A Vehicle Occupant Counting System Based on Near-Infrared Phenomenology and Fuzzy Neural Classification

Ioannis Pavlidis, *Senior Member, IEEE*, Vassilios Morellas, *Member, IEEE*, and Nikolaos Papanikolopoulos, *Member, IEEE*

***Abstract*—We undertook a study to determine if the automatic detection and counting of vehicle occupants is feasible. An automated vehicle occupant 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. We also propose a fuzzy neural network classifier to operate upon the fused near-infrared imagery and perform the occupant detection and counting function. We demonstrate experimentally that the combination of fused near-infrared phenomenology and fuzzy neural classification produces a robust solution to the problem of automatic vehicle occupant counting. We substantiate our argument by providing comparative experimental results for vehicle occupant counters based on visible, single near-infrared, and fused near-infrared bands. Interestingly, our proposed solution can find a more general applicability as the basis for a reliable face detector both indoors and outdoors.**

***Index Terms*—Fuzzy neural network, near-infrared fusion, vehicle occupant detection.**

I. Introduction

THERE are compelling reasons for the existence of an automatic vehicle occupant 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. 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 commuting periods.
3) It will enable the state agencies 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 an occupant detector and a license plate reader. Although substantial work has been reported in the technical literature regarding license plate readers [1]–[3], work for automated vehicle occupant detectors and counters is still in its infancy. There are three major technical challenges in the development of an automatic vehicle occupant detector/counter.

1) The imaging signal should provide a clear picture of the interior of the vehicle. 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 vehicle occupant detection and counting 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 accuracy, real-time operation, and protection from the weather elements.

In earlier publications [4], [5], we reported preliminary results regarding only the first technical challenge (sensor phenomenology). In this paper, we address all three technical challenges. We describe a novel near-infrared fusion system that provides high-quality imaging signal both during the day and at night and in certain adverse weather conditions. Various sensor fusion methods that increase the quality of the imaging signal and boost system performance have been reported in the literature [6]–[8]. The novelty of our sensor fusion method lies in the exploitation of the unique reflectance properties of the human skin in the near-infrared. In particular, in Section II we give an overview and justification of our imaging approach. Then, we describe in detail the theoretical computations that support our imaging assertions. We also present the experimental validation of our imaging hypotheses. In Section III we outline

Manuscript received March 6, 2000; revised August 29, 2000. This work was supported by the Minnesota Department of Transportation under Contract Q5216211101.The Guest Editor for this paper was Dr. Katsushi Ikeuchi.

I. Pavlidis and V. Morellas are with the Honeywell Technology Center, Minneapolis, MN 55418 USA (e-mail: ioannis.pavlidis@honeywell.com; vassilios.morellas@honeywell.com).

N. Papanikolopoulos is with the Department of Computer Science, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: npapas@cs.umn.edu).

Publisher Item Identifier S 1524-9050(00)10229-7.

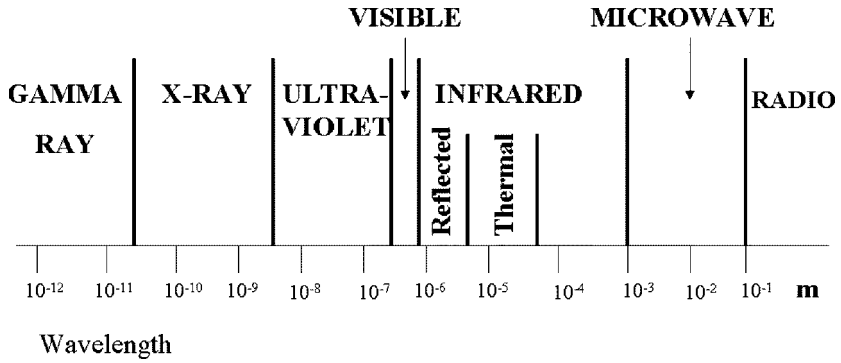


Fig. 1. The electromagnetic (EM) spectrum.

the fuzzy neural network algorithm for occupant detection and counting and provide test results of its performance on single band and fused near-infrared imagery as well as on visible band imagery. In Section IV, we describe the system architecture of our prototype vehicle occupant counter. Finally, in Section V, we conclude the paper and briefly mention our ongoing and future work.

II. The Imaging Problem

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 electromagnetic (EM) spectrum that can penetrate through the vehicle's window glass, during day and at night and in adverse weather conditions? Do the objects of interest (vehicle occupants) 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?

Fig. 1 shows the EM spectrum. We have limited our sensor phenomenology investigation into the infrared and visible spectrum regions. Nature constraints 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 [9]. 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 this particular application. A visible range sensor (like the human eye) cannot easily see at night unless it 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

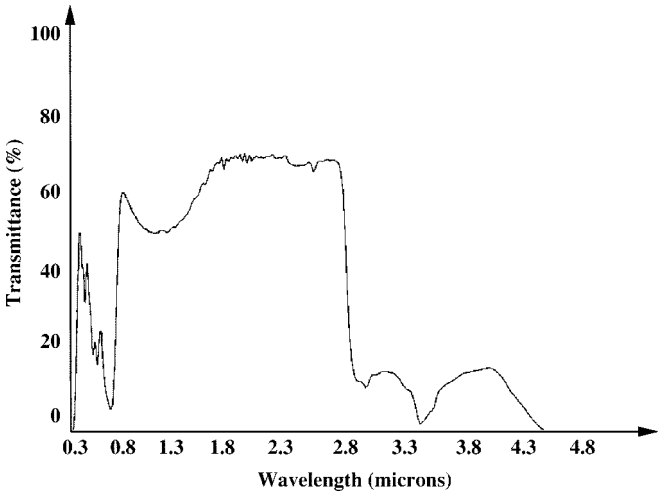


Fig. 2. Transmittance of a typical tinted vehicle window. Measurements were carried out with a Varian Cary 2400 Spectrophotometer.

(now common in certain vehicle types) prohibits a clear view of the vehicle's interior to visible range sensors (see Fig. 2). Also, visible range sensors are incapacitated during foul weather conditions. Finally, vehicle occupants 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 holds promise for a solution to the problem. Within the infrared range two bands of particular interest are the reflected-infrared (0.7–3.0 μm) and the thermal-infrared (3.0–5.0 μm and 8.0–14.0 μm) 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. The opacity of far-infrared (8.0–14.0 μm) is well documented in the literature [10]. We soon found that the mid-infrared (3.0–5.0 μm) was also difficult to be exploited for HOV purposes because vehicle glass severely attenuates EM radiation beyond 2.4 μm (see Figs. 2 and 3).

Fortunately, a major portion of the reflected-infrared range, the so-called near-infrared range (0.7–2.4 μm), appeared very suitable for the application at hand. In particular, we found the following.

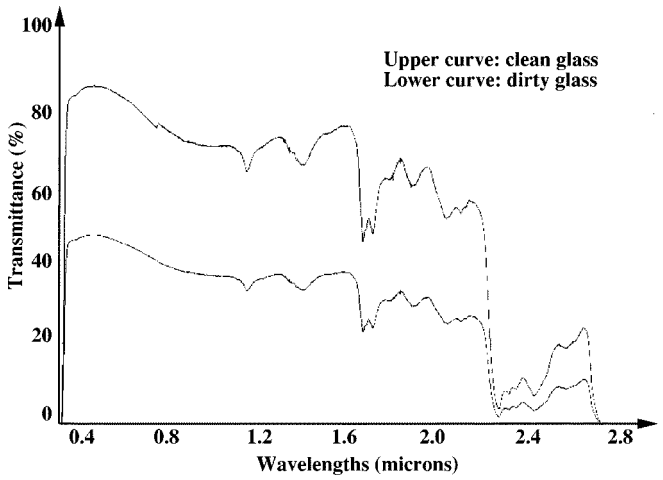


Fig. 3. Transmittance of a typical nontinted vehicle window. Measurements were carried out with a Varian Cary 2400 Spectrophotometer.

