

Automatic Detection of Vehicle Passengers Through Near-Infrared Fusion

I. Pavlidis¹ P. Symosek¹ B. Fritz¹ N. Papanikolopoulos² K. Schwartz³

¹Honeywell Technology Center
3660 Technology Dr.
MN65-2500
Minneapolis, MN 55418

²University of Minnesota
Dept. of Computer Science
200 Union St. S.E.
Minneapolis, MN 55455

³MnDOT
Metropolitan Division
1500 W. County Rd. B-2
Roseville, MN 55113

Abstract

We undertook a study to determine if the automatic detection and counting of vehicle passengers is feasible. An automated passenger counting system would greatly facilitate the operation of freeway lanes reserved for buses, car-pools, and emergency vehicles (HOV lanes). In the present paper we report our findings regarding the appropriate sensor phenomenology and arrangement for the task. We propose a novel system based on fusion of near-infrared imaging signals and we demonstrate its adequacy with theoretical and experimental arguments.

1 Introduction

There are compelling reasons for the existence of an automatic passenger counting system in the HOV lane. In particular, such a system will be useful in the following respects:

1. It will facilitate the gathering of statistical data for road construction planning. The gathering of usage statistics in the HOV lane is mandated by the U.S. Federal Highway Administration. These statistics are crucial for construction planning. Currently, the gathering of data is performed manually. This is obviously laborious, inefficient, and prone to error.
2. It will facilitate law enforcement in the HOV lane. Currently, HOV lane enforcement requires substantial commitments of State Highway Patrol personnel and equipment. HOV lane enforcement has other costs as well. These include the risks of high-speed pursuit in lanes adjacent to stop-and-go traffic and the deterioration of traffic flow when tickets are issued during peak commute periods.
3. It will enable the States to offer the option to single drivers to use some HOV lanes for a nominal monthly fee.

A complete HOV monitoring system suitable for the above applications will consist of a passenger detector and a license plate reader. Although, substantial work has been reported to the technical literature regarding license plate readers, work for automated passenger detectors is still in its infancy. There are three major technical

challenges in the development of an automatic passenger detector:

1. The imaging signal should provide a clear picture of the interior of the car. The contrast between the human silhouettes and the background should be sufficient to provide for reliable image processing.
2. The pattern recognition algorithm that performs the passenger detection should exhibit high recognition rates and robust behavior. Of course its performance, depends to a significant degree on the quality of the imaging signal. Even the best pattern recognition algorithm cannot perform reliably when the imaging signal is corrupted with noise.
3. The system architecture should be designed in such a way that will ensure real-time operation, passenger detection in both the front and back seats, and protection from the weather elements.

In this paper we address the first from the above three technical challenges. We describe a novel near-infrared fusion system that provides high quality imaging signal both at day and night and in adverse weather conditions. In particular, in Section 2 we give an overview and justification of our approach. In Section 3 we describe in detail the theoretical computations that support our assertions. In Section 4 we present the experimental validation of our hypotheses. Finally, in Section 5 we conclude the paper and briefly mention our ongoing and future work.

2 Overview

Our research for a solution to the imaging aspect of an HOV system (sensor phenomenology) was guided by the following questions:

1. Is there a band in the EM spectrum that can penetrate through the vehicle's window glass, during day and night and, in adverse weather conditions? Do the objects of interest (vehicle passengers) have a consistent appearance in this EM band, irrespectively of their physical characteristics?

2. If there is more than one band, can we fuse the multiple bands in a meaningful way to increase the detecting power and reliability of the system?
3. Are there appropriate cameras for these bands that have the necessary resolution and speed to live up to the requirements of the problem?

