot with the test vehicle either stationary or moving at low speed (up to 20 mph). We performed one such experiment in winter, even before the official start of the program and one in spring. The side view images were very clear (see Fig. 1.5) except in one case. That was in winter time when the vehicle's defogger was on for more than half an hour. Then, the thermal signature of the air in the interior of the vehicle was becoming stronger than the thermal signature of the passengers. The result was a cluttered imaging signal.

In a third experiment, in spring, we took our testing to an actual freeway site ($I394$). This time we were shooting from the side of the freeway at vehicles moving at speeds of 65 mph. As always, we were using the infrared camera side by side with

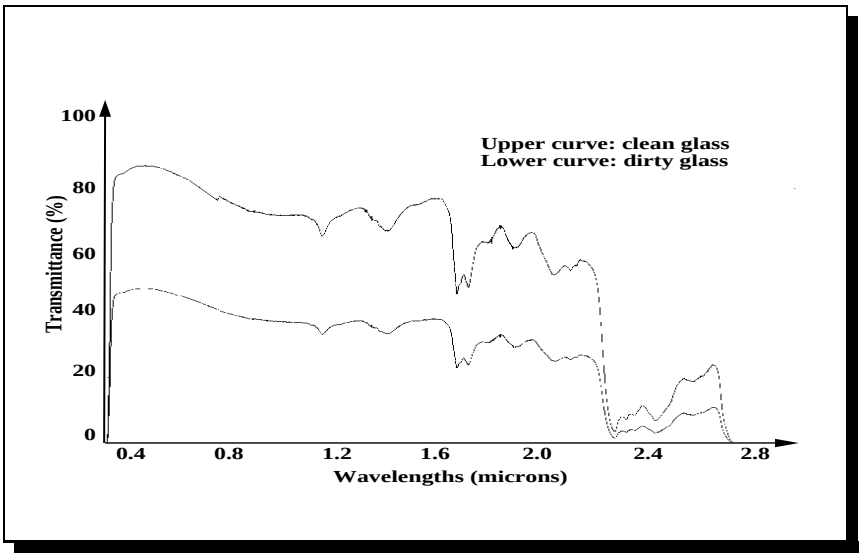


Figure 1.2: Transmittance of a typical windshield from $0.3 - 3.0 \quad \mu m$. Upper curve corresponds to a clean windshield. Lower curve corresponds to the same windshield when it gets dirty.

our digital visible spectrum Sony $DSR \quad 200$ camera. The results showed that the mid-infrared camera was not capable of capturing clear images of such fast moving targets (see Fig. 1.6). The Mitsubishi $IR - 700$ operateds at a frequency of $30 \quad Hz$. In comparison, the visible spectrum camera, was operating at $1000 \quad Hz$ to achieve clear images of such fast moving targets. Unfortunately, mid-infrared technology could not afford such a high frequency rate at the present time. The situation would be better if the infrared camera were able to penetrate the windshield, because in frontal view the car remains for more time in the camera's field of view. Consequently, this would impose less severe speed capture demands. But, placing the mid-infrared camera in frontal view is not an option since it cannot penetrate the front windshield. That left us in a deadlock and we turned our attention to the reflected infrared band and particularly the range $1.0 - 2.0 \quad \mu m$.

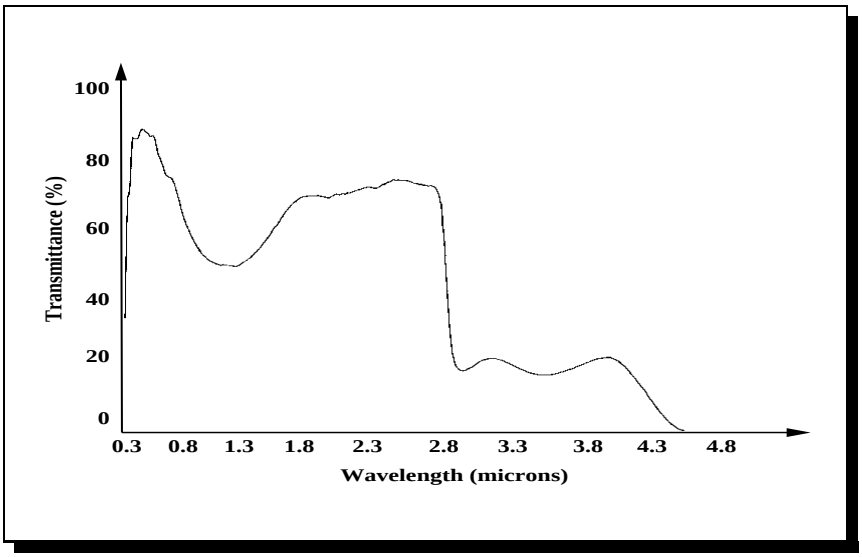


Figure 1.3: Transmittance of a lightly tinted side window from $0.3 - 3.0 \quad \mu m$. It is evident the graceful drop in transmittance even after the critical threshold of $2.8 \quad \mu m$. This spectral behavior allows for penetration of some thermal radiation. Compare this with the windshield transmittance diagram in Fig. 1.2.

1.4 Near-Infrared Approach

The first concern was to find a state of the art camera sensitive in the reflective near-infrared band that can be consigned to us for experimentation, much the same way as the Mitsubishi $IR - 700$ did. Sensors Unlimited Inc. agreed to consign us their $SU - 320$ near-infrared camera that fulfilled our specification. We performed two preliminary rounds of experiments in HTC's parking lot with the $SU - 320$. The first round took place in spring and the second in summer. Based on this experience and subsequent theoretical investigation we determined that the $SU - 320$ camera can live up to the challenges of the HOV requirements if certain steps are taken. Specifically:

1. We found theoretically and experimentally a unique differentiator for the human face in the range $1.4 - x \quad \mu m \quad (where \quad x > 1.4 \quad \mu m)$ that substitutes the

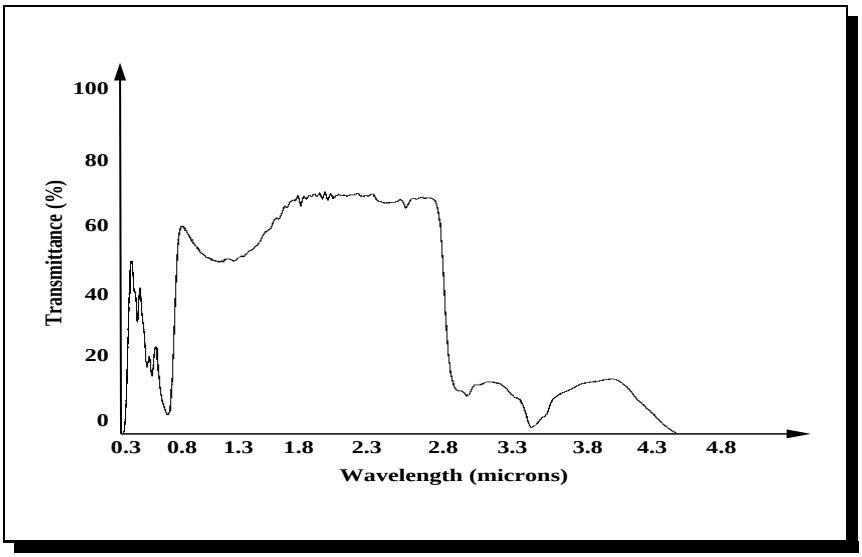


Figure 1.4: Transmittance of a heavily tinted side window from $0.3 - 3.0 \quad \mu m$. Spectral transmittance is worse than the case of an untinted side window (see Fig. 1.3) But, it still drops gracefully after $2.8 \quad \mu m$ comparatively to the transmittance of the windshield (see Fig. 1.2).

corresponding thermal differentiator of the mid-infrared range. Above $1.4 \quad \mu m$ human skin appears consistently dark irrespectively of the person's physical characteristics (Figs. 1.7 - 1.10) [8]. The phenomenon is exemplified in Figs. 1.11 and 1.12. In Fig. 1.11 a Caucasian male is pictured next to a dummy head when the near-infrared camera is equipped with a band pass filter in the range $1.1 - 1.4 \quad \mu m$. They both appear in a relatively lighter color than the background close to the way they would appear in the visible spectrum. In Fig. 1.12 the same Caucasian male and dummy head show different when the camera is equipped with a band pass filter in the range $1.4 - 1.7 \quad \mu m$. In fact, the face of the Caucasian male appears dark (darker than the background). The face of any other human, would exhibit irrespectively of its physical characteristics In contrast, the dummy head appears as a light colored object (lighter than the

Figure 1.5: Side snapshot of a low speed car with a mid-infrared camera.

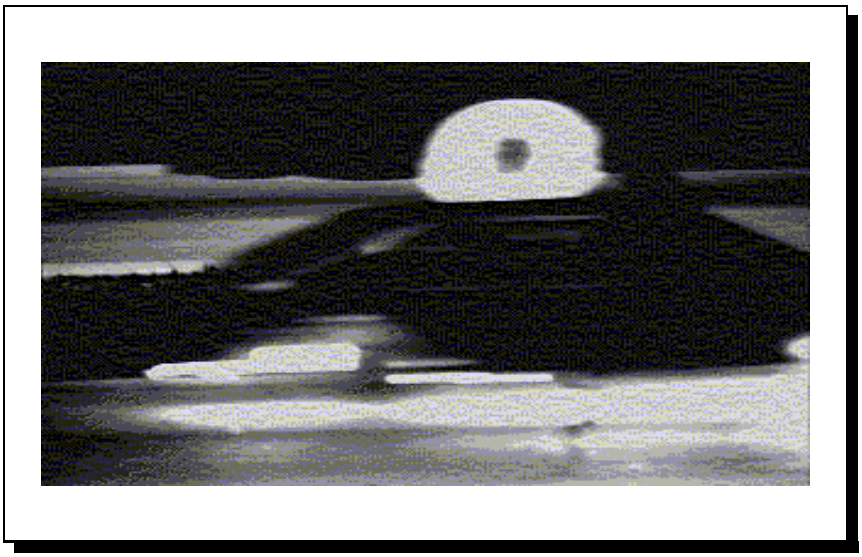


Figure 1.6: Side snapshot of a fast moving car (65 mph) with a mid-infrared camera. The image appears heavily blurred since the speed of the camera cannot keep up with the speed of the target.

