COLOR BASED RETRIEVAL AND RECOGNITION

Nicu Sebe and Michael S. Lew

Leiden Institute of Advanced Computer Science,
Niels Bohrweg 1, 2333 CA, Leiden, The Netherlands
{nicu mlew}@liacs.nl

ABSTRACT

In content based retrieval, color indexing is one of the most prevalent retrieval methods. Two key problems in color indexing are (1) determination of the color space and (2) finding the best distance measure. Most of the attention from the research literature has been focussed on the color model with little or no consideration of the noise models. By focusing on the noise model, we showed how to find a distance measure which is optimal in the sense that it maximizes the similarity probability. Our results include experiments in color based retrieval and object recognition. Furthermore, we implemented and tested several promising algorithms from the research literature as benchmarks.

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. Most image retrieval systems adopt a two step approach to search image databases. First, for each image in the database, a feature vector characterizing some image properties is computed and stored in a featurebase. Second, given a query image, its feature vector is computed, compared to the feature vectors in the featurebase, and images most similar to the query image are returned to the user.

In color indexing, given a query image, the goal is to retrieve all the images whose color compositions are similar to the color composition of the query image. Typically, the color content is characterized by color histograms, which are compared using the histogram intersection distance measure [8]. Hafner, et al. [3] suggested the usage of a more sophisticated quadratic form of distance measure, which tries to capture the perceptual similarity between any two colors. Another approach which incorporates the spatial correlation of colors in a color correlogram, was introduced in [4]. Gevers et. al [2] proposed new color models that are invariant to a change in viewing direction, object geometry and illumination. In [7] and [1], comparisons have been made between different distance metrics. However, their results did not explain why a particular metric would provide better results. Here in our work, we show that the maximum likelihood paradigm explains why one metric will outperform another one based upon the underlying noise model. Furthermore, we show how to derive a better distortion measure based upon the real noise distribution.

One of the common assumptions is that the real noise distribution should fit either the Gaussian or the Exponential distributions. This justifies using the L_2 or L_1 distance measures from a maximum likelihood perspective. One of the questions we examined in this paper is what if there is another distribution which fits the real noise distribution better than the Gaussian or the Exponential?. Toward answering this question, we used international test sets and promising algorithms from the research literature. Furthermore, one of the canonical measures of similarity from the field of information theory, the Kullback relative information, was also implemented and compared to the metrics based on maximum likelihood.

Section 2 introduces the color features and methods used in our experiments. Section 3 describes the mathematical support for the maximum likelihood approach. In Section 4 we apply the theoretical results to determine the influence of the real noise model on the accuracy of retrieval methods in color image databases. Conclusions are given in Section 5.

2. COLOR BASED RETRIEVAL AND RECOGNITION

If we map the colors in the image I into a discrete color space containing n colors, then the color histogram $H(I)$ is a vector $(h_{c_1}, h_{c_2}, \cdots, h_{c_n})$, where each element h_{c_j} represents the probability of having the color c_j in the image I. Two widely used distance metrics between two color histograms are L_1 (equivalent to the histogram intersection measure [8]) and L_2. Hafner, et al. [3] introduced a more sophisticated method of comparing histograms which more closely corresponds to human judgment of color similarity. They used a quadratic distance measure L_q which allows for similarity matching between different colors (represented by the color histograms bins):

$$L_q(x, y) = (x - y)^t A(x - y) \tag{1}$$

where x and y are two color histograms and $A = [a_{ij}]$, with a_{ij} denoting similarity between bins (colors) i and j.

Observing the fact that the color histograms lack information about how color is spatially distributed, Huang et al. [4], introduced a new color feature for image retrieval called color correlogram. This feature characterizes how the spatial correlation of pairs of color changes with distance in an image. A color correlogram of an image is a table indexed by color pairs, where the k-th entry for $\langle c_i, c_j \rangle$ specifies the probability of finding a pixel of color c_j at distance k from a pixel of color c_i in the image:

$$Cg^k_{c_i,c_j}(I) = Pr_{p_1 \in I_{c_i}, p_2 \in I}[p_2 \in I_{c_j} || p_1 - p_2| = k] \quad (2)$$

where p_1 and p_2 are pixels in the image and $I_c = \{p|I(p) = c\}$ with $I(p)$ denoting the pixel color. Usually, because the size of color correlogram in quite large, the color auto-correlogram is often used instead. This feature only captures spatial correlation between identical colors.