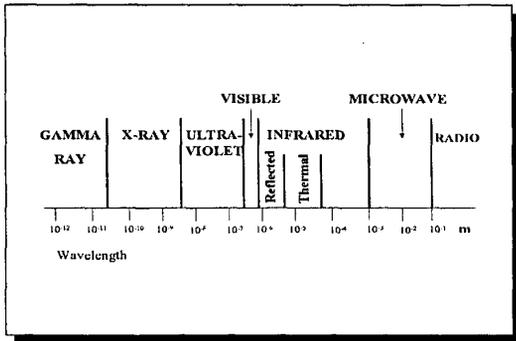


Figure 1: Electro-Magnetic (EM) spectrum.

Fig. 1 shows the Electro-Magnetic spectrum. We have limited our sensor phenomenology investigation into the infrared and visible spectrum regions. Nature constrains our choices below the visible spectrum, since, gamma rays, X-rays and, ultraviolet radiation are harmful to the human body. Therefore, the typically active systems in these ranges cannot be employed in the HOV lane. Technology constrains our choices beyond the infrared region, since millimeter wave and radio wave imaging sensors are very expensive, bulky and, with insufficient resolution [3]. Still, the visible plus the infrared range is a huge area of the EM spectrum and we had to identify narrow bands within this area that are appropriate for the task.

We know from experience as humans that the visible spectrum has certain disadvantages for the purpose of the particular application. A visible range sensor (like the human eye) cannot easily see at night unless is aided by an artificial illumination source. Employing a visible range flashlight to illuminate the passing vehicles is definitely not an option since it will distract the drivers with probably fatal results. Tinted window glass (now common in certain vehicle types) prohibits a clear view of the vehicle's interior to visible range sensors (see Fig. 3). Also, visible range sensors are incapacitated during foul weather conditions. Finally, passengers produce variable patterns in the visible range, depending on their physical characteristics, time of day, and illumination conditions. This variability makes the machine vision task much more difficult.

From the above discussion, it is apparent that only the infrared range held promise for a solution to the problem. Within the infrared range two bands of particular interest are the reflected infrared ($0.7 - 3.0 \mu m$) and

the thermal infrared ($3.0 - 5.0 \mu m$, $8.0 - 14.0 \mu m$) 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. We soon found that the thermal infrared band was difficult to be exploited for HOV purposes because vehicle glass severely attenuates EM radiation beyond $2.4 \mu m$ (see Figs. 2 and 3).

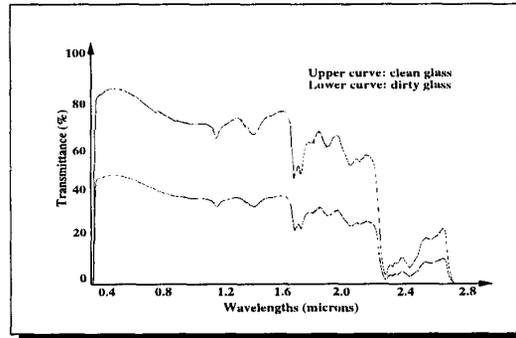


Figure 2: Transmittance of a typical windshield.

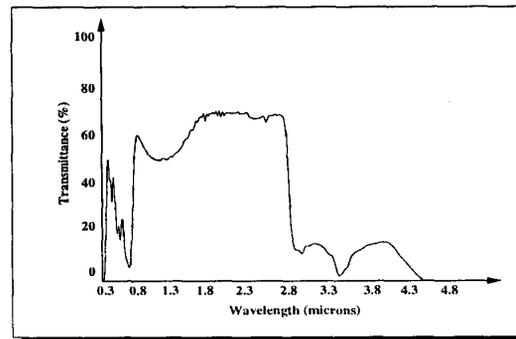


Figure 3: Transmittance of a typical tinted side-window.

Fortunately, a major portion of the reflected-infrared range, the so-called near-infrared range ($0.7 - 2.4 \mu m$), appeared very suitable for the application at hand. In particular, we found that:

1. A camera in this range can safely operate in the HOV lane both day and night. During nighttime we would need a matching near-infrared illumination source to enhance the scene. Provided that the spectral signature of the illumination source is deep into the near-infrared range, the light will be invisible to the human eye. Therefore, no danger to distract the attention of the driver exists.

