Color machine vision for autonomous vehicles

Shashi D. Buluswar*
Dept. of Computer Science
University of Massachusetts
Amherst, MA, U.S.A.
buluswar@cs.umass.edu

Bruce A. Draper
Dept. of Computer Science
Colorado State University
Ft. Collins, CO, U.S.A.
draper@cs.colostate.edu

Keywords: Color, Autonomous vehicles, Machine learning in computer vision.

Abstract

Color can be a useful feature in autonomous vehicle systems that are based on machine vision, for tasks such as obstacle detection, lane/road following, and recognition of miscellaneous scene objects. Unfortunately, few existing autonomous vehicle systems use color to its full extent, largely because color-based recognition in outdoor scenes is complicated, and existing color machine vision techniques have not been shown to be effective in realistic outdoor images.

This paper presents a technique for achieving effective real-time color recognition in outdoor scenes. The technique uses Multivariate Decision Trees for piecewise linear non-parametric function approximation to learn the color of a target object from training samples, and then detects targets by classifying pixels based on the approximated function. The method has been successfully tested in several domains, such as autonomous highway navigation, off-road navigation and target detection for unmanned military vehicles, in projects such as the U.S. National Automated Highway System (AHS) and the U.S. Defense Advanced Project Agency - Unmanned Ground Vehicle (DARPA-UGV). MDT-based systems have been used in stand-alone mode, as well as in conjunction with systems based on other sensor configurations.

*Supported by the Advanced Research Projects Agency through Rome Labs under contract F30602-94-C-0042.

1 Introduction

Machine vision techniques are increasingly being used in intelligent autonomous vehicle systems [9, 11, 24, 30, 40, 46]. Most of these systems (with a few exceptions [9, 40, 48]) do not utilize color, despite the fact that color can be a useful feature for detecting objects such as lanes, obstacles and traffic signs, and even though color cameras are becoming an increasingly inexpensive part of autonomous vehicle platforms. Although gray-scale vision-based autonomous vehicle systems have been shown to work well in highway and off-road scenarios, their capabilities can potentially be vastly increased by combining them with color-based techniques. The reason color has not been used much in this domain is the lack of effective color-based recognition methods for outdoor images. This work presents a technique for achieving reliable real-time color recognition in outdoor imagery, that has been used for tasks such as highway lane/road detection, obstacle detection for on- and off-road vehicles, and automatic target recognition for unmanned military vehicles.

The color (or rather, the *apparent* color) of an object depends on illuminant color, the reflectance of the object, illumination geometry (orientation of the surface normal with respect to the illuminant), viewing geometry (orientation of the surface normal with respect to the sensor), and sensor parameters [21]. In outdoor images, the color of the illuminant (i.e., daylight) varies with the time-of-day, cloud cover and other atmospheric conditions [22]; the illuminant and viewing geometry vary with changes in object and camera position and orientation. In addition, shadows and inter-reflectances [19], and certain sensor response parameters [34], all of which can be difficult to model in outdoor scenarios, may also affect the apparent color of objects. Consequently, at different times of the day, under different weather conditions, and at various positions and orientations of the object and camera, the apparent color of an object can be different.

Figure 1 shows the variation in the apparent color of two simple matte surfaces (white and green) under different lighting and viewing conditions from about 50 images; the figure also shows the color of each surface from one sample (represented by a single point), and the overall distribution in RGB space over the 50 images. In this example, the overall variation for each surface is about 250% of the distance between the centroids of the two clusters. In other words,

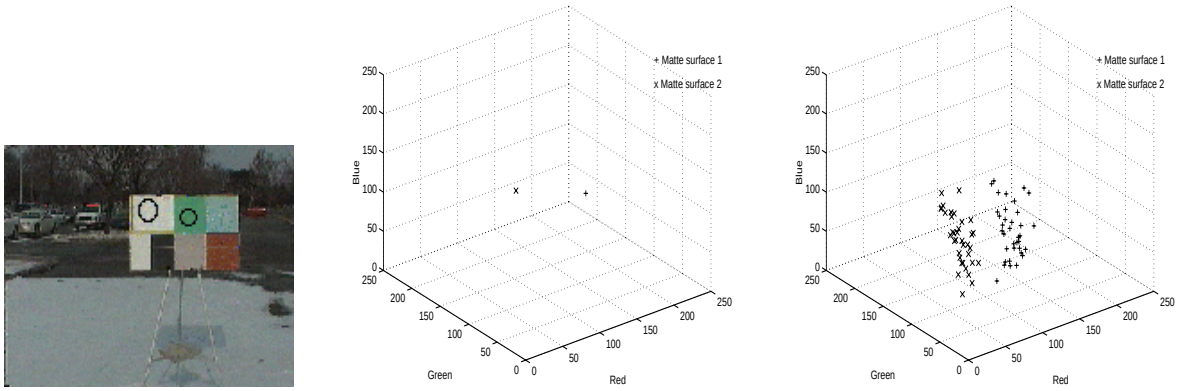

Figure 1: Variation of apparent color in outdoor images: (left to right) samples from two matte surfaces (extracted from circles), the RGB color from a single image, and the variation over 50 images.

the variation in the apparent color of a single surface can be greater than the difference (in color-space distance) between two distinct colors (white and green, in this case). The variation in the apparent color of more realistic objects, such as a road surface and a camouflaged military vehicle (figure 2), can be even greater.

Human beings have an adaptive mechanism called color constancy that compensates for this color shift. Unfortunately, no corresponding adaptive mechanism exists in machine vision systems, and the notion of a color associated with an object is precise only within the context of scene conditions. Previous approaches (described later) have attempted to recognize object color without context or sufficiently robust models, and consequently have produced methods for color recognition that are effective only in highly constrained imagery.

This paper analyzes variations in the apparent color of objects with respect to existing models of daylight and surface reflectance, and shows that the shift in apparent color under outdoor conditions can be represented by characteristic distributions in RGB space. It is then shown that such distributions can be "learned" from training samples using Multivariate Decision Trees (MDT's) [4] for non-parametric approximation of decision boundaries around the training samples. Image pixels are then classified according to their location with respect to the learned decision boundaries. MDT-based classification is then demonstrated in a number of domains, such as highway and off-road navigation (including lane-finding and obstacle-detection), and target detection for autonomous military vehicles.

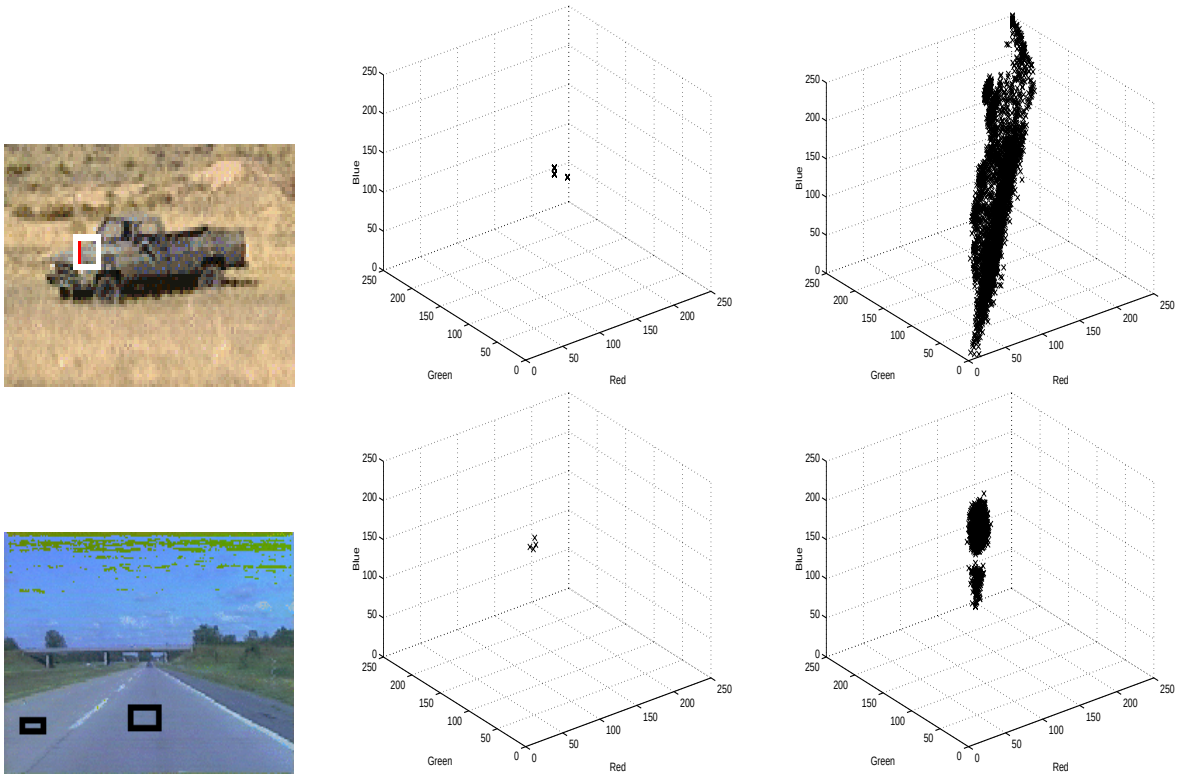

Figure 2: Samples of objects in real outdoor applications - highway road surface (top) and camouflaged military vehicle (bottom), the (RGB) color from a single image (sample color extracted from drawn boxes), along with the variation over about 100 images.

2 Previous work

Past work in color-based recognition under varying illumination can be divided into two categories: computational color constancy and non-parametric (sample-based) classification. In addition, there has been work on lane-finding and obstacle-detection for autonomous vehicles in gray-scale images based on edge-detection or stereo [11, 24, 29, 30]; since this paper is concerned with the use of color information, gray-scale techniques will not be discussed in the literature review.

2.1 Computational color constancy

Most of the work in computational color recognition under varying illumination has been in the area of color constancy, the goal of which is to match object colors under varying illumination without knowing the spectral composition of either the incident light or the surface reflectance;

the general approach is to recover an illuminant-invariant measure of surface reflectance by first determining the properties of the illuminant.

Depending on their assumptions and techniques, color constancy algorithms can be divided into the following six categories [17]: (1) those which make assumptions about the statistical distribution of surface colors in the scene, (2) those which make assumptions about the types of reflection and illumination, (3) those assuming a fixed image gamut, (4) those which obtain an indirect measure of the illuminant, (5) those which require multiple illuminants, and finally, (6) those which require the presence of surfaces of known reflectance in the scene.

Among the algorithms that make assumptions about the statistical distributions of surface colors in the scene, Buchsbaum [6] assumes that the average of the surface reflectances over the entire scene is gray (the gray-world assumption); Gershon [19] assumes that the average scene reflectance matches that of another known color; Vrhel [47] assumes knowledge of the general covariance structure of the illuminant, given a small set of illuminants, and Freeman [16] assumes that the illumination and reflection in a scene follow known probability distributions. These methods are effective when the distribution of colors within the scene follows the assumed model or distribution. In outdoor scenes, the CIE daylight model [22] suggests that the gray-world assumption will not be valid; at the same time, as later sections will show, no general assumptions can be made about the distribution of surface colors even if the distribution of daylight color is known. Consequently, these methods are too restrictive for all but very constrained scenes.

The second set of color constancy algorithms make assumptions about the dimensionality of spectral basis functions [45] required to accurately model illumination and surface reflectance. For instance, Maloney [28] and Yuille [52] assume that the linear combination of two basis functions is sufficient. Under the assumption, the variation in surface color in a three-dimensional color space would follow a plane. Daylight, however, follows a parabolic surface in three dimensions (RGB) [7]; hence, the assumptions of these methods are true only under specifically controlled illumination.

Among the algorithms that make assumptions about image gamuts is Forsyth's CRULE (coefficient rule) algorithm [15], which maps the gamut of possible image colors to another gamut of colors that is known a-priori, so that the number of possible mappings restricts the set of pos-

sible illuminants. In a variation of this algorithm, Finlayson [13] applies a spectral sharpening transform to the sensory data in order to relax the gamut constraints. The assumptions about gamut-mapping restrict the application of CRULE to matte Mondrian surfaces under controlled illumination and fixed orientation. Ohta [37] assumes a known gamut of illuminants (controlled indoor lighting that lies on some points along the CIE model), and uses multi-image correspondence to determine the specific illuminant from the known set. By restricting the illumination, this method is applied only to synthetic or highly constrained indoor images.

Another class of algorithms uses an indirect measure of the illumination. For instance, Shafer [44], Klinker [23] and Lee [26] use surface specularities (Sato [43] uses a similar principle, but not for color constancy); similarly, Funt [18] uses inter-reflections to measure the illuminant. These methods are based on the assumption of a single point-source illuminant; this assumption is not valid for an extended or non-point-source illuminant such as daylight.

In yet another approach, D'Zmura [53]and Finlayson [14] require light from multiple illuminants incident upon the multiple instances of a single surface in the same scene. The problem with these approaches is that they require identification of the same surface in two spatially distinct parts of the image that are subject to different illuminants. Once again, the approaches have been shown to be effective only on Mondrian or similarly restricted images.

The final group of color constancy algorithms assumes the presence of surfaces of known reflectance in the scene and then determine the illuminant. For instance, Land's Retinex algorithm [25] and its many variations require the presence of a surface of maximal (white) reflectance within the scene. Similarly, Novak's supervised color constancy algorithm [36] requires surfaces of other known reflectances. Such assumptions, while applicable to controlled settings, are not generally applicable to unconstrained images.

The assumptions made by the aforementioned algorithms are such that most of them perform only on highly restricted images (such as Mondrians), under mostly constrained lighting. Forsyth [15] aptly states, *"Experimental results for [color constancy] algorithms running on real images are not easily found in the literature... Some work exists on the processes which can contribute to real world lightness constancy, but very little progress has been made in this area."*

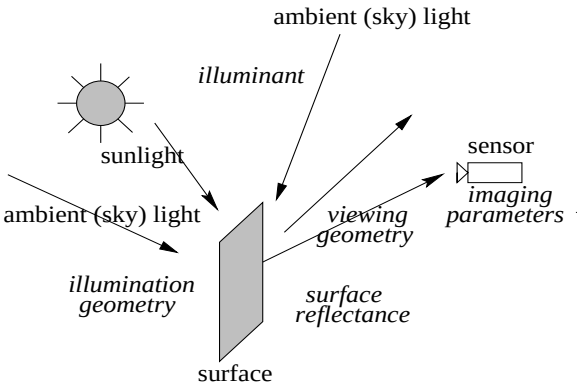

Figure 3: Image formation in outdoor scenes, along with the various processes involved.

2.2 Non-parametric (sample-based) approaches

The emergence of road-following as a machine vision application has spawned several methods that use color for road-following without specific parametric models. Crisman's SCARF algorithm [9] approximates an "average" road color from samples, and models the variation of the color of the road under daylight as Gaussian noise about an empirically derived "average" road color; pixels are then classified based on minimum-distance likelihood. This technique was successfully applied to road-following, but cannot be applied for general color-based recognition of road-scene objects. For instance, in the case of the examples in figures 1 and 2, this approach would calculate an average color for each surface from the corresponding distribution, and use that average as the most likely color of the object under any set of conditions.

Pomerleau's ALVINN road-follower [40] uses color images of road scenes along with user-induced steering signals to train a neural network to follow road/lane markers. Although the ALVINN algorithm made no attempt to explicitly recognize lanes or roads, it showed for the first time, that a complex visual domain with unmodeled variation can be approached as a non-parametric learning problem. This approach represents a significant advance in road-following methodology; however, it is designed specifically for road-following and is hence not applicable to color-based recognition of road-scene objects.

3 Color shift in outdoor scenes: Causes and analysis

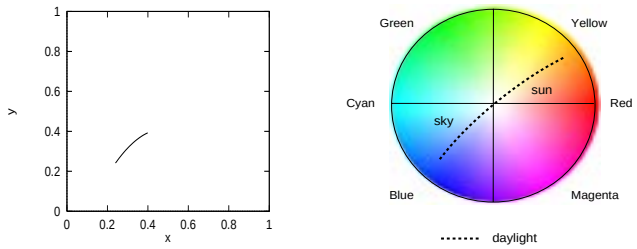

Figure 4: The CIE parametric model of daylight in the chromaticity space (left) and the color circle (right). The regions of the color circle representing the colors of sunlight and skylight (empirically determined) are shown.

The standard model of image formation [21] describes the observed color of objects in an image as a function of (i) the color of the incident light (daylight, in the case of outdoor images), (ii) the reflectance properties of the surface of the object (iii) the illumination geometry, (iv) the viewing geometry, and (v) the imaging parameters. Theoretical parametric models exist for all the phases of this process. Unfortunately, these models have not been proven effective in unconstrained color imagery for model-based color recognition; still, they provide an approximate qualitative description of the variation of apparent color. Figure 3 shows a pictorial description of the various processes involved in the formation of outdoor images. The pertinent models are described below, and a general hypothesis about RGB distributions representing apparent object under daylight is developed thereafter.

3.1 Illumination

Daylight is a combination of sunlight and (ambient) skylight; the variation in the color of daylight is caused by changes in the sun-angle, cloud cover and other weather conditions. The CIE daylight model [22] describes the variation in daylight color as a parabola in the CIE chromaticity space[1] (figure 4).

$$y = 2.87x - 3.0x^2 - 0.275, \tag{1}$$

where $0.25 <= x <= 0.38$. In RGB space, the parabola stretches out into a thin paraboloid surface [7].

[1] RGB is a linear transform of the CIE chromaticity space [20]

3.2 Illumination geometry and viewing geometry

Illumination geometry, i.e., the orientation of the surface normal with respect to the illuminant, affects the composition of the light incident upon the surface. The surface orientation determines how much light from each of the two components of daylight (sun and sky), is incident upon the surface. For instance, a surface that faces the sun is illuminated mostly by sunlight, whereas one that faces away is illuminated by the ambient light. The viewing geometry, which is the orientation of the surface with respect to the camera, determines the composition and magnitude of the light reflected onto the camera; this is primarily a function of the reflectance properties of the surface. For instance, a matte (Lambertian) surface has uniform reflection in all directions, whereas a shiny (specular) surface reflects light only along that angle of reflection which equals the angle of incidence.

3.3 Surface reflectance

The effect of illumination geometry and viewing geometry depends on the reflectance properties – upon the strength of the specular component, to be precise. Most realistic surfaces have components of both Lambertian and specular reflection. A number of models have been applied with varying degrees of success to such surfaces, most notably Phong's shading model [38], Shafer's Dichromatic model [44], and Nayar's hybrid reflection model based on photometric sampling [32].

The Dichromatic reflection model (originally proposed by Shafer [44] and subsequently extended by Novak [35], Lee [26] and Klinker [23]) models the net reflection of a surface as the linear combination of the specular and Lambertian reflection components.

$$L(\lambda, i, e, g) = m_s(i, e, g)c_s(\lambda) + m_b(i, e, g)c_b(\lambda) \tag{2}$$

where $L(\lambda, i, e, g)$ is the intensity of light at wavelength λ, angle of incidence i, angle of reflection e and phase angle g (angle between direction of incident light and viewing direction); $m_s(i, e, g)$ is the geometric scale factor (determined by the illumination and viewing geometry) of $c_s(\lambda)$, the spectral power distribution of the specular component of the reflected light, and $m_b(i, e, g)$ and $c_b(\lambda)$ are the same quantities for the Lambertian reflection component. Specularities in the

image are used to determine the weights for each component.

The Phong shading model [38] approximates the falloff in brightness of specular reflection as $\cos^n(\alpha)$, where α is the difference between the viewing angle and the angle of maximal specular reflectance. The value of n is determined empirically for a given surface, and varies from 1 (for matte surfaces) to 200 (for highly specular surfaces). At $\alpha = 0$, the brightness is the maximum (i.e., 1), and falls off as the surface is rotated, to the minimum (i.e., 0) at -90^o and 90^o. The Phong model has been very widely applied by the computer graphics community as an effective method of achieving shading effects in rendering grey-scale and color images.

Nayar [32] describes the brightness of surface reflectance as a linear combination of the Lambertian and specular components (a concept similar to the Dichromatic Model [44]):

$$I = IL + IS, \tag{3}$$

where I is the total intensity at a given point in the surface, and IL and IS the intensities of the specular and Lambertian components. $IL = Acos(\theta_s - \theta_n)$ (Lambert's law), where A is the constant representing the weight of the Lambertian component, and θ_s and θ_n are the directions of the illumination source and the surface normal. IS, modeled by the delta function [32], is $B\delta(\theta_s - 2\theta_n)$, where B is the weight of the specular component. Hence, $I = Acos(\theta_s - \theta_n) + B\delta(\theta_s - 2\theta_n)$. The model is adapted for an extended light source to determine the weights of the reflection components by photometric sampling, a method by which brightness samples are obtained for multiple angles of illumination and viewing.

The above reflectance models show how the strength of the specular and Lambertian components of surface reflectance determines the effect of the illuminant color and the illumination and viewing geometry on the apparent color of a surface. Evidently, different types of surfaces exhibit different color shifts, depending on the combination of the aforementioned factors.

3.4 Shadows and inter-reflections

Inter-reflections and shadows can cause a further variation in color by altering the color of the light incident upon the surface [19]. Inter-reflections, for instance, cause light reflected off other surfaces in the scene to be incident upon the surface being examined. Shadowing can cause the

elimination of incident sunlight (if the surface is self-shadowed), and further inter-reflection (if the surface is shadowed by another surface).

3.5 Imaging parameters

There are a number of imaging parameters, effective between the lens and the image plane, that may change the apparent color of objects in a scene. Clipping occurs when pixel values that are too high (i.e., too bright) or too low (i.e, too dark) are not registered beyond the limited response range of the camera, thus resulting in a loss of information and possible color skewing (if all three color bands are not clipped at the same time). Clipping is easily detected but not easily avoided, especially in outdoor images, where it is difficult for imaging hardware to adapt to the variation in the range of intensities. Any software approach to interpreting clipped pixels is bound to be ad-hoc and domain-dependent [34]; hence, until improvements in sensor design take place, machine vision methods may be forced to simply detect clipped pixels and discard those points in the image.

Blooming is a related phenomenon, where sensor cells saturated due to clipping "bleed" into neighboring cells. Blooming is harder to detect, except through finding clipped pixels and probabilistically tracing pixel values in the direction most likely to cause blooming [34]. On one hand, blooming is a much more serious problem than clipping because it is harder to detect; on the other, inter-cell bleeding is a simpler problem to prevent from a hardware design point of view. The method in this study does not present new approaches to classifying clipped or bloomed pixels; instead it assumes that clipping and blooming are localized phenomena and uses region-level heuristics (such as morphological operations and connected-components-based extraction of bounding boxes) to compensate for pixel-level errors introduced by the two phenomena.

Nonlinear response results in an inconsistent mapping between spectral power distributions and corresponding digital color values across the sensor range, and consequently a disproportional skewing in each of the color bands. For instance, the response to a surface highly saturated in the red channel (such as a red "Stop" sign) may be in the linear response range along the green and blue channels, but in the nonlinear range in the red channel. The effect of nonlinear

response is virtually impossible to detect, except with careful calibration [34].

Another problem that has been shown to cause color skewing in calibration studies is chromatic aberration [2]. This phenomenon occurs because the focal length of a lens is a function of the wavelength of the light incident upon the lens. Hence, different colors may focus at different points with respect to the image plane and the optical axis. There are two types of displacement caused by chromatic aberration, lateral and longitudinal. Lateral chromatic aberration can cause light of a certain wavelength to focus on a cell neighboring the intended cell, causing color mixing. Experiments [2, 34] indicate that this type of color mixing occurs mostly along surface boundaries, thus leaving the non-boundary pixels unaffected. Longitudinal displacement of light, i.e., along the optical axis can cause unequal blurring of different wavelengths. The same experiments [2, 34] indicate that parametric methods sensitive to small perturbations in the assumed physics-based models are far more likely to be affected by such blurring than are the empirical methods used in this study, given the relative magnitude of color shifts due to the other complicating factors.

3.6 Overall distribution in RGB space

Assuming, from the standard image formation model [21], that apparent color is determined by the product of the incident light and the surface reflectance, and then somewhat altered by shadows, inter-reflections and imaging parameters, it can be deduced that the RGB distributions can be arbitrarily shaped, depending on the nature of the surface. The aforementioned reflectance models suggest that the distribution for a Lambertian surface will form a single thin region; that for a specular surface forms two clusters (one cluster near the color of the illuminant, due to the specular spike, and the other cluster near the color of lambertian component); surfaces with mixed reflectance form a continuous blob. Figure 5 shows the distributions for two surfaces; a piece of matte paper (top) forms a continuous blob in RGB, and a shiny red traffic "Stop" sign (bottom) forms two distinct clusters.

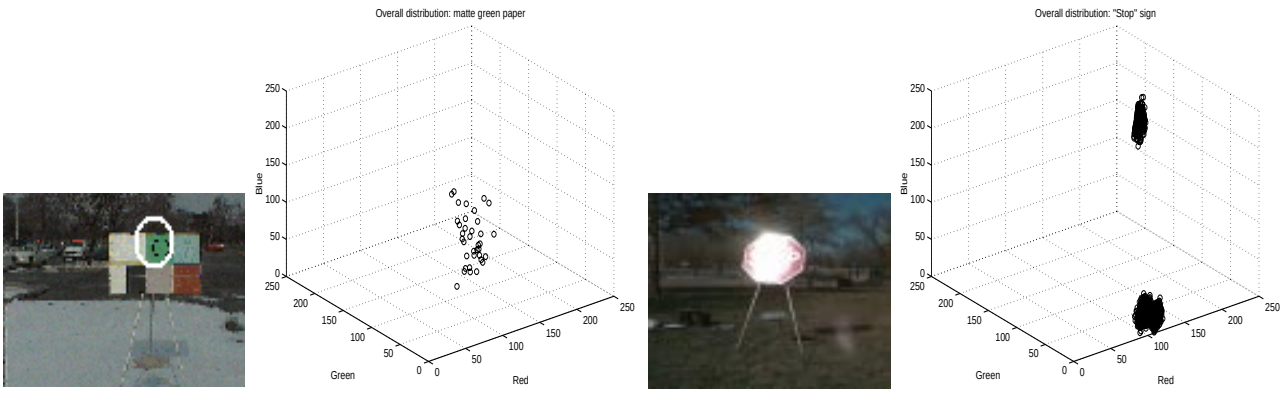

Figure 5: Images and RGB distributions under daylight for different types of reflectances – matte paper (left pair) and shiny "Stop" sign (right pair).

4 Proposed solution: Nonparametric classification

In principle, it should be possible to predict the apparent color of a surface in outdoor images, given the (i) sun-angle, (ii) weather conditions, (iii) surface orientation with respect to the sun and the camera, and (iv) robust models of surface reflectance. Since existing reflectance models have not been shown to be robust in unconstrained outdoor imagery, this approach assumes no knowledge of the aforementioned parameters; rather, the goal is to learn a function that maps RGB values from training samples of an object to particular classes. Thereafter, image pixels are classified into the separate classes based on the learned function associated with a given object. To classify pixels in outdoor color images, we need to select a non-parametric classification scheme that can approximate arbitrarily shaped functions in feature space.

There are two phases in the non-parametric approach to color recognition: training and classification. The training phase approximates a function (or a set of functions) representative of a distribution from samples of the distribution. The approximated function constitutes a mapping between a (training) set of RGB values and the surface (class) it represents. The classification phase determines the class of a given image pixel from the mapping function for the training set that pixel represents. Every pixel in a color image is classified, resulting in a gray-scale image where pixels belonging to a particular class will have the same gray value. There are two issues to consider: (1) the ability of the technique to generalize the function so as to adequately represent the distribution in color space without being too loose (resulting

in the inclusion of samples from a different class), or too tight (resulting in the exclusion of samples belonging to the class); (2) the number of training samples required to approximate the distribution.

There are a number of techniques that have been used in other domains for function approximation and classification. In nearest-neighbor classification [51, 5], given a set $X_n = \{x_0, x_1, ..., x_n\}$ of n independent samples, a new instance x_{n+1} is classified according to the distances between x_{n+1} and each element of the set $\{x_0, x_1, ..., x_n\}$. The class assigned to x_{n+1} is the class of the training sample with the shortest distance. The problem with using this approach on color images is that pixels forming a thin distribution will not be correctly classified using three-dimensional Cartesian distance in RGB.

Gaussian maximum-likelihood classification (discussed in earlier sections) approximates an "average" feature value from samples and models the entire distribution as Gaussian noise about the average value; subsequently, pixels are classified based on the probability that they are noisy instances of the set represented by average. This technique cannot be applied to general outdoor color recognition because the variation of apparent color under daylight is not well-modeled as Gaussian noise.

Another way of classifying pixels is to segment the feature space and classify pixel instances based on their position in the segmented feature space. This can be done in a number of ways: by drawing explicit piecewise-linear boundaries in RGB space (decision trees [41, 4]); by learning a nonlinear function (genetic algorithms [31] and radial basis functions [39]) that maps RGB values explicitly to numerical values which are then thresholded to find decision boundaries; and by learning a mapping function that uses RGB as the input feature space but maps the input feature space to an intermediate feature space, so as the facilitate boundary fitting (neural networks with a hidden layer [42, 10]).

(Univariate) Decision Trees [41] approximate a boundary by fitting hyperplanes around the samples, orthogonal to the axes of the feature space. Multivariate Decision Trees [4] are more general, and fit hyperplanes of arbitrary orientation around the distributions. Genetic algorithms (GA's) use principles from evolutionary biology to converge on optimal parameters of a fixed-dimensional nonlinear polynomial function. Radial basis functions (RBF's) approximate a

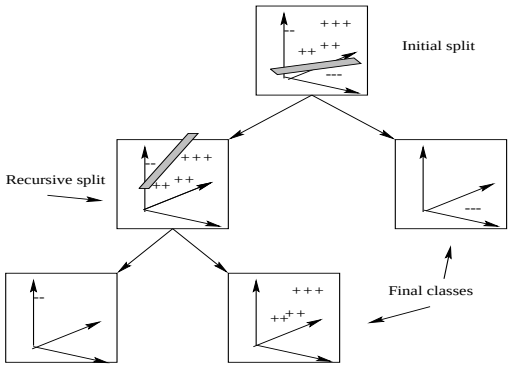

Figure 6: Recursive discriminants of an MDT, separating the '+'s from the '-'s.

function as a weighted sum of Gaussians. Neural Networks (NN's) are more general than GA's or RBF's, and approximate a function of arbitrary dimensionality (determined by the number of hidden units) as a weighted sum of nonlinear squashing functions. Although NN's can be expected to perform well for RGB distributions in this application, the arbitrary nature of the hidden layer feature space makes analysis difficult; consequently, the work presented here uses Multivariate Decision Trees.

5 Multivariate Decision Trees

Multivariate Decision Trees (MDT's) [4] create piecewise-linear approximations of regions in feature space by recursively dividing feature space with hyperplanes (figure 6). MDT's recursively subdivide the feature space by linear threshold units (LTU's) [33, 12]. The LTU's are binary tests, represented by linear combinations of feature values and associated weights. Each division attempts to separate, in a set of known instances (the training set), target instances from non-targets. If two subsets are linearly separable, a single LTU will separate them and the multivariate decision tree consists of the single node. If not, the LTU linearly divides the feature space so as to separate the instances to the extent possible, and the MDT recursively creates and trains new LTU's on the two subsets of instances. The result, therefore, is a tree of LTU's recursively dividing the feature space into multi-dimensional polygons so as to perform a piecewise linear approximation of the region in color-space consisting of the positive samples. The terminal nodes in the tree correspond to inseparable sets, which are labeled as individual classes. Thus, each node in a decision tree is either a decision or a class. Figure 6 shows a

decision-tree operating in a three-dimensional feature space.

Several methods exist for learning the weights in a linear threshold unit; this implementation uses the Recursive Least Squares (RLS) algorithm [50]. The RLS method is recommended for dual-class (target vs. non-target) classification, and is a recursive version of Gauss' Least Squares algorithm, which minimizes the mean squared error between the estimated $\overline{y}_i$ and true y_i values, $\Sigma(y_i - \overline{y}_i)^2$ of the selected features over a number of training instances. RLS incrementally updates the weight vector W according to

$$W_k = W_{k-1} - K_k(X_k^T W_{k-1} - y_k) \tag{4}$$

where W_k is the weight vector for the instance k, of size n, W_{k-1} is the weight vector for instance $k-1$, X_k is the instance vector; X_k^T is X_k transposed, and y_k is the class of the instance. $K_k = P_k X_k$, where P_k is the $n \times n$ covariance matrix for instance k, reflecting the uncertainty in the weights, and

$$P_k = P_{k-1} - P_{k-1} X_k [1 + X_k^T P_{k-1} X_k]^{-1} X_k^T P_{k-1} \tag{5}$$

The weights are initialized randomly, and the matrix consists of 0 values everywhere except along the diagonal, which is set to a very large value: 10^6 according to Young's recommendation [50].

If at any level, the LTU results in a non-negative value, the corresponding set of pixels is labeled as belonging to the object (target), otherwise, it is labeled negative (non-target). Figure 7 shows the structure of a multivariate decision tree. In this tree, the non-terminal nodes represent the LTU tests, and the leaf nodes the classes; the '+' leaf nodes correspond to the inseparable sets classified as one class, and the '-' nodes, the other.

Like other non-parametric learning techniques, decision trees are susceptible to over-training. In order to correct for over-fitting, a fully grown tree can be pruned [3, 41, 4] by determining the classification error for each non-leaf subtree, and then comparing it to the classification error resulting from replacing the subtree with a leaf-node bearing the class label of the majority of the training instances in the set. If the leaf-node results in better performance, the subtree is replaced by it.

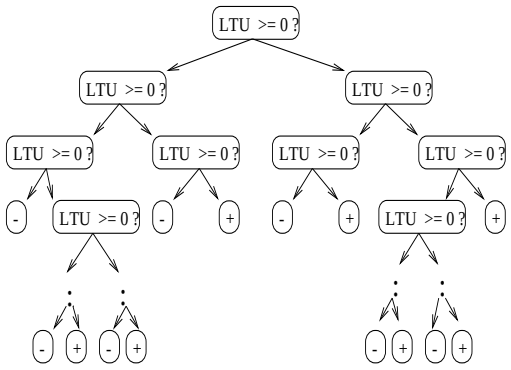

Figure 7: LTU's and targets of an MDT: target (+) & background (-).

The discrete nature of the RGB color space for digital images makes real-time classification possible through the use of lookup tables that are constructed off-line. After a decision tree is built for a given target, every possible RGB color value is classified into target and background (non-target) classes. Thereafter, given a color image, each pixel can be classified from the lookup table in near-real-time. In the results shown here, the result of pixel classification is a binary image in which all suspected target pixels are white, and the background pixels are black. Multiple lookup tables can be combined for multi-class classification.

6 Results

Implementations of MDT-based classification have been tested in several domains, such as automated highway systems, off-road obstacle-detection, military target detection, wildlife detection in aerial images, and skin finding. The results from the first three applications are discussed below. In each case, the system has been (or will be) used independently or in conjunction with systems based on other sensor configurations, such as stereo or infra-red cameras. The following tests were conducted using cross-validation, where half (or fewer) of the images were used in training, and the others for testing.

6.1 MDT's for highway scenes

MDT-based classification is currently being used in the National Automated Highway System (AHS) project for detecting lanes and obstacles in highway scenes for autonomous vehicles. Figure 8 shows a sample image from a highway scene; the goal in this application is to find

Figure 8: Representative results for MDT-based classification for lane-markers and obstacles (left to right) – original color image, classification for lane-markers, classification for road, obstacles and lanes extracted.

the lane-markers and obstacles. There are two lookup tables constructed, one for lane-markers and one for the road surface. Pixels classified as non-road are either lane-markers or obstacles; lane-marker pixels are classified separately, thus identifying the objects on the road that are potential obstacles. The vehicle heading is determined by fitting lines to the lane-marker pixels, and the potential obstacles are extracted by clustering connected pixels and using region-level heuristics. In this system, stereo and motion techniques are used to further prune the obstacle map by identifying the potential obstacles that lie above the ground plane. Representative results of classification for lane-markers and obstacles is shown in figure 8. The color-based component of the obstacle detection system has been tested on thousands of images of hundreds of sequences, and tests conducted in the AHS project have found the system to be sufficiently "reliable" for practical application. For the purposes of this paper, tests were conducted on 10 sequences of 100 images each of highway scenes from the U.S. Midwest. At the pixel level, about 83% of the lane-marker pixels were correctly classified, and about 64% of the obstacle pixels were correctly classified (as non-road and non-lane-marker pixels). The false positive classification percentage was less than 2% for lane markers and about 14% for obstacles. Out of the 1000 images, 100% of the lane-markers were detected, with the 1480 out of 1497 obstacles detected. Obstacles included vehicles on the highway in the current and adjacent lanes, as well as miscellaneous objects cluttering the highway. The obstacles that were not detected were portions of black rubber tires that were almost the color of the highway tarmac.

6.2 Ground-level terrain detection for off-road navigation

While the goal of the AHS project is to provide highway-based autonomous vehicles, the U.S. military is interested in autonomous off-road driving systems. Toward this end, the Unmanned

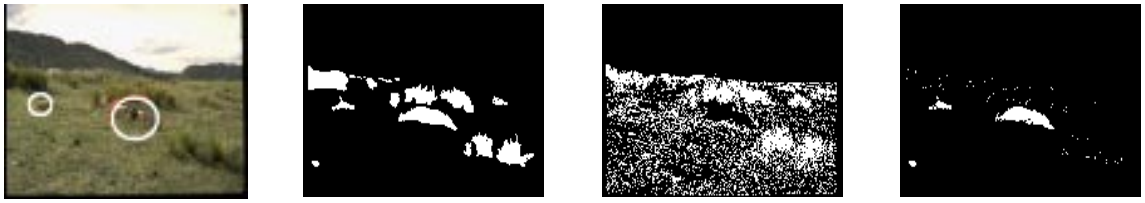

Figure 9: Results from MDT-based classification for yucca bushes (left to right) – original color image (rocks/obstacles marked with circles), simulated depth-based obstacle map, classification for yucca, final obstacle map.