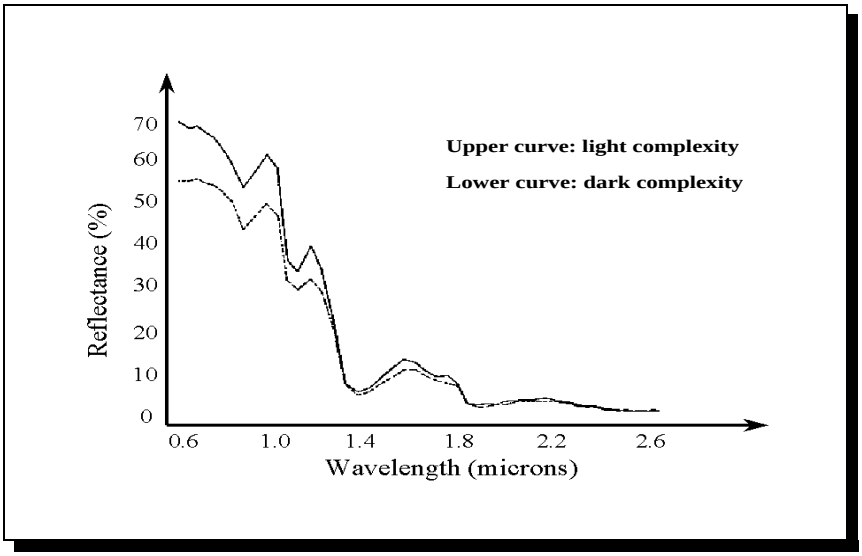


Figure 1.7: Skin reflectance of Caucasian males. Upper curve corresponds to light complexion while lower curve to dark complexion.

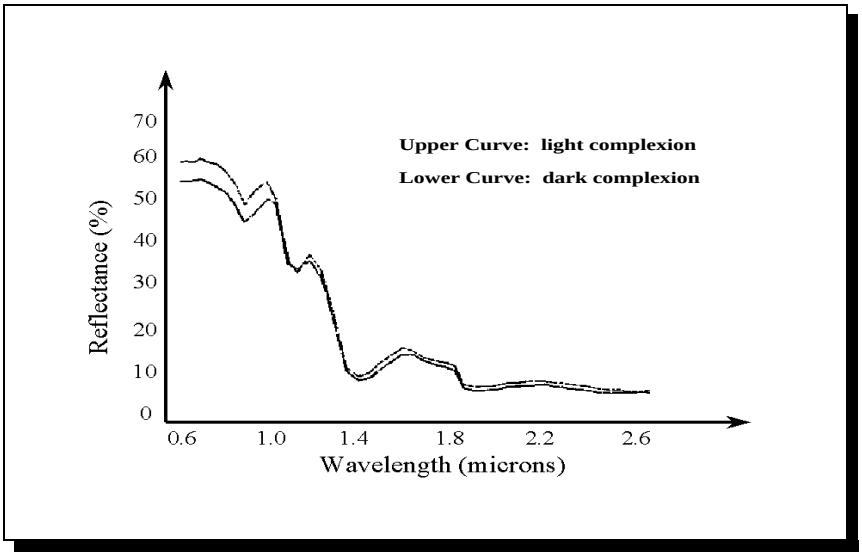


Figure 1.8: Skin reflectance of asian males. Upper curve corresponds to light complexion while lower curve to dark complexion.

background), easily distinguishable from the human head. This sort of response is shared by many other inanimate objects that can be found inside a vehicle like for example upholstery, dashboard, fabrics (see Fig. 1.13). The low reflectivity of human flesh for the $1.4 - 1.7 \quad \mu m$ range can be explained if we notice the spectral response of the water in the same region. Beyond $1.4 \quad \mu m$ the water absorbs substantially infrared radiation and appears in the image as a dark body (see Fig. 1.14). Since the composition of the human body consists of 70% water, naturally, its its spectral response is very similar to that of the water. Hence, the camera should be equipped with a $1.4 - x \quad \mu m \quad (where \quad x > 1.4 \quad \mu m)$ band pass filter to capture this unique passenger differentiator.

2. The solar illumination in the range $1.4 - x \quad \mu m \quad (where \quad x > 1.4 \quad \mu m)$ creates a lot of glare effects that lessen the quality of the imaging signal. A polarizing filter is needed during day time to improve the quality of the image.

3. The operating range $1.4 - x \quad \mu m \quad (where \quad x > 1.4 \quad \mu m)$ is quite apart from the visible band and we can quite safely employ during night time a matching near-infrared illumination source to improve the quality of the image. The light will not only be invisible to the drivers but also completely harmless to their eyes since its wavelength is above the safe threshold of $1.4 \quad \mu m$.

4. Since we operate at a lower band than the mid-infrared band, glass penetration is less of a problem and we can easier see through the windshield. This makes the speed requirements for the camera less stringent. In the actual freeway site, where shooting would be performed from a distance, probably a zoom lens would be required. In general, a complete optical design seemed to be in order that would verify mathematically the feasibility of the approach.

1.4.1 Radiometric Calculation

We undertook a complete optical design of the near-infrared experimental setup that proved theoretically the feasibility of the approach. Fig. 1.15 shows the layout of the proposed near infrared system during day time. We assume that a vehicle is moving down a freeway with velocity v and is observed in frontal view with the near-infrared $SU-320$ camera at a distance d and from a height h. We also assume that the $SU-320$ camera is equipped with the following accessories:

1. A telephoto lens.

2. Alternate band-pass filters either in the range $1.4-x \quad \mu m \quad (where \quad x > 1.4 \quad \mu m)$ or in the range $y-1.7 \quad \mu m \quad (where \quad y < 1.4 \quad \mu m)$ for the reasons explained in Section 1.4.

3. A polarizing filter to reduce the glare effect from the sun illumination during daytime.

During daytime the system uses the illumination of the sun. The objective is to determine if there is any appropriate geometric arrangement for the $SU-320$ camera so that the signal to noise S/N ratio and the camera speed are kept in acceptable levels even under adverse conditions. An acceptable S/N ratio is considered anything above 35. The speed quality is considered acceptable when the image smearing does not exceed the width of one pixel.

The first step in a radiometric computation is to determine the amount of radiation that falls upon the objects of interest [7] - in our case the vehicle passengers. As we stated, we consider two spectral bands, one above the $1.4 \quad \mu m$ threshold point and one below it. Because of constraints due to the quantum efficiency of the $SU-320$ camera we limit the upper band in the range $1.4-1.7 \quad \mu m$. For symmetry reasons we limit the lower band in the range $1.1-1.4 \quad \mu m$. We will demonstrate our computation

for the upper band only, since very similar things apply also to the lower band. The spectral radiance of the sun (our illumination source) on a clear day at sea level is approximately $R_{sunny} = 0.015 \quad watts/cm^2$ in the $1.4 - 1.7 \quad \mu m$ wave-band. In our computation, however, we consider the worst case scenario of an overcast day. For an overcast day the radiance value is reduced by 10^{-3} giving a radiance at the vehicle of approximately

$$R_{overcast} = 10^{-3} * R_{sunny} = 10^{-3} * 0.015 \quad watts/cm^2 = 15 \quad \mu watts/cm^2. \tag{1.1}$$

The transmittance of the windshield of a common vehicle in the spectral band of interest is approximately 0.4. We assume the worst case scenario of a dirty window. This results in an irradiance on the vehicle occupants of

$$I_{passenger} = 0.4 * R_{overcast} = 0.4 * 15 \quad \mu watts/cm^2 = 6 \quad \mu watts/cm^2. \tag{1.2}$$

The second step in a radiometric computation is to determine how much of the incident radiation on the objects of interest is reflected back to the sensor (the $SU - 320$ near-infrared camera in our case). The radiance into a hemisphere assuming a reradiate of 0.4 would be

$$\begin{aligned} R_{passenger} &= 0.4 * I_{passenger} & (1.3)\\ &= 0.4 * 6 \quad \mu watts/cm^2 - ster & (1.4)\\ &= 2.40 \quad \mu watts/cm^2 - ster. & (1.5) \end{aligned}$$

This represents the reflected portion of the passenger irradiation. The rest is absorbed by the passenger's body. The reflected radiation has to pass through the windshield, the camera lens, the band-pass filter, and the polarizer to reach the near-infrared sensor array. As we did earlier, we assume a 0.4 windshield transmittance in the spectral band of interest. We also assume a $f/2$ camera lens (14.32^o cone angle) with 0.8 transmittance, a polarizer with 0.4 transmittance, and a band-pass filter with 0.6 transmittance. Assuming a $f/2$ camera lens (14.32^o cone angle) with 0.8

transmittance Then, the irradiance at the sensor array of the $SU-320$ camera will be

$$\begin{aligned} I_{camera} &= 0.4*0.8**0.4*0.6*\pi*R_{passenger}*\sin^2(14.32^o) && (1.6)\\ &= 0.4*0.8*0.4*0.6*\pi*2.4 \quad \mu watts/cm^2-ster*\sin^2(14.32^o) && (1.7)\\ &= 0.035 \quad \mu watts/cm^2-ster. && (1.8) \end{aligned}$$

The SU-320 camera has square pixels with a side of $37.5*10^{-4} \quad cm$ or an area

$$A = 37.5*10^{-4}*37.5*10^{-4} = 1.40*10^{-5} \quad cm^2. \tag{1.9}$$

Consequently, the radiant power on the camera pixel will be

$$\begin{aligned} P_{pixel} &= A*I_{camera} && (1.10)\\ &= 1.4*10^{-5} \quad cm^2*0.035 \quad \mu watts/cm^2-ster && (1.11)\\ &= 0.49*10^{-12} \quad watts. && (1.12) \end{aligned}$$

The camera's detectivity D^* is $D^* = 10^{12} \quad \sqrt{}-cmHz/watts$. The Noise Equivalent Power NEP is related to detectivity D^*, pixel area A, and electronic bandwidth Δf by the following equation:

$$NEP = (A/\Delta f)^{1/2}/D^*. \tag{1.13}$$

The bandwidth Δf is determined by the exposure time of the camera. The exposure time depends on the velocity, range, and field of view of the camera such that the images smear is less than 1 pixel. Assuming a vehicle traveling at a speed of $65 \quad mph$, a distance d of $40 \quad m$ apart from the camera, and a field of view of $1.6m$, the 320×240 pixel array of $SU-320$ gives a maximum exposure time of $1 \quad msec$ or a bandwidth of $\Delta f = 1 \quad KHz$. Substituting the values for A, Δf, and D^* in the formula of NEP (see Eq. (1.13)) we get

$$NEP = 1.18*10^{-16} \quad watts. \tag{1.14}$$

Therefore, the camera signal to noise ratio S/N will be

$$S/N = P_{pixel}/NEP = 4152. \tag{1.15}$$

In conclusion, assuming a worst case scenario (overcast day, dirty windshield) we determined that the $SU-320$ camera, equipped with a $f/2$ lens, a $1.4-1.7 \quad \mu m$ filter, and a polarizer, if it is positioned at a distance of $d = 40m$ from the incoming vehicle and with a field of view of $1.6 \quad m$ (or height $h = 7 \quad m$ at the specified distance), will give an acceptable signal to noise ratio $S/N = 4152 > 35$ and an acceptable smear of up to one pixel. Also, the required exposure time of $1 \quad msec$ is within the nominal speed specification of the $SU - 320$ camera. During night time, the same computation holds, and the sun illumination is substituted by the illumination of an artificial near-infrared light source (see Fig. 1.16).