Gevers, et al. [2] analyzed and evaluated various color features for the purpose of image retrieval by color-metric histogram matching under varying illumination environments. They introduced a new color model l:

$$l_1(R,G,B) = \frac{(R-G)^2}{(R-G)^2 + (R-B)^2 + (G-B)^2} \quad (3)$$

$$l_2(R,G,B) = \frac{(R-B)^2}{(R-G)^2 + (R-B)^2 + (G-B)^2} \quad (4)$$

$$l_3(R,G,B) = \frac{(G-B)^2}{(R-G)^2 + (R-B)^2 + (G-B)^2} \quad (5)$$

where R, G, B are the color values in the RGB color space and showed that is invariant to viewing direction, object geometry, highlights and illumination direction and intensity.

The authors [2] concluded that this color model is the most appropriate color model to be used for image retrieval by color-metric histogram matching under the constraint of a white illumination source. This conclusion was drawn using histogram intersection (L_1) as the comparison metric between the color histograms.

3. MAXIMUM LIKELIHOOD APPROACH

Consider M image pairs (or more generally, feature vectors) from the database (D): $(x_i, y_i) \in D$, with $i = 1, \cdots, M$ which according to the ground truth (G) are similar: $x_i \equiv y_i$. Considering n_i as the "noise" image obtained as the difference between the other two images (x_i and y_i), the similarity probability can be defined:

$$P(G) = \prod_{i=1}^{M} \{\exp[-\rho(n_i)]\} \quad (6)$$

where function ρ is the negative logarithm of the probability density of the noise. According to (6) we have to find the probability density function of the noise that maximizes the similarity probability: *maximum likelihood* estimate for the noise distribution [5]. Taking the logarithm of (6) we find that we have to minimize the expression:

$$\sum_{i=1}^{M} \rho(n_i) \quad (7)$$

Note that when the Exponential and Gaussian distributions are used in equation (7), we arrive at the L_1 and L_2 metrics, respectively. A distribution with more extensive tails is the Cauchy distribution, and the corresponding metric L_c is given by the expression:

$$L_c(X,Y) = \sum_{i=1}^{M} \log(\mathbf{a}^2 + (x_i - y_i)^2) \quad (8)$$

where **a** is a parameter which determines the height and the tails of the distribution.

For a general noise distribution, considering ρ as the negative logarithm of the probability density of the noise, the corresponding metric is given by equation (7). In practice, the probability density of the noise can be estimated from the normalized histogram of the absolute differences.

4. RESULTS

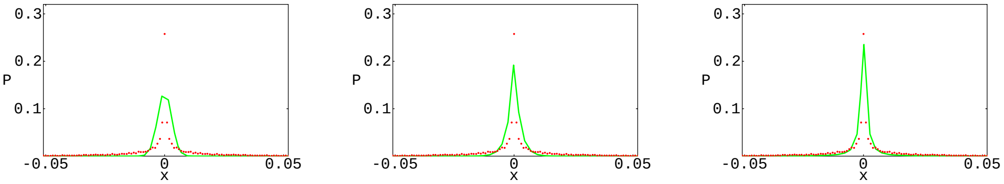
The setup of our experiments is the following. First, we assume that representative ground truth is provided. The ground truth is split into two non-overlapping sets: the training set and the test set. Second, the training set is converted to a feature vector which is then normalized to what we denote the real noise distribution. The Gaussian, Exponential, and Cauchy distributions are fitted to the real distribution. The Chi-square test is used to find the fit between each of the model distributions and the real distribution. We select the model distribution which has the best fit and its corresponding metric is used in ranking. The ranking is done using only the test set. For benchmarking purposes in all of the experiments we compare our results with the ones obtained using the Kullback relative information (K) [6]. We chose the Kullback relative information because it is the most frequently used information theory similarity measure.

It is important to note that for real applications, the parameter in the Cauchy distribution is found when fitting this distribution to the real distribution from the training set. This parameter setting would be used for the test set and any future comparisons in that application.

As noted in the previous section, it is also possible to create a metric based on the real noise distribution using maximum likelihood theory. Consequently, we denote the maximum likelihood (ML) metric as (7) where ρ is the negative logarithm of the normalized histogram of the absolute differences from the training set. This normalized histogram is our approximation for the probability density function.

We applied the theoretical results described in Section 3 in two experiments. We determined the influence of the similarity noise model on finding similar images which differ due to either printer-scanner noise or change of viewpoint. We used two color image databases. The first one was the Corel Photo database and the second one consisted of 500 reference images of domestic objects, tools, art artifacts, etc.

For benchmarking purposes in both experiments we compared our results with the ones obtained using Hafner's quadratic distance measure (L_q) and the corel auto-correlogram (Cg). Since both L_q and C_g were meant to be benchmarks, they were implemented as described in the original papers. For L_q Hafner used a 256 bin histogram in the RGB color space. In computing the auto-correlograms, there were used 64 colors in RGB color space and $\{1,3,5,7\}$ for



(a) Gauss (b) Exponential (c) Cauchy (**a**=1.32)

Figure 1: Noise distribution in Corel database compared with the best fit Gaussian (a) (approximation error is 0.106), best fit Exponential (b) (approximation error is 0.082) and best fit Cauchy (c) (approximation error is 0.068)

spatial distances, resulting in a 256 feature vector. The comparison was made using L_1.