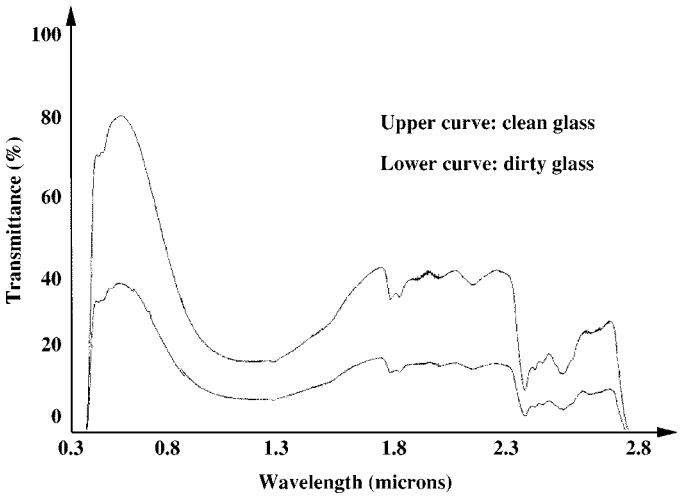
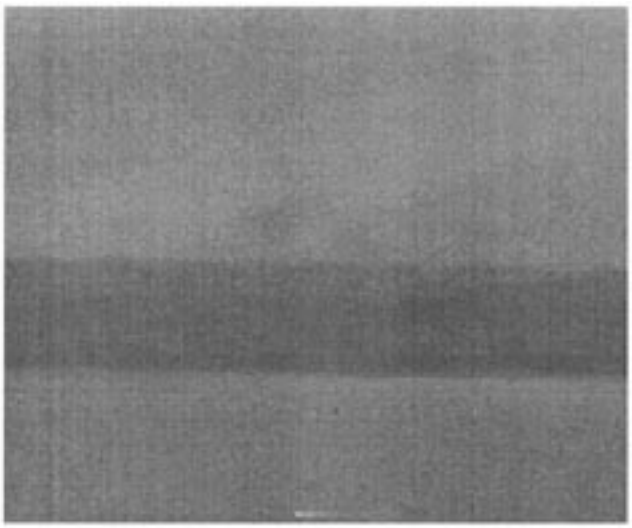


Fig. 4. Transmittance of an EZKOOL class of vehicle window. Measurements were carried out with a Varian Cary 2400 Spectrophotometer.

1) A camera in this range can safely operate in the HOV lane both during day and at night. In low-light conditions (nighttime, overcast skies) 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" the vehicle's windows. The transmittance of typical vehicle windows in the near-infrared spectrum is at least 40% (see Figs. 2 and 3). Transmittance remains high across the near-infrared band and, therefore, only a portion of the band could provide sufficient energy for the operation of an HOV system. There is an exotic category of vehicle windows, however, the so-called EZKOOL class that attenuates near-infrared illumination unevenly and more severely (see Fig. 4). Currently, only a very small number of luxury cars, like the Oldsmobile Aurora, feature EZKOOL windows. To image effectively the interior of these vehicles the full length of the near-infrared spectrum should be employed.
3) A camera in this range can operate in certain adverse weather conditions. For example, it has been established that the near-infrared spectrum is particularly good in penetrating haze (see Fig. 5). The explanation for this phenomenon lies in the size of the droplets in haze, which is smaller than the near-infrared wavelengths. This property is particularly useful in metropolitan areas where haze conditions are endemic (e.g., San Francisco).
4) If the near-infrared range is split into two bands around the threshold point of 1.4 μm, the *lower band* (0.7–1.4 μm) and the *upper band* (1.4–2.4 μm), 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



(a)

(b)

Fig. 5. Natural scene bathed in haze. (a) Image captured with a visible band camera. (b) Image captured with a near-infrared camera (Sensors Unlimited *SU-320*).

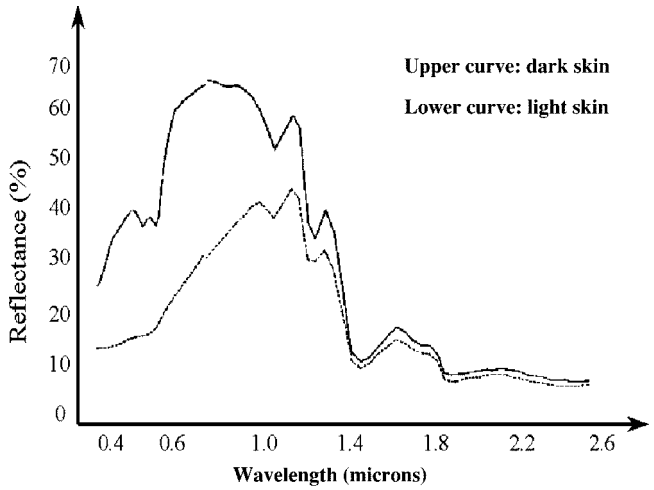


Fig. 6. Reflectance of dark skin versus light skin. Measurements were carried out with a Varian Cary 2400 Spectrophotometer.

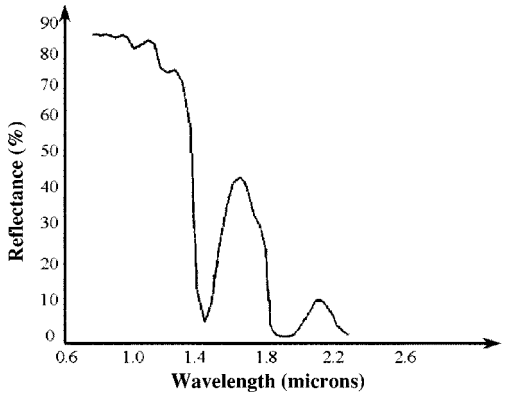


Fig. 7. Reflectance of distilled water. Measurements were carried out with a Varian Cary 2400 Spectrophotometer.

reflectance just before 1.4 μm but very low reflectance just after 1.4 μm (see Fig. 6) [11], [12].

We found that the intriguing phenomenon of the abrupt change in the reflectance of human skin around 1.4 μm is due to the water content of the human body. Water absorbs heavily near-infrared radiation above 1.4 μm (see Fig. 7) 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 μm [13]–[15]. For example, see Fig. 8 for the reflectance diagrams of some fabric materials commonly found in the interior of vehicles. Fig. 9 is especially interesting because it depicts the reflectance diagram of a special material: leather upholstery. The reflectance of leather upholstery also remains stable in the near-infrared despite its superficial affinity to animal skin. It is not the skin per se that produces the singular reflectance behavior but what is hidden below the skin, that is, the water content of the human body. 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

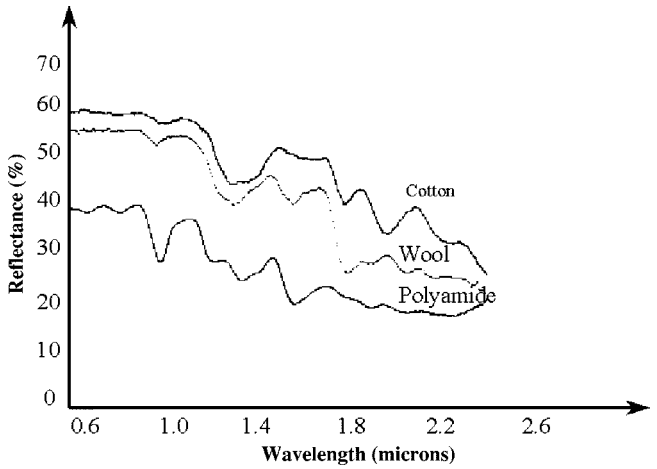


Fig. 8. Reflectance of different fabric materials. Measurements were carried out with a Varian Carry 2400 Spectrophotometer.

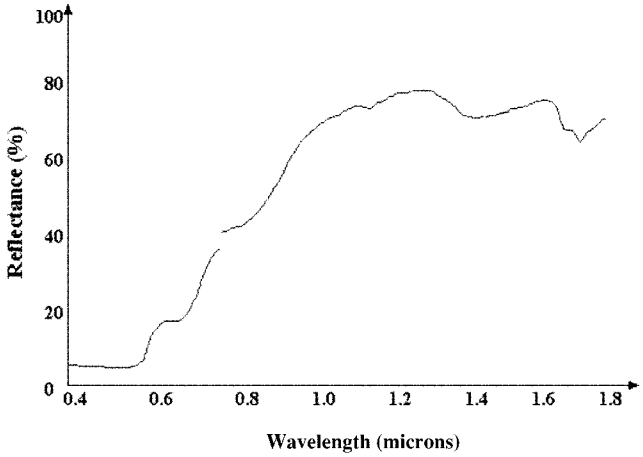


Fig. 9. Reflectance of typical leather upholstery. Measurements were carried out with a Varian Carry 2400 Spectrophotometer.

an image from the lower band from its matching coregistered image in the upper band we can produce a fused image where:

1) the silhouettes of the vehicle occupants' 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 vehicle occupants 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.