1.4.2 Mn/ROAD Experiment

The above theoretical scenario was put into testing in the Mn/ROAD research facility outside Monticello, MN. The experiment lasted for a week in late September 1998. One test lane of freeway $I94$, one mile long, was released from traffic and given to us for exclusive use. We set up the $SU-320$ camera with all its accessories above the test lane of freeway $I94$. In the absence of a permanent installation device we installed the sensor suite in the basket of a cherry-picker. We implemented the experiment exactly as it was specified in Subsection 1.4.1. We used two testing cars that made successive passes through the field os view of the near-infrared camera. The passes were done at speed increments of $10mph$, ranging from 10 - $50mph$. One of the testing cars was representative of the compact category (Mitsubishi Mirage) and the other of the luxury category (Oldsmobile Aurora). The experiment had both day and night sessions. During night time we used an artificial near-infrared illumination source in the range $1.0 - 2.0 \quad \mu m$ that covered the illumination needs of both the lower and upper band.

The experiments confirmed our theoretical predictions. The most interesting result was the experimental verification of skin appearance in the near-infrared region of choice. Above the 1.4 μm threshold human skin appeared consistently dark in the imagery while below the 1.4 μm threshold it appeared consistently light (see Figs. 1.17 and 1.18. Remarkably, everything else in the scene (e.g. upholstery, dashboard, car frame) appeared more or less unaffected in the imagery of both bands. Based on this observation, we tried to co-register in the lab matching images from the two bands, subtract them, and threshold them. Ideally, if everything but the human signatures remains the same in the two bands, the image subtraction should produce an image where only the silhouettes of the vehicle occupants remain. Everything else would be cancelled out. Fig. 1.19(b1) shows the result of fusion from the images in Fig. 1.17(a) and 1.17(b). We fused the images by first co-registering them through a warping operation and then subtracting them from each other. We then thresholded the fused image to get the final processed result (Fig. 1.19(b2)). Somebody may notice that along with the face of the driver a small amount of noise is also present at the final image. We determined that this is mostly due to the imperfections of the fusion process we applied. In particular, because we had only one camera, we first shot a scene with the camera equipped with the lower band filter. After some time, we shot the same scene with the camera equipped with the upper band filter. Because, however, the two shootings took place half an hour apart the illumination of the scene was different (diurnal cycle) and therefore the image subtraction operation produced less than the ideal theoretical results. Even under these adverse circumstances, our approach is clearly superior than the traditional visible spectrum approach. Fig. 1.19(a2) shows the thesholded image that resulted from the visible spectrum image of Fig. 1.19(a1). The visible spectrum image represents the same scene as the near-infrared images 1.17(a) and 1.17(b). It was shot with the $SONYDSR-200$ digital camera that was standing side by side with the $SU-320$ the day of the experiment. It is evident that the visible spectrum thresholded image

has a couple of order of magnitudes more noise than the thresholded fused image.

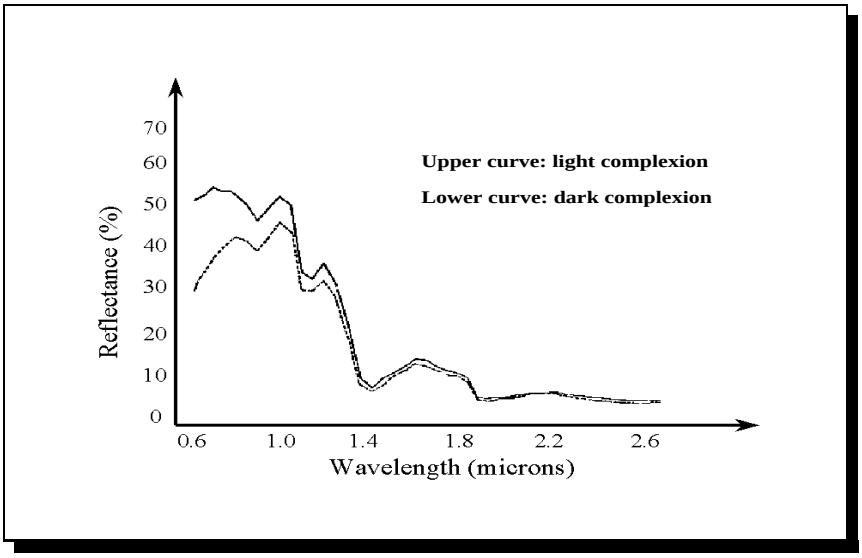


Figure 1.9: Skin reflectance of black males. Upper curve corresponds to light complexion while lower curve to dark complexion.

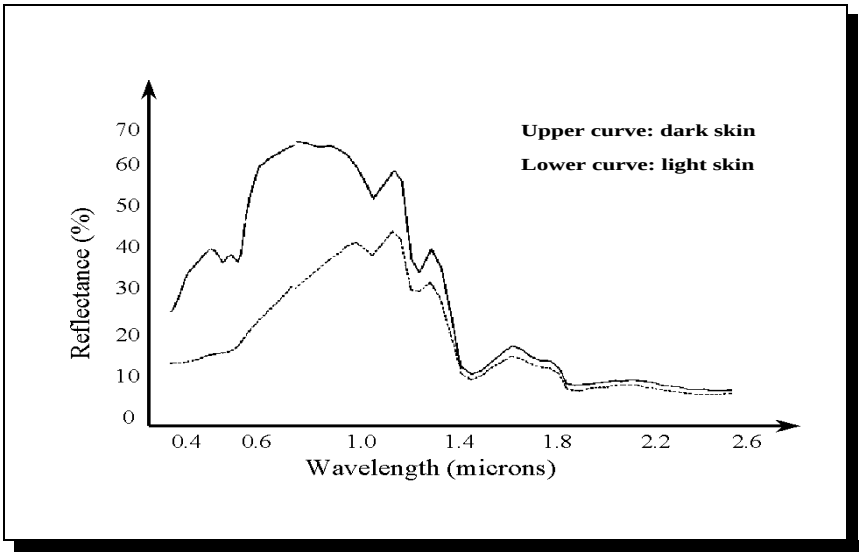


Figure 1.10: Reflectance of dark skin versus light skin. The lower curve corresponds to dark skin while the upper curve to light skin. Up to 1.4 μm the discrepancy between the two curves is substantial and it explains why to the human eye white people appear white and black people black. After the 1.4 μm threshold point, however, the two curves are almost coincident. They both feature very low reflectance values in this range, which explains why everybody appears dark in the near-infrared camera operating in this band.

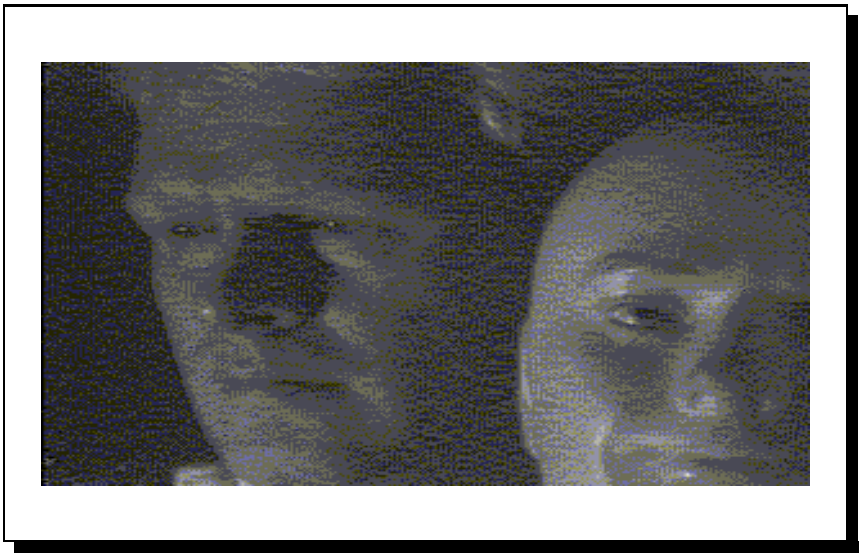


Figure 1.11: A Caucasian male and a dummy head in the range $1.1 - 1.4 \quad \mu m$.

Figure 1.12: A Caucasian male and a dummy head in the range $1.4 - 1.7 \quad \mu m$.

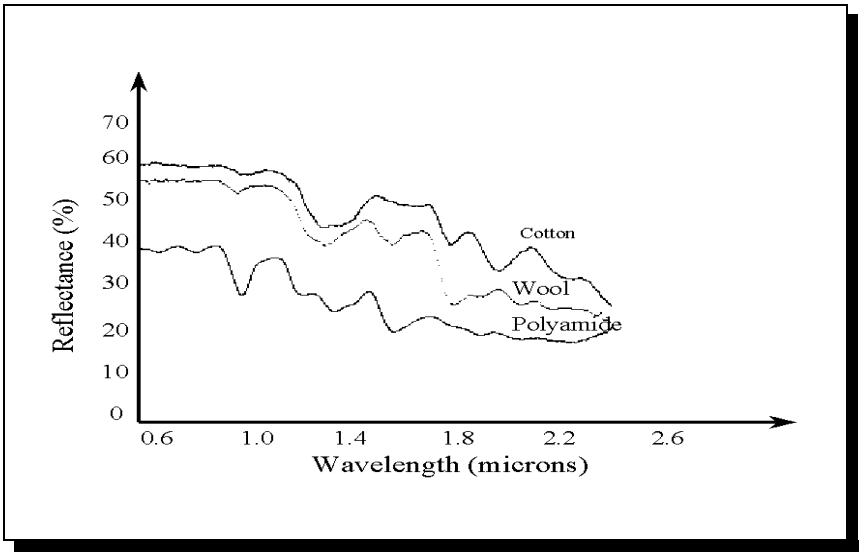


Figure 1.13: Reflectance of different fabric materials. The drop in reflectance after the 1.4 μm threshold point is relatively minor.

