Color Recognition by Learning: ATR in Color Images

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

Abstract

Traditional methods for ATR (Automatic Target Recognition) use infrared (IR) sensors for detecting heat emanating from targets. IR-based ATR techniques are susceptible to sensor-induced errors; for instance, targets may not be detected if they are cold (when vehicle engines are turned off), or when the background is hot (on a hot day).

This work presents an approach to real-time color-based ATR which uses multivariate decision trees for recursive non-parametric function approximation to learn the color of a target from training samples, and then detects targets by classifying pixels based on the approximated function. Tests of the color-based system, sanctioned by the U.S. Defense Advanced Research Projects Agency - Unmanned Ground Vehicle Project (DARPA-UGV), have resulted in a 90% target detection rate (compared to the 45% detection rate of the IR-based system developed for the same tests). When the color system was used in conjunction with the IR-based system, 100% of the targets were detected.

1 Introduction

Traditional military ground-level Automatic Target Recognition (ATR) systems analyze IR images for the signatures of potential targets. Although such systems have proven quite successful in wide-spread use, they fail in certain predictable scenarios, notably when the targets are colder than expected, or when the background is hotter than expected (see figure 1). One approach to this problem is to develop more sophisticated target recognition algorithms for IR images (Schachter [17] contains a review of several methods). It is our belief, however, that the gains possible through this line of research are limited due to problems inherent to the data. A more promising approach, we believe, is to collect additional data using non-IR sensors, and to look for target signatures there. The issues with this approach are cost and independence (in the sense that ATR on the additional data should succeed in scenarios where the IR-based system fails).

*Supported by DARPA through Rome Labs under contract F30602-94-C-0042.

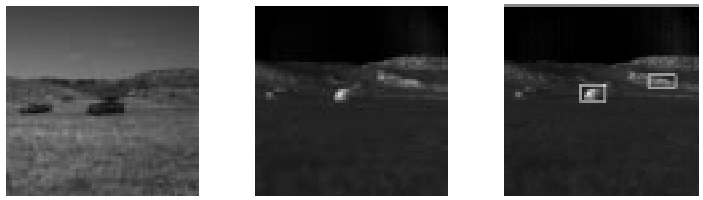

Figure 1: Visible light (left) and IR (middle) images of targets. In the IR image, not all targets are visible, and some parts of the background are as bright as the targets. The results (right) from applying the DARPA-UGV Demo-C IR ATR system show one missed target and one false positive.

This paper presents an alternative approach to ATR that uses color imagery. There are several advantages to using color (described later) which enable our system to be used either in stand-alone mode, or with systems based on other sensors. We should emphasize that we are suggesting supplementing – not replacing – IR-based ATR systems. IR systems work well in many scenarios and are already in wide-spread use; color-based systems (or any other method based on visible spectrum data), on the other hand, cannot ordinarily be used at night. However, at least one of the scenarios in which IR systems fail (i.e., due to background heat) is an typically daytime scenario, when color-based systems should be most reliable.

Color-based target recognition is inherently difficult, due to (i) the camouflage on targets, and (ii) variation in the apparent color of objects under outdoor imaging conditions. Camouflage, is, of course, the standard counter-measure against detection in visible light, and it forces any color-based ATR system to make very fine distinctions in order to separate target from background. However, the color of background vegetation continually changes, so it is difficult, if not impossible, for camouflage color to perfectly match the background; furthermore, mismatches in color between target and background are made even more common by the multiple colors used.

The apparent color of a given target (or object) varies under outdoor conditions due to a number of factors, namely the color of the incident daylight, surface reflectance properties of the target, illumination geometry (i.e., the position and orientation of the target surface w.r.t. the illuminant) and viewing geometry (the position and orientation of the camera w.r.t. the target surface). The color of daylight changes significantly due to the sun-angle and weather conditions, and the position and orientation of the target are also expected to vary. Consequently, the apparent color of a target varies under realistic conditions. Previous methods in computational color recognition, such as color constancy algorithms [18, 7, 6], have dealt with varying color in highly constrained environments, and are generally not applicable to outdoor imagery.