The near-infrared camera we found most appropriate to use for testing our ideas was the Sensors Unlimited *SU-320* ($25 000 at 1999 prices). In terms of spectral response it was less than perfect because it did not cover the entire spectral band we were interested in (0.7–2.4 μm). Instead, it covered the subband 1.1–1.4 μm for the lower band and 1.4–1.7 μm for the upper band. These subbands are sufficient for all but the EZKOOL class of vehicle windows. An alternative camera solution, the *SYS256RM* by Santa Barbara Focalplane, that covers the entire near-infrared band costs three times as much ($75 000 at 1999 prices). Because the current number of

vehicles featuring EZKOOL glass is very small, we opted for a significantly more economical solution at the expense of some accuracy. Our choice was influenced by practical budgetary constrains of Transportation Departments for this kind of applications. The imaging 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 subsection.

A. Theoretical Computations

Our hypotheses, as described in the previous section, held great promise. Nevertheless, before we could 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 mi/h (freeway speed).

As we stated earlier, we consider two spectral bands, one above the 1.4-μm threshold point and one below it. We assume that two *SU-320* cameras would film simultaneously the same scene. One camera would be equipped with an upper band filter and one with a lower band filter. Both cameras would be equipped with a polarizer during daytime to reduce solar glare. They would also be equipped with a tele-photo lens. Because of constrains due to the quantum efficiency of the *SU-320* camera we limit the upped band to the range 1.4–1.7 μm and the lower band to the range 1.1–1.4 μm. As we will prove, these truncated near-infrared ranges allow the acquisition of usable imagery for vehicle windshields that feature at least 40% transmittance (dirty, non-EZKOOL windows). We will demonstrate our S/N computation for the lower band only, since very similar results also apply to the upper band.

We assume that the camera is pointed at the vehicle's windshield, not at the side-window. For all practical purposes, the Department of Transportation is primarily interested if passing vehicles carry at least one more person ("the passenger") in addition to the driver. The passenger usually sits in the front of the vehicle. The radiant power on the camera pixel is given from the following relation:

$$\begin{aligned} P_{\text{pixel}} &= A * I_{\text{camera}} \\ &= 0.084 * 10^{-12}\,\text{W} \end{aligned} \tag{1}$$

where A is the area of the *SU-320* Focal Plane Array (FPA) and I_{camera} is the irradiance at the camera's FPA. The value of I_{camera} can be computed from the radiometric factors listed in Table I. The factor values were chosen to reflect a typical worst case scenario.

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

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

TABLE I
RELEVANT RADIOMETRIC FACTORS

Sun Irradiance (Overcast)	$8\ \mu W/cm^2$
Windshield Trasmittance (Dirty)	40%
Camera Lens	$\frac{f}{2}$
Lens Transmittance	40%
Polarizer Transmittance	40%
Band-Pass Filter Transmittance	40%
Focal Plane Array Area	$1.40 * 10^{-5}\ cm^2$

We already know the values of D^* and A. In order to compute the NEP we need to also know the value of Δf. The bandwidth Δf is determined by the exposure time (speed) of the camera. In turn, the exposure time depends on the vehicle speed (v), the camera's Instantaneous Field Of View (IFOV), the range (r), and the footprint of the horizontal translation (f_{ht}). Based on the parameters u, IFOV, r, and f_{ht} we compute the required exposure time (speed) of the camera such that the image smear is less than 1 pixel. Then, we check if the exposure time value falls within the operational range of the *SU-320* camera. If it does, the *SU-320* camera is adequate for the HOV task in terms of speed. We can substitute the corresponding value for the bandwidth Δf in (2) and continue the process of computing the S/N ratio.

Fig. 10 shows the configuration of the *SU-320* camera relative to the oncoming traffic. The camera is located $c_{\text{ground}} = 3.6$ m above the ground, $c_{\text{freeway}} = 7.5$ m off the edge of the freeway, and at a distance of $r = 40$ m from the oncoming traffic. This arrangement ensures that the camera is located in a safe place and has the appropriate field of view. We assume that the camera focuses at the centerline of the incoming vehicle, at the level of the occupants' faces ($p_{\text{ground}} = 1.2$ m). The half width of a standard freeway lane is $w_{\text{lane}} = 1.8$ m. We assume that the vehicle travels in the middle of the freeway lane. Therefore, the lateral distance of the vehicle's centerline from the camera is

$$c_{\text{vehicle}} = c_{\text{freeway}} + w_{\text{lane}} = 7.5 + 1.8 = 9.3\ \text{m}. \tag{3}$$

Finally, we assume that, for a typical vehicle's windshield, the average width and height are $w_{\text{win}} = 1.5$ m and $h_{\text{win}} = 0.9$ m, respectively.

The IFOV is the camera's field of view with respect to a single pixel (see Fig. 11). We assume that the distance r' is approximately equal to the camera's range r ($r' \approx r = 40$ m). Then, the IFOV can be computed from the following equation:

$$\begin{aligned} \text{IFOV} &= \frac{\arctan\left[(h_{\text{win}}/2)/r'\right]}{(h_{\text{FPA}}/2)} \\ &= 0.0001\ \text{rad} \end{aligned} \tag{4}$$

where $h_{\text{FPA}} = 240$ *pixels* is the vertical dimension of the *SU-320* Focal Plane Array (FPA).

At time t, the camera's IFOV sees a small portion of the occupant's face of diameter D. This small face area is what is imaged into a single pixel. We can determine the value of D from the following equation:

$$\begin{aligned} D &\approx \text{IFOV} * r \\ &= 0.004. \end{aligned} \tag{5}$$

The angle θ in Fig. 11 is the angle between a horizontal plane and the optical axis of the camera. Because we have assumed

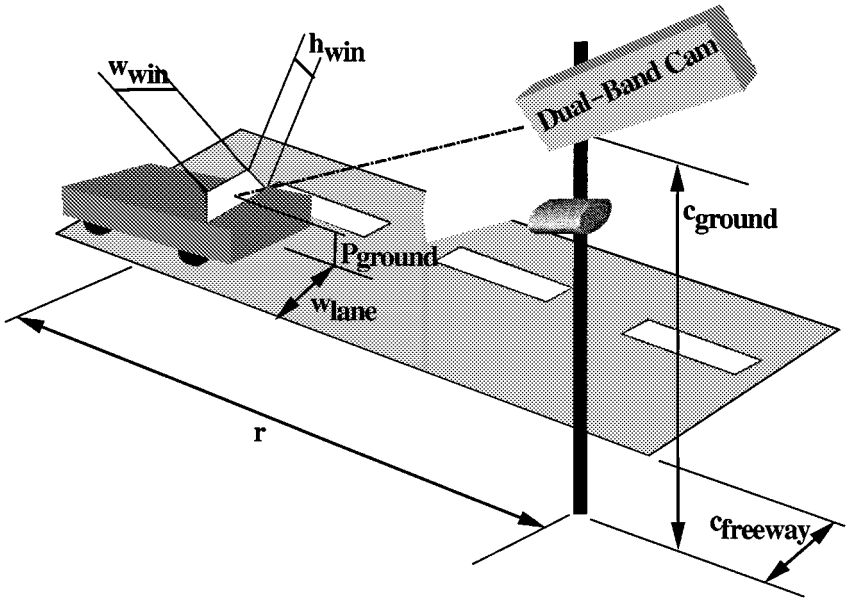


Fig. 10. Configuration of the camera set.