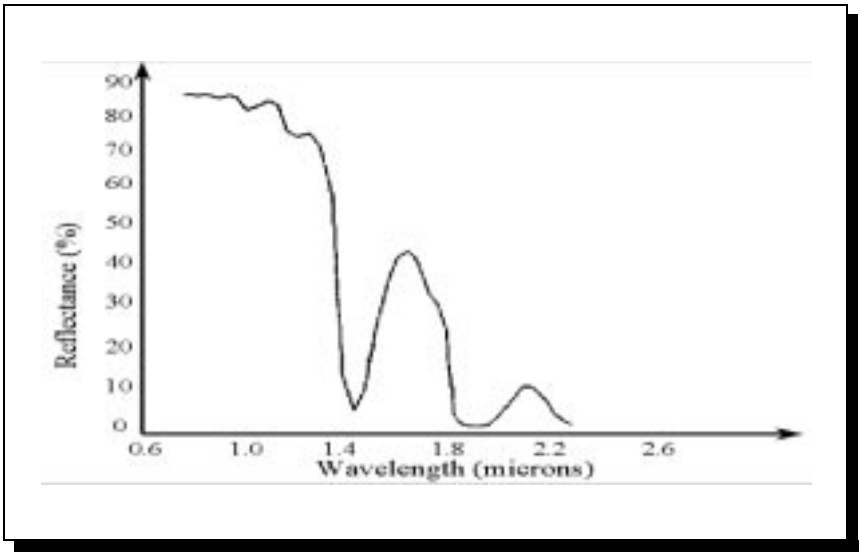


Figure 1.14: Reflectance of distilled water. The drop in reflectance after the 1.4 μm threshold point is substantial.

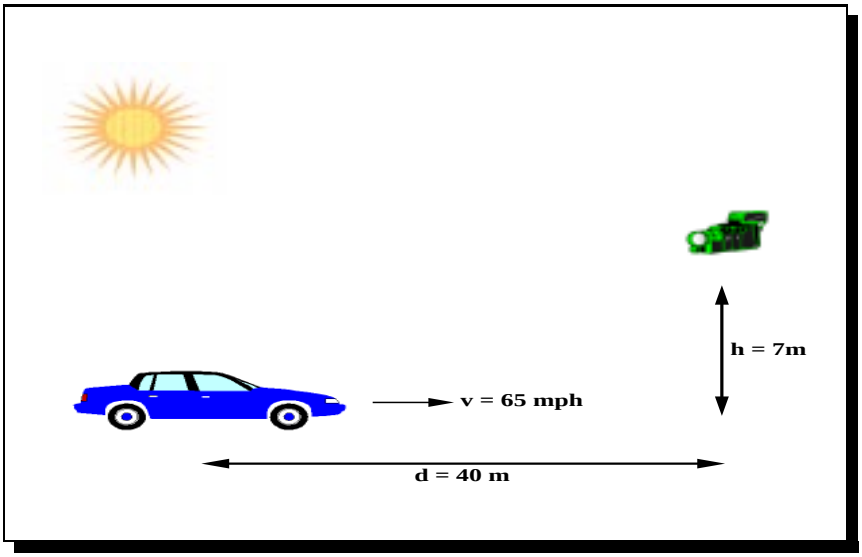


Figure 1.15: Sensor arrangement for day time scenario.

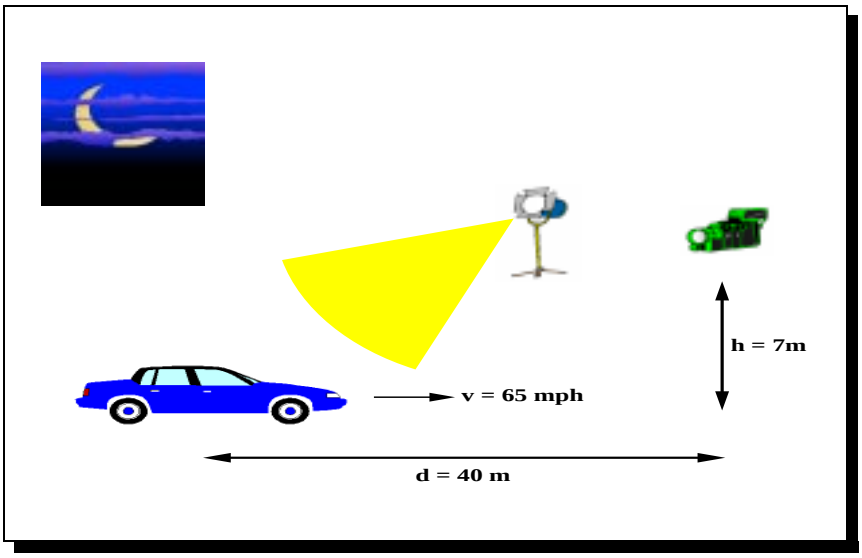


Figure 1.16: Sensor arrangement for night time scenario.

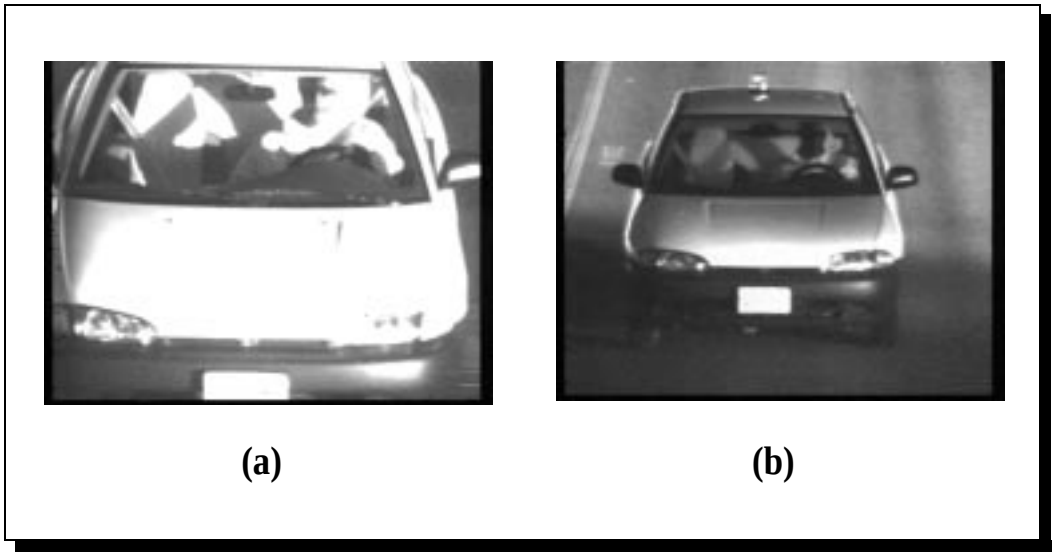


Figure 1.17: Near-infrared day time results. (a) Image in the band $1.1 - 1.4 \quad \mu m$. Image in the band $1.4 - 1.7 \quad \mu m$.

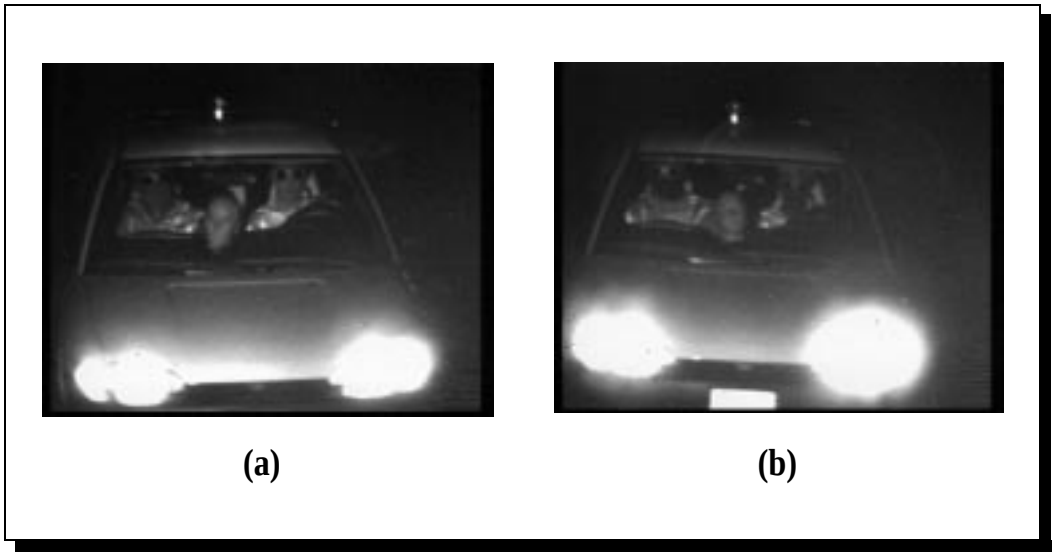


Figure 1.18: Near-infrared night time results. (a) Image in the band $1.1 - 1.4 \quad \mu m$. Image in the band $1.4 - 1.7 \quad \mu m$.

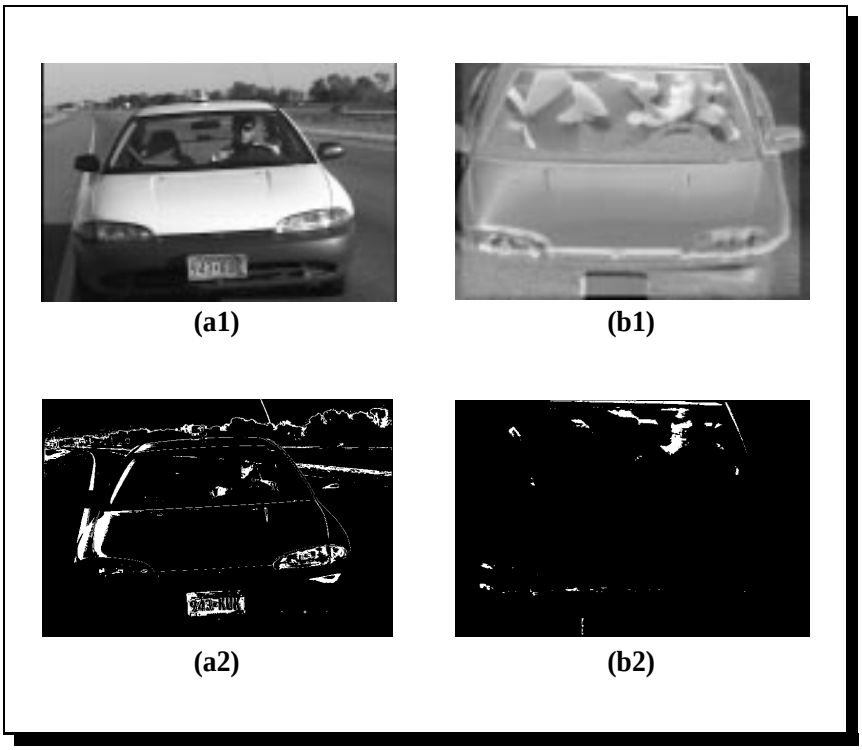


Figure 1.19: Comparative results between the visible spectrum approach and the near-infrared fusion approach.

Chapter 2

PRIMARY ALGORITHM FOR AUTOMATIC DETECTION OF VEHICLE OCCUPANTS

We chose to perform the detection of vehicle occupants with a neural network. In particular, we opted for a fuzzy neural network that implements the Adaptive Resonance Theory (ART). This type of neural network features a series of appealing properties for the application at hand.

