A Maximum Likelihood Investigation Into Color Indexing

Nicu Sebe Michael S. Lew

Department of Computer Science, Leiden University

Postbus 9512, 2300 RA, Leiden, The Netherlands

{nicu mlew}@wi.leidenuniv.nl

Abstract

In content based image retrieval, color indexing is one of the most prevalent retrieval methods. In the previous literature, most of the attention has been focussed on the color model with little or no consideration of the noise models. In this paper we investigate the problem of color indexing from a maximum likelihood perspective. We take into account the color model, the noise distribution, and the inter-dependence of the color features. Our investigation concludes with results on a real stock photography database consisting of 11,000 color images.

1 Introduction

As the world enters the digital age, visual media is becoming prevalent and easily accessible. Factors such as the explosive growth of the World Wide Web, terabyte disk servers, and the digital versatile disk reveal the growing amount of visual media which is available to society. With the availability of visual media comes the associated problem of searching for it, and consequently the focus of researchers toward providing automatic content based retrieval systems. Of the visual media retrieval methods, color indexing is one of the dominant methods because it has been shown to be effective in both the academic and commercial arenas. In color indexing, histogram methods are often used because they are feasible in terms of memory usage and provide sufficient accuracy. The histogram methods quantize each image into a feature vector based on a color model such as RGB [3] or HSV [3], and then compare the query image feature vector to the database image feature vectors using a minimum distance classifier.

1.1 Color Models

Color models describe different aspects of the color space of an image. Two frequently used models are RGB [3] and HSV [3]. RGB refers to the intensity of 3 additive color primaries, red, green, and blue. Each primary is typically quantized into 256 levels and then combined to create 256*256*256 possible colors. The HSV model separates the color components from the luminance component. The hue and saturation of a color are represented by H and S, and the luminance is represented by V.

When we create a color histogram, we must quantize each component of the color model using a number of bits. We define quantization X:Y:Z for color model ABC as quantizing color component A using X bits, B using Y bits, and C using Z bits. In the case of HSV, a 4:2:2 quantization refers to quantizing H using 4 bits, S using 2 bits, and V using 2 bits.

We chose to use 11,000 images from the Corel Photo database because it represents a widely used set of photos by both amateur and professional graphical designers. Furthermore, it is available on the Web at http://www.corel.com.

1.2 Color Indexing

The paradigm of color indexing into an image database works as follows: Given a query image, we want to retrieve all the images whose color compositions are similar to the color composition of the query image. Color indexing is based on the observation that often color is used to encode functionality: grass is green, sky is blue etc.

If we map the colors in the image $\mathcal{Q}$ into discrete color space containing n colors, then the color histogram ([12, 10]) $H(\mathcal{Q})$ is a vector $(h_{c_1}, h_{c_2}, \cdots, h_{c_n})$, where each element h_{c_j} represents the number of pixels of color c_j in the image $\mathcal{Q}$.

Two widely used distance metrics are L_1 ([4]) and L_2 ([1]). For example the L_1 distance applied to two color histograms H and I is defined as

$$d_{L_1}(H,I) = \sum_{i=1}^{n} |h_{c_i} - i_{c_i}|$$

Similarly, the L_2 distance will be

$$d_{L_2}(H,I) = \sqrt{\sum_{i=1}^{n} (h_{c_i} - i_{c_i})^2}$$

Other criterion functions that have been used in previous literature are (1) histogram intersection ([12]), (2) average color distance ([5]), (3) the quadratic distance measure form ([8]).