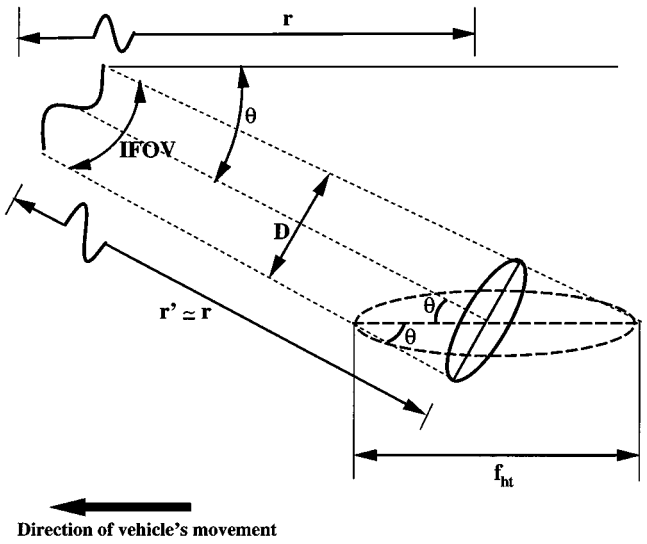


Fig. 11. Geometry for the computation of the footprint of a single pixel in a horizontal plane.

that the camera is focused at the level of the occupants' faces (see the geometry in Fig. 10), the angle θ is

$$\theta = \arctan\left[\frac{c_{\text{ground}} - p_{\text{ground}}}{r'}\right] = 3.43^{\circ}. \quad (6)$$

Since we know the values for D and θ, from the geometry of Fig. 11 we can compute the footprint f_{ht} of a single pixel's horizontal translation

$$f_{ht} = D/\sin(\theta) = 0.067 \text{ m}. \quad (7)$$

We assume that the vehicle occupants travel at the nominal freeway speed of $v = 65$ mi/h or $v = 29.3$ m/s. At this freeway speed the footprint f_{ht} is covered during time t_f

$$t_f = f_{ht}/v = 2.28 \text{ ms}. \quad (8)$$

Therefore, the exposure time t_{exposure} of the camera should be $t_{\text{exposure}} < 2.28$ ms if we would like to have image smear of not more than 1 pixel. The operational range of the *SU-320* camera in terms of exposure time is 127 μs–16.3 ms. Therefore, the required exposure time of $t_{\text{exposure}} < 2.28$ ms is within the camera's operational range or, in other words, the speed of the *SU-320* is up to the HOV task. We choose to set the exposure time of the camera to 1 ms ($t_{\text{exposure}} = 1$ ms < 2.28 ms) which corresponds to a bandwidth of $\Delta f = 1$ kHz.

Now, that we have addressed the speed question and we know the value of Δf, we can substitute the values for A, Δf, and D^* in (2) and calculate the NEP

$$\text{NEP} = 1.18 * 10^{-13} \text{ W}. \quad (9)$$

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

$$S/N = P_{\text{pixel}}/\text{NEP} = 0.7. \quad (10)$$

In conclusion, assuming a typical worst case scenario (overcast day, dirty windshield) we determined that the *SU-320* camera, equipped with an *f/2* lens, a 1.1–1.4-μm filter, and a polarizer, if it is positioned at a distance of $r = 40$ m from the incoming vehicle and at a height of $c_{\text{ground}} = 3.6$ m above the ground, will achieve the following.

1) An acceptable smear of less than one pixel because the required exposure time of 1 ms 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 on overcast days we need to

TABLE II
S/N RATIOS FOR DIFFERENT DAY CONDITIONS AND SPECTRAL BANDS

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

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 μm. Since this range is deep into the near-infrared spectrum there is no danger of distracting the driver and the illuminator 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 II). Therefore, in clear day conditions the system can work passively without any use of artificial light source.

B. Experimental Validation

Based on the above theoretical scenario we designed and implemented a prototype HOV counting system in one of the Department of Transportation's traffic monitoring and research facilities (Mn/Road) in Minneapolis. During experimentation we found that lighting conditions change continuously even in relatively stable weather conditions. Consequently, S/N ratios fluctuate between the extreme values listed in Table II all the time. We also found that S/N ratios above 250 are essential to the flawless operation of the system. We addressed this problem by outfitting the HOV counter with a sophisticated light management system that senses the near-infrared light levels in the scene and automatically adjusts the power levels of the artificial illumination sources, so that a minimum $S/N > 250$ is maintained at all times. Establishing the minimum acceptable illumination levels is very important. It allows us not to overpower the scene with excessive illumination and either thwart the image or cause harm to the eyes of people that for some strange reason look toward the illuminator for prolonged periods of time. Although, near-infrared light is not sensed by the human eye, it still enters the retina and its raw energy should be regulated. Also, maintaining relatively steady scene illumination levels $(250 < S/N < 500)$ by continuously adjusting the illumination source, simplifies the image processing task in general.

An additional implementation challenge was presented by the theoretical requirement for perfect coregistration between the lower and upper band cameras. This coregistration is essential for the performance of the image fusion (weighted subtraction). We solved this problem by providing an optical signal splitter between the two cameras. The optical splitter functions simultaneously as a bandpass optical filter funneling the lower band light signal of the scene into one camera and the upper band into the other. More details about the architecture of the HOV counter are provided in Section IV.

The prototype HOV counting system became fully functional in February 2000 and it has undergne regular testing since then. Fig. 12(a) and (b) shows the images from a particular scene in the upper and lower bands. The scene is quite interesting because it features as a passenger a mannequin and not a real human. One can observe the significant brightness difference on the face of the driver between the two images. In contrast, the face of the mannequin and all the other inanimate objects in the scene maintain about the same brightness levels in both bands.

The image in the upper band is subtracted (fused) from the image in the lower band and produces the image in Fig. 12(c). The weighting factor (coefficient) of the subtraction is determined on-line based on the relative readings of the HOV photometers in the upper and lower band. Because the illumination source of the HOV system continuously adjusts its power to maintain relatively steady and even illumination levels in the scene, the subtraction coefficient is usually close to 1 and rarely takes values above 2. The fused image is cropped to the approximate area of the windshield only. This area is computed based on the camera geometry and the information provided by a radar sensor regarding the position and speed of the incoming vehicle. The cropping ensures that the image subtraction operation applies only to objects found in the interior of vehicles (such as upholstery, faces, and clothes). We have studied the reflectance properties of such objects well and we anticipate excellent behavior for our dual-band imaging method.

The image in Fig. 12(c) demonstrates clearly the increase of the brightness levels in the face of the driver (real human) and the diminution of the mannequin's head and the rest of the background. This allows for nearly perfect segmentation of the driver's face in Fig. 12(d). The thresholded image has been normalized in terms of position (face blob is aligned to top row) and size to facilitate subsequent processing by a pattern recognition algorithm. The mannequin has been eliminated altogether since it is an inanimate object with no water content in its body similar to that of a human being. Therefore, like the other inanimate objects in the image the mannequin does not change abruptly its reflectivity around 1.4 μm. One should also notice the glimpse of the driver's hands in the final thresholded image. In general, in this final stage, live, uncovered human skin is the only object that appears in the image. Fig. 12(d) demonstrates the primary advantage that our imaging method provides. It seals away the tremendous variability that would otherwise be introduced because of light changes and skin color. Instead, it provides the pattern recognition algorithm with a simple and consistent binary pattern featuring a face blob for each vehicle occupant.

The images in Fig. 13 show typical scenes in the visible band during day- and nighttime. Although, the image quality is rather fair during daytime (provided the vehicle's window is not tinted), it is unacceptable during nighttime where the S/N ratio is almost zero. The worst is that this deficiency cannot be rectified since the employment of a visible illuminator would cause safety hazards.