1. *Self-Organization.* This is a property that characterizes the learning process of the neural network. In contrast to supervised learning neural networks (i.e., back-propagation), Fuzzy ART networks do not need any external guidance but find automatically the regularities interwoven in the input data. This translates to easier and less expensive ground-truthing, an important factor in a cost critical endeavor such as ours.

2. *Stable Categorization.* This property is related to the degree that a neural network forgets categories (patterns), which it had encountered far into the past. This is the so called *stability-plasticity* dilemma. The ART network features a feedback mechanism between the layers that help solving the stability-plasticilty problem. This feedback mechanism facilitates the learning of new information without destroying old information. Most important, stable categorization is maintained even at a fast learning pace. Learning stabilizes after just one presentation of each input pattern. This was very important in our case because the size of our image database was rather small and we could not afford going through an expensive learning cycle.

3. *Broad and Narrow Classification.* ART networks have an explicit parameter called *vigilance* that controls its generalization capability. In other words, vigilance controls the formation of broad and narrow classifications. This control is very useful in the presence of highly variable patterns of vehicle occupants.

4. *Fuzzy Classification.* The incorporation of fuzzy set theory into the operation of ART networks addresses the problem of disambiguating overlapping categories with minimum risk.

2.1 The Fuzzy Neural Network Algorithm

Fuzzy ART neural networks are comprised of an input layer F_0 and an output layer F_1. The typical structure of an ART neural module is shown in Fig. 2.1. The input layer consists of N nodes (neurons) which encode the input vector $\overrightarrow{I} = (I_1, I_2, ..., I_N)$. In our application each input node represents the gray level intensity of a pixel

$$I_j \in [0,1], \forall j \in (1, 2, \cdots, N). \tag{2.1}$$

The input vector is augmented to achieve input normalization through a process that is called complement coding. The complement coded input vector $\overrightarrow{P}$ becomes a $2N$-dimensional vector

$$\overrightarrow{P} = (\overrightarrow{I}, \overrightarrow{I^c}) \equiv (I_1, \cdots, I_N, I_1^c, \cdots, I_N^c), \tag{2.2}$$

where $I_j^c \equiv 1 - I_j$. One may observe that the complement-coded input $\overrightarrow{P}$ is normalized since

$$|\overrightarrow{P}| = |(\overrightarrow{I}, \overrightarrow{I_c}| = \sum_{i=1}^{N} I_i + \left(N - \sum_{i=1}^{N} I_i\right) = N. \tag{2.3}$$

The M nodes in the output layer represent the classification categories (e.g. the activation of the leftmost output neuron denotes the presence of one passenger, the activation of the neighboring neuron denotes the presence of two passengers, and so on). Each output neuron j is associated with a vector $\overrightarrow{w_j} = (w_{j1}, w_{j2}, \cdots, w_{j2N})$ of

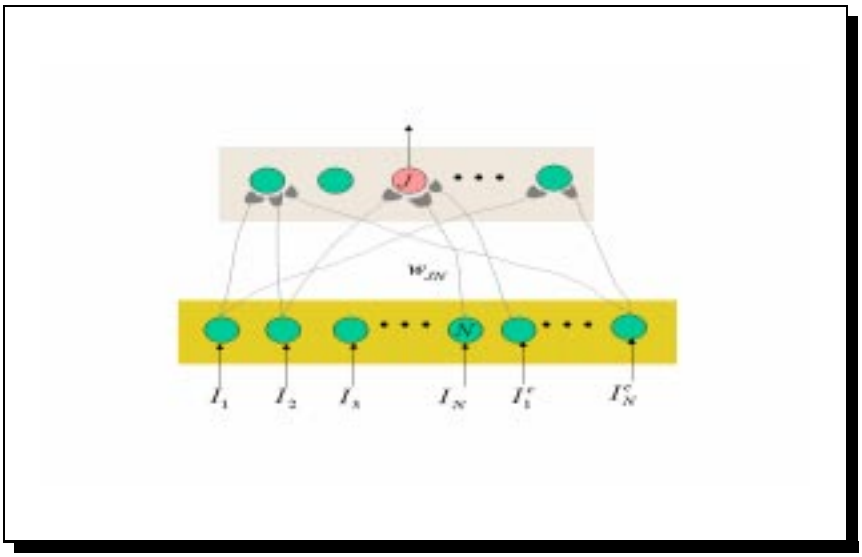


Figure 2.1: ART networks are two-layer neural modules. There exists a complete set of bottom up weights from the input layer (red box) neurons to the output layer (light blue box) neurons. The size of the adaptive weights, which change through learning, is graphically denoted by the different size of the blobs that surround the output neurons. The pink colored output is the category selected for the present input.

adaptive weights that represent the knowledge that the neural network retains at the current time. The values of the elements of this vector change during the neural network operation. Initially, they all have unit values.

For a typical input $\overrightarrow{P}$, a choice function T_j is computed for every output neuron as

$$T_j(\overrightarrow{P}) = \frac{|\overrightarrow{P} \wedge \overrightarrow{w_j}|}{|\overrightarrow{w_j}|}, \tag{2.4}$$

where the fuzzy AND operator $\wedge$ is defined by $(\overrightarrow{x}, \overrightarrow{y})_j \equiv min(x_j, y_j)$ and $|\bullet|$ represents the Hamming distance norm.

The choice function measures the degree to which the weight vector w_j is a fuzzy subset of the input $\overrightarrow{P}$. There is only one neuron that is activated for a particular

input (i.e., an image) that is presented in the input layer. For this reason fuzzy ART networks belong to the class winner-take-all networks. The output node J is the chosen candidate for classifying the current input for which

$$T_j(\overrightarrow{P}) = max\{T_j|j = 1, \cdots, M\}. \tag{2.5}$$

The chosen candidate neuron J finally classifies correctly the present input if it meets the *vigilance* criterion. The vigilance criterion is mathematically described by the following equation:

$$\frac{|\overrightarrow{P} \wedge \overrightarrow{w_j}|}{|\overrightarrow{P}|} > \rho, \tag{2.6}$$

where ρ is the vigilance parameter. If Eq. (2.6) is met we say that *resonance* occurs. Hence, resonance occurs when the degree to which the input $\overrightarrow{P}$ is a fuzzy subset of $\overrightarrow{w_j}$ exceeds the vigilance parameter ρ, which takes values in the interval $(0, 1]$. The vigilance parameter defines the lower bound of the degree of dissimilarity of disparate inputs that are classified under the same category. If the vigilance criterion is not met, the choice function associated with the chosen neuron is reset to $-1(T_j(\overrightarrow{P}) = -1)$ until the presentation of a new input. The same process for choosing a different neuron j is then repeated until one is found that meets the vigilance criterion. When such a category j has been found we say that it is a fuzzy subset choice for input $\overrightarrow{P}$. For this selected output neuron j learning occurs as follows:

$$\overrightarrow{w}_j^{(new)} = \beta(\overrightarrow{w}_j^{(old)}) + (1 - \beta)\overrightarrow{w}_j^{(old)} \tag{2.7}$$

where, the learning parameter β can take values in the interval $(0, 1]$.

2.2 Geometric Representation of the Fuzzy Neural Network Classification

There is an interesting geometric interpretation of the category formation process when the Fuzzy-ART networks are employed. In order to make our point clear,

we will assume that our inputs represent $2-D$ vectors instead of the 300×110 - dimensional pixel vectors that were used in our application. The results from the $2-D$ case can easily be generalized to the N - dimensional case.

The formation of classification categories is shown in the space of input vectors (see Fig. 2.2). When an output node is chosen for the first time we say that the neuron commits to a new class. Since this input is the only point in the class, this point represents the respective class. The second time this committed output neuron is selected to represent another input different from the previous one, the smallest rectangle that will contain those two points will be formed. This is the rectangle that will represent the class from now on. The same process will be repeated for new inputs. The maximum size of the rectangles (represented by its perimeter) is determined by the vigilance parameter. In a similar fashion other classes are formed. One may see that classes (color-coded rectangles) overlap due to fact that fuzzy concepts are incorporated into the neural network.

2.3 Performance of the Algorithm

The neural network described above was tested on 90 different images with the camera operating at the $1.1\mu m - 1.4\mu m$ and $1.4\mu m - 1.7\mu m$ wavelength ranges. These images were taken at different times during the day and at car speeds of $0-40mph$. Table 2.1 describes in detail the characteristics of the sample images. The network performed at the fast learning mode ($\beta = 1$) and scored perfectly in classifying the 90-sample image set. In order to test the stability of the algorithm, images from the same set were presented in an arbitrary order to the network after the first epoch and again it scored perfectly.

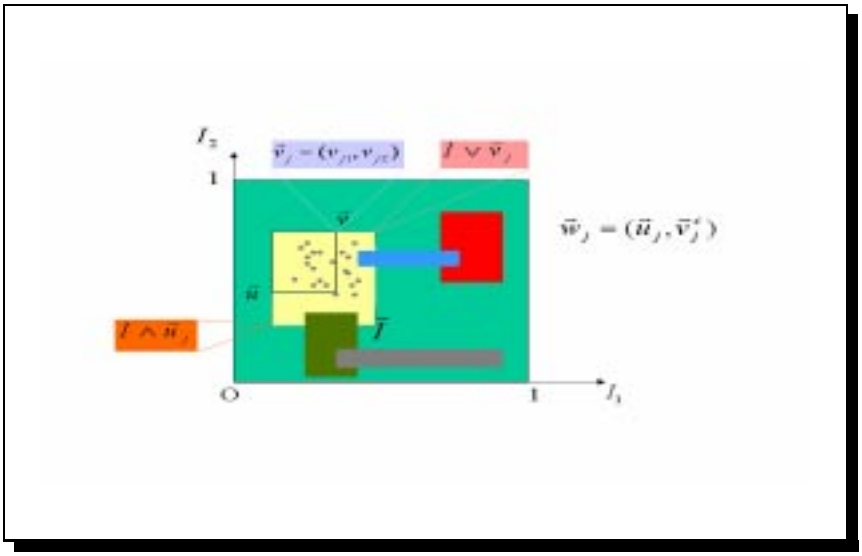
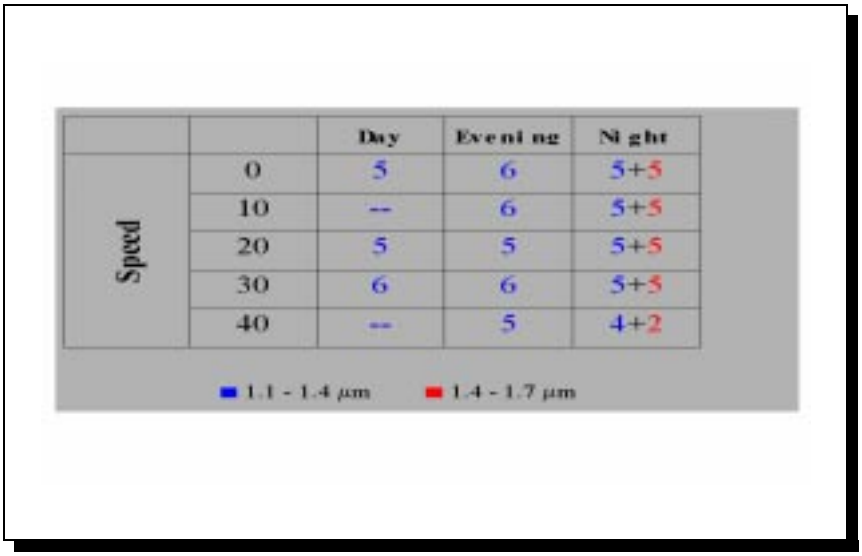


Figure 2.2: Classes in Fuzzy-ART networks are represented as color coded rectangles. Inputs that fall within a particular rectangle are classified by the output neuron associated with the respective class.



