THE ROLE OF MASSIVE COLOR QUANTIZATION IN OBJECT RECOGNITION

Signe Redfield and John G. Harris

University of Florida
Electrical and Computer Engineering Department
PO Box 116200
Gainesville, FL 32611

ABSTRACT

Psychophysical experiments inspire a more complete analysis of the effect of quantization on a modified version of the histogram indexing method of object recognition. We derive an equation that describes how the amount of quantization and number of features kept affects recognition accuracy. The equation shows that quantization from 2^{24} colors to 15 colors has a negligible effect on accuracy. A simulation shows that large numbers of objects cause a corresponding decrease in accuracy, but that keeping more features can increase the accuracy even for massive quantization. An object recognition experiment with real data shows dramatically better results when quantization is used, indicating that massive color quantization can provide some invariance to lighting conditions.

1. INTRODUCTION

Psychophysical experiments provide insight to color memory and its possible uses in recognition. Berlin and Kay [1] determined that there are eleven unique color categories. Experiments on human memory [2] show that people remember colors as members of an internal color category. Our massive quantization is intended to be the engineering analogue to human color categorization.

In this paper, we address three questions. First, how far can we quantize the initial color space before recognition is seriously impaired? Second, how many colors are sufficient to describe a given size set of objects without appreciable reduction in object recognition accuracy? Third, is massive quantization effective in reducing the effects of lighting variations on real images? These three questions are posed from within the framework of the color histogram indexing method of object recognition.

Color histogram indexing [3] was proposed as a method for object recognition by Swain and Ballard. In this method an object is represented by a histogram of the colors present in one or more images of the object. When a given object appears, its histogram is compared with the histograms in the database. The object is assumed to be segmented from its environment, and the colors of a given object are assumed constant. Their research claimed that "the range of colors that occur in the world need only be split into about 200 different discrete colors to distinguish a large number of objects."[3] Their database had 66 objects in it, they kept all the values of the histogram, and obtained only "small changes in match effectiveness" even for 64 bins.

To derive an equation describing the effect of the number of bins on accuracy, we use a slightly different approach. Instead of storing values for all possible colors, we keep only the colors of the p largest areas. We are splitting the number of bins in the original algorithm into two parts: the number of possible colors and the largest values of the histogram. The simulation (used to generate a plot for varying the number of objects in the database) also used this approach. However, we used the original algorithm when testing real data with segmented images.

2. THEORY, SIMULATION AND METHODS

Our computer experiment is based on a psychophysical experiment to determine the capabilities of human memory. In the psychophysical experiment, the observer is first shown an object. Next, the object is taken away, and replaced with a set of distinguishable objects, one of which is the original object. The observer is then asked to identify the original object. If the set contained multiple objects from the same category as the original, humans were not capable of reliably picking out the specific object they had previously seen. In our computer experiment, the observer is a computer and although the computer quantizes the colors present in each object to store them, no other changes are made and no degradation over time occurs. Quantization takes the place of degradation through memory. The computer only makes a mistake if two or more objects in the set quantize to the same colors as the original object, in which

0-7803-6297-7/00/$10.00 © 2000 IEEE

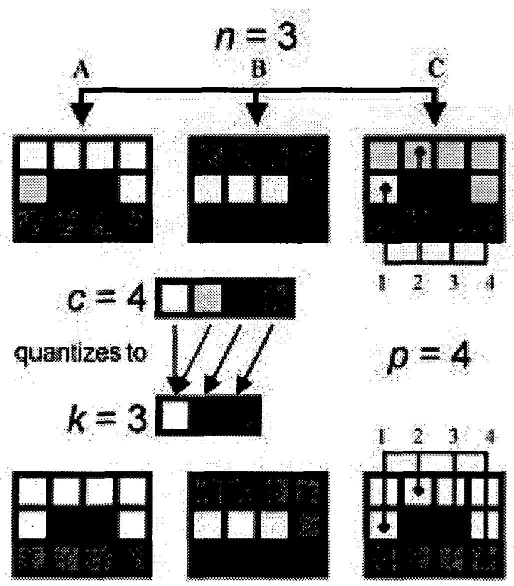

Fig. 1. *Diagram showing examples of each variable used in the theoretical equations.*

case it chooses between the (perceptually) identical objects randomly.

As shown in Figure 1, several variables affect the accuracy of recognition. Let c denote the number of possible colors for the original image of the object, and k represent the number of possible colors that the object can have in memory. If the c initial colors are quantized to 10 colors for storage, k would be 10. For our feature vector, we take the colors of the p largest areas. In Swain and Ballard's work [3], this would be equal to the number of possible colors k. Let n denote the number of objects in the group. All objects in a group must be different in the original space c. Our theoretical results produce the expected value of the accuracy with 3 objects per group. It is possible to derive similar equations for other values of n, but they quickly become intractable. This derivation assumes that if there were m identical objects in the set, the chance of choosing the correct one would be $1/m$. In addition, each case is assumed to occur equally often. For a given set of objects, each individual object would be the test object equally often, and no set of objects could occur more often than any other.