The problem is formulated as follows: Let $\mathcal{Q}_1, \cdots, \mathcal{Q}_n$ be the query images and for the i-th query $\mathcal{Q}_i$, $\mathcal{I}_1^{(i)}, \cdots, \mathcal{I}_m^{(i)}$ be the images similar with $\mathcal{Q}_i$ according to the ground truth. The retrieval method will return this set of answers with various ranks. As an evaluation measure of the performance of the retrieval method we used precision versus recall at different scopes: For a query $\mathcal{Q}_i$ and a scope $s > 0$, the recall r is defined as $|\{\mathcal{I}_j^{(i)} | rank(\mathcal{I}_j^{(i)}) \leq s\}|/m$, and the precision p is defined as $|\{\mathcal{I}_j^{(i)} | rank(\mathcal{I}_j^{(i)}) \leq s\}|/s$.

4.1. Experiments with Corel Database

The first experiments were done using 8,200 images from the Corel database. We used this database because it represents a widely used set of photos by both amateur and professional graphical designers.

Before we can measure the accuracy of particular methods, we first had to find a challenging and objective ground truth for our tests. The idea of our experiments was to measure the effectiveness of a retrieval method when trying to find a copy of an image in a magazine or newspaper. In order to create the ground truth we printed 82 images using an Epson Stylus 800 color printer at 720 dots/inch and then scanned each of them at 400 pixels/inch using an HP IIci color scanner. Note that we purposely chose a hard test set. The query image is typically very different from the target image. The copy pairs typically differ by color shifts, quantization artifacts, and dithering noise.

We used the HSV color model and quantized H using 4 bits, S using 2 bits and V using 2 bits. The first question we asked was, ”Which distribution is a good approximation for the real color model noise?” To answer this we needed to measure the noise with respect to the color model. The real noise distribution (Figure 1) was obtained as the normalized histogram of differences between the elements of color histograms corresponding to copy-pair images from the training set (50 image pairs).

The best fit Exponential had a better fit to the noise distribution than the Gaussian. Consequently, this implies that L_1 should have better retrieval accuracy than L_2. The Cauchy distribution is the best fit overall, and the results obtained with L_c reflect this. Figure 2 shows the precision/recall graphs. L_c gave a significant improvement in accuracy as compared to L_2, L_1 and L_q. The Kullback relative information, L_q, and the correlograms Cg performed better than L_2 and L_1. Overall, the ML metric gave best accuracy.

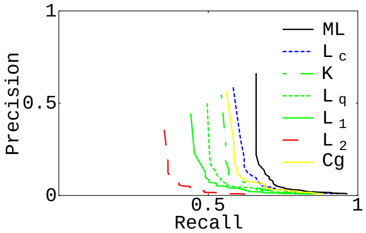


Figure 2: Precision/Recall in Corel database; for L_c, **a**=1.32

Another interesting performance evaluation is to display the percentage of correct copies found within the top n matches. These results are shown in Table 1.

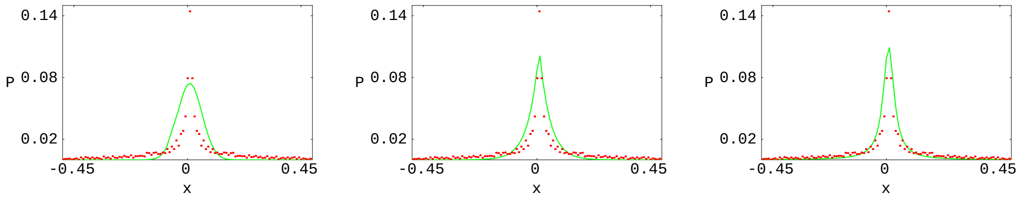
Top	20	40	100
L_2	48.78	54.87	67.07
L_1	62.19	68.29	84.14
L_q	66.34	73.66	88.29
K	68.29	75.60	86.58
L_c **a**=1.32	71.95	79.26	92.68
ML	75.60	82.92	96.34
Cg	71.09	79.63	88.17

Table 1: Retrieval accuracy (%) in the Corel database

4.2. Experiments with Objects Database

In the second experiment we used a database consisting of 500 images of color objects such as domestic objects, tools, toys, food cans, etc. As ground truth we used 48 images of 8 objects taken from different camera viewpoints (6 images for a single object). For this experiment we chose to implement a method designed for indexing by color invariants. Our goal was to study the influence of the similarity noise on the retrieval results.