1.3 Early Experiments

Before we can measure the accuracy of particular methods, we first had to find a challenging and objective ground truth for our tests. We perused the typical image alterations and categorized various kinds of noise with respect to finding image copies. Copies of images were often made with images at varying JPEG qualities, in different aspect ratio preserved scales, and in the printed media. We defined these as JPEG noise, Scaling noise, and Printer-Scanner noise. JPEG noise was created by coding and then decoding a JPEG image using varying JPEG-quality values. Using HSV 4:2:2 and JPEG quality 30, we were able to recover the exact image copy as the top rank with 100% accuracy from our large image database. In Scale noise, we made the copy by reducing the image in size so that the image was aspect ratio preserved with maximum size 32x32. Using HSV 4:2:2, the copy was found within the top 10 ranks with 100% accuracy. We concluded that JPEG noise and Scaling noise were not sufficiently challenging to separate the different color indexing methods.

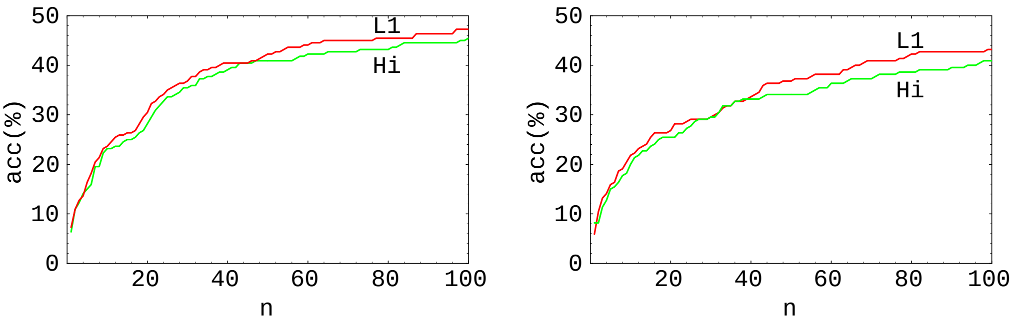
In Printer-Scanner noise, the idea was to measure the effectiveness of a retrieval method when trying to find a copy of an image in a magazine or newspaper. We printed each image using an Epson Stylus 800 color printer at 720 dots per inch, and then scanned it using an HP IIci color scanner. The noise from this copy process was the most significant in that the copy was found in the top 10 ranks using HSV 4:2:2 with less than 20 % accuracy. For benchmarking purposes, the exact test set can be found at the Web page: http://www.wi.leidenuniv.nl/home/lim. Examples of the test copy pairs are shown in Figure 1.

(a) (b) (c) (d)

Figure 1: Two examples of test copy pairs used (a)-(c) the original image; (b)-(d) copy image

1.4 Comparision with histogram intersection

Another early question was whether to use the well known comparison criterion, histogram intersection ([12]). In our tests, we graph the retrieval accuracy of the system with respect to finding the image copy in the top 1 to 100 ranks. As is evident from Figure 2 the histogram intersection criterion does not give better retrieval accuracy than a simple criterion such as L_1.



(a) (b)

Figure 2: L_1 (L1) vs. histogram intersection (Hi) (a) HSV (b) RGB

1.5 Usability Issues

In creating a system for users, it is important to take into account the way in which users will interact with the system. Two important issues are the total response time of the system and the number of results pages which the user must look at before finding the image copy. We make the following assumptions. First, in order to have an interactive experience, the total system response time should be less than 2 seconds. Furthermore, the typical user will only look at the first few results pages, however, in order to view the global retrieval accuracy, we occasionally show the results regarding the top 1 to 6000 ranks. We also avoid methods which require more than a few seconds of response time.

Section 2 describes the mathematical support for maximum likelihood approach. In Section 3 we examine the modeling of the noise distribution along with the retrieval accuracy with respect to the color model, quantization and inter-feature dependence. Conclusions are given in Section 4.

2 Maximum likelihood estimator

From the mathematical-statistical point of view, the problem of finding the right model for the similarity noise comes down to the maximization of the similarity probability.

Consider first, two subsets of M images from the database (D) : X ⊂ D, Y ⊂ D which according to the ground truth are similar:

$$X \equiv Y \tag{1}$$

This can be written:

$$x_i \equiv y_i, \qquad i = 1, ..., M \tag{2}$$

where x_i and y_i represent the feature vectors associated with the images in the corresponding subsets.

