

# Efficient Object Recognition Using Color

Signe Redfield, Michael Nechyba, John G. Harris and Antonio A. Arroyo

March 8, 2001

## Abstract

A real-time prototype system for identifying soda cans uses histogram indexing with only 16 colors, with excellent results. The final implementation identified cans almost perfectly, even on public display with many people attempting to fool it. Results from preliminary versions are presented, with a complete description of the circumstances used in training and testing the system. The final system required only 1.8 kilobytes of storage space for the database, 37 kilobytes of space for the program, and 70 milliseconds to process each image, proving its suitability for implementation on an autonomous mobile robot.

Keywords: histogram indexing, quantization, object recognition, color categories, computer vision

## 1 Introduction

Object recognition using images is generally inefficient. The vast quantities of input data tend to overwhelm systems attempting to extract only useful information. Systems that recognize objects based on shape alone require moderate databases and impose a substantial computational burden. Systems that rely on color have to deal with triple the input information required for shape-based systems, and have difficulty when the lighting changes. However, under controlled conditions, or when the lighting changes are somewhat predictable, efficient object recognition using color is possible. Furthermore, if a truly efficient object recognition algorithm using color can be derived, more expensive techniques can be used on only the subset of data produced by the color recognition process. This could dramatically speed up the recognition process, even when multiple object properties are used as features.

Objects can be identified using many possible features. Here we look exclusively at color, used as the only feature. Obviously, not all databases are suited for this approach. Soda cans, however, form a very good data set for this experiment. They are uniform in shape and thus easily segmented from their background, and they come in a wide variety of colors. In addition, there is a potential use for the system, in the form of a butler robot. If you send your robot to get you an ice cold Coca-Cola, the robot should be able to identify it reliably and not bring you Fresca instead. Our goal was to show that color can

be used to identify objects reliably and with minimal storage and processing requirements.

Shape-based techniques involving images generally require relatively large amounts of storage space and substantial processing time. Using color as the only feature dramatically reduces both storage and processing requirements. For this implementation, we use a slightly modified version of Swain and Ballard's histogram indexing technique[5]. Swain and Ballard showed that under strictly controlled lighting conditions objects could be reliably recognized using color histograms. They used 512 colors to prove their point, but showed little degradation for as few as 64 colors. However, our previous experiments[1, 3, 2, 4] have shown that quantization to approximately 10 colors can produce improved recognition accuracy under varying lighting conditions. As a continuation of that research, we implemented prototype systems using 10, 11, 14, and 16 colors with varying results.

## 2 Overall Structure

The complete prototype consisted of a camera connected via an S-video cable to a workstation running Linux. The computer captured images from the camera, via an OmniMedia Sequence P1S frame grabber. These images were cropped to a specified range of pixels and quantized. The computer generated the corresponding histogram, and then matched that histogram to the histograms in its database. Finally, the soda name corresponding to the closest database histogram was displayed on the screen. Various combinations of soda cans were used to test the system, beginning with 6 cans and gradually expanding to our current total of 14. Two cameras were used, a Sony TRV-310 and a Sharp VL-H860. As long as the database was generated with the same camera used for testing, the results were equivalent.

The camera and the can were assumed to be arranged such that the can occupied a given region of the image. Each 640 by 480 image was cropped to a rectangular region of columns 283 to 389 and rows 170 to 343. When the database was generated, each can was placed in the same location with respect to the camera. Figure 1 shows an image taken with this arrangement, with the cropped region used to generate the histogram outlined in red, and the remainder of the image shaded. This arrangement allows some freedom on the part of the tester with regard to can placement; roughly half an inch to either side and an inch or more of forward/backward movement.

HLS space was used for the quantization step. In HLS space, hue is an angular measurement around the circumference of a cylinder, with lightness along the axis and saturation determined radially. All values vary from zero to one. On the lightness axis, this corresponds to black (zero) to white (one). The saturation axis varies from achromatic (zero) to fully saturated (one). Usually, the hue axis measurement has zero at red, with 0.5 corresponding to an angular measure of 180 degrees, and 1 corresponding to 360 degrees, or red again. Because there are few pink sodas and many red ones, and to simplify visualization



Figure 1: *Sample image of input to system. Region used to create histogram is outlined in red; remainder of image is shadowed.*

of the results, the hue axis was offset by one-sixth. This moved red to 60 degrees and put pink at zero and one.

Quantization was done using a data-independent tree method. This method is shown in Figure 2, for 11 chromatic regions. First, a saturation threshold of 0.15 determined whether the pixel was chromatic or achromatic. If the pixel was achromatic (saturation below 0.15), it was put into one of three achromatic bins. Below a lightness of 0.2, the pixel was put in the black bin. Above a lightness of 0.9, the pixel was put in the white bin. Between the values, the pixel was assumed gray. If the pixel was chromatic, it was placed in one of n chromatic bins. The boundaries for the chromatic bins were determined by uniformly quantizing the hue axis into n regions.

### 3 Off-Line Implementation

The first implementation did not operate in real time. Each image was captured by one program and stored to disk as a file. A second program performed the object recognition component of the system, reading in the information from the image file and displaying the name of the soda corresponding to the matching histogram. For example, if a can of Country-Time Lemonade was placed in the appropriate location in front of the camera, the user would run both programs and the computer would respond This soda is Country-Time Lemonade.

The database consisted of thirty-six histograms representing six cans. Each can was represented by histograms generated from six images, taken in two different orientations (pull-tab towards the camera and pull-tab away from the camera) and three different lighting conditions (overhead fluorescent lighting on

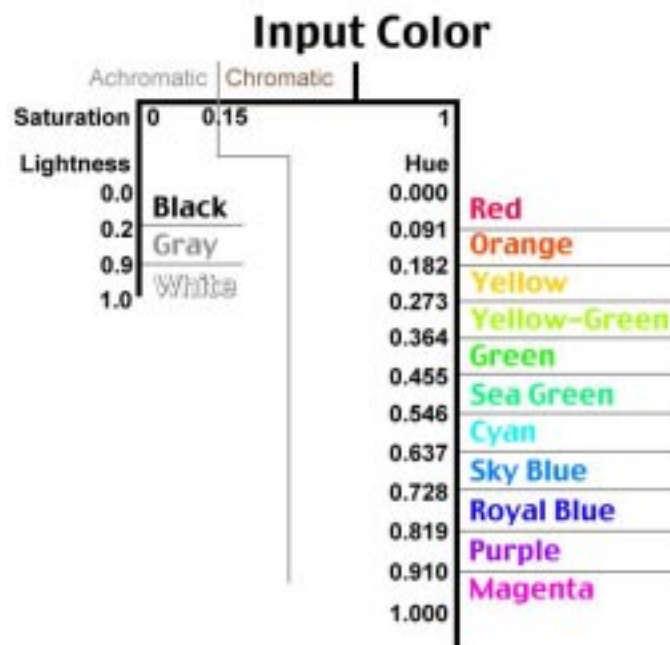


Figure 2: *Diagram for tree quantization.*

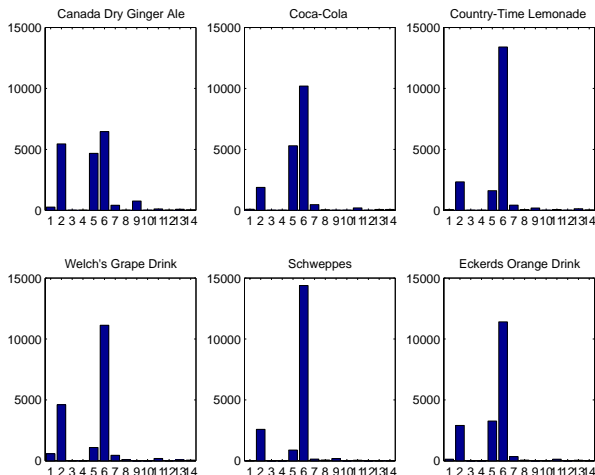


Figure 3: *Sample histograms from database.*

and blinds closed, overheadfluorescent lighting on and blinds open, and overhead fluorescent lighting off and blinds open). Each cropped image was quantized to 14 values ( $n$  above equal to 11), so each histogram consisted of 14 values. Because every histogram was generated from the same number of pixels, the histograms did not need to be normalized, so each value in a given histogram was the number of pixels that fell into that bin during quantization.

The cans used in this initial implementation were Welch's Grape Drink, Eckerds Orange Soda, Country-Time Lemonade, Coca-Cola, Sprite and Canada Dry Ginger Ale. Representative histograms from the database are shown in Figure 3.

### 3.1 Results

The final database was generated in early afternoon and tested in the early evening. The system was robust to combinations of lighting represented in the database. For example, the database contained images under fluorescent lighting with the blinds fully open and fully closed. The system accurately identified cans regardless of whether the blinds were partially open, fully open, or closed. The system also correctly recognized cans upside down or on their sides. When two cans were placed side by side in the imaging area, the system reliably chose one of the two cans as the correct can. The system was less reliable when cans were placed so their bottom or top faced the camera, but still managed to correctly identify the can roughly half the time. However, because the white balance on the camera was on, changing the background reduced the systems accuracy.

The system perfectly identified five of the six cans in the database. How-

Soda	1	Err	2	Err
Sch	21/30	7U	21/30	7U
CT	27/30	CD	0/30	CD
EO	30/30		30/30	
WG	30/30		30/30	
C	15/30	EO	0/30	EO
D7	13/30	CD	0/30	21/30 CD 9/30 7U
CD	1/30	D7	13/30	D7
7U	28/30	Sch	13/30	Sch

Table 1: Number correct when identification of sodas is tested using orientation midway between those in the database. Eight cans with two orientations in database, fourteen colors. The numbers 1 and 2 correspond to orientations; Err indicates what was chosen instead of the correct soda. Codes are: Sch = Schweppes, CT = Country-Time Lemonade, EO = Eckerds Orange Drink, WG = Welch’s Grape Drink, C = Coca-Cola, D7=Diet 7-Up, CD = Canada Dry, and 7U = 7-Up.

ever, Canada Dry Ginger Ale proved exceptionally difficult to recognize. The default orientations images contained no green and few gold pixels. Thus, when green or gold was the predominant color the system routinely misidentified the can as Sprite. Only the exact orientations included in the database were correctly identified. The other cans were always identified correctly, regardless of orientation.

## 4 Real Time Implementation

### 4.1 First Implementation

A second, real-time, system also produced satisfactory results. For this implementation, only seven chromatic bins were used, although the three achromatic bins remained the same. The seven bins, chosen by hand, were pink, red, orange, yellow, green, blue, and purple. All testing and database generation were done under as close to a single lighting condition as possible (blinds fully closed). Initially 8 cans were used with the two orientations as defined in Section /ref-sec:impon. Unfortunately, the system did not perform as well as before, even when no people were allowed into the image area while data was being taken.

### 4.2 Results on First Implementation

Results with two test orientations are shown in Table 1. The test orientations used were the midpoints between the database orientations, corresponding to 90 and 270 degrees with respect to the pull tab (database orientations correspond to 0 and 180 degrees). While the system identified Eckerds Orange Soda and

Soda	(1)	Err	(2)	Err	(3)	Err	(4)	Err
Sch	20%	7U	98%	7U	100%		54%	7U
CT	100%		100%		100%		100%	
EO	100%		100%		100%		100%	
WG	100%		100%		100%		100%	
C	100%		100%		100%		100%	
D7	100%		58%	CD	46%	CD	82%	CD
CD	100%		100%		62%	D7	100%	
7U	44%	D7	10%	Sch	2%	4% D7	0%	Sch
						94% Sch		

Table 2: Number correct when identification of sodas is tested using orientation midway between those in the database. Eight cans with four orientations in the database, eleven colors. The numbers 1, 2, 3, and 4 correspond to orientations; Err indicates what was chosen instead of the correct soda. Codes are: Sch = Schweppes, CT = Country-Time Lemonade, EO = Eckerds Orange Drink, WG = Welch’s Grape Drink, C = Coca-Cola, D7=Diet 7-Up, CD = Canada Dry, and 7U = 7-Up.

Welchs Grape Drink without error, the results on the remaining sodas were remarkably poor. A few seconds delay was introduced to allow the camera time to adjust the white balance, without improvement.

Accuracy improved dramatically when four orientations were included in the database. Again, the intermediate orientations were used to test the database. The four orientations included were at 0, 90, 180 and 270 degrees (with respect to the pull tab). The test orientations were at 45 degrees from each of these. One additional chromatic category was added to compensate for consistently large values in the blue bin over all cans. The results are shown in Table 2. Each number in the table corresponds to the results over 50 trials without changing the environment or the camera. Obviously Schweppes, Diet 7-Up, and 7-Up had the worst accuracy, and Schweppes, Diet 7-up, 7-Up, and Canada Dry were responsible for the most errors. Diet 7-Up and 7-Up were removed from the database for the next trial, replaced by Pepsi, Slice, Mountain Dew and Lipton Diet Lemon Brisk Iced Tea. This was intended to reduce errors resulting from misclassification of green pixels and determine whether the blue category was subject to the same problems. Results on this new arrangement are shown in Table 3, with an average accuracy of 86%. Again, Canada Dry (now with Mountain Dew) were responsible for many errors. Other than Mountain Dew and Canada Dry, Eckerds Orange Drink produced the only major error. This occurred when the nutritional information was facing the camera. The orange from this soda and the red from Coca-Cola were largely in the same histogram bin, so when no blue was present, the system incorrectly identified the can. When Canada Dry and Mountain Dew were replaced by 7-Up, Diet Pepsi, and Wild Cherry Pepsi, the results were unreliable.

Soda	(1) Err	(2) Err	(3) Err	(4) Err
Sch	100%	94% MD	100%	100%
CT	100%	100%	100%	100%
EO	100%	100%	16% C	100%
WG	100%	100%	100%	100%
C	100%	100%	100%	100%
CD	58% MD	10% MD	100%	0% MD
MD	16% CD	100%	68% CD	0% CD
P	100%	100%	98% WL	100%
SL	100%	100%	100%	100%
LDL	98% P	100%	100%	100%

Table 3: Number correct when identification of sodas is tested using orientation midway between those in the database. Ten cans with four orientations in the database, eleven colors. The numbers 1, 2, 3, and 4 correspond to orientations; Err indicates what was chosen instead of the correct soda. Codes are: Sch = Schweppes, CT = Country-Time Lemonade, EO = Eckerds Orange Drink, WG = Welch’s Grape Drink, C = Coca-Cola, CD = Canada Dry, MD = Mountain Dew, P = Pepsi, SL = Slice, and LDL = Lipton Diet Lemon Brisk Iced Tea.

### 4.3 Second Implementation

A third version of the system was created for public display. This system was also real-time, although the display function became completely automatic. A one second delay between each reading made viewing easier. This system used 16 histogram bins, 3 achromatic and 13 chromatic. At first only the eleven cans from the final second implementation were used, with results shown in Table 4. The error was significantly reduced, with Eckerds Orange nutritional information producing the only major remaining error. The average accuracy increased to 95%. Finally, Eckerds was replaced with Minute Maid Orange Soda, and Sprite was added. Results on the Minute Maid soda were far below those for the Eckerds soda (2% in one orientation and 66% in the other, both mistaken for Wild Cherry Pepsi). Average accuracy on the Minute Maid was 67%, compared to 78% for the Eckerds soda. Minor additional errors were introduced for Wild Cherry Pepsi and Coca-Cola, resulting in an increase in error of 1-2% each. Sprite was always recognized correctly, and introduced no new errors.

A new database was generated with the white balance off and fixed background. With the blinds closed but some light leaking in, the database was robust to outdoor lighting changes (overcast, time of day, sunlight). In Table 5, the results for this database show that overall accuracy was very good. Other than Diet Pepsi, the only errors occurred when the nutritional information on a given can was facing the camera. Average accuracy (including Diet Pepsi) was 92.2%. Excluding the results from Diet Pepsi, accuracy increased to 94.73%.



Soda	(1)	Err	(2)	Err	(3)	Err	(4)	Err
Sch	100%		100%		100%		100%	
CT	100%		100%		100%		100%	
EO	100%		100%		12%	C	100%	
WG	100%		100%		100%		100%	
C	100%		100%		98%	EO	100%	
P	100%		100%		100%		100%	
SL	100%		100%		100%		100%	
LDL	92%	DP	96%	DP	96%	DP	98%	DP
7U	100%		100%		100%		100%	
WCP	96%	EO	100%		10%	EO	100%	
DP	92%	LDL	100%		100%		100%	

Table 4: Number correct when identification of sodas is tested using orientation midway between those in the database. Eleven cans with four orientations in the database, sixteen colors. The numbers 1, 2, 3, and 4 correspond to orientations; Err indicates what was chosen instead of the correct soda. Codes are: Sch = Schweppes, CT = Country-Time Lemonade, EO = Eckerds Orange Drink, WG = Welch’s Grape Drink, C = Coca-Cola, P = Pepsi, SL = Slice, LDL = Lipton Diet Lemon Brisk Iced Tea, 7U = 7-Up, WCP = Wild Cherry Pepsi, and DP = Diet Pepsi.

#### 4.4 Fixed Lighting

The system was on display in a room with fixed lighting. A new database was generated at the display location, consisting of four histograms each for Country-Time Lemonade, Schweppes, Welch’s Grape Drink, Minute Maid Orange Soda, Coca-Cola, Diet Coca-Cola, Pepsi, Slice, Lipton Diet Lemon Brisk Iced Tea, 7-Up, Wild Cherry Pepsi, Sprite, Mountain Dew, and Dr. Pepper. Passers-by were encouraged to place sodas in front of the camera and test the system. The camera was placed close to the desired can location. This made it less likely that people would stand in the image and throw off the cameras white balance (the Sony TRV-310’s white balance cannot be turned off).

The system identified all the cans without error, with one exception. When Mountain Dew’s nutritional information was the only thing visible, it was mistaken for 7-Up. The images used to generate their histograms are shown in Figure 4. Even to a human, these images are virtually indistinguishable.

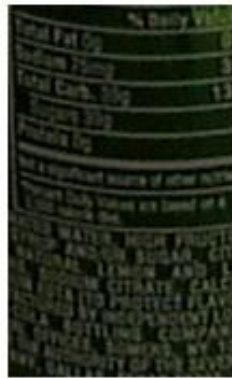
## 5 Conclusion

This implementation requires 128 bytes per object, assuming a generous maximum of 2 bytes per histogram element, up to 65536 pixels per bin. Our implementation had a maximum possible value of 18338. The 128 bytes per object is based on 16 colors. Assuming that an additional byte will be required to note which histogram goes with which object (allows databases up to 128 objects),

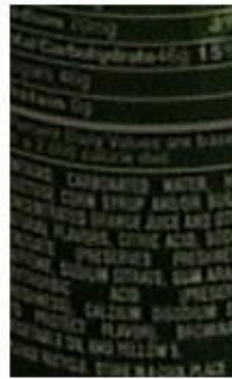
Soda	(1) Err	(2) Err	(3) Err	(4) Err
Sch	100%	100%	100%	100%
CT	100%	100%	100%	100%
MMO	56% C	100%	100%	100%
WG	100%	100%	100%	100%
C	100%	100%	100%	100%
P	100%	100%	100%	100%
SL	100%	100%	0% LDL	100%
LDL	100%	36% DP	100%	100%
7U	100%	100%	98% Sch	100%
WCP	100%	100%	82% MMO	96% DP
DP	96% MMO	74% 8% WCP 18% LDL	68% 6% SL 26% WCP	24% 22% LDL 54% WCP
SP	100%	100%	100%	100%

Table 5: Number correct when identification of sodas is tested using orientation midway between those in the database. Twelve cans with four orientations in the database, sixteen colors. The numbers 1, 2, 3, and 4 correspond to orientations; Err indicates what was chosen instead of the correct soda. Codes are: Sch = Schweppes, CT = Country-Time Lemonade, MMO = Minute Maid Orange Soda, WG = Welch’s Grape Drink, C = Coca-Cola, P = Pepsi, SL = Slice, LDL = Lipton Diet Lemon Brisk Iced Tea, 7U = 7-Up, WCP = Wild Cherry Pepsi, DP = Diet Pepsi, and SP = Sprite.

the total bytes per object needed will be 129. Our 14 object database requires only 1,806 bytes of storage space for the entire database. In theory, with four images of each object, a database with 100 objects would require only 12,900 bytes of storage space. Of course, this will change as more varied lighting conditions are introduced and more database histograms are necessary, but the size of the database should increase linearly, not exponentially, with the number of lighting conditions. A database of 100 objects under 80 different lighting conditions would fit into 1 MB of RAM. The program, incorporating the camera drivers and database, takes up 37 kilobytes of space. However, this code has not been optimized in any way. The program analyzes a single frame in 70 milliseconds, including the time to crop the image. This method results in extremely accurate object identification, as long as the lighting doesn’t vary beyond the limits imposed by the database. Accuracy of almost 100% was achieved under fixed lighting conditions. Future work includes experimenting with more lighting conditions and larger databases. Eventually, we hope to have a system capable of recognizing almost 100 soda cans, functioning under the lighting conditions typical of our laboratory.



**7-Up**



**Mountain Dew**

Figure 4: *Sample histograms from database.*

## References

- [1] Signe Redfield and John G. Harris. The role of color categorization in object recognition [arvo abstract]. *Invest Ophthalmol Vis Sci.*, 41(4):B635. Abstract nr 1260, 2000.
- [2] Signe Redfield and John G. Harris. The role of extreme color quantization in object recognition. In *Proceedings of CGIP 2000*, pages 225–230, 2000.
- [3] Signe Redfield and John G. Harris. The role of massive color quantization in object recognition. *Proceedings of the International Conference on Image Processing*, 33(2):137–186, 2000.
- [4] Signe Redfield and John G. Harris. Bit-depth of color memory [arvo abstract]. Submitted to ARVO 2001, 2001.
- [5] Michael J. Swain and Dana H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.