III. VEHICLE OCCUPANT DETECTION AND COUNTING ALGORITHM

After having solved the imaging problem we concentrated our attention to the algorithmic aspect of the HOV system. We

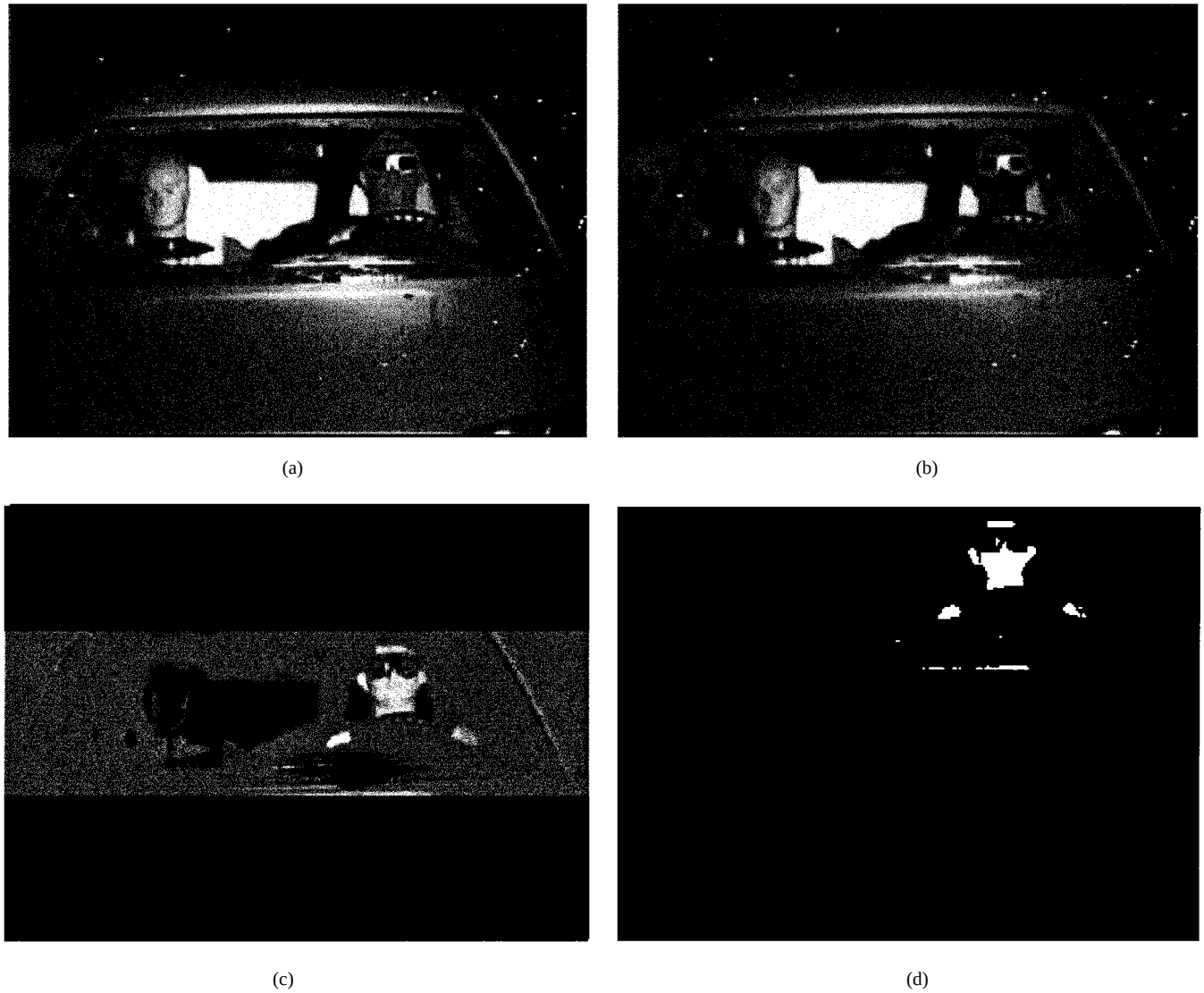
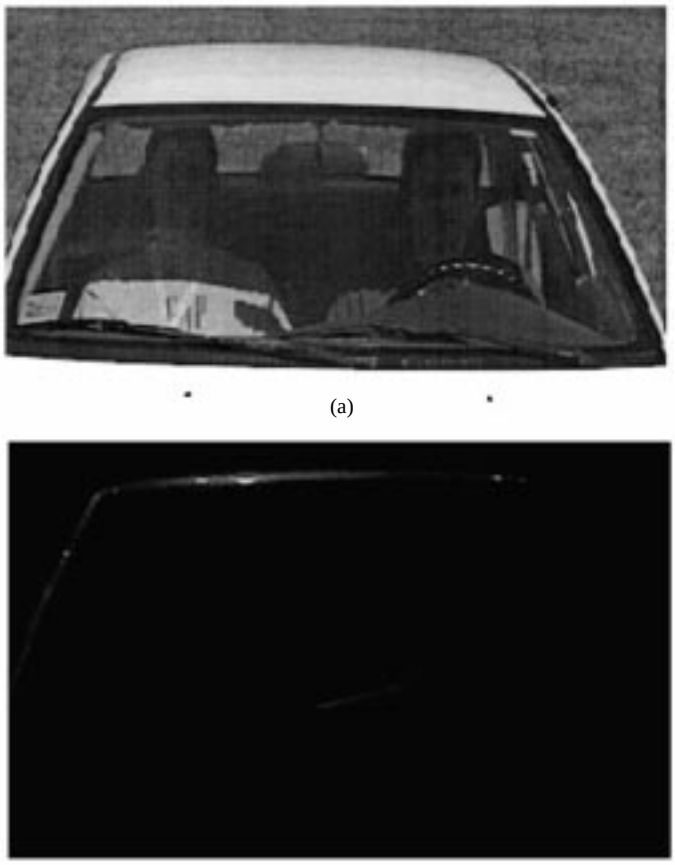


Fig. 12. Scene with a driver and a mannequin. (a) Lower band image. (b) Upper band image. (c) Weighted difference (fusion) between images (a) and (b). (d) Thresholded outcome of image (c).

chose to perform the counting of vehicle occupants with a neural network. During neural network operation, output neurons (see Fig. 14) are assigned symbolic meaning by encoding classes of images. In our case, each class corresponds to a different number of vehicle occupants. 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 operation of the neural network. In self-organized networks there are *no distinct training and performance phases*. Instead, a certain metric (i.e., fuzzy metric for Fuzzy ART networks) is used for measuring similarity of inputs in the feature space and a learning procedure enables the clustering of inputs into classes. Therefore, in contrast to supervised learning networks (i.e., back-propagation), Fuzzy ART networks do not need external guidance for training on specific input sets. This translates into 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 in the past. This is the so-called *stability–plasticity* dilemma. The ART network features a feedback mechanism between the layers that helps solve the stability–plasticity 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.
3) *Broad and Narrow Classification.* ART networks have an explicit parameter called *vigilance* that controls their 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



(a)

(b)

Fig. 13. Visible band images. (a) Daytime. (b) Nighttime.

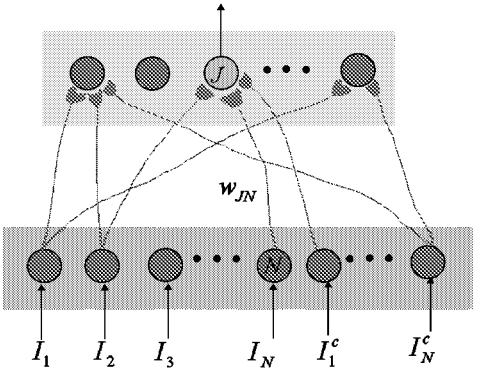


Fig. 14. ART networks are two-layer neural modules. There exists a complete set of bottom up weights from the input layer (dark box) neurons to the output layer (light box) neurons. The size of the adaptive weights, which change through learning, is graphically denoted by the different size of the blobs that surround the output neurons. The light colored output neuron J is the category selected for the present input.

problem of disambiguating overlapping categories with minimum risk.