2. A camera in this range can “see through” both the vehicle’s windshield and its side-windows. The transmittance of typical vehicle windows in the near-infrared spectrum is at least 40% (see Figs. 2 and 3).
3. A camera in this range can operate in adverse weather conditions. For example, it has been established that the near-infrared spectrum is particularly good in penetrating haze.
4. If the near-infrared range is split into two bands around the threshold point of $1.4 \mu\text{m}$, the *lower-band* ($0.7 - 1.4 \mu\text{m}$) and the *upper-band* ($1.4 - 2.4 \mu\text{m}$), then vehicle occupants will produce consistent signatures in the respective imagery. In the upper-band imagery, humans will appear consistently dark irrespectively of their physical characteristics and the illumination conditions. In the lower-band imagery, humans will appear comparatively lighter. This is because human skin appears to have very high reflectance just before $1.4 \mu\text{m}$ but very low reflectance just after $1.4 \mu\text{m}$ (see Fig. 4) [2].

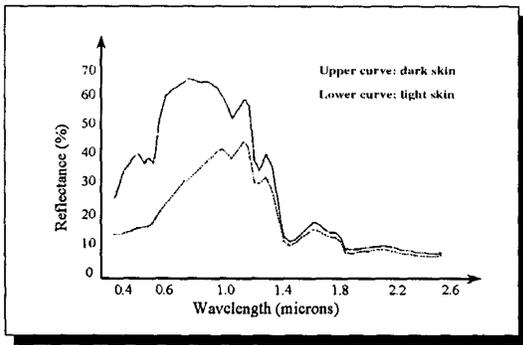


Figure 4: Reflectance of dark skin versus light skin.

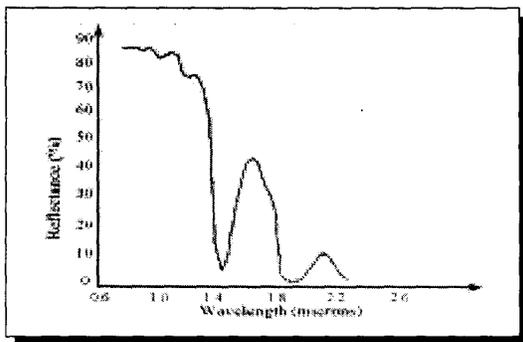


Figure 5: Reflectance of distilled water.

We found that the intriguing phenomenon of the abrupt change in the reflectance of human skin around

$1.4 \mu\text{m}$ is due to the water content of the human body. Water absorbs heavily near-infrared radiation above $1.4 \mu\text{m}$ (see Fig. 5) and thus appears as black body in the respective imagery. Humans consist 70% of water and therefore they exhibit spectral behavior very similar to the water. Interestingly, other inanimate objects in the vehicle scene maintain their reflectance levels almost unchanged, below and above the threshold point $1.4 \mu\text{m}$. For example, see Fig. 6 for the reflectance diagrams of some fabric materials commonly found in the interior of vehicles. This observation provoked the following line of thought: Ideally, everything but the human skin signature should appear proportionally the same in the HOV imagery from the two bands. Therefore, by subtracting an image from the lower-band from its matching co-registered image in the upper-band we can produce a fused image where:

1. The silhouettes of the passengers’ faces will be reinforced (big difference) and more clearly stand out.
2. The background will dim away (small difference).

This increased contrast will facilitate a clean-cut segmentation of the fused image. The thresholded result will be an image where only the face blobs of the passengers remain and everything else is eliminated. A good classifier will always classify fast and accurately such a simple binary pattern, ensuring the reliable real-time operation of the HOV system.