		Day	Evening	Night
Speed	0	5	6	5+5
	10	--	6	5+5
	20	5	5	5+5
	30	6	6	5+5
	40	--	5	4+2

1.1 - 1.4 μm 1.4 - 1.7 μm

Table 2.1: Types of samples used for testing the performance of the neural network algorithm. Images taken at two wavelengths are shown in the blue and red numbers respectively. The total number of sample images was 90.

Chapter 3

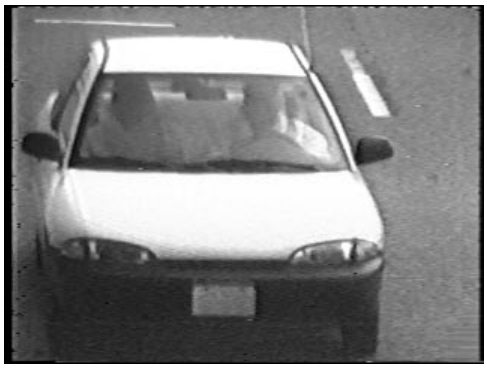
ALTERNATIVE ALGORITHMS FOR AUTOMATIC DETECTION OF VEHICLE OCCUPANTS

The first stage of our alternative detection algorithms seeks to locate the *focus of attention* where the occupants may be seated i.e. the windshield of the automobile. At the second stage, we threshold the image to a binary version such that the interval of pixel values for human skin is colored white and all other pixel values are colored black. Linear regression models are created in order to predict the expected values of this interval. The binarized image is then segmented into regions of adjacent white pixels. In the final stage, the windshield region is split into two sides (passenger, driver) each of which is the input for classification. We explore two classification systems, the first is a hand-coded version and the second is an artificial neural net. Both classify each blob into one of two groups: passenger, not passenger.

Figure 3.1 shows representative examples for the near-infrared configuration that are used throughout the rest of this paper.

3.1 Focus of Attention: the Windshield

In this first processing stage, we assume there is an automobile in view and seek to locate its windshield. Since windshields are rectangular in shape, we search for rectangles in the image. These are determined by inspecting the vertical and horizontal edges, combining them into rectangular regions, and finally selecting the most reasonable one for further processing. We begin with smoothing the image in order to reduce speckling noise and then sub-sampling to improve efficiency.



(a)

(b) (c)

Figure 3.1: Examples of input images: (a) zoom lens on a fairly bright day with passenger and driver, (b) very bright afternoon with passenger and driver, (c) zoom lens at dusk with driver and no passenger.

3.1.1 Locating the Windshield: a Rectangle with Special Properties

Using the smoothed and sub-sampled image, we construct a set of potential rectangular regions in the image. We use image projection to determine where lines are. This method accumulates all pixel values along one dimension (e.g. horizontally or vertically). One benefit of this technique is broken edges still contribute to the overall count. One disadvantage is that excessive noise such a mechanical malfunctions or painted lane markings would create short lines that might be misinterpreted as

broken edges of the car.

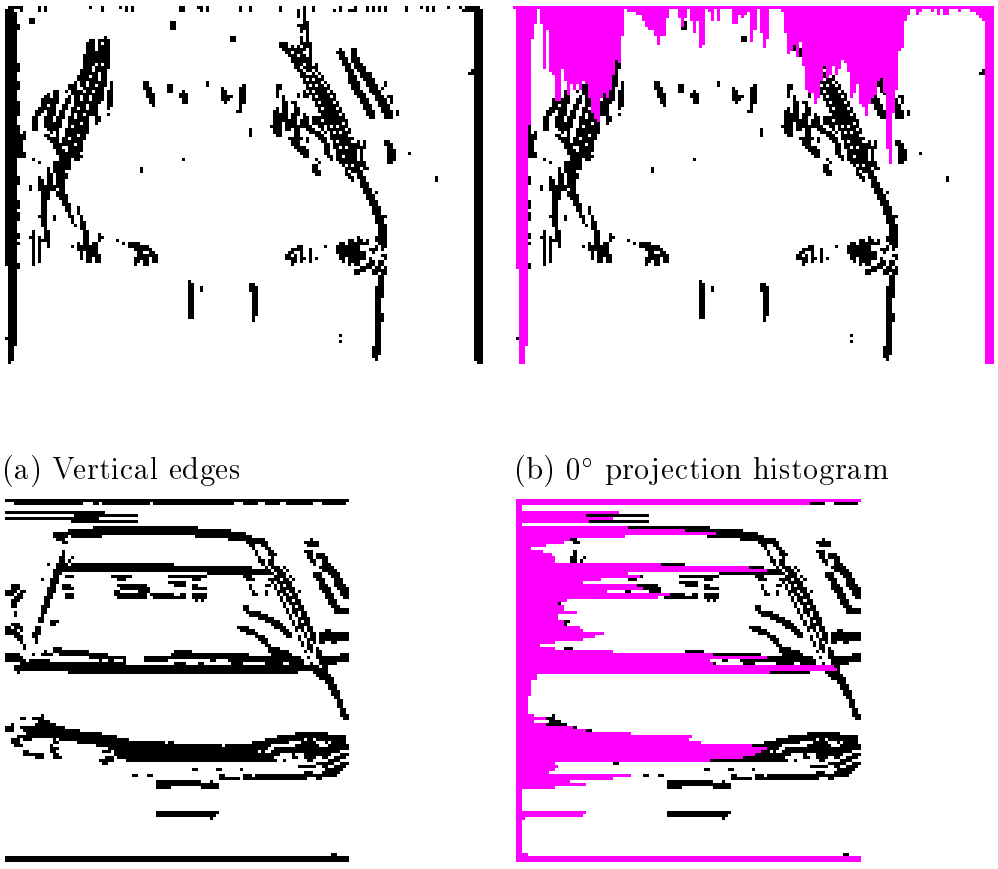
First, we find the two vertical lines that correspond to the left and right sides of the car. This is done with the application of a vertical edge kernel followed by the noise-reducing thresholding. Next, a 0^o projection sums the number of lit pixels for each column (see Figs. 3.2(a)(b)). While there may be many vertical lines, we search recursively from the left and right outsides of the image toward then center until two similarly-sized vertical lines are found. The range of allowable vertical line length extends from one-half to the entire frame height

To form the horizontal line set the image is first cropped to the outside vertical edges. Next we apply a horizontal edge kernel, threshold to remove noise, and then perform a 90^o projection that sums the number of lit pixel along each row (see Figs. 3.2(c)(d)).

Edge thinning combines adjacent lines into a single representative line. Since the projections are linear this task is accomplished quickly in one pass. We limit the length of allowable lines to the range of $75 - 100\%$ of the (narrowed) image width. To form the set of rectangular regions we combine the pairs of vertical lines with the set of horizontal lines to form a grid. The grid entries do not inscribe any other rectangles. Figure 3.3(b) shows an example.

Given the set of rectangular regions, we determine which one bounds a region that is likely to be a windshield. Two properties are used to differentiate them from the other regions. The windshield is a rectangle whose $length : breadth$ ratio is around $1 : 4$. Excessively wide entries fall below $1 : 4$. Excessively tall entries are above $1 : 2$.

The windshield is also a region whose pixel color variance is higher than its other surfaces, e.g. the hood is "smooth" (low variance) while the window area is "rough" (higher variance). We select the remaining rectangle with the highest variance as the candidate for further processing (see Fig. 3.3).



(a) Vertical edges (b) 0° projection histogram

(c) Horizontal edges (d) 90° projection histogram

Figure 3.2: Edges and their projections of Fig. 3.3(a)

3.2 Thresholding Using Statistical Models

In this stage of processing, the windshield is split into two halves: the passenger and the driver sides. Each will be thresholded prior to being segmented into *blob* regions which are later classified as being a passenger or not. This section concerns the thresholding the image into two colors: pixels whose values are those of skin will be colored white, and the rest will be black. Thresholding is a necessary step for segmentation, although, thresholding the images we encounter is not easy. [1, 2] discuss thresholding approaches for bimodal images. This method assumes that

(a) (b)

Figure 3.3: Windshield location results: (a) shows the original image, (b) shows the horizontal and vertical lines that form a grid from which the windshield is identified (marked in an alternate color).

background pixels fall into the darker of strictly two classes. However, in our case the occupant's face is colored in a limited grayscale range that such an approach could not accommodate.

3.2.1 Creation of Models Using Linear Regression

Observation of the sample input images shows that the theoretical expectations are accurate in ideal settings when certain assumptions are true (Fig. 3.1(b)). Near-infrared reflectance requires sufficient illumination, passively from the sun and actively from an illumination device. Figs. 3.1(a,c) demonstrates that insufficient reflectance from the human skin falls close to 0% (seen as black in the images). Purely theoretical calculations are not sufficient for the real world situation we expect to encounter.

Statistical linear regression [11, 4, 10] has the potential to help us develop an algebraic expression that transforms image characteristics into an accurate range of values for the human skin as detected by our camera. We experimented with sample

images in order to get some idea about the threshold range. The results indicated that the threshold ranges vary in both the mean and width. Therefore two models were designed: one to estimate the expected mean, and the other to estimate the expected interval width. The domain inputs to these models would be chosen from among statistical characteristics of the two windshield sides.

Results

The creation of a regression model requires a representative sample data set. Thirty-one random samples were collected from among two hundred images of the three different signal conditions of Fig. 3.1. Each image was manually processed to determine a thresholding interval that distinguished the facial region from the rest of the image. Statistical data was gathered and stored for each window side and its corresponding threshold interval.