A. The Fuzzy Neural Network Algorithm

Fuzzy ART neural networks are comprised of an input layer F_0 and an output layer F_1 [16]. The typical structure of an ART neural module is shown in Fig. 14. The input layer consists of N nodes (neurons) which encode the input vector $\vec{I} = (I_1, I_2, \ldots, I_N)$. In our application each input node represents the gray level intensity of a pixel

$$I_j \in [0, 1], \quad \forall j \in (1, 2, \ldots, N). \tag{11}$$

Specifically, $N = 320 \times 240$ and represents the number of pixels in a standard subimage we are extracting from each frame. The subimage is centered on the window area of the incoming vehicle. The vehicle window is located based on the camera geometry, information regarding the speed and position of the incoming vehicle provided by the HOV radar, and the Hough transformation. The input vector $\vec{I}$ is augmented to achieve input normalization through a process that is called complement coding. The complement coded input vector $\vec{P}$ becomes a $2N$-dimensional vector

$$\vec{P} = \left(\vec{I}, \vec{I}^c\right) \equiv (I_1, \ldots, I_N, I_1^c, \ldots, I_N^c) \tag{12}$$

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

$$|\vec{P}| = \left|\left(\vec{I}, \vec{I}_c\right)\right| = \sum_{i=1}^{N} I_i + \left(N - \sum_{i=1}^{N} I_i\right) = N. \tag{13}$$

The M nodes in the output layer represent the classification categories. In our application we have two image classes: the one-occupant class (single driver) and the two-occupant class (driver and front passenger). To have these classes established we initially present to the network one representative image of each class. The presentation order is important *only* for the assignment of symbolic meanings to the output neurons. The first image we present depicts a single occupant and for that reason the leftmost output neuron symbolizes the single-occupant class. The second image we present depicts two occupants and, consequently, the next output neuron symbolizes the two-occupant class. If the initial presentation order was reversed so would the meaning of the output neurons. If during network operation, output neurons besides the first two are activated, that means that the neural network mistakenly did not classify the incoming image as one of the only two possible cases. Instead, it started forming new unknown categories (clusters) where it assigns the misclassified patterns.

Each output neuron j is associated with a vector $\vec{w}_j = (w_{j1}, w_{j2}, \ldots, w_{j2N})$ of 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 $\vec{P}$, a choice function T_j is computed for every output neuron as

$$T_j(\vec{P}) = \frac{\left|\vec{P} \wedge \vec{w}_j\right|}{|\vec{w}_j|} \tag{14}$$

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

The choice function measures the degree to which the weight vector $\vec{w}_j$ is a fuzzy subset of the input $\vec{P}$. There is only one neuron that is activated for a particular input (320×240 image) that is presented in the input layer. In other words, 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

$$T_J(\vec{P}) = \max\{T_j | j = 1, \ldots, M\}. \tag{15}$$

Then, as a final step, the chosen candidate neuron J classifies correctly the present input if it meets the *vigilance* criterion. The vigilance criterion is mathematically described by the following equation:

$$\frac{|\vec{P} \wedge \vec{w}_J|}{|\vec{P}|} > \rho \tag{16}$$

where ρ is the vigilance parameter. If (16) is met we say that *resonance* occurs. Hence, resonance occurs when the degree to which the input $\vec{P}$ is a fuzzy subset of $\vec{w}_J$ 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 $-1 (T_J(\vec{P}) = -1)$ 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 $\vec{P}$. For this selected output neuron J learning occurs as follows:

$$\vec{w}_J^{(\text{new})} = (1 - \beta)\vec{w}_J^{(\text{old})} + \beta\left(\vec{P} \wedge \vec{w}_J^{(\text{old})}\right) \tag{17}$$

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

B. Geometric Representation of the Fuzzy Neural Network Classification

There is an interesting geometric interpretation of the category formation process when fuzzy-ART networks are employed at the fast learning mode ($\beta = 1$). In order to make our point clear, we will assume that our inputs represent two–dimensional (2-D) vectors instead of the 320×240-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. 15). When an output node is chosen for the first time we say that the neuron commits to a new class. For example, by presenting to the network an image of one occupant as the very first image, the leftmost output node is committed to the "one occupant" 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 throughout classification. The maximum size of the rectangles (represented by its perimeter) is determined by the vigilance parameter. In a similar fashion other classes beyond the initial two are formed during classification if the hyperspace points fall outside the greatest hyperrectanagle determined by the vigilance parameter. One can see that classes (grey-level-coded rectangles) may overlap due to the fact that fuzzy concepts are incorporated into the neural network.

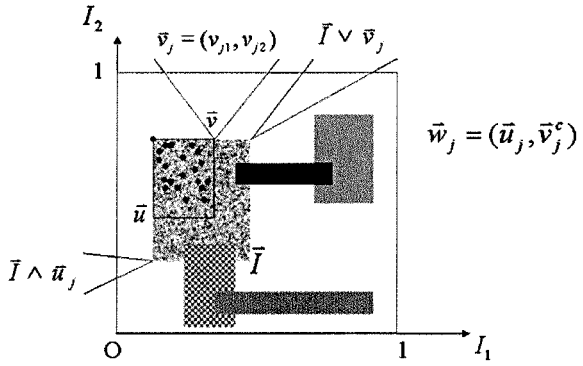


Fig. 15. Classes in fuzzy-ART networks are represented as color-coded rectangles. Inputs that fall within a particular rectangle are classified by the output neuron associated with the respective class.

C. Performance of the Algorithm

The neural network described above was tested comparatively in four different experiments. Each experiment included imagery from a particular EM band or processed category, that is:

1) experiment with visible spectrum imagery;
2) experiment with lower band near-infrared imagery;
3) experiment with upper band near-infrared imagery;
4) experiment with thresholded imagery.

For each experiment we used 100 images. The corresponding images of the three EM bands (visible, upper near-infrared, and lower near-infrared) were captured simultaneously by our prototype HOV system installed in the Mn/Road experimental facility. The images of the visible band were acquired with a professional grade digital camera (SONY DSR-200), which is part of the HOV prototype system. The visible band camera is used for comparative evaluation purposes only and is not a critical part of the HOV system. The upper and lower band near-infrared images were captured with the dual-band *SU-320* camera apparatus. The images are accurately coregistered since they were acquired through an optical signal splitter/filter that splits the same scene information into two bands and funnels it to the corresponding camera FPAs. The thresholded images were produced from the corresponding lower and upper band images after fusion and thresholding. The fusion (subtraction) coefficient was determined on-line at each case based on the instantaneous readings from the upper and lower band photometers in the scene.

The 100 4-tuple image sets (upper band, lower band, fused, and thresholded) were selected randomly among thousands of archived sets. They represent typical scenes during day and night over a period of four months (February–May 2000). They also represent various weather conditions ranging from overcast skies to clear days. Some scenes were shot during light rain. No scenes exist with heavy rain or snow due to the surprisingly mild winter and spring of 2000 in Minnesota. Nevertheless, our theoretical prediction is that the performance of the system will degrade in downpour conditions.

TABLE III
CONFUSION MATRIX FOR THE VISIBLE BAND EXPERIMENT

		Classes		
		$1P$	$2P$	Other
Images	$1P$	10	10	40
	$2P$	10	10	20
	Other	0	0	0

TABLE IV
CONFUSION MATRIX FOR THE LOWER NEAR-INFRARED EXPERIMENT

		Classes		
		$1P$	$2P$	Other
Images	$1P$	20	30	10
	$2P$	0	10	30
	Other	0	0	0

TABLE V
CONFUSION MATRIX FOR THE UPPER NEAR-INFRARED EXPERIMENT

		Classes		
		$1P$	$2P$	Other
Images	$1P$	20	30	10
	$2P$	0	20	20
	Other	0	0	0

TABLE VI
CONFUSION MATRIX FOR THE THRESHOLDED EXPERIMENT

		Classes		
		$1P$	$2P$	Other
Images	$1P$	60	0	0
	$2P$	0	40	0
	Other	0	0	0

For each experiment, we selected one image with a single vehicle occupant and one with two occupants as the initial input set. The classification results for the four experiments are shown in the respective confusion matrices (see Tables III–VI). In all these tables $1P$ stands for a single vehicle occupant and $2P$ stands for two vehicle occupants.

The lowest correct classification performance (20%) is scored in the case of the visible band imagery. This is rather expected since the network cannot classify all the nighttime images in this band, which account for almost half of the total image population. This fact also reflects to the large number of images (60%) classified in the inconclusive category "Other."

The second worst performance (30% correct classification) is scored in the case of the lower band experiment. The relatively improved performance in comparison with the visible band case owes to the employment of near-infrared illumination. Nevertheless, variability is still high, which keeps the overall recognition score in low levels.