The equation (2) can be further written as:

$$x_i = y_i + n_i, \qquad i = 1, ..., M \tag{3}$$

where n_i represent the “”noise” vector obtained as the difference between the two vectors.

In this context the similarity probability can be defined:

$$P(X,Y) = \prod_{i=1}^{M} \{\exp[-\rho(x_i, y_i)]\} \tag{4}$$

where function ρ is the negative logarithm of the probability density of the noise.

According to (4) we have to find the probability density function of the noise that maximizes the similarity probability: *maximum likelihood* estimator for the noise distribution ([7]).

We can further suppose that this noise distribution is valid for all the database, so using it for all the images in the database one obtains the best possible ranking results.

Taking the logarithm of (4) we find that we have to minimize the expression:

$$\sum_{i=1}^{M} \rho(x_i, y_i) \tag{5}$$

In this case, according to (3), the function ρ depends not independently on its two arguments, query vector x_i and the predicted one y_i, but only on their difference. We have thus a *local* estimator and we can replace (5) with:

$$\sum_{i=1}^{M} \rho(z_i) \tag{6}$$

where $z_i \equiv x_i$-y_i and the operation “-” denotes difference between corresponding values in the feature vectors.

To analyze the behavior of the estimator we take the approach described in [6] and [9] based on *influence function*. The influence function characterizes the bias that a particular measurement has on the solution and is proportional to the derivative, ψ, of the estimator ([2]).

$$\psi(z) \equiv \frac{d\rho(z)}{dz} \tag{7}$$

In case the noise is Gaussian distributed:

$$Prob\{x_i - y_i\} \sim \exp([x_i - y_i]^2) \tag{8}$$

then

$$\rho(z) = z^2 \qquad \psi(z) = z \tag{9}$$

If the errors are distributed as a *double* or *two-sized exponential*, namely

$$Prob\{x_i - y_i\} \sim \exp(-|x_i - y_i|) \tag{10}$$

then, by contrast,

$$\rho(z) = |z| \qquad \psi(z) = sgn(z) \tag{11}$$

In this case, the maximum likelihood estimator is obtained by minimizing the *mean absolute deviation*, rather than the *mean square deviation.* Here the tails of the distribution, although exponentially decreasing, are asymptotically much larger than any corresponding Gaussian.

One can easily notice that equation (8) resembles the L_2 metric while equation (10) resembles the L_1 metric. Thus, *maximum likelihood* gives a direct connection between the noise distribution and the comparison metrics.

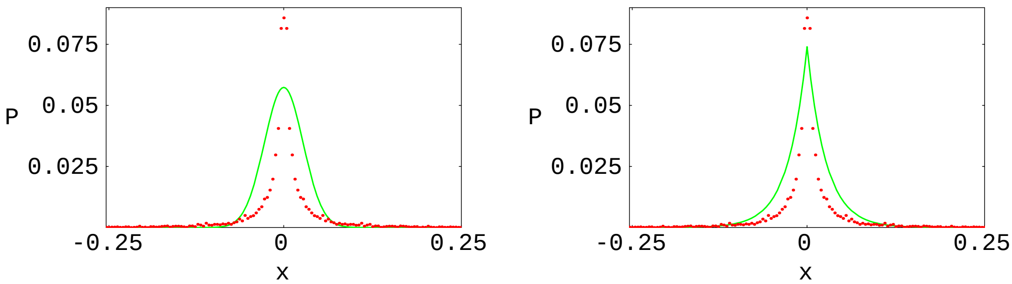
For normally distributed errors, equation (9) says that the more deviant the points, the greater the weight. By contrast, when tails are somewhat more prominent, as in (10), then (11) says that all deviant points get the same relative weight, with only the sign information used.

3 Experiments

The first question we asked was, "Which distribution is the closest to the real color model noise?" To answer this we needed to measure the noise with respect to each color model and then we could choose the color model and noise which had the best accuracy.