It will be shown that as imaging conditions vary, the apparent color of objects forms characteristic types of clusters in color *RGB* space, depending on the surface properties. The method presented here uses multivariate decision trees (MDT's) for recursive, non-parametric function approximation to estimate the clusters in *RGB*, based on training samples of targets. Given samples of a target under different lighting conditions, MDT's construct a piece-wise linear approximation of the boundary of the region in color space. After the training phase is complete, every image pixel can be classified as target or background according to whether or not it lies within the learned boundary. The result is a binary region-of-interest image that marks all the pixels that lie within the region in color space occupied by the object's representation; the target pixels in the binary images are then grouped to produce bounding rectangles around the targets. The RGB representation of color makes it possible to use a lookup table for real-time classification on standard hardware.

This method has been implemented in a system for ATR of camouflaged military vehicles in real-time, and has been tested in a DARPA-sanctioned study [19] on the Ft. Carson data set [1] and at the DARPA UGV Demo-C [11]. In each test, over 90% of the targets were detected (compared to a 45% detection rate by the IR-based system). A combination of color and IR systems resulted in detection of nearly 100% of the targets.

2 IR-based ATR

IR-based ATR systems detect targets based on the heat emanated in the 800-1200 nm range. They offer the clear advantage of being useful at any time, day or night, and can be used in many types of smoke and fog. Most IR-based ATR systems assume that the targets are warmer than the background [17] (or have characteristic heat signatures w.r.t. the background), and can therefore fail when the target heat, relative to the background, varies unpredictably. This can happen when the engines of a target vehicle have been shut off (possibly making the target as cool as the background), or when objects in the the background (such as rocks on a hot day) are also warm. Figure 1 shows both these situations encountered in a single image from the Ft. Carson data. In this scene, there are two targets; however, only one is clearly visible in the IR image. In addition, part of the background appears almost as bright as the target. Such problems are not uncommon in IR imagery; furthermore, when vehicle structural design is similar, there is no easy way to distinguish between military and civilian vehicles or enemy and friendly vehicles, since they are likely to generate similar IR signatures.

The IR-based ATR system used at the DARPA UGV Demo-C is based on double-window detection [11, 17]. Using this method on 25 randomly chosen images from the Ft. Carson IR data set, only 22 out of 50 targets were detected, with 5 false alarms; in addition, the only civilian vehicle in the image set was mistaken for a target. A representative result is shown in figure 1. While other IR-based techniques have been proposed [17], there is no strong evidence to indicate that these techniques can overcome the problems inherent to IR data.

3 Color imagery for ATR

This paper advocates using color to enhance ATR systems. Color imagery offers a number of advantages: (i) the data is inexpensive to obtain (color cameras are cheap and freely available, and many prototype research vehicles are equipped with them), (ii) we have developed methods for real-time target detection shown

Figure 2: (From left) Sample target; Target color (RGB) in a single outdoor image (sample extracted from vehicle hood); Variation of apparent color over several hundred images in a single day; Variation distribution rotated.

to be effective under most naturally occurring daytime conditions, and (iii) the system can be easily combined with systems based on IR (or other sensory) data for even more reliable performance.

3.1 Problems with using color for ATR: variation of apparent color

While color can be a useful feature for target detection, there are several issues that complicate the use of color for recognition, especially in outdoor images. One clear disadvantage of using color (or any other feature from the visible spectrum) for ATR is that it cannot be used at night, in thick smoke or fog, or any conditions under which the targets are not visible. Additionally, there are other problems inherent to outdoor color imagery that further complicate color-based recognition.

The apparent color of an object is a function of the color of the incident light, surface reflection, illumination geometry, viewing geometry and imaging parameters [9]. Each of these factors can vary in outdoor conditions; in addition, the effect of a host of unmodeled phenomena, such as shadows and inter-reflections, is unpredictable. 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 2 shows a camouflaged military vehicle, with its apparent color in RGB space in a single image (which is a single point in RGB) and the variation over 100 images in one day.