Using 24 images with varying viewpoint as the training set, we calculated the real noise distribution and studied the influence of different distance measures on the retrieval results. We used the l color model introduced before and we quantized each color component with 3 bits resulting in color histograms with 512 bins. The Cauchy distribution was the best match for the measured noise distribution. The Expo-



(a) Gauss (b) Exponential (c) Cauchy (**a**=2.88)

Figure 3: Noise distribution in color objects database compared with the best fit Gaussian (a) (approximation error is 0.123), best fit Exponential (b) (approximation error is 0.088) and best fit Cauchy (c) (approximation error is 0.077)

nential distribution was a better match than the Gaussian (Figure 3). Table 2 shows the precision and recall values at various scopes. The results obtained with L_c were consistently better than the ones obtained with the other measures.

	Precision			Recall		
Scope	5	10	25	5	10	25
L_2	0.425	0.258	0.128	0.425	0.517	0.642
L_1	0.45	0.271	0.135	0.45	0.542	0.675
L_q	0.46	0.280	0.143	0.46	0.561	0.707
K	0.466	0.279	0.138	0.466	0.558	0.692
L_c	0.525	0.296	0.146	0.525	0.592	0.733
ML	0.533	0.304	0.149	0.533	0.618	0.758
Cg	0.5	0.291	0.145	0.5	0.576	0.725

Table 2: Recall/Precision vs Scope; for L_c, **a**=2.88

Figure 4 shows the precision-recall graphs. The curve corresponding to L_c is above the others showing that the method using L_c is more effective. Note that the Kullback relative information performed better than L_1 or L_2.

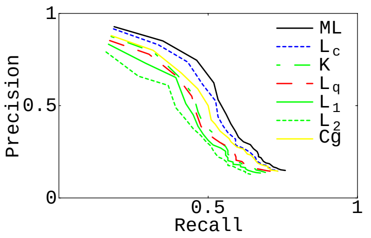


Figure 4: Precision/Recall for color objects database; for L_c, **a**=2.88

In summary, L_c performed better than the analytic distance measures, and the ML metric performed best overall. It is interesting that the Kullback relative information performed consistently better than the well-known histogram intersection (L_1), and roughly the same as L_q. The correlogram (Cg) performed better that L_1, L_2, L_q and K.

5. CONCLUSIONS AND DISCUSSION

In summary, we examined two applications from computer vision which were color based retrieval and color object recognition. Regarding color based retrieval, the first application was finding copies of images which had been printed and then scanned. For this application we used the Corel stock photo database and a color histogram method for finding the copies. The second application dealt with object recognition using color invariance. Both the ground truth and the algorithm came from the work by Gevers, et al. [2]. Note that in their work, they used the L_1 metric. Furthermore, for both applications, we implemented Hafner's quadratic perceptual similarity measure and Huang's correlogram as benchmarks.

For both applications in our experiments, the ML metric consistently outperformed all of the other metrics including the algorithms by Hafner and Huang. Minimizing the ML metric is optimal with respect to maximizing the likelihood of the difference between image elements when the real noise distribution is representative. Therefore, the breaking points occur when there is no ground truth, or when the ground truth is not representative.

Our main contributions are in showing that the prevalent Gaussian distribution assumption is often invalid, and in proposing the Cauchy metric as an alternative to both the L_1 and Kullback relative information. In the case where representative ground truth can be obtained for an application, we provide a method for selecting the appropriate metric. Furthermore, we showed how to create a maximum likelihood metric based on the real noise distribution, and in our experiments we found that it consistently outperformed all of the analytic metrics.

6. REFERENCES

[1] M. Flicker, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by image and video content: The QBIC system. *IEEE Computer*, 28(9):23–32, 1995.

[2] T. Gevers and A. Smeulders. Color-based object recognition. *Pattern Recognition*, 32(3):453–464, 1999.

[3] J. Hafner, H. Sawhney, W. Equitz, M. Flicker, and W. Niblack. Efficient color histogram indexing for quadratic form distance functions. *IEEE Trans. on PAMI*, 17(7):729–736, 1995.

[4] J. Huang, S. Ravi Kumar, M. Mitra, W. Zhu, and R. Zabih. Image indexing using color correlogram. *CVPR*, 1997.

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

[6] S. Kullback. *Information theory and statistics*. 1968.

[7] J. Smith. *Integrated Spatial and Feature Image Systems: Retrieval, Compression and Analysis*. PhD thesis, Columbia University, February 1997.

[8] M. Swain and D. Ballard. Color indexing. *IJCV*, 7(1):11–32, 1991.