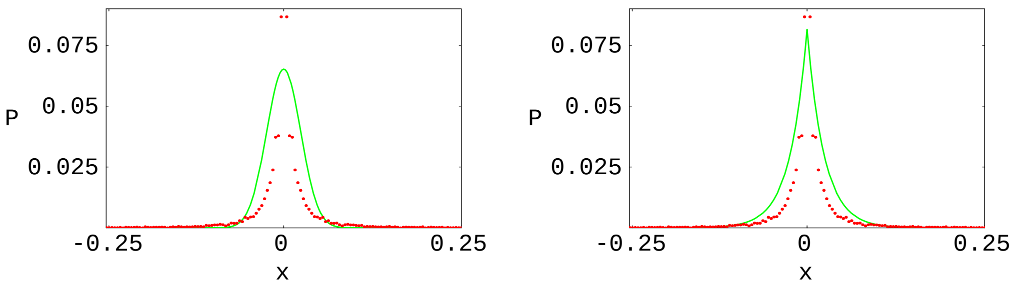
3.1 Distribution Analysis

In Figures 3 and 4 we display the real noise distribution in RGB and HSV respectively. Note that the best fit exponential has a better fit to the noise distribution than the Gaussian for both color models. Consequently, this implies that the L_1 metric will give better retrieval accuracy than



(a) (b)

Figure 3: Similarity noise distribution in RGB compared to best fit Gaussian (a) (modeling error is 0.11) and best fit exponential (b) (modelling error is 0.085)



(a) (b)

Figure 4: Similarity noise distribution in HSV compared to best fit Gaussian (a) (modeling error is 0.106) and best fit exponential (b) (modelling error is 0.082)

the L_2 in both cases. For the retrieval accuracy we choose to display percentage of correct copies found within the top n matches. From the tests as shown in Figure 5 it is clear that the L_1 metric gives a significant improvement in retrieval accuracy as compared to L_2.

3.2 Color Model

The second question we asked was, "Which color model gives better retrieval accuracy?". As shown in Figure 6 using the L_1 metric we obtained an improvement in retrieval accuracy by up to 8% when using the HSV color model.