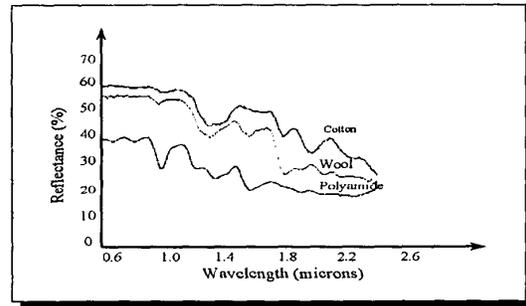


Figure 6: Reflectance of different fabric materials.

The near-infrared camera we found most appropriate to use for testing our ideas was the Sensors Unlimited *SU - 320*. In terms of spectral response it was less than perfect because it didn’t cover the entire spectral range we were interested in ($0.7 - 2.4 \mu\text{m}$). Instead, it covered the subrange $1.1 - 1.4 \mu\text{m}$ for the lower-band and $1.4 - 1.7 \mu\text{m}$ for the upper-band. The question we had to address, given our hypotheses and the particular camera model available, was if the signal to noise (S/N) ratio and the speed of the camera would live up to the task. The complete set of calculations and the interpretation of their results are described in the next section.

3 Theoretical Computations

Our hypotheses, as described in the previous section, held great promise. Before, however, proceed with any experimentation we had to determine if given the particular *SU - 320* camera specifications:

1. We would have had an imaging signal with sufficient *S/N* ratio.
2. The speed of the camera would have been sufficient to capture the vehicle passengers moving at an average speed of *65 mph* (freeway speed).

As we stated earlier, we consider two spectral bands, one above the $1.4 \mu m$ threshold point and one below it. We assume that two *SU - 320* cameras would film simultaneously the same scene. One camera should be equipped with an upper-band filter and one with a lower-band filter. Both cameras should be equipped with a polarizer during day time to reduce solar glare. They should also be equipped with a tele-photo lens. Because of constraints due to the quantum efficiency of the *SU - 320* camera we limit the upper-band in the range $1.4 - 1.7 \mu m$ and the lower-band in the range $1.1 - 1.4 \mu m$. We will demonstrate our *S/N* computation for the lower-band only, since very similar things also apply to the upper-band.

The first step in a *S/N* radiometric computation is to determine the amount of radiation that falls upon the objects of interest [1] - in our case the vehicle passengers. The spectral irradiance of the sun (our illumination source) on a clear day at sea level is approximately $I_{sunny} = 0.008 \text{ Watts/cm}^2$ in the $1.1 - 1.4 \mu m$ wave-band. In our computation, however, we consider the worst case scenario of an overcast day. For an overcast day the irradiance value is reduced by 10^{-3} giving a radiance at the vehicle of approximately

$$\begin{aligned} I_{overcast} &= 10^{-3} * I_{sunny} \\ &= 10^{-3} * 0.008 \\ &= 8 \mu\text{Watts/cm}^2. \end{aligned} \quad (1)$$

We assume that the camera is pointed at the vehicle's windshield, not at a side-window. 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 (see Fig. 2). This results in an irradiance on the vehicle passengers of

$$\begin{aligned} I_{passenger} &= 0.4 * I_{overcast} \\ &= 0.4 * 8 \\ &= 3.2 \mu\text{Watts/cm}^2. \end{aligned} \quad (2)$$

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 the worst case reradiate of 0.4 (see Fig. 4), would be

$$\begin{aligned} R_{passenger} &= 0.4 * I_{passenger} / \pi \\ &= 0.4 * 3.2 \pi \\ &= 0.4 \mu\text{Watts/cm}^2 - \text{ster}. \end{aligned} \quad (3)$$

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° cone angle) with 0.8 transmittance, a polarizer with 0.4 transmittance, and a band-pass filter with 0.6 transmittance. Then, the irradiance at the sensor array of the *SU - 320* camera will be

$$\begin{aligned} I_{camera} &= 0.4 * 0.8 * 0.4 * 0.6 * \pi * \\ &\quad R_{passenger} * \sin^2(14.32^\circ) \\ &= 0.4 * 0.8 * 0.4 * 0.6 * \pi * \\ &\quad 0.4 * \sin^2(14.32^\circ) \\ &= 0.006 \mu\text{Watts/cm}^2. \end{aligned} \quad (4)$$