Each of the two models has the form $y = x_0 + x_1 t^{(1)} + x_2 t^{(2)}$ where y is the expected mean $\hat{\mu}$ in the case of the first model and the expected width $\hat{w}$ for the second model. x_i are constant coefficients. $t^{(i)}$ are the input variables: the input image's mean $\overline{x}$ and mean squared $\overline{x}^2$ for the first model, and the gradient image's standard deviation s and its square $s^2$for the second model. To determine the x_i, multiple linear regression solves the equation $\mathbf{y} = \mathbf{A}\mathbf{x}$ for the coefficients $\mathbf{x}$. $\mathbf{y}$ contains the actual experimental target values (the threshold interval mean for the first model, the threshold interval width for the second model). $\mathbf{A}$ contains the two measured values (either $\overline{x}$ and $\overline{x}^2$,

or s and s^2). $\mathbf{x}$ and $\mathbf{A}$ have the forms:

$$\mathbf{x} = \begin{matrix} x_0 \\ x_1 \\ . \\ . \\ . \\ x_3 0 \end{matrix}$$

$$\mathbf{A} = \begin{matrix} 1 & a_{1,1} & a_{1,2} \\ . & a_{2,1} & a_{1,2} \\ . & \dots & \dots \\ 1 & a_{31,1} & a_{31,2} \end{matrix}$$

The first attempt at building a model was unsuccessful. Using raw image input, we sampled swatches of skin and the overall window side, and calculated the standard deviation and mean of each type. We were able to model the mean, however the width could only be over-fit in order to account for the majority of variability in the data.

The second attempt was successful. First we sought to heighten the contrast of the images. The raw images appear very dark and weak signal strength appears gray to black, which unfortunately is close to the color of the camera filter's representation of skin. Experimentation revealed that the contrast of the image was improved with a histogram equalization of uniform distribution of pixel values. In other words, contrast was heightened as shown in Figs. 3.5(a,b).

Using the equalized histogram of the smoothed image we manually experimented with binary threshold rages for the population sample of images in order to determine an interval that maximized the distinctiveness of facial regions among the overall image. We stored the interval as well as the variance and mean of the window side. Our attempt at a regression yielded a successful model for the mean that captured more than 90% of the variance.

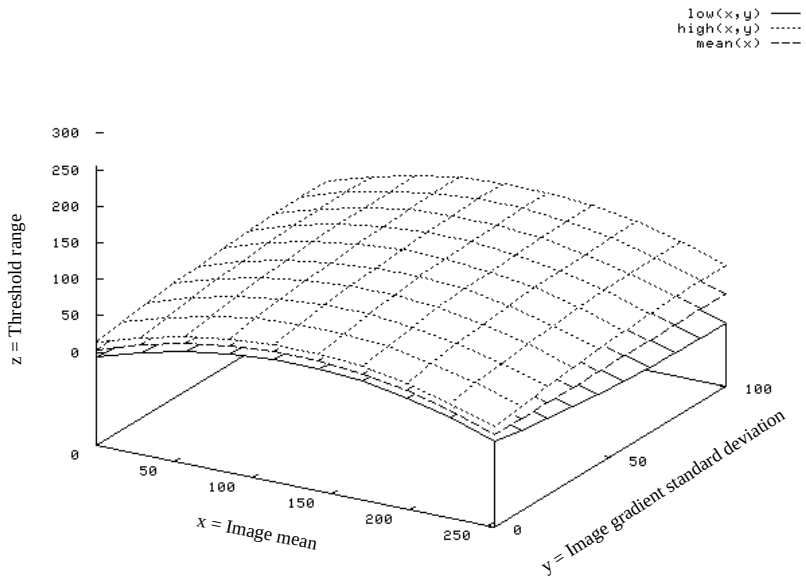
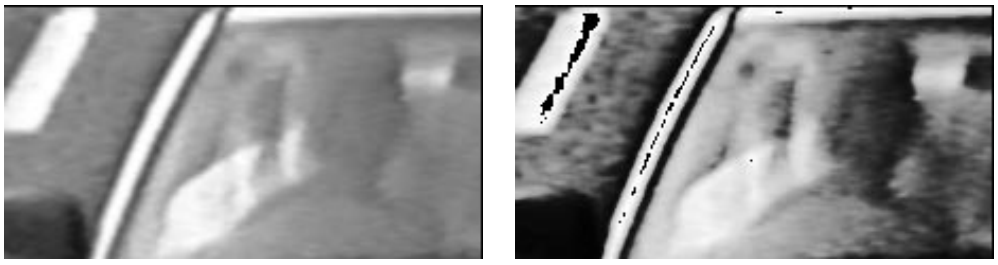


Figure 3.4: Threshold model surface of interval upper bound, mean, lower bound as functions of the image mean pixel value and image gradient standard deviation.

Unfortunately the model for width over-fit the data. We decided to explore fitting the model utilizing the variance and standard deviation of the magnitude intensity gradient image. It was anticipated that this transformation would provide a more regular description of data and thereby produce a suitable model. The Sobel operator provided an excellent choice for estimating the partial derivatives ($\frac{\delta I}{\delta x}$, $\frac{\delta I}{\delta y}$). The result was a second degree model that captured over 70% of the variance.

A plot of the combined thresholding models' surfaces is shown in Fig. 3.4 and an example of its use is shown in Fig. 3.5. The models we use for mean estimate $\hat{\mu}$ and width $\hat{w}$ of the window side are:

$$\hat{\mu} = 3 + 0.75\overline{x} - 0.003\overline{x}^2 \qquad \hat{w} = 20 + 1.2s - 0.006s^2$$



(a) Passenger side original (b) Uniform histogram equalization

(c) Thresholded Image

Figure 3.5: Results of Threshold Model

3.3 Classification of a Person

Another principal algorithmic concern is the classification of the input image's segments. This process takes a binarized image (each half of the windshield region) that is segmented into regions of neighboring pixels (blobs). Given some inaccuracies in sensing, the image can contain noise in the form of extraneous regions that we attempt to filter out. In this section we present two approaches to classification: the first is a hand-coded version and the second is an artificial neural network.

Classification maps a feature vector containing mathematical characteristics of a region to a boolean value: *is a face*, *is not a face*. The hand-coded version outputs a strict *true* or *false* output, while the neural net outputs a probabilistic range $[0, 1]$ where 0.0, 0.5, 1.0 represent, respectively, *false*, *uncertain*, *true*. *Feature selection* creates a vector of values which determines characteristics that describe

the uniqueness of the person's face. Fig. 3.6 shows these features as they relate to our work. Additional features include: *elongation* ($Length/Breadth$), *roughness* ($Perimeter/Perimeter_convex$) to measure how jagged the blob is, *compactness* ($Perimeter^2/(4\pi Area)$) to measure how close pixels are to each other.

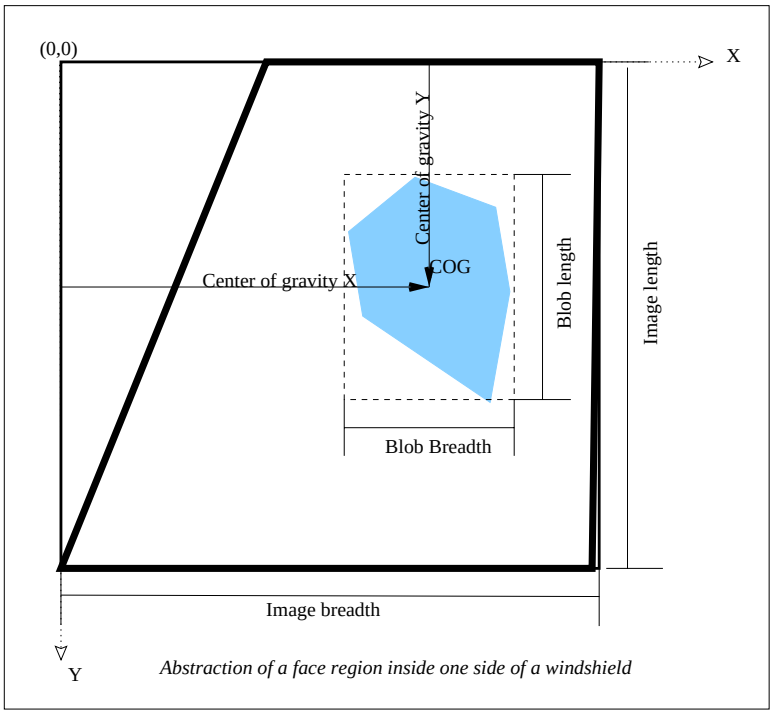


Figure 3.6: Feature Vector Composition

3.3.1 Hand-coded Version

In the *hand-coded* version we manually selected parameterizations of a feature vector until we achieved a reasonably consistent percentage of correct classifications from another sampling of the data. The goal is to exclude blobs that are not consistent with the following features and their respective constraints.

The elimination of blobs that are not centered in the image uses the blob's center of gravity COG and the image $Width$ and $Height$ through the exclusion of those

which do not satisfy either $0.4\,Width < COG_x < 0.9\,Width$ or $0.6\,Width > COG_y < 0.1\,Width$. The elimination of blobs that are too large or small excludes violations of $Area_{image}/8 < Area_{blob} < Area_{image}/4$. The elimination of blobs that are too horizontally thin excludes violations of $Elongation > 2.0$. The elimination of blobs that are too spread out excludes violations of $Compactness > 10$; Results are shown in Fig. 3.7 and 3.11(b).

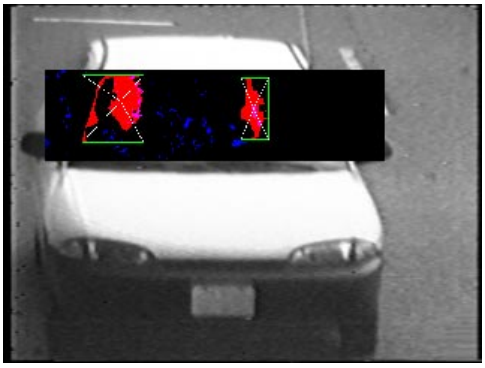


Figure 3.7: Results of Hand-coded Classification

3.3.2 Neural Networks

Another way to classify a blob is with an artificial neural network [6, 5]. Conceptually, the network behaves like primitive brain that has no knowledge of how to classify. Using a technique called *supervised learning* we incrementally present a *training set* of the image segments that were seen with our cameras during the data collection phase. Some of these may be passengers (*positive examples*) and others may not (*negative examples*). In the *forward phase*, the network classifies each sample and outputs a number that represents its confidence that this is a person. At the subsequent *backward phase* we tell the network whether is was correct or not, and the network learns by modifying its structure slightly. We measure the success rate using a *test set* of similar images. This process of providing positive and negative examples, querying

the network, improving performance continues as the network converges when the success rate reaches a satisfactory level.

The inputs to the networks represent *features* that describe the image. Some examples of these features include: shape, length, elongation, roughness, mean color values, etc. It is important that we represent our knowledge of the input image in a manner that is invariant to transformations such as scale. For example, consider that we use the length feature: since the focal length of the camera may vary causing the length of a person's head to vary as well. Therefore a more suitable solution would be to use the relative length of an object inside the window to the entire window length.