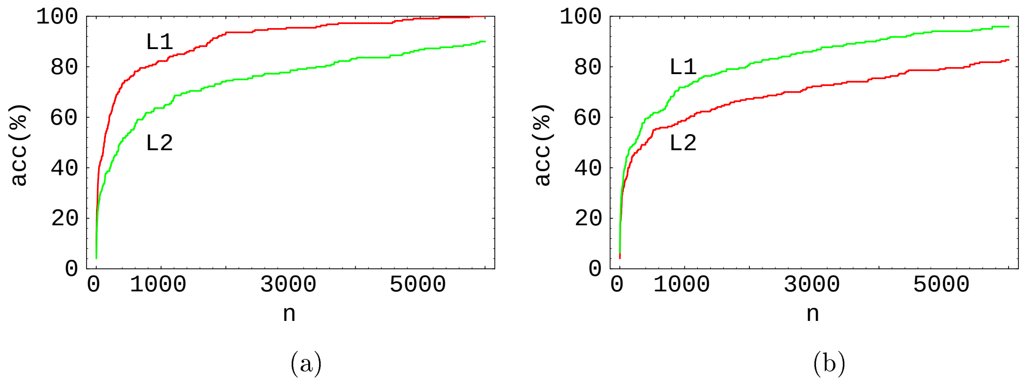


Figure 5: Retrieval accuracy for the top 6000 matches (a) HSV (b) RGB

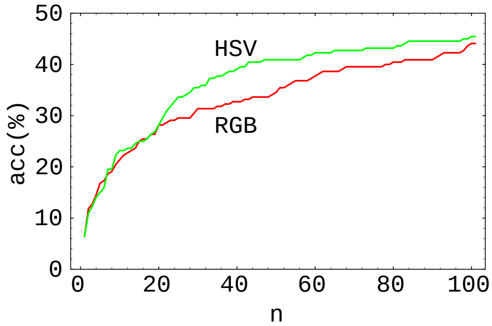
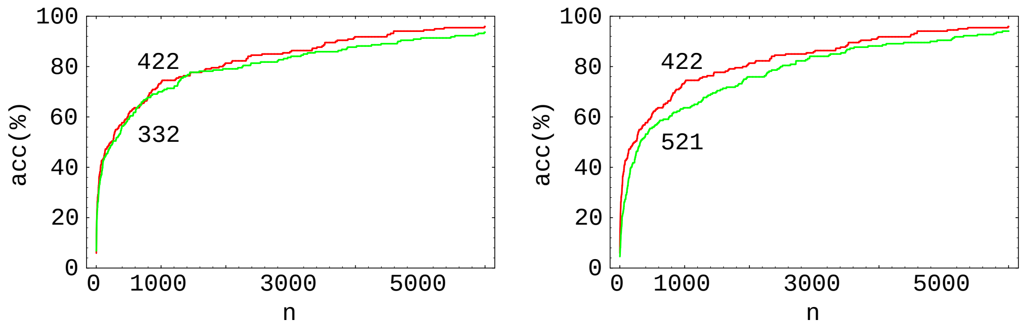


Figure 6: Retrieval accuracy for the top 100 using L_1 - RGB vs. HSV

3.3 Quantization

Based upon the improvement in the retrieval accuracy it is clear that the best choice is to use the HSV color model with the L_1 metric. So, the next question is, "How does the quantization scheme affect the retrieval accuracy?". In Figure 7(a) it appears that increased resolution in H may be the cause of increased accuracy. This leads us to ask whether further H resolution will give even better results. Figure 7(b) shows that this is not the case.

In summary, the experiments in this section showed that the choice of color model, noise distribution, and quantization can affect the accuracy by up to 8%, 15%, and 5%, respectively.



(a) (b)

Figure 7: Retrieval accuracy for HSV using different quantization models (a) 4:2:2 - 3:3:2 and (b) 4:2:2 - 5:2:1

3.4 Examining the Independence Assumption

In the maximum likelihood approach it is essential to consider both the shape of the distribution and the independence or inter-dependence of the features. In using an L_1 or L_2 metric, there is an assumption that the features are independent ([11]). However, from the process of creating a histogram, we know that color values may fall into nearby bins in the color space, which violates the independence assumption. If we wanted to optimally solve the problem regarding histogram bin dependencies, it would be necessary to calculate the joint probability distribution of N features, which would be an N dimensional histogram. From a theoretical perspective it is the right solution, however, from a practical viewpoint, it is difficult to estimate, represent, and store the N dimensional distribution accurately.

As a practical alternative, we address the bin dependence problem by taking a weighted average of the values at varying quantization levels. In the computer vision literature, these are often called pyramidal, scale space, or hierarchical methods. These methods have the advantage that they integrate the notion of local to global scale. From a maximum likelihood perspective, we examined the shape of the noise distribution between using a single HSV quantization level of 4:2:2, and using multiple HSV quantization levels of 4:2:2 (256 bins), 3:2:2 (128 bins), 2:2:2 (64 bins), 2:2:1 (32 bins), 2:1:1 (16 bins), and 1:1:1 (8 bins). Figure 8 shows the relationship between the noise distributions. Note that the variance of the noise has decreased from 0.0014 to 0.0008. This results

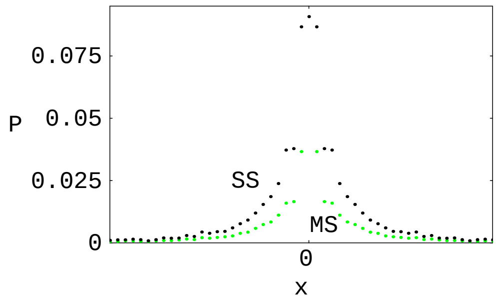


Figure 8: Similarity noise distribution for HSV : single scale (SS), multi-scale (MS)