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

$$\begin{aligned} A &= 37.5 * 10^{-4} * 37.5 * 10^{-4} \\ &= 1.40 * 10^{-5} \text{ cm}^2. \end{aligned} \quad (5)$$

Consequently, the radiant power on the camera pixel will be

$$\begin{aligned} P_{pixel} &= A * I_{camera} \\ &= 1.4 * 10^{-5} * 0.006 \\ &= 0.084 * 10^{-12} \text{ Watts}. \end{aligned} \quad (6)$$

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

$$NEP = (A * \Delta f)^{1/2} / D^*. \quad (7)$$

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 image smear is less than 1 pixel. Assuming a vehicle traveling at a speed of *65 mph*, a distance *d* of 40 *m* apart from the camera, and a field of view of $1.6m$, the 320×240 pixel array of *SU - 320* needs to set its exposure time to 1 *msec* or a bandwidth of $\Delta f = 1 \text{ KHz}$. Substituting the values for *A*, Δf , and D^* in the formula of *NEP* (see Eq. (7)) we get

$$NEP = 1.18 * 10^{-13} \text{ Watts}. \quad (8)$$

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

$$S/N = P_{pixel}/NEP = 0.7. \quad (9)$$

In conclusion, assuming a worst case scenario (overcast day, dirty windshield, dark passenger skin) we determined that the $SU - 320$ camera, equipped with a $f/2$ lens, a $1.1 - 1.4 \mu m$ 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 $1.6 m$ (or height $h = 7 m$ at the specified distance), will achieve:

1. An acceptable smear of less than one pixel because the required exposure time of $1 msec$ is within the camera's speed capabilities.
2. A poor signal to noise ratio $S/N = 0.7$. To boost the S/N ratio to a higher value in overcast days we need to employ an illumination source. This illumination source will also be useful during nighttime. If we operated in the visible spectrum the use of illuminator in the HOV lane would be prohibitive. Fortunately, in our case the spectral signature of the illuminator should match the range $1.1 - 1.7 \mu m$. Since this range is deep into the near-infrared spectrum there is no danger of distracting the driver and can be safely employed in the HOV lane. Contrary to the overcast sky scenario, the S/N ratio is exceptionally good in the case of a clear day in both bands (see Table 1). Therefore, in clear day conditions the system can work passively without any use of artificial light source.

4 Experimental Validation

The above theoretical scenario was put into testing in a Department of Transportation traffic monitoring and research facility. The experiment lasted for a week in late September 1998. One test lane of the freeway, 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 the freeway. 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 Section 3. The only deviation was that we used one $SU - 320$ camera instead of two for budgetary reasons. To simulate the operation of the two-camera system, we equipped the camera first with the lower-band filter. We filmed some carefully arranged scenes. Then, we changed the lower-band filter with the upper-band filter and filmed again

	$1.1 - 1.4 \mu m$	$1.4 - 1.7 \mu m$
Clear Day	700.0	500.0
Overcast Day	0.7	0.5

Table 1: S/N ratios for different day conditions and spectral bands.

the exact same scenes. Later in the lab, we co-registered the imagery from the two bands using a warping transformation. This way, we simulated to a certain degree of accuracy the operation of two co-registered cameras in the lower- and upper-band respectively.

We used two testing cars that made successive passes through the field of view of the near-infrared camera. The passes were done at speed increments of $10 mph$, ranging from $10 - 50 mph$. 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 $1.0 - 2.0 \mu m$ that covered the illumination needs of both the lower- and upper-band camera versions.

