ELSEVIER

Pattern Recognition Letters 22 (2001) 223–230

Pattern Recognition Letters

www.elsevier.nl/locate/patrec

Color-based retrieval

Nicu Sebe *, Michael S. Lew

Leiden Institute of Advanced Computer Science, Department of Computer Science, Niels Bohrweg 1, Leiden, CA 2333, The Netherlands

Received 21 February 2000; received in revised form 14 July 2000

Abstract

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 show how to find a distance measure which is optimal in the sense that it maximizes the similarity probability. Our results include experiments which involve distortions derived from changes in viewpoint and the process of printing and scanning. Furthermore, we implemented and tested several promising algorithms from the research literature as benchmarks. © 2001 Elsevier Science B.V. All rights reserved.

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 feature database. Second, given a query image, its feature vector is computed, compared to the feature vectors in the feature database, 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 (Swain and Ballard, 1991). Hafner et al. (1995) 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 (Huang et al., 1997). Gevers and Smeulders (1999) proposed new color models that are invariant to a change in viewing direction, object geometry, and illumination. In (Smith, 1997; Flicker et al., 1995), 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

---

[*] Corresponding author. Tel.: +31-71-527-7050; fax: +31-71-527-6985.

E-mail addresses: nicu@liacs.nl (N. Sebe), mlew@liacs.nl (M.S. Lew).

PII: S 0 1 6 7 - 8 6 5 5 (0 0) 0 0 0 9 2 - 1

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 (sum of the squared differences) or L_1 (sum of the absolute differences) 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 (K), 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

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 $(h_{c_1}, h_{c_2}, \ldots, h_{c_n})$, 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 (Swain and Ballard, 1991)) and L_2. Hafner et al. (1995) 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):

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

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. (1997), 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 kth 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:

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

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 is quite large, the color auto-correlogram is often used instead. This feature only captures spatial correlation between identical colors.

Gevers and Smeulders (1999) 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:

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

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

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

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

The authors Gevers and Smeulders (1999) 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 $i = 1, \ldots, M$ 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 two images (x_i and y_i), the similarity probability can be defined:

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

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 (Huber, 1981). Taking the logarithm of (6), we find that we have to minimize the expression:

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

Note that when the Exponential and Gaussian distributions are used in Eq. (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:

$$L_c = \sum_{i=1}^{M} \log\left(1 + \left(\frac{x_i - y_i}{\boldsymbol{a}}\right)^2\right), \tag{8}$$

where $\boldsymbol{a}$ 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 Eq. (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. 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 as shown in Fig. 1. Note that L_k is a notation for all possible metrics that can be used, e.g., L_1, L_2, L_c. First, for each image in the training set a feature vector (color histogram) is extracted. Second, the real noise distribution is computed as the normalized histogram of differences from the corresponding elements in feature vectors taken from similar images according to the ground truth. The Gaussian, Exponential, and Cauchy distributions are compared 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

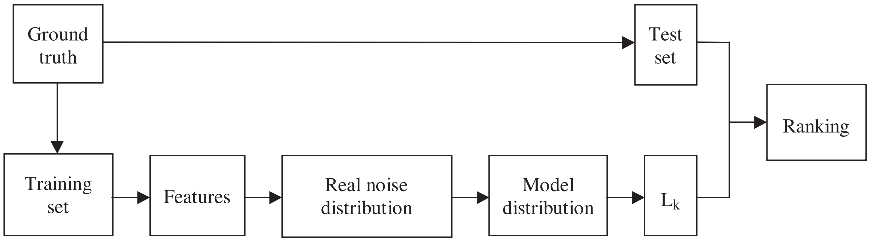

Fig. 1. An overview of the similarity matching algorithm.

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) (Kullback, 1968). 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.

In summary, our algorithm for choosing an analytic metric can be described as follows:

Step 1: Compute the feature vectors from the training set.

Step 2: Compute the real noise distribution from the differences between corresponding elements of the feature vectors.

Step 3: Compare each of the model distributions to the real noise distribution using the chi-square test.

Step 3.1: For a parameterized metric such as L_c compute the value $\boldsymbol{a}$ of the parameter that minimizes the chi-square test.

Step 4: Select the corresponding L_k of the best fit model distribution.

Step 4.1: Use the value $\boldsymbol{a}$ found from Step 3.1 in the parameterized metrics.

Step 5: Apply the L_k metric in ranking.

As noted in Section 3, 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 color auto-correlogram (Cg). Since both L_q and Cg 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 spatial distances, resulting in a 256 feature vector. The comparison was made using L_1.

In a typical image retrieval application the result of a query is a ranking list of images that are hopefully interesting to the user (a group of images similar to the query image). From this list only a limited number of the retrieval results are showed to the user. This is because in general a user will not want to browse through a large number of retrieval results to find the image(s) