Non-parametric Classification of Pixels Under Varying Outdoor Illumination

Shashi D. Buluswar and Bruce A. Draper
Computer Science Department
University of Massachusetts, Amherst

Keywords: **Pixel Classification, Decision Trees, Color, Chromatic Variation, Focus of Attention.**

Abstract

Using color for visual recognition outdoors has proven to be a difficult problem, chiefly due to varying illumination. Attempts to classify pixels or image patches in outdoor scenes often fail, partly because of the paucity of the data, but partly because color shifts due to changes in illumination are not well modeled as random noise. Approaches which attempt to recover the "true color" of objects by calculating the color of the incident light (i.e. color-constancy approaches) appear to work in constrained environments, but are not yet applicable to outdoor scenes.

We present a technique that uses training images of an object under daylight to learn the shift in color of an object. Our method uses multivariate decision trees for piecewise linear approximation of the region corresponding to the object's appearance in color space. We then classify pixels in outdoor scenes depending on whether they fall within this region, and group clusters of target pixels in to regions of interest (ROIs) for a model-based RSTA system. The techniques presented are demonstrated on a challenging task: recognizing camouflaged vehicles in outdoor military scenes.

1. Introduction

Classifying objects based on their color (and in some cases, texture) is one of the oldest and best-studied problems in computer vision and pattern recognition. In indoor settings, where lighting and other imaging conditions can be controlled, maximum-likelihood classifiers recognize objects by modeling variations in color as Gaussian noise in feature space, and assigning every pixel (or image patch) the label with the highest probability. Although these systems have been around for decades, recently Swain has rejuvenated interest in indoor color-based object classification by showing impressive (and real-time) results using color histogram matching [Swain 90].

Unfortunately, using color in outdoor imagery has proven more difficult, enough so that almost all existing military or remote-sensing classification systems rely on non-visual sensors (e.g. FLIR or LADAR). The main problem with (visible) color has been the unpredictability of an object's reflectance in daylight; the color shift between sunny and cloudy days (or between morning and afternoon) is simply not well modeled as Gaussian noise in RGB. Researchers working on *color constancy* attempt to resolve this problem by reconstructing from the image the color of the incident light, and then adjusting the observed reflectances accordingly (e.g. [Forsyth 88]). Unfortunately, their techniques so-far can only be applied in highly-constrained contexts.

Using the (admittedly crude) standard models of the color reflectance and earlier quantitative studies of natural lighting, we argue that the observed reflectances of an object in outdoor scenes should trace out a smooth (sometimes piecewise-linear) surface in RGB. The best approach to classifying objects in outdoor images is therefore to learn this daylight reflectance surface for each object, and classify pixels according to their distance from this surface. As a first approximation to this strategy, we use multivariate decision trees to learn a piecewise-linear volumetric approximation in RGB to the reflectance curve of an object, and classify pixels according to whether they fall inside of outside this volume. The technique is applied to recognizing

camouflaged vehicles in the Ft. Carson image set [Beveridge 94].

2. The Standard Model of Color Variation

The most significant deterrent to the use of color in outdoor settings has been the natural chromatic variation in daylight, which results in a significant shift in an object's apparent color. Our approach to this problem is to use multivariate decision trees (MDTs) [Brodley] (a type of recursive, non-parametric classifier) to learn the volume in color space that corresponds to the possible (observed) colors of an object under natural light. Before proceeding, however, we will first review the standard color reflectance models and argue that this approach is reasonable given the underlying physics.

The observed irradiance of an object in a color image depends on 1) the color and intensity of the incident light; 2) the reflectance properties of the object; and 3) the camera parameters. Fortunately, standard models exist for all three phases of this process. (It should be noted that we are not proposing new physics-based models, nor even using the most sophisticated models available. Our intent here is simply to give a qualitative understanding of the underlying physics as it relates to MDTs.) In outdoor images, the incident light is daylight. Extensive studies done in the 60's confirmed that the color of daylight varies along a characteristic curve, called the CIE daylight curve, defined by the following equation [Judd 64]:

$$y = 2.87\,x - 3.0\,x^2 - 0.275$$

Figure 1 shows the daylight curve in a chromaticity diagram. The equation shows changes in the color of daylight to be changes of a single variable, called the *temperature* of the daylight, which is independent of intensity. The incident light in outdoor images therefore lies on a surface in RGB, formed by sweeping out the CIE curve along the intensity axis.

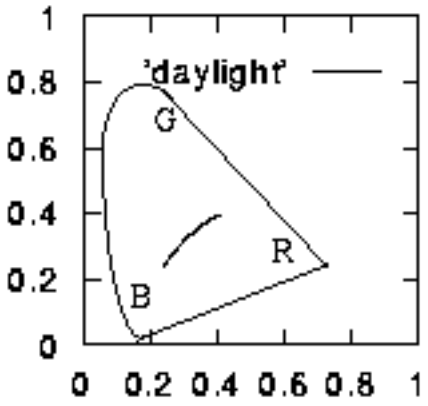

Figure 1: The CIE daylight curve.

Researchers who model the reflectance properties of objects generally assume a lambertian color model, in which the surface reflectance of an object is strictly a function of the wavelength λ of the incident light, given as:

$$s(\lambda) = e(\lambda)\rho(\lambda)$$

where $e(\lambda)$ is the intensity of the incident light at wavelength λ, and $\rho(\lambda)$ is the object's albedo function giving the percent of light reflected at each wavelength. It should be noted that this model is highly restrictive and unrealistic for most circumstances. For example, it does not account for inter-reflectance effects, shadowing or specularities. Nonetheless, it is the best available working model of color reflectance.

This basic reflectance model can be made simpler or more sophisticated, depending on the purpose for which it is used. Forsyth, for example, models the sensitivity of the receptors to different wavelengths of light, and adds another term $s^*(\lambda)$ which specifies a residual reflectance (although in practice, he sets this term to zero). The basic properties of the model do not change significantly, however, and for our purposes it will be sufficient to model an object as reflecting a fixed percentage of the incident light in each of the three color bands.

The light reflected off an outdoor lambertian object is therefore a function of the temperature of the daylight and the object's albedo function. To predict the intensity response in each color band

(i.e. the RGB pixel values) we also have to know the intensity of the daylight and the camera parameters. Given the radiance of an object L, the observed intensities depend on the lens diameter **d** and focal length **f** of the camera, and the image position of the object measured as an angle **a** off the optical axis, as given by the standard irradiance equation [Horn 87]:

$$E = L(\pi/4)\ (d/f)^2 \cos 4a$$

The irradiance equation applies equally to all three color bands, so it does not affect the observed color, only its intensity. (This is an approximation: the focal length of a lens is a function of the wavelength λ, creating an effect known as *chromatic aberration*, as discussed in [Boult 92], for example. As a result, the image irradiance equation does imply a slight color shift as a function of image location, but for our purposes we will assume that chromatic aberration is negligible.)

For our purposes, we are not interested in the details of these three equations so much as qualitative aspects of their solutions. The CIE daylight curve predicts that the incident light in an outdoor scene will fall somewhere along a smooth, curved surface in RGB space. The lambertian reflectance model predicts that the object's reflectance will be a distorted version of this surface, where the extent and direction of the distortion depend on the object's albedo. Finally, the observed irradiance is the reflected light surface scaled by the irradiance equation, assuming that the camera parameters are fixed. If we add a small amount of Gaussian noise, we find that the observed reflectances should lie in a thin, smooth, curved volume in color space.

3. Selecting a Classifier

If we completely believed the standard color model outlined above, then we would know all the form of the observation function (a function of temperature, albedo, etc.), and training a maximum-likelihood classifier would be a parameter estimation task. Unfortunately, we know that these models are not that accurate. Inter-reflectance effects, specularities and other unmodeled phenomena will introduce further distortions into this volume. When digital CCD cameras are used, digitizer parameters, pixel saturation effects, and non-linearities in the NTSC signal standard will warp the volume of possible irradiances still further. It therefore seems premature to approach the outdoor classification problem as an explicit parameter estimation problem for a known equation. Nonetheless (and we are admittedly hand-waving here), it seems reasonable to assume that the observed irradiances from a single object under natural lighting will lie in a smooth, thin, irregularly curved volume.

This assumption is by no means trivial. Work done by several researchers (and summarized in [Novak 93]) suggests that dielectric materials with specular reflections produce "T-shaped" or "dog-legged" distributions in color histograms (with no variation in lighting). Where the lambertian model predicts a straight line distribution in color space, dielectrics produce piecewise-linear curves with two segments, one corresponding to diffuse reflection, the other to specular. The curves are not discontinuous, however. If we assume the same effect holds outdoors, the observed reflectance volume should have a point at which it changes direction, but it could still be qualitatively described as a thin, irregularly shaped volume.

Our approach is to assume that we do not know the exact form of the equation governing the observed irradiance of colored objects in outdoor scenes, but that we know have qualitative description of its shape. To recognize objects in outdoor scenes, we therefore need to select a classification scheme that performs well on smooth, thin, irregularly curved volume in feature space. By definition, parametric classifiers are out, since we don't know the underlying equation. Instance-based classifiers were not considered, because intuitively they should perform poorly on thin, curved surfaces. Neural networks (i.e. feed-forward backpropagation nets) were considered and would presumably perform well, but require large numbers of samples for training. Instead, we chose multivariate decision trees (MDTs) [Brodley], which create piecewise-linear approximations to surfaces in feature space by recursively dividing feature space with hyperplanes. For classes characterized by smooth, continuous volumes in feature space, MDTs generally produce good classification results

from fewer training samples than are required by neural networks.

4. Multivariate Decision Trees (A Quick Overview)

We use multivariate decision trees, an approach that recursively subdivides feature space with hyperplanes. Multivariate decision trees are well known in the machine learning community, but have been used less often in computer vision (although, see [Draper 94]).

The basic goal of a multivariate decision tree is to divide feature space into regions such that all the training samples in a region have the same label. In our case, there are only two labels (camouflage and other), so the methodology is to test if all the instances in the current region of feature space have the same label. If so, label the region; if not, find the hyperplane(s) that maximally separates instances of the two labels, divide feature space into two regions using this hyperplane, and recurse on each region.

The hyperplanes used to divide feature space are represented as linear threshold units (LTUs) [Nilsson 65, Duda 73]. Several methods exist for learning the weights in a linear threshold unit. Brodley and Utgoff [Brodley] discuss four such methods: the Recursive Least Squares (RLS) algorithm [Young], the Pocket algorithm [Gallant], Thermal Training [Frean], and CART's coefficient learning method [Breiman 84]. Because we are concerned only with two-class classification in this domain, the RLS training method is used in this paper (see [Young 84] for a description of training LTUs for two-class classification, and [Draper 94] for a description of multi-class classification using Frean's thermal training rule [Frean 90])

Like other non-parametric learning techniques, decision trees are susceptible to overtraining. In decision trees, overfitting manifests itself in the form of nodes which classify noise training pixels, resulting in deteriorated overall performance. In order to correct for overfitting, a fully grown tree can pruned 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 [Breiman 84; Quinlan 86; Brodley].

5. Finding Camouflaged Vehicles

The color model described in sections 2 and 3 suggests that multivariate decision trees should be a good technique for classifying objects in outdoor images. To test this hypothesis, we tested MDTs ability to identify military vehicles in the Ft. Carson data set [Beveridge 94]. This is obviously a difficult task: not only are these outdoor images with natural lighting, but the vehicles have been intentionally camouflaged! In addition, we are classifying pixels from only their RGB values.

The motivation for attempting such a challenging task is the need for a rapid focus of attention (FOA) mechanism for RSTA [Beveridge]. Any classification scheme that labels pixels based on only their RGB values can be loaded into a 24-to-1 bit lookup table, and applied in real time. Obviously, the data is so impoverished that some errors are inevitable, but if most pixels can be accurately classified then contiguous groupings of target pixels can be used as regions of interest (ROIs) for the RSTA identification and matching algorithms to process. The vital question is whether pixels can be classified accurately enough to produce "useful" ROIs.

Tests were conducted on a set of 45 images from the Fort Carson data set. (This is essentially all the RGB images except for calibration images and a few images that are excessively dark; [Beveridge 94] describes the complete data set and how it can be obtained. The image numbers in all tables below are in reference to Table 1, 3, 5, 7, 10 and 12 in that report.) The objects being sought were camouflaged military vehicles in a natural setting. There are four different army vehicles in these images: two armored personal carriers, a pickup truck, and an M60 tank; all four bear slightly different colors and patterns of camouflage.

Quantitative evaluation of the performance of the classifier was done at two levels: pixel and ROI. At the pixel level, classification accuracy is a poor evaluation criterion; since 99.6 % of the pixels are

Image	Target Pixel Count	% Correct	Background Pixel Count	% Correct
04	603	56.55	279738	98.98
05	615	86.50	313059	99.96
06	644	84.93	307382	99.88
07	764	78.40	307266	99.92
08	873	66.32	307995	99.95
09	1397	46.67	282897	99.98
10	1266	46.45	288263	99.91
11	1022	65.75	277300	99.89
12	1102	65.61	277886	99.81
14	934	61.35	278061	99.98
15	870	58.05	301750	99.84
17	779	45.83	286634	99.82
18	732	53.01	289372	99.84
19	802	60.72	287322	99.81
20	846	63.12	287231	99.84
21	793	46.03	287432	99.93
22	905	50.39	298621	99.90
23	964	50.21	298532	99.92
26	924	50.62	299367	99.73
28	988	47.06	298864	99.77
40	1521	40.89	228325	99.87
41	1375	42.47	273360	99.92
42	1980	51.77	239424	99.87
43	1710	77.49	239424	99.92
44	1299	44.57	239611	99.90
45	1343	69.91	230347	99.75
46	1567	82.39	230787	99.69
47	1167	88.43	259595	99.90
49	1075	81.58	247355	99.92
50	1568	68.05	276292	99.89
51	868	63.48	297963	99.84
52	948	49.05	284632	99.89
54	923	68.58	295735	99.90
60	1273	76.75	265536	99.92
61	1354	79.32	267464	99.92
62	1305	84.21	267464	99.92
63	1273	56.95	267599	99.91
66	750	73.20	267663	99.97
67	936	69.76	218846	99.82
68	936	40.10	218846	99.82
69	973	65.16	218988	99.56
70	874	62.13	218353	99.47
78	213	72.30	232212	99.81
84	245	83.27	232212	99.98
92	1021	88.44	209882	99.99
Totals	46320	69.39	12083157	99.84

Table 1. Ability of Multivariate Decision Trees to separate target pixels (of camouflaged vehicles) from background pixels for the Ft. Carson images. Image numbers are with respect to Table 1, 3, 5, 7, 9, 10, and 12 in [Beveridge 94].

background, a classifier that labeled nothing as target would be highly accurate. Instead, we evaluate our system in terms of its accuracy on target pixels ("true positives" vs. "false negatives") and its accuracy on background pixels ("true negatives" vs. "false positives").

The pixel-wise performance of the MDT classifier on the Ft. Carson images is given in Table 1, broken out by image as well as by the evaluation measures listed above. On the average, 69.39% of the target pixels and 99.84% of the target pixels were correctly classified. These results shown were obtained using cross-validation, where the decision tree was trained on approximately half the images and tested on the other half. In order to train the decision trees, approximately 35,000 sample pixel values were extracted from targets and background from the training images. (Unfortunately, a bias was discovered in how background pixels were sampled for training and for compiling Table 1, which selected against background pixels on the same column as a target. Although we do not believe this will significantly alter the results, the tests are being rerun.) Figure 2 shows images 47, 04 and 66 with the corresponding classified binary images; these figures show the best, average and worst results (in terms of qualitative accuracy), respectively.

Table 1 provides a measure of the multivariate decision tree as a classifier for this domain, testing the argument about MDT's ability to compensate for natural chromatic variation made in Sections 2 & 3. Another way to test the validity of this argument is to look at the distributions of target and background pixels in the Ft. Carson data, and the region of color-space associated with target pixels learned by MDT. Figures 3 and 4 show the target and background pixels in a chromaticity diagram; Figure 5 shows the region of space learned by the MDT, also projected onto a chromaticity diagram. Note that large portions of the space are occupied by neither target not background pixels, so the label associated with these parts of color space is arbitrary. In this example, MDT assigns much of this space the class "target".

A region-level evaluation, on the other hand, certifies how well MDTs perform at generating regions of interest for a RSTA system. ROIs were generated from the pixel-label data by finding connected groups of target pixels that were at least ten pixels high and wide, and that had a maximum aspect ratio (height to width or width to height) of 10:1. ROIs whose bounding rectangles overlapped were then merged. For ROIs, there are two relevant evaluation criteria: the probability that a ROI will be generated for any target, and the probability given an ROI that it contains a target. For the Ft. Carson data, there were 116 occurrences of targets in the 45 test images. The MDT classifier produced 149 ROIs, of which 113 contained targets; 3 targets were not recognized. Table 2 shows the performance of MDT at generating ROIs on a per-image basis.

Image	Targets	Targets Found	False alarms
04	4	4	0
05	4	4	0
06	4	4	1
07	4	4	0
08	4	3	3
09	4	4	0
10	4	4	2
11	4	4	0
12	4	4	0
14	4	4	0
15	4	4	0
17	4	4	0
18	4	4	0
19	4	3	2
20	4	4	1
21	4	3	0
22	4	4	0
23	4	4	3
26	4	4	0
28	4	4	0
40	2	2	0
41	2	2	0
42	3	3	0
43	3	3	0
44	2	2	0
45	2	2	1
46	1	1	0
47	1	1	1
49	1	1	0
50	1	1	0
51	1	1	1
52	1	1	1
54	1	1	1
60	1	1	1
61	2	2	1
62	2	2	0
63	1	1	0
66	1	1	4
67	1	1	2
68	1	1	3
69	1	1	3
70	1	1	3
78	1	1	0
84	1	1	0
92	2	2	2
Totals	116	113	36

6. Future Work and Conclusions

The experiments shown here are preliminary but very promising. The methods described here appear to work well on a difficult set of outdoor scenes. We are studying other function approximation and classification schemes such as neural networks, and are gathering extensive data in order to further test the robustness of our approach.

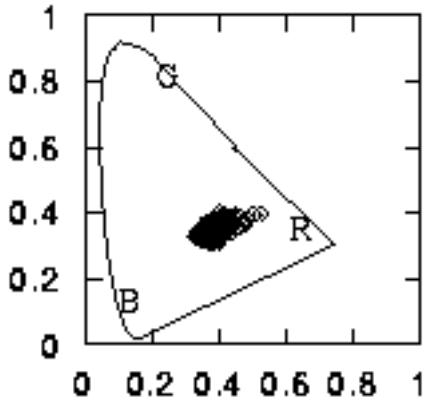

Figure 3: Target pixel distribution

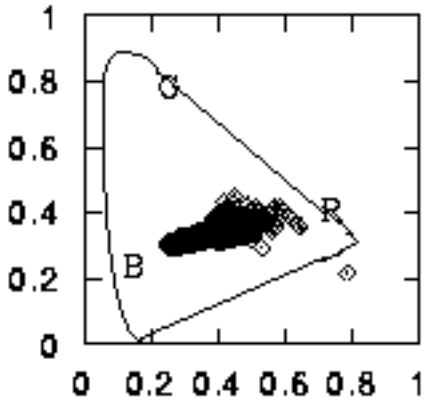

Figure 4: Background pixel distribution.

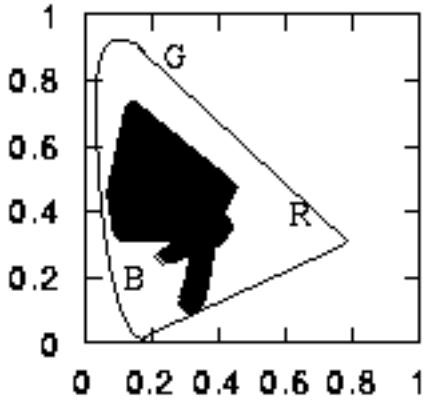

Figure 5: The region of color space (projected onto a chromaticity diagram) labeled "target" by the MDT.

References

Beveridge, J.R., Panda, D. and Yachik, T., *November 1993 Fort Carson RSTA Data Collection*, Colorado State Univeristy Dept. of Computer Science technical report CSS-94-118, 1994.

Boult, T.E. and Wolberg, G. "Correcting Chromatic Aberrations Using Image Warping," *DARPA Image Understanding Workshop*, San Diego, CA., Jan. 1992. pp. 363-377.

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

Brodley, C.E. and Utgoff, P.E., "Multivariate Decision Trees," *Machine Learning*, to appear.

Draper, B.A., Brodley, C. and Utgoff, P.E., "Goal-directed Classification Using Linear Machine Decision Trees," *IEEE Pattern Analysis and Machine Intelligence*, to appear.

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

Forsyth, D.A., *Colour Constancy and its Applications in Machine Vision*, PhD thesis, University of Oxford, 1988.

Frean, M. Small Nets and Short Paths: *Optimizing Neural Computation*, PhD thesis, University of Edinburgh, 1990.

Gallant, S.I., "Optimal Learning Discriminants," *Proceedings of the International Conference on Pattern Recognition* (pp. 849-852), 1986.

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

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

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

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

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

Swain, M. *Color Indexing*, Doctoral dissertation, University of Rochester, 1990.

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

(a) Image 47: best performance

(b) Image 40: average performance

(c) Image 58: worst performance

Figure 5: Source and classification images