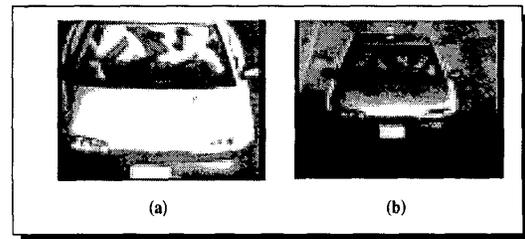


Figure 7: Near-infrared day time results. (a) Image in the band $1.1 - 1.4 \mu m$. (b) Image in the band $1.4 - 1.7 \mu m$.

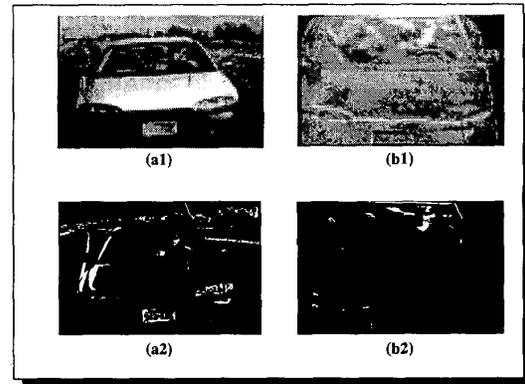


Figure 8: Comparative results between the visible spectrum and the near-infrared fusion approach.

Fig. 7 shows the images from a particular scene in the upper- and lower- bands. The image in the upper-band (Fig. 7(b)) looks in general darker than the image in the lower-band (Fig. 7(a)) because the energy of the sun illumination in the upper-band is less than that of the lower-band. Proportionally, however, the face of the vehicle occupant looks much darker because of the marked difference in reflectance the human skin features between

the two bands. Fig. 8(b1) shows the result of the fusion operation among the images of Fig. 7. Fig. 8(b2) shows the result of the thresholding operation upon the fused image. As someone can observe, only the face of the vehicle occupant has remained and a small amount of noise. The noise would have not been there if we had two cameras accurately co-registered and operating at once. This simple experiment demonstrates the enormous potential of data fusion from the two near-infrared bands we were using. Comparatively, someone can observe in Fig. 8(a2) how much more extraneous information remains after a thresholding operation in the visible range image of Fig. 8(a1).

5 Conclusions and Future Work

We have described a novel method to provide high quality imaging signals to a system that will perform passenger detection and counting in the HOV lane. The method calls for two co-registered near-infrared cameras with spectral sensitivity above (upper-band) and below (lower-band) the $1.4 \mu\text{m}$ threshold point respectively. The quality of the signal remains high even during overcast days and nighttime, because we can safely illuminate the scene with an eye-safe near-infrared illuminator [4]. The near-infrared cameras can also provide clear imaging signals even in certain foul weather situations, like for example in hazy conditions.

The crown jewel of the method is the fusion of the co-registered imaging signals from the lower- and upper-band cameras. Because of an abrupt change in the reflectance of the human skin around $1.4 \mu\text{m}$, the fusion has as a result the intensification of the passenger face silhouettes and the diminution of the background. This increased contrast allows for perfect segmentation that leaves in the final processed image only the face blobs of the passengers. Evidently, such a clean-cut binary imagery will ensure the reliable and fast operation of the pattern classifier that will perform in the future the passenger detection task.

We are currently developing the passenger detection algorithm and results of its performance will be reported in another forum. We are also planning for the future to study the systems aspects of an HOV prototype.

References

- [1] B.K.P. Horn. *Robot Vision*, pages 202–277. The MIT Press, Cambridge, Massachusetts, 1986.
- [2] J.A. Jacquez, J. Huss, W. McKeenan, J.M. Dimitroff, and H.F. Kuppenheim. The spectral reflectance of human skin in the region $0.7 - 2.6 \mu\text{m}$. Technical Report 189, Army Medical Research Laboratory, Fort Knox, April 1955.
- [3] F.E. Sabins. *Remote Sensing, Principles and Interpretation*. W.H. Freeman and Company, New York, third edition, 1997.
- [4] D.H. Sinley. "Laser and Led Eye Hazards: Safety Standards". *Optics and Photonics News*, pages 32–37, September 1997.