Neural networks may be designed in various ways. They may vary according to connectivity, input and output representations, the manner of error correction, and the rate at which they learn. These variable parameters must be manually fine-tuned with careful experimentation. This process may be time-consuming but worthwhile. Neural networks are capable of classifying complex patterns such as the ones we face in this study. In addition, the run-time performance of a well-tuned neural network is very efficient

3.3.3 Experimental Design and Results

We have developed a *multi-layer feed-forward network* using back-propagation error correction. The sigmoid function was used for activation and a bias node was attached to each internal node. Learning was performed in *pattern mode* where each example was presented followed by a weight update. We used two learning rates: one for the input to hidden node links and another for the hidden to output node links. A momentum constant was used to accelerate convergence. The complete structure along with its feature vector of inputs is shown in Figure 3.8 and a detailed listing of training parameters is contained in Table 3.1.

We trained and tested the network with a total of 240 examples, half were negative examples (blobs that were not human faces) and half were positive examples. 70% were used for training and 30% for testing. The learning curve is shown in Fig. 3.9. Examples

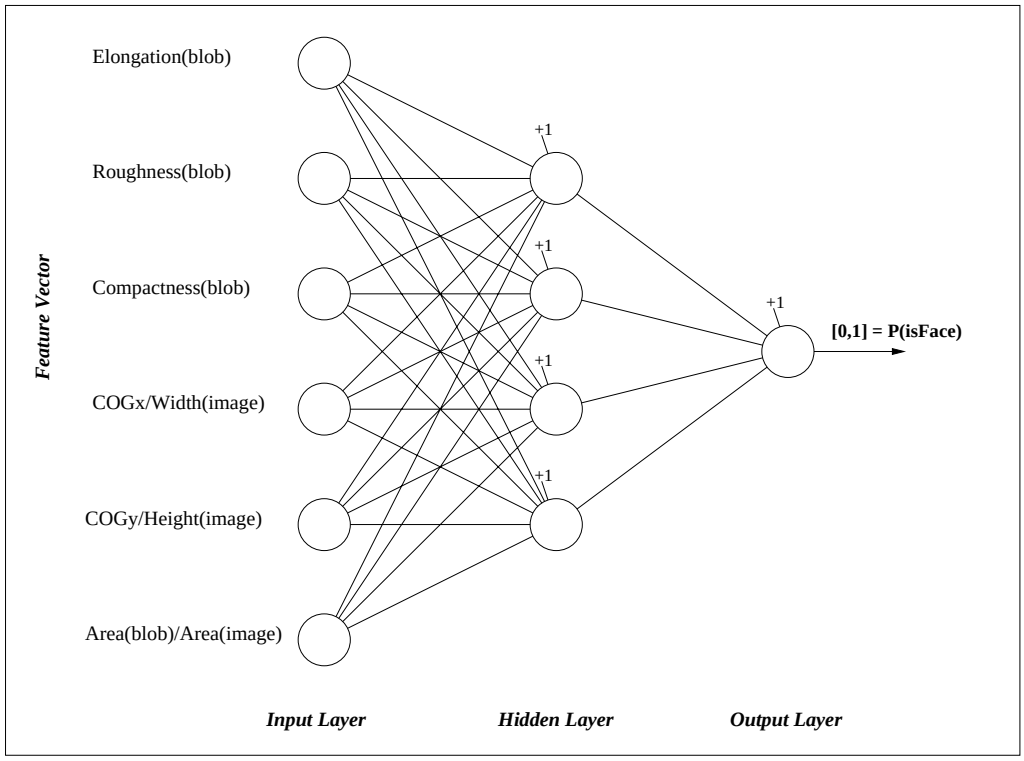


Figure 3.8: Neural Net Structure

of classification results are shown in Figs. 3.10, 3.11(c).

3.4 Software and Hardware

The software design consists of a user-interface that served as both a development and experimental tool project processing algorithms. A user-friendly front-end was developed coded in *C++* under **Windows NT**. The interface issues commands to the Matrox imaging system [9], a set of parallel vector processors. The regression analysis were performed with the *MacAnova* statistical package [12].

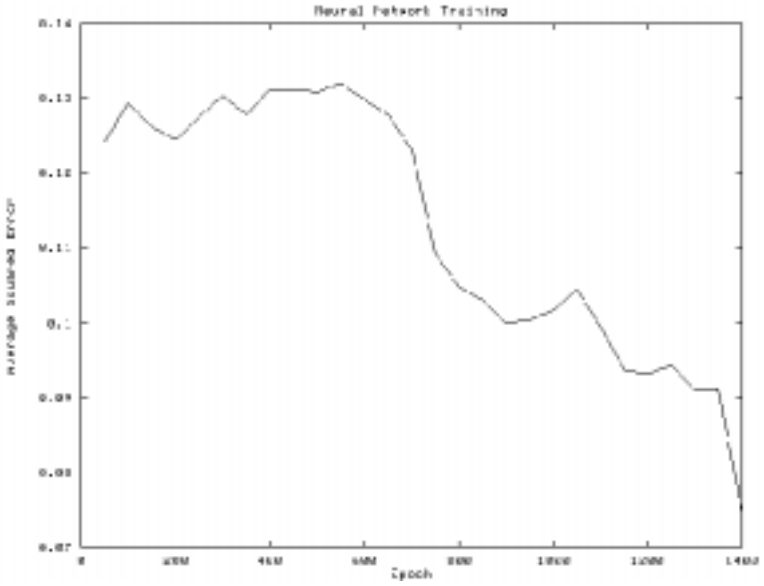


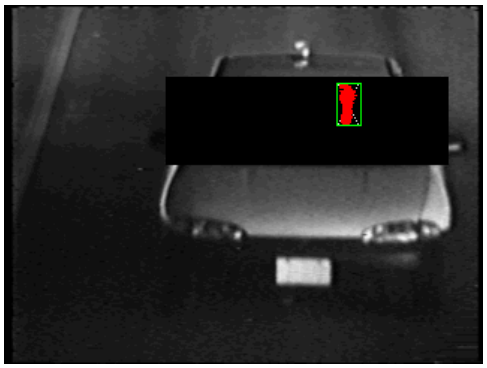
Figure 3.9: Neural network training convergence

Figure 3.10: Neural network classification (passengers marked with boxes)

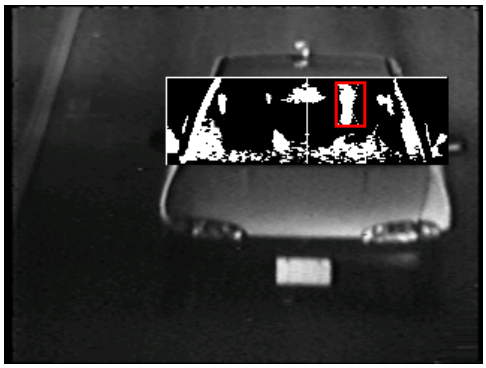
N	number of training epochs	
n	training epoch	
i	input node	
j	hidden node	
k	output node	
C	all neurons in output layer	
x_i	input vector	
$d_k(n)$	output node k (desired value)	
$w_{ij}(n)$	link weight from input i to hidden j	
$w_{jk}(n)$	link weight from hidden j to output k	
$\Delta w_{ij}(n-1)$	momentum of ij weight change	
$\Delta w_{jk}(n-1)$	momentum of jk weight change	
η_{ij}	input to hidden link learning rate	0.10
η_{jk}	hidden to output link learning rate	0.15
α	momentum constant	0.15
$\varphi(\cdot)$	node activation function	$\frac{1}{1+\exp(-x)}$
$\varphi'(\cdot)$	activation function derivative	$\varphi(\cdot)(1-\varphi(\cdot))$
$y_j(n)$	hidden node activation	$\varphi(v_j(n))$
$y_k(n)$	output node activation	$\varphi(v_k(n))$
$\delta_j(n)$	hidden node local gradient	$\varphi'(v_j(n))e_j(n)$
$\delta_k(n)$	output node local gradient	$\varphi'(v_k(n))e_k(n)$
$v_j(n)$	hidden node j input	$\sum w_{ij}(n)x_i(n)$
$v_k(n)$	output node k input	$\sum w_{jk}(n)y_j(n)$
$\mathcal{E}(n)$	instantaneous sum of squares error	$\frac{1}{2}\sum_{j\in C} e_j^2(n)$
$\mathcal{E}_{avg}$	averaged squared error	$\frac{1}{N}\sum_{n=1}^{N}\mathcal{E}(n)$
$e_j(n)$	hidden node j error	$\sum_k \delta_k(n)w_{k,j}(n)$
$e_k(n)$	output node k error	$d_k(n)-y_k(n)$
$\Delta w_{ij}(n)$	w_{ij} correction	$\alpha\Delta w_{ij}(n-1)+\eta_{ij}\delta_j(n)y_i(n)$
$\Delta w_{jk}(n)$	w_{jk} correction	$\alpha\Delta w_{jk}(n-1)+\eta_{jk}\delta_k(n)y_k(n)$

Table 3.1: Neural Network Parameters

(a) Input



(b) Hand-coded classification



(c) Neural network classification

Figure 3.11: Results from both classification methods

BIBLIOGRAPHY

[1] D. H. Ballard and C. M. Brown. *Computer Vision.* Prentice-Hall, 1982.

[2] E. R. Davies. *Machine Vision.* Academic Press, 1990.

[3] J.A. Freeman. *Simulating Neural Networks with Mathematica.* Addison-Wesley, 1994.

[4] J. E. Freund. *Mathematical Statistics.* Prentice-Hall, 2nd edition, 1971.

[5] E. Gose, R. Johnsonbaugh, and S. Jost. *Pattern Recognition and Image Analysis.* Prentice Hall, 1996.

[6] S. Haykin. *Neural Networks.* Macmillan, 1994.

[7] B.K.P. Horn. *Robot Vision*, pages 202–277. The MIT Press, Cambridge, Massachusetts, 1986.

[8] J.A. Jacquez, J. Huss, W. McKeenan, J.M. Dimitroff, and H.F. Kuppenheim. The spectral reflectance of human skin in the region $0.7 - 2.6 \quad \mu m$. Technical Report 189, Army Medical Research Laboratory, Fort Knox, April 1955.

[9] Matrox Electronic Systems. *Genesis Native Library*, 1997.

[10] J. McClave, F. Dietrich II, and T. Sinich. *Statistics.* Prentice-Hall, 1997.

[11] B. Noble and J. W. Daniel. *Applied Linear Algebra.* Prentice-Hall, 2nd edition, 1977.

[12] G. Oehlert. *MacAnova User's Duide.* School of Statistics, Univeristy of Minnesota, 1998.

[13] F.E. Sabins. *Remote Sensing, Principles and Interpretation.* W.H. Freeman and Company, New York, third edition, 1997.

[14] D.H. Sinley. "Laser and Led Eye Hazards: Safety Standards". *Optics and Photonics News*, pages 32–37, September 1997.