In the case of the upper band experiment the correct classification score is improved even further (40%). The reflectance of the human skin in the upper band is more stable comparatively to the lower band. Also, there is a sharper and more consistent contrast between the human skin that reflects almost nothing and the other objects in the scene that usually feature significant reflectance.

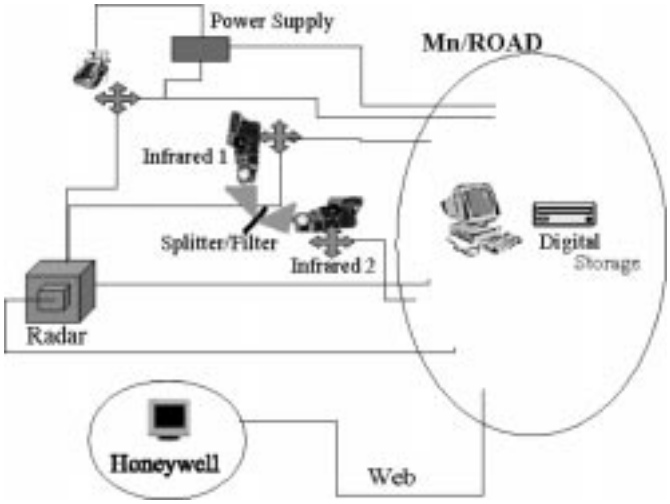


Fig. 16. Architecture of the prototype HOV counting system.

Finally, in the experiment with the thresholded images that were produced through dual-band fusion and thresholding, the score is perfect (100%). This is the experimental confirmation for the superiority of our proposed method. It seals away all the variability that affects the performance of the other three methods. Instead, it provides the neural network with simple and consistent binary patterns day or night, with or without clouds.

A live demonstration for all four experiments can be found in the HOV project web site [17].

IV. ARCHITECTURE OF A PRACTICAL HOV COUNTING SYSTEM

Based on the results of our sensor phenomenology and algorithmic study, a prototype HOV counting system was designed and built by February 2000. The system employs two near-infrared cameras, one in the lower band and one in the upper band. The cameras are coregistered and operate in sync (genlocked). Coregistration is achieved through an optical signal splitter/filter. Since vehicles are passing by the system only occasionally and not continuously we have the cameras working in a discrete mode. The cameras take snapshots of the road scene only when they are triggered by a radar device (see Fig. 16). The radar device issues a trigger signal when it senses the presence of an incoming vehicle. The radar also communicates to the computer that controls the HOV system the position and speed information of the incoming vehicle. There are certain advantages to having the cameras operate in discrete mode, including savings in computational power as well as reduced storage for image archival.

The dual-band camera system rests upon a computer-controlled pan-tilt device, so that accurate aiming is feasible through remote operations. The camera system is accompanied by an artificial near-infrared light source. The light source is powered by a computer-controlled power supply. The computer automatically adjusts the illumination level of the light source to an optimum value based on the readings of two external near-infrared photometers (upper and lower band). The goal is to maintain at all times S/N ratios above 250, the minimum requirement for a clear imaging signal.

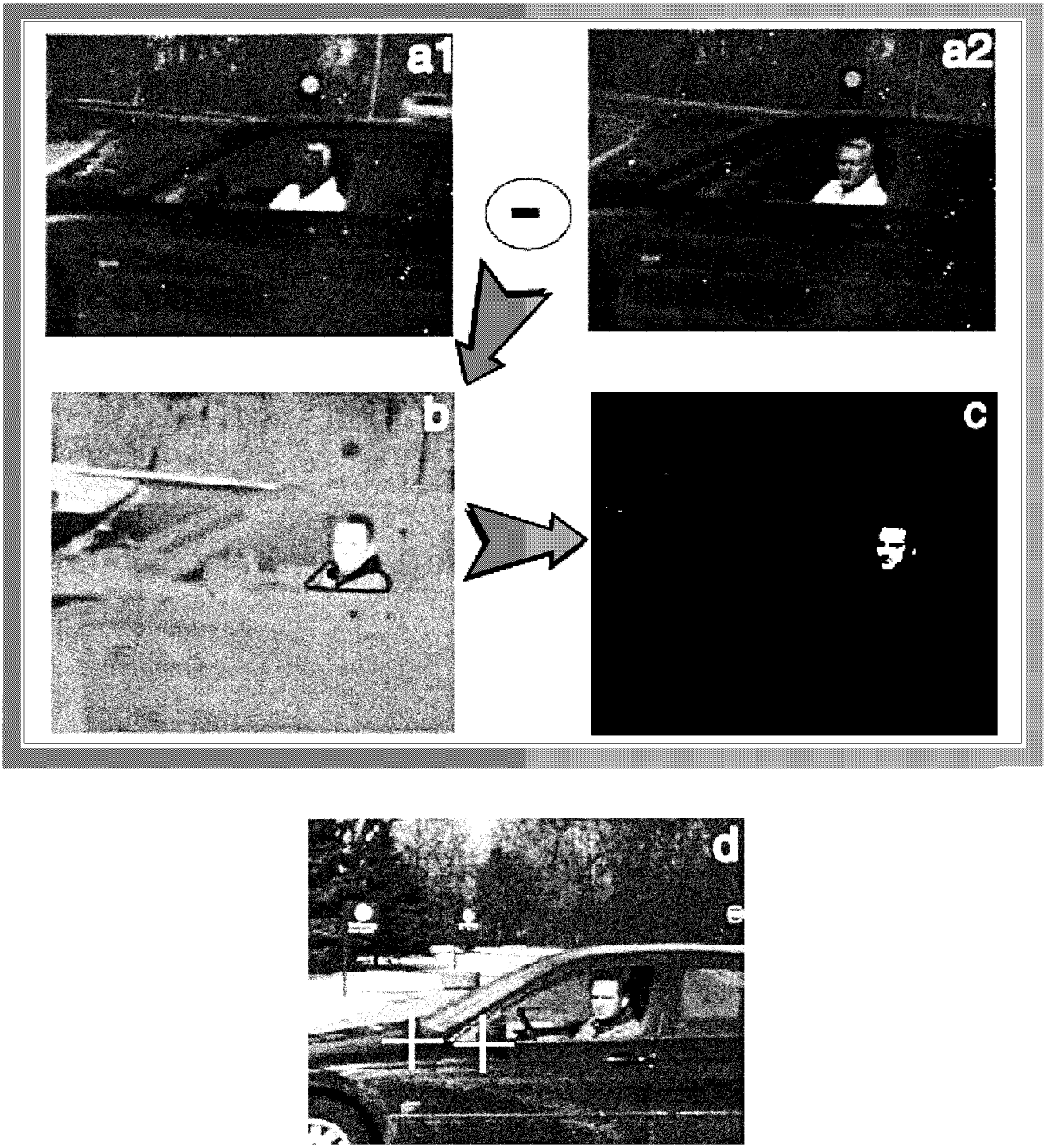


Fig. 17. Image of a Caucasian male outdoors in (a1) the upper band and (a2) the lower band of the near infrared. The vehicle's window is open and is not a factor. (c) The result of fusion between (a1) and (a2). (d) The final thresholded result. (e) The Visionics FaceIt alignment operation fails to locate the face of the subject as evidenced by the location of the white crosses. FaceIt is a state-of-the-art face recognition system marketed by the Visionics Corporation [20].

The subtraction of the upper band image from the lower band image is performed with a weighting factor that is determined by the readings of the photometers. Thresholding and neural processing follow in the computational pipeline. The original images along with the processing results are stored locally for archival purposes. Every incoming piece of data contributes toward the update of a global statistical measure (vehicle occupancy on the HOV lane). Future HOV systems may even provide the capability for law enforcement if they are bundled with a license plate reader. Locally, the computing and digital storage equipment is encased in a weather-proof cabinet. The local HOV system communicates with our lab through a slower web link (see Fig. 16). The web link provides the means to control remotely the equipment, to get up-to-date global statistics from the system, and to download at a relaxed pace the locally archived data for permanent storage purposes.