Ground Vehicle (UGV) project developed vehicles that used stereo cameras to detect obstacles by marking all objects over a fixed height above the ground plane (corresponding to the ground clearance of the vehicle) as obstacles. In the off-road tests in Colorado, this strategy proved excessively conservative, in that it forced the vehicle to meticulously avoid yucca bushes and other "obstacles" it could easily drive over. In this scenario, MDT-based classification was used to detect yucca bushes and eliminate them from the obstacle map. In 45 test images that contained 212 identifiable yucca bushes, 176 of the bushes were successfully detected; there were many false positives at the pixel level, mostly from grassy regions, which did not affect the performance of the system because they were not in the initial obstacle map. Figure 9 shows results from one image with a simulated obstacle map; the yucca bushes (pixels) are detected, and those pixels are removed from the obstacle map, leaving only the rocks in the final obstacle map.

6.3 Military target detection using MDT's

The most challenging and comprehensive application of color-based classification has been in domain of camouflaged military target detection using autonomous vehicles. This task is particularly difficult because the goal of camouflage patterns and colors on military targets is precisely to blend the targets into the background vegetation. However, it is not always possible to get a perfect match between the background color and camouflage because the color of vegetation is not constant. Consequently, the hyperplanes of the MDT can make fine distinctions between target color and the background. The MDT-based system was tested on the Ft. Carson data set [1] by a DARPA-sanctioned study by LGA, Inc. [49], and at UGV Demo-C.

Figure 10 shows the results from two color images from the Ft. Carson images. Targets

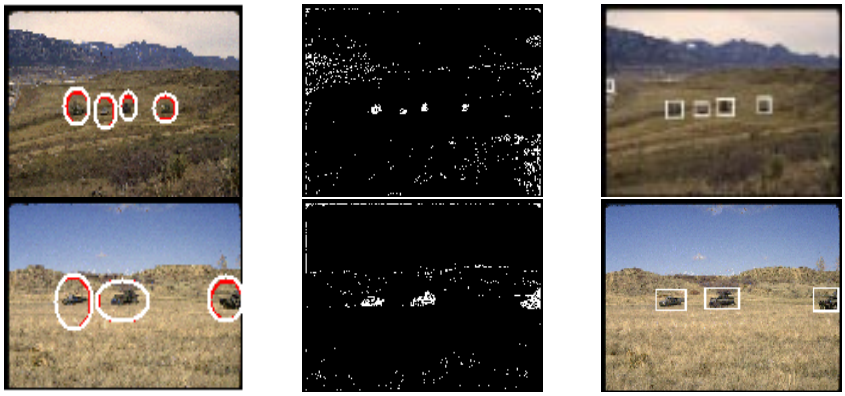

Figure 10: Results from MDT-based classification for camouflaged target detection - original color images (left, targets marked with circles), binary classification (middle), targets extracted (right).

are extracted from the binary classification image by using clustering target pixels and applying region-level heuristics such as (the range of) expected vehicle size(s) and aspect ratio.

Out of 96 images in the Ft. Carson set, 176 out of 211 targets were detected, along with 180 false positives. In the DARPA-UGV Demo-C tests, 100% of 74 the targets were detected over 50 images, with 32 false positives. In both tests, between 50% and 95% of the on-target pixels were correctly classified, enough to form clusters approximately the expected size of the targets. In further UGV tests, the color-based system was combined with an infra-red system [27] to further improve the performance.

7 Future work and conclusions

In all of the tests, there has been a large number of false positives. There are two reasons for the proliferation of false positives: first, the apparent color of background objects can sometimes be very close to the color of the target; second, the region in color space being approximated can be large – the larger the region, the greater the likelihood of an intersection between the region representing the target and that representing another object. This suggests that although MDT-classification has been used successfully in different applications, it serves more as a focus-of-attention mechanism than as a method for full-fledged object recognition. Clearly, providing larger amounts of training data can reduce both the false positive and the false negative rates. At the same time, a tighter threshold on the training error can reduce the number false positives. The false positives have not reduced the usefulness of the method, since false positives can be

eliminated (or reduced) by combining the color-based approach with other sensors (e.g., stereo and infra-red), and using region-level heuristics, such as expected target size. Another issue being explored is automatic training, thereby eliminating the need for user input in the initial training phase.

Overall, the MDT approach appears to be an effective way of achieving color recognition for various applications of autonomous intelligent vehicles, and proves that color can serve as a useful feature for a number of different tasks in outdoor machine vision.

References

[1] J.R. Beveridge, D. Panda and T. Yachik, *November 1993 Fort Carson RSTA Data Collection*, Colorado State University Technical Report CSS-94-118, 1994.

[2] T.E. Boult and G. Wolberg, "Correcting Chromatic Aberrations Using Image Warping", *DARPA Image Understanding Workshop*, 1992.

[3] L. Breiman, J.H. Friedman, R.A. Olshen and C.J. Stone, *Classification and Regression Trees*, Belmont, CA: Wadsworth International Group, 1984.

[4] C.E. Brodley and P.E. Utgoff, "Multivariate decision trees", *Machine Learning*, 1995.

[5] T.A. Brown and J. Koplowitz, "The weighted nearest neighbor rule for class dependent sample sizes", *IEEE Transactions of Information Theory*, 25:617-619, 1979.

[6] G. Buchsbaum, "A Spatial Processor Model for Object Colour Perception", *Journal of the Franklin Institute*, 310:1-26, 1980.

[7] S. Buluswar, *Trichromatic model of Daylight Variation*, University of Massachusetts Computer Science Department, technical report, UM-CS-1995-012.

[8] H.R. Condit and F. Grum, "Spectral Energy Distribution of Daylight" *Journal of the Optical Society of America*, 54(7):937-944, 1964.

[9] J. Crisman and C. Thorpe, "Color Vision for Road Following", *Vision and Navigation: The Carnegie Mellon NAVLAB*, Kluwer, 1990.

[10] J.E. Dayhoff, *Neural Network Architectures*, Van Nostrand Reinhold, New York, 1990.

[11] E. D. Dickmanns and B. D. Mysliwetz, "Recursive 3-D road and relative ego-state recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):199-213, 1992.

[12] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*, New York: Wiley & Sons, 1973.