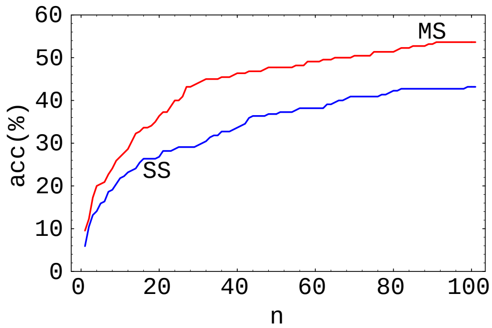


Figure 9: Retrieval accuracy in HSV using L_1 - (SS) single scale (MS) multi-scale

in improved accuracy as shown in Figure 9. The average system response time on an SGI INDY R5000, 150 MHZ was 0.793 seconds for a single image query on the 11,000 image database.

4 Conclusions

In this paper we investigated the problem of color indexing for content based retrieval using the maximum likelihood paradigm. The maximum likelihood theory provides us with a direct connection between the noise distribution and the retrieval accuracy of the system. Furthermore, it directed us toward improving the modeling of the noise distribution and reducing the variance of the noise by using a pyramidal color histogram. We tested the maximum likelihood based methods on an 11,000 stock image database and found the following results

- Choice of color model affect the accuracy by up to 8%
- Choice of noise distribution affects the accuracy by up to 15%
- Choice of color model quantization affects the accuracy by 5%
- Using a pyramidal color histogram reduces the variance from 0.0014 to 0.0008
- Using a pyramidal color histogram increases the accuracy by up to 10%
- Using the histogram intersection criterion function does not outperform the L_1 criterion.

Although using L_1 as a criterion is attractive from a computational efficiency perspective, it does not perfectly match the noise distribution. Future work will examine the benefits from optimal noise distribution modeling.

Acknowledgements

This research was funded by a grant from Philips Research, Eindhoven. We would also like to thank Nies Huijsmans for useful comments and discussions.

References

[1] A. Berman and L.G. Sapiro. Efficient image retrieval with multiple distance measures. *Proc. SPIE, Storage and Retrieval for Image/Video Databases*, 3022:12–21, 1997.

[2] M.J. Black. *Robust Incremental Optical Flow.* PhD thesis, Yale University, September 1992.

[3] J. Foley, A. van Dam, S. Feiner, and J. Hughes. *Computer Graphics - principles and practice.* Addison-Wesley, 1990.

[4] A. Gupta, S. Santini, and R. Jain. In search of information in visual media. *Communications ACM*, 12:34–42, 1997.

[5] J. Hafner. Efficient color histogram indexing for quadratic form distance functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(7):729–736, 1997.

[6] F.R. Hamper, E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel. *Robust Statistic: The Approach Based on Influence Functions.* John Wiley and Sons, New York, 1986.

[7] P.J. Huber. *Robust Statistic.* NewYork: Wiley, 1981.

[8] W. Y. Ma, Y. Deng, and B.S. Manjunath. Tools for texture/color based search of images. *Proc. SPIE, Human Vision and electronic imaging II*, 3106:496–507, 1997.

[9] P.J. Rousseeuw and A.M. Leroy. *Robust Regression and Outlier Detection.* John Wiley and Sons, New York, 1987.

[10] H.S. Sawhney and J.L. Hafner. Efficient color histogram indexing. In *Proc. of 1994 IEEE International Conference on Image Processing*, volume 2, pages 66–70, 1994.

[11] N. Sebe, M. Lew, and N. Huijsmans. Which ranking metric is optimal? with applications in image retrieval and stereo matching. *Proceedings of the Int. Conf. on Pattern Recognition*, 1998.

[12] M.J. Swain and D.H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.