V. Conclusions and Future Work

We have described an innovative method and system for performing automatic counting of vehicle occupants in the freeway (HOV system). We identified three aspects to the problem: a) the imaging aspect, b) the algorithmic aspect, and c) the engineering aspect. Accordingly, we first invented a method to provide high-quality imaging signals to the HOV system. The method calls for two coregistered near-infrared cameras with spectral sensitivity above (upper band) and below (lower band) the 1.4-μm 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 [18]. The near-infrared cameras can also provide clear imaging signals even in certain foul weather situations, such as in hazy conditions. This is very important for HOV purposes because haze is endemic in certain metropolitan areas (e.g., San Francisco).

The hallmark of the method is the fusion of the coregistered imaging signals from the lower and upper band cameras. Because of an abrupt change in the reflectance of the human skin around 1.4 μm, the fusion has as a result the intensification of the occupant 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 vehicle occupants.

Second, we designed and tested a fuzzy neural classifier to perform the vehicle occupant detection in the near-infrared imagery. The classifier scored perfectly on a random sample of 100 thresholded images. The same classifier scored below 50% in classification experiments with corresponding images from the lower and upper near-infrared band, and the visible band.

Third, we designed and implemented a prototype HOV counting system based on the previous two results of our research. We are currently in the process of designing and implementing a license plate reader to work in tandem with the baseline HOV prototype. We have also expanded our research in the face detection area. Face detection is a crucial part of a face recognition system. There are no reliable face detectors for outdoor environments and this is one of the primary reasons that keep face recognition technology restricted to indoor applications at the time being. Our dual-band method promises to change that. Preliminary comparative experiments are extremely encouraging (see Fig. 17).

Acknowledgment

The authors wish to extend their deep appreciation to K. Schwartz, the Mn/DOT HOVL Program Manager for his generous help and support. Many thanks go to B. Worel and J. Herndon for accommodating our needs in the Mn/ROAD facility. The authors would also like to thank J. Keller, P. Reutiman, and P. Tsiamyrtzis for their help during HOV testing. Finally, we would like to thank S. Nelson, P. Symosek, and B. Fritz for a valuable discussion regarding some technical issues in this project.

References

[1] T. Naito, T. Tsukuda, K. Yamada, K. Kozuka, and S. Yamamoto, "License plate recognition method for inclined plates outdoors," in *Proc. 1999 Int. Conf. Information Intelligence and Systems*, 1999, pp. 304–312.

[2] L. Salgado, J. M. Menendez, E. Rendon, and N. Garcia, "Automatic car plate detection and recognition through intelligent vision engineering," in *Proc. 1999 IEEE Int. Carnahan Conf. Security Technology*, 1999, pp. 71–76.

[3] M. Shridhar, J. W. V. Miller, G. Houle, and L. Bijnagte, "Recognition of license plate images: Issues and perspectives," in *Proc. 5th Int. Conf. Document Analysis and Recognition*, 1999, pp. 20–22.

[4] I. Pavlidis, P. Symosek, B. Fritz, R. Sfarzo, and N. P. Papanikolopoulos, "Automatic passenger counting in the high occupancy vehicle lanes (hovl)," in *Proc. 1999 Annu. Meet. Intelligent Transportation Soc. America*, 1999.

[5] I. Pavlidis, P. Symosek, B. Fritz, and N. P. Papanikolopoulos, "Automatic detection of vehicle passengers through near-infrared fusion," in *Proc. 1999 IEEE/IEEJ/JSAI Int. Conf. Intelligent Transportation Systems*, 1999, pp. 304–309.

[6] A. Balasuriya and T. Ura, "Multisensor fusion for autonomous underwater cable tracking," in *MTS/IEEE Oceans 1999 Proc.*, 1999, pp. 209–215.

[7] G. P. Lemeshewsky, "Multispectral multisensor image fusion using wavelet transforms," *1999 SPIE Proc.*, vol. 3716, pp. 214–222, 1999.

[8] T. Peli, E. Peli, K. Ellis, and M. A. Walthman, "Multispectral image fusion for visual display," *1999 SPIE Proc.*, vol. 3719, pp. 359–368, 1999.

[9] F. E. Sabins, *Remote Sensing, Principles and Interpretation*, 3rd ed. New York: W. H. Freeman, 1997.

[10] W. L. Wolfe and G. J. Zissis, *The Infrared Handbook*. Ann Arbor, MI: Environmental Res. Inst., 1985.

[11] R. R. Anderson, B. S. Parrish, and J. A. Parrish, "The optics of human skin," *J. Investigative Dermatology*, vol. 77, no. 1, pp. 13–19, 1981.

[12] 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 μm," Army Medical Res. Lab., Fort Knox, Tech. Rep. 189, Apr. 1955.

[13] J. Graham and P. J. Hendra, "Rapid identification of plastics components recovered from scrap automobiles," *Plastics, Rubber, and Composites Processing and Applications*, vol. 24, no. 2, pp. 55–67, 1995.

[14] H. E. Howell and J. R. Davis, "Qualitative identification of fibers using nir spectroscopy," *Textile Chemist and Colorist*, vol. 23, no. 9, pp. 69–73, 1991.

[15] M. Papini, "Analysis of the reflectance of polymers in the near- and mid-infrared regions," *J. Quantitative Spectrosc. Radiation Transfer*, vol. 57, no. 2, pp. 265–274, 1997.

[16] G. A. Carpenter, S. Grossberg, and D. B. Rosen, "Fuzzy art: Fast stable learning and categorization of analog patterns by an adaptive resonance system," *Neural Networks*, vol. 4, pp. 759–771, 1991.

[17] HOV home page. [Online]: Available: www.htc.honeywell.com/projects/hov

[18] D. H. Sinley, "Laser and led eye hazards: Safety standards," *Opt. Photon. News*, pp. 32–37, Sept. 1997.

[19] B. K. P. Horn, *Robot Vision*. Cambridge, MA: MIT Press, 1986, pp. 202–277.

[20] Visionics home page. [Online]: Available: http://www.visionics.com/

Ioannis Pavlidis (S'85–M'87–SM'00) received the B.S. degree in electrical engineering from the Democritus University, Greece, the M.S. degree in robotics from the Imperial College of the University of London, U.K., and the M.S. and Ph.D. degrees in computer science from the University of Minnesota, Minneapolis.

He joined the Honeywell Technology Center, Minneapolis, immediately upon his graduation in January 1997. His expertise is in the areas of computer vision beyond the visible spectrum and pattern recognition of highly variable patterns. He published extensively in these areas in major journals and refereed conference proceedings over the past several years.

Dr. Pavlidis is the Cochair of the IEEE series of Workshops in Computer Vision Beyond the Visible Spectrum and serves as a Program Committee member in several other major conferences. He is a Fulbright Fellow and a member of ACM.

Vassilios Morellas (M'98) received the B.S. degree in mechanical engineering from the National Technical University of Athens, Greece, the M.S. degree in mechanical engineering from Columbia University, New York, and the Ph.D. degree in mechanical engineering from the University of Minnesota, Minneapolis.

He has been with the Honeywell Technology Center, Minneapolis, since 1998. His expertise is in the areas of computer vision, sensor integration, and learning theories as they apply to enhancing robot autonomy and advancing machine intelligence. Prior to his current position, he pioneered the SAFETRUCK Research Project while working at the University of Minnesota as a Research Associate. SAFETRUCK successfully demonstrated the use of differential GPS (Global Positioning System) and radar sensing technologies to enhance safety of semitractor-trailers by developing lane departure detection and collision avoidance systems. SAFETRUCK won the second prize in the 1997 ITS World GPS Showcase competition.

Nikolaos Papanikolopoulos (S'88–M'92) received the Ph.D. degree in robotics and computer vision from Carnegie Mellon University, Pittsburgh, PA, in 1992.

He is currently the McKnight Associate Professor at the Computer Science Department of the University of Minnesota, Minneapolis. He has authored more than 25 articles in refereed journal papers and he is the recipient of multiple research contracts from the U.S. Government. Currently, he is the Principal Investigator in a $5 million research project funded by the Defense Advanced Research Projects Agency (DARPA). The project aims to develop distributed robotics technology to be employed in the future battlefield.

Dr. Papanikolopoulos is the Chairman of the IEEE Robot Vision Technical Committee. He also serves regularly as Committee Member in IEEE Conferences and as Guest Editor in international journals.