The variation in the color of daylight is caused by changes in the sun-angle, cloud cover, atmospheric haze and other weather conditions. The illumination geometry in a scene determines the orientation of the surface with respect to the two components of the illuminant, sunlight and (ambient) skylight, and hence the color of the incident light. Viewing geometry, i.e., the position and orientation of the camera with respect to the surface, determines the amount and composition of the light reaching the camera, depending on the specular content of the surface. Shadows and inter-reflections also affect the color of the light incident upon a surface [8]. Shadowing occurs either when the surface is facing away from the sun (self-shadowing), or when a second object blocks the sunlight. Inter-reflections are caused when other surfaces reflect light incident upon them, onto the surface in question. In both cases, the color of the incident light (and hence the apparent color of the surface) is affected. A number of imaging parameters cause further color shifts. For instance, wavelength-dependent displacement of light rays by the camera lens onto the image plane due to chromatic aberration can cause color mixing and blurring [14]. Nonlinear camera response and digitization errors can skew the ratio of the values in the three color bands (red, green and blue), and the dynamic range of intensity in outdoor scenes accentuates the possibility of blooming and clipping [14].

3.2 Previous approaches to color vision under varying illumination

In the past, color recognition under varying illumination has generally been addressed as a color constancy problem, where the goal is to match object colors under varying illumination without knowing the spectral composition of the incident light or surface reflectance. An illuminant-invariant measure of surface reflectance is recovered by first determining the properties of the illuminant from variations across images. Unfortunately, in order to separate illumination conditions from surface reflectance effects, most color constancy algorithms make strong assumptions about the nature of the world. For example, Forsyth [7] assumes a Mondrian world with constant illumination without inter-reflections or multiple light sources; Finlayson [6] assumes that surfaces with the same reflectance have been identified in two spatially distinct parts of the image, and that the unknown illumination falls within the gamut of known artificial illuminants; Ohta [16] assumes artificial illumination constrained by the CIE model to reduce performance errors; Novak and Shafer [15, 18] assume a point light source and pure specular reflection; Buchsbaum [3] assumes that the surface reflectance averaged over the entire image is grey; Maloney's work [12] is a refinement of Buchsbaum's but has been applied only under the constraints of an indoor world with Munsell color chips. While many of these constancy algorithms are quite sophisticated and perform impressively within the specified constraints, Forsyth [7] 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."*

3.3 The nature of the variation of apparent object color in outdoor scenes

According to the standard model of image formation [9], the observed color of objects in images is a function of (i) the color of the incident light (daylight), (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 the various phases of the image formation process [9, 10, 13, 18], although these models appear too restrictive to be used in unconstrained imagery; still, they provide an approximate qualitative description of the variation of apparent color. The CIE model [10] states that the color of daylight varies along a characteristic curve, defined by the following equation in the CIE chromaticity space (of which RGB is a linear transform).

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

where $0.25 <= x <= 0.38$. In RGB space, the parabola stretches out into a thin curved surface [4].

The effect of illumination geometry and viewing geometry depend on the reflectance of the surface. Most realistic surfaces have reflectances that have a mixture lambertian and specular components. Existing reflectance models of mixed reflection surfaces [9, 13, 18] are yet be applied to unconstrained imagery in the context of color-based recognition. We can, however, deduce from the CIE model and the reflection models, that the RGB distribution representing apparent color variation of a surface under daylight will lie along (a) a thin continuous curved volume if the surface is purely lambertian or purely specular, (b) a single blob if the surface has mixed reflection with a dominant lambertian component, and (c) two distinct clusters if the surface has mixed reflection with a dominant specular component.