[13] G.D. Finlayson, "Color Constancy in Diagonal Chromaticity Space", *Proceedings of the Fifth International Conference on Computer Vision*, 1995.

[14] G.D. Finlayson, B.V. Funt and K. Barnard, "Color Constancy Under Varying Illumination", *Proceedings of the Fifth International Conference on Computer Vision*, 1995.

[15] D. Forsyth."A Novel Approach for Color Constancy", *International Journal of Computer Vision*, 5:5-36, 1990.

[16] W. Freeman and D. Brainard, "Bayesian Decision Theory: the maximum local mass estimate", *Proceedings of the Fifth International Conference on Computer Vision*, 1995.

[17] B.V. Funt, G.D. Finlayson, "The State of Computational Color Constancy", *Proceedings of the First Pan-Chromatic Conference*, Inter-Society Color Council, 1995.

[18] B.V. Funt and M.S. Drew, "Color Space Analysis of Mutual Illumination", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:1319-1326, 1993.

[19] R. Gershon, A. Jepson and J. Tsotsos, *The Effects of Ambient Illumination on the Structure of Shadows in Chromatic Images*. RBCV-TR-86-9, Dept. of Computer Science, University of Toronto, 1986.

[20] F.S. Hill, *Computer Graphics*, Macmillan, New York, 1990.

[21] B.K.P. Horn, *Robot Vision*, MIT Press, Cambridge, MA, 1987.

[22] D. Judd, D. MacAdam and G. Wyszecki, "Spectral Distribution of Typical Daylight as a Function of Correlated Color Temperature", *Journal of the Optical Society of America*, 54(8):1031-1040, 1964.

[23] G.J. Klinker, S.A. Shafer and T. Kanade, "Color image analysis with an intrinsic reflection model", *Proceedings of the International Conference on Computer Vision*, 1988.

[24] K. Kluge and C. Thorpe, "Representation and recovery of road geometry in YARF", *Intelligent Vehicles*, 1992.

[25] E.H. Land, "Lightness and Retinex Theory", *Scientific American*, 237(6):108-129, December 1977.

[26] S.W. Lee, *Understanding of Surface Reflections in Computer Vision by Color and Multiple Views*, Ph.D. Dissertation, University of Pennsylvania, 1992.

[27] Lockheed-Martin Corp., from DARPA UGV DEMO-C, 1995.

[28] L.T. Maloney and B.A. Wandell, "Color Constancy: A Method for Recovering Surface Spectral Reflectance", *Journal of the Optical Society of America*, A3:29-33, 1986.

[29] I. Masaki, *Vision-based Vehicle Guidance*, Springer-Verlag, 1992.

[30] L. Matthies, A. Kelly, T. Litwin and G. Tharp, "Obstacle detection for unmanned ground vehicles: A progress report", *Intelligent Vehicles*, 1995.

[31] M. Mitchell, *An Introduction to Genetic Algorithms*, MIT Press, 1996.

[32] S.K. Nayar, K. Ikeuchi and T. Kanade, "Determining Shape and Reflectance of Hybrid Surfaces by Photometric Sampling", *IEEE Transactions on Robotics and Automation*, 6:418-431, 1990.

[33] N.J. Nilsson, *Learning Machines*, New York: McGraw Hill, 1965.

[34] C. Novak, S. Shafer and R. Wilson, "Obtaining Accurate Color Images for Machine Vision Research", *Proceedings of the SPIE*, v 1250, 1990.

[35] C. Novak and S. Shafer, *A Method for Estimating Scene Parameters from Color Histograms*, Carnegie Mellon University School of Computer Science, technical report, CMU-CS-93-177, 1993.

[36] C. Novak, "Supervised Color Constancy for Machine Vision", *Proceedings of the SPIE: Conference on Visual Processing and Digital Display*, 1991.

[37] Y. Ohta and Y. Hayashi, "Recovery of Illuminant and Surface Colors from Images Based on the CIE Daylight", *Proceedings of the Third European Conference on Computer Vision*, 1994.

[38] B.T. Phong, "Illumination for Computer Generated Images", *Communications of the ACM*, 18:311-317.

[39] T. Poggio, and F. Girosi, "Regularization algorithms for learning that are equivalent to multilayer networks" *Science*, 247:978-982, 1990.

[40] D.A. Pomerleau, *Neural Network Perception for Mobile Robot Guidance*, Kluwer Academic Publishers, Boston, 1993.

[41] J.R. Quinlan, "Induction of Decision Trees", *Machine Learning*, 1:81-106, 1986.

[42] D.E. Rumelhart, G.E. Hinton and J.L. McLelland, "A general framework for parallel distributed processing", *Parallel Distributed Processing: Explorations in the microstructures of cognition*, Bradford Books/ MIT Press, Cambridge, MA, 1986.

[43] Y. Sato and K. Ikeuchi, "Reflectance analysis under solar illumination", *Proceedings of the IEEE Workshop for Physics-based Modeling in Computer Vision*, 1995.

[44] S.A. Shafer, "Using Color to Separate Reflection Components", *Color Research Application*, 10:210-218, 1985.

[45] J.L. Simonds, "Application of characteristic vector analysis to photographic and optical response data", *Journal of the Optical Society of America*, 53(8), 1963.

[46] C.E. Thorpe, M. Herbert, T. Kanade and S. Shafer, "Vision and navigation for the Carnegie-Mellon NAVLAB", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(3):362-373.

[47] M.J. Vrhel and H.J. Trussell, "Filter considerations in color correction" *IEEE Transactions on Image Processing*, 3:147-161, 1994.

[48] A.M. Waxman, J.J. LeMoigne, L.S. Davis, B. Srinivasan, T.R. Kushner, E. Liang, T. Siddalingaiah, "A Visual Navigation System for Automonous Land Vehicles", *IEEE Transactions on Robotics and Automation* A(3):124-141, 1987.

[49] T. Yachik, "Status of Evaluation, RSTA Workshop", *DARPA Image Understanding Workshop*, 1995.

[50] P. Young, *Recursive Estimation and Time-Series Analysis*, New York: Springer-Verlag, 1984.

[51] T. Young and T. Calvert, *Classification, Estimation and Pattern Recognition*, Elsevier, 1974.

[52] A. Yuille, "A method for computing spectral reflectance", *Biological Cybernetics*, 56:195-201, 1987.

[53] M. D'Zmura, and G. Iverson, "Color Constancy: Basic theory of two stage linear recovery of spectral descriptions for lights and surfaces" *Journal of the Optical Society of America*, A 10:2148-2165, 1993.