Figure 1 shows the effects of quantization. The original database (the first row) in colorspace c contains three objects ($n = 3$), each of which has four regions ($p = 4$). When the objects are stored, only $k = 3$ colors are used, which results in the second row of objects. When one of the original objects is presented, it is first quantized and then compared to all objects in memory. Object B has no duplicates after quantization. Objects A and C, however, quantize to the same object in the new colorspace. Therefore, the computer will confuse objects A and C, resulting in reduced accuracy. On average, with this set of objects, the computer will guess correctly $2/3$ of the time.

As we wished to test the accuracy of this method for large n, we also implemented a Monte Carlo simulation. We averaged 10,000 sets of n objects each to obtain reasonably smooth results.

In the theoretical equation, each area of uniform color is assigned an index, and so it is possible to derive accuracy values for $k < p$. For real data, an area was designated as the sum of all pixels of a given color, so k must be greater than or equal to p. Fewer colors than areas was not possible.

To test the effectiveness of quantization for object recognition, we constructed a database from images of 9 different soda cans under 4 different lighting conditions. We quantized all the images twice. One database of quantized images used a 256 color set. As in Swain and Ballard's work, these colors were generated using uniform quantization of 8x8x4 in RGB space. The other database consisted of the same images quantized to a set of 3 achromatic and 11 chromatic values generated by a data-independent algorithm in HLS space. The 3 achromatic colors each contained a third of the possible lightness values and the 11 chromatic colors were generated by uniformly quantizing the hue axis. A saturation threshold was arbitrarily set to distinguish the chromatic values from the achromatic values. For the 256 color set ($k = 256$), the 14 highest histogram values were kept and the rest were set to zero, and for the 14 color set ($k = 14$), all 14 values were kept. Thus the results are for 14 histogram values ($p = 14$) in each case.

We used four lighting conditions likely to be found in our lab: one with the flourescent overhead lights on and the blinds closed, one with the lights off and the blinds open, one with the left-hand lights on and the blinds open, and one with the right-hand lights on and the blinds open. The speculars under each lighting condition were different, and the images with the blinds open were noticeably bluer than the images with the blinds closed.

Template images for the database were defined as the subset of images taken under a single lighting condition. The accuracy was calculated as the set of correctly identified objects out of the set of images not being used as templates. Each lighting condition was used as the template in turn, generating four accuracy values under each quantization condition.

3. RESULTS

Theoretically, for $n = 3$, we know that there are only three possible cases. When there are no duplicated objects in the quantized space, the computer observer cannot make a mistake. When two of the objects quantize to the same object, the test object is either the unique object, in which case the observer is correct, or one of the duplicated objects, in which case the observer will be correct half the time. In the third case, all three objects quantize to be the same. No matter which object is the test object, the observer will have to guess. To determine the percentage of correct scores, we calculate the number of wrong answers out of all possible cases, and subtract this from the perfect score. The number of all possible cases is

$$N_{all} = \frac{c^p!}{(c^p - n)!} \tag{1}$$

where c^p is the total number of possible objects in the unquantized space. The number of instances where two of the three objects quantize to the same object is given by

$$N_2 = \sum_{j=1}^{k^p} s_j \left((s_j - 1)(c^p - s_j) + \sum_{t \neq j} s_t (s_t + s_j - 2) \right) \tag{2}$$

where k^p is the number of distinguishable objects in the quantized space. When i is a number from 1 to k^p, s_i represents the number of objects from the set of c^p possible original objects that map into the i^{th} object out of k^p possible distinguishable objects in the quantized space. The third case is represented by a much simpler equation. When all the objects are the same, once the first object is chosen the rest must come from the same set of s_i objects. Therefore we get

$$N_3 = \sum_{j=1}^{k^p} s_j (s_j - 1)(s_j - 2) \tag{3}$$

This gives our final equation for the theoretical accuracy given sets consisting of only 3 objects.

$$Accuracy = 1 - \frac{1}{N_{all}}[\frac{1}{3}N_2 + \frac{2}{3}N_3] \tag{4}$$

Figure 2 shows what happens when k is swept from 1 to 25 and c is held to 2^{24}. For values of k greater than 25, the accuracy continues to approach 1. When $k = 1$, the accuracy is $1/n$, in this case $1/3$. By $k = 5$ the system is correct over 95% of the time, and by $k = 10$ the system is making mistakes less than one percent of the time. When p is changed to 3, accuracy remains near one until k is even smaller. Varying p produces a strong impact on accuracy. For every incremental increase in p, the number of possible objects in both the quantized and unquantized spaces increases exponentially, and thus the chances of a fixed size set incorporating two objects similar enough to be indistinguishable in the quantized space are reduced. As expected,

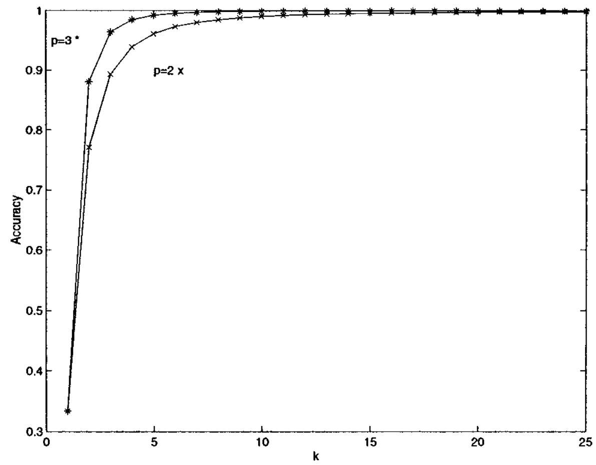

Fig. 2. *Theoretical results for $c = 2^{24}$, $p = 2$ and $p = 3$, $n = 3$, and k varying.*

increasing the number of features used to represent each object increases the recognition accuracy.

Holding p and k constant, and varying c, produces very little effect on accuracy until c is very close to k. When $c = k$, accuracy is one, and no quantization occurs. If the number of colors in the original space is larger than the number of colors in the quantized space, the original number of colors becomes irrelevant. In fact, varying c from 2^{24} to 2^8 produces an insignificant difference in accuracy ($p = 2$, $n = 3$, $k = 10$), and reducing it still further, to only six more levels than k, 16, produces a rise of only .002%

The results of the simulation are shown in Figure 3. For $k = 10$ and $p = 2$, when $n = 100$ there are as many objects in the set as there are distinguishable objects in the world. This is reflected in the accuracy, which decreases to about 63 percent. As expected, this cue breaks down for large n. When p is increased, the accuracy suffers less. As the number of objects that must be distinguished increases, the accuracy falls proportionately. Increasing the number of features alleviates this problem.

Real data from the object recognition experiment described above produced convincing results, shown in Table 1. Under the lighting condition that provided the best results with 14 colors, 78% of the 14-color objects were identified correctly. Only 37% of the 256-color objects were identified correctly. In this case, introducing quantization increased the accuracy by 41%. Under the lighting condition that produced the best results with 256 colors, 52% of the 256-color objects were identified correctly. With 14 colors, under the same lighting condition, 63% were identified correctly. At worst, the difference between the accuracies was only 7%, but in all cases, quantization improved object recognition.

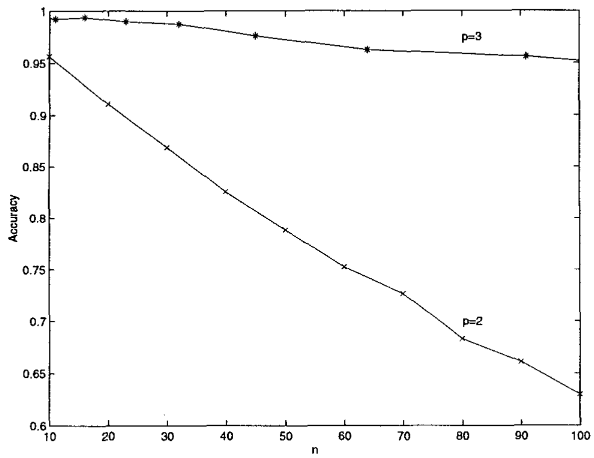

Fig. 3. *Simulation (averaged) results for $c = 2^{24}$, $p = 2$ and $p = 3$, $k = 10$, and n varying.*

Lighting Condition	256 Colors	14 Colors
1	41%	63%
2	52%	63%
3	41%	48%
4	37%	78%

Table 1. *Object recognition accuracy generated from database of 9 soda cans under 4 different lighting conditions. Lighting condition in the table indicates the images that were used as the templates in the database, while the test data consisted of the remaining 27 images.*

4. CONCLUSIONS

In theory, we can reduce the number of colors from 2^{24} to 15 before the accuracy is substantially reduced. Massive color quantization has very little effect on accuracy where database size is small compared to the number of possible objects. The simulation showed that if enough features of each object are kept, 15 or fewer colors may be sufficient. If the feature vectors are stored in a computer, only $np(log_2 k)$ bits need to be stored. Trading off p and k would allow manipulation of the database size without loss of accuracy. Real data shows the benefits of massive quantization for object recognition. Even without a color constant pre-processor, taking the best results under each quantization condition, accuracy increased by 28% under a variety of lighting conditions.

Acknowledgments

This work was supported by an NSF Minority Engineering Doctoral Initiative Fellowship.

5. REFERENCES

[1] Brent Berlin and Paul Kay, *Basic Color Terms: their Universality and Evolution,* University of California Press, Berkeley and Los Angeles, 2 edition, 1991.

[2] K. Uchikawa, H. Ujike, and T. Sugiyama, "Categorical characteristics of multiple-color memory," in *John Dalton's Color Vision Legacy*, Christine Dickinson, Ian Murray, and David Carden, Eds. Taylor and Francis, 1997.

[3] Michael J. Swain and Dana H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.