The goal of imaging systems is to preserve the color of objects as they appear in the scene, depending on a few imaging parameters (focal length, response function, etc.). Unfortunately, phenomena such as clipping, blooming and nonlinearities will introduce distortions to the appearance of objects in color space [14].

Even if we assume that the distortions to the RGB distributions due to unmodeled parameters are not drastic, in the absence of precise and robust models of the various processes involved (as is the case with outdoor color images), the only assumption that can be made is that the RGB distributions representing the color of objects can be arbitrarily shaped.

4 Multivariate Decision Trees (MDT) for learning target color

Our approach is to assume that we do not know the exact form of the equation governing the observed color of objects in outdoor scenes. To recognize targets in outdoor scenes, we therefore need to select a classification scheme that performs well on arbitrarily shaped clusters in feature space. By definition, parametric classifiers (such as minimum-distance classifiers, as used by Crisman [5]) can be ruled out, since the underlying equations are unknown. Based on their success in other areas of non-parametric approximation, neural networks (i.e., feed-forward back-propagation nets) and multivariate decision trees were considered. Neural nets would presumably perform accurate nonlinear function approximation, but are difficult to analyze because of the arbitrary nature of the function approximated by the hidden layer. Multivariate decision trees create piecewise-linear approximations to surfaces in feature space by recursively dividing feature space with hyperplanes, and have been shown to produce good classification results from relatively few training samples.

Multivariate decision trees [2] recursively subdivide the feature space by linear threshold units (LTU's). Each LTU is a binary test 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 the two resulting subsets are linearly separable, a single LTU will separate them and the multivariate decision tree consists of the single node. If not (as is generally

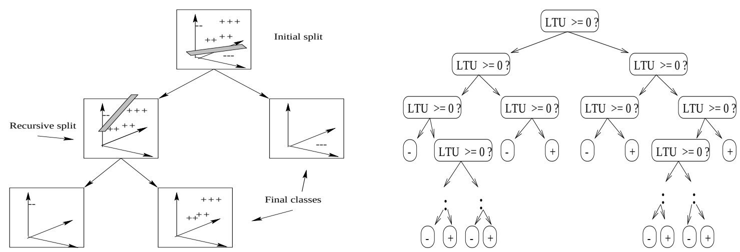

Figure 3: Recursive discriminants of an MDT, separating the '+'s from the '-'s (left), and MDT structure with LTU's and final classes (right).

the case with realistic images and objects), the LTU linearly divides the feature space so as to separate the instances to the extent possible, and the MDT creates and trains new LTU's on the two divisions of the instances. The result, therefore, is a tree of LTU's recursively dividing the feature space into 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 3 (left) shows a decision-tree operating in a three-dimensional feature space, where the two classes being separated are the '+'s and the '-'s; Brodley [2] describes further details.

The LTU weights are approximated using the Recursive Least Squares (RLS) 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 $W_k = W_{k-1} - K_k(X_k^T W_{k-1} - y_k)$, 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 $P_k = P_{k-1} - P_{k-1}X_k[1 + X_k^T P_{k-1} X_k]^{-1} X_k^T P_{k-1}$. The weights are initialized randomly, and the matrix consists of 0 values everywhere except along the diagonal, which is set to 10^6 (empirically determined).

If at any level, the LTU results in a non-negative value, the corresponding set of pixels is labeled as belonging to the object (i.e., positive), otherwise, it is labeled negative. If the set of instances at any level can be further divided, the tree is recursively grown; if no further division is possible, that set of instances is represented by a terminal node or a class (with the LTU determining whether the class is positive or negative). Figure 3 (right) 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 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

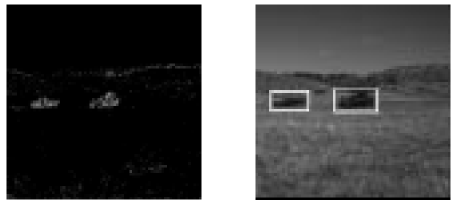

Figure 4: Post-classification binary image (left), with target boundaries extracted (right).

leaf-node results in better performance, the subtree is replaced by it [2].

5 ATR system using MDT

A decision tree for the camouflaged targets is built by providing sample pixels of the targets and background (e.g., vegetation, sky, rocks, etc.) from images taken under various conditions. After the decision tree is built, the next step is to build a lookup table for real-time ATR classification. This is accomplished by classifying (off-line) every possible *RGB* color value into target and background classes. Thereafter, given a color image, each pixel can be classified from the lookup table in real-time. The result of pixel classification is a binary image, in which all suspected target pixels are on (white), and the background pixels off (black). Figure 4 (left) shows the binary post-classification image for the scene from figure 1.

From the binary image, the clusters of target pixels are grouped, and bounding rectangles then extracted. Finally, overlapping bounding rectangles are merged, to produce a region-of-interest image, with the boxes drawn around the targets; figure 4 shows the result of grouping and extracting target regions from the corresponding binary image.

6 Results

The Ft. Carson data set [1], collected in a DARPA-sanctioned study, consists of about 150 color and IR images of camouflaged military vehicles under conditions that vary from bright (and hot) daytime to dark (and cool) dusk; the distance to the targets ranges from 100 to about 500 meters. The two independent systems (color and IR) were tested on corresponding images of 25 randomly chosen scenes from the Ft. Carson set.

The color-based system was applied by cross-validation, where half the images were used for training and the other half for testing (with rotation, so that all 25 images were used). In this test, 47 out of 50 targets were detected, with 39 false alarms. The false alarms were all due to background foliage which was very close in color to the camouflage of the vehicles; in two images with extremely poor lighting conditions the system missed the targets. In addition, the system was tested live at the UGV Demo-C, with similar results (the exact numbers from Demo-C are not available).

By comparison, the IR-based system [11] detected 22 of the 50 targets, with 5 false alarms. Four of the false alarms were from background foliage, and one was a civilian vehicle. Two issues must be noted, however: (a) the failure of the IR system can be attributed to the image quality – the fact that such images were collected in a realistic DARPA exercise goes to show that IR images cannot always be relied upon, even with sophisticated detection techniques; (b) when the color system failed due to poor lighting conditions, the IR system successfully detected the targets. When the two systems were combined, 100% of the targets were detected.

7 Conclusions

This paper describes a method for using color images for highly effective ground-level ATR. Although extensive tests on the Ft. Carson data and at the live UGV Demo-C have been successful, and in some instances, better than IR-based ATR, we do not intend to recommend that color be used in exclusion of other ATR technologies. This work demonstrates that the color-based method described can be used as an effective, yet inexpensive addition to existing systems. The number of false alarms indicates that the method is more useful as a focus-of- attention mechanism, than for full-fledged recognition. The learning and classification method described in this paper has been applied, with similar success, to other problems such as road/lane detection, skin recognition, detection of wildlife in aerial imagery, ground-level terrain detection and landmark recognition. The images and results from the Ft. Carson tests are available at the following world-wide-web address:

`http://vis-www.cs.umass.edu/projects/learning/mdt.html.`

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] C.E. Brodley and P.E. Utgoff, "Multivariate decision trees", *Machine Learning*, 1995, pp 45-57.

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

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

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

[6] G.D. Finlayson, B.V. Funt and K. Barnard, "Color Constancy Under Varying Illumination", *Proceedings of the Fifth International Conference on Computer Vision*, pp 720-725, 1995.

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

[8] 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.

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

[10] 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.

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

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

[13] S.K. Nayar, K. Ikeuchi, and T. Kanade, "Determining shape and reflectance of hybrid surfaces by photometric sampling", *IEEE Transactions on Robotics and Automation* pp 418-431, 1990.

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

[15] 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.

[16] 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*, pp 235-246, 1994.

[17] B.J. Schachter, "A Survey and Evaluation of FLIR Target Detection/Segmentation Algorithms", *DARPA Image Understanding Workshop*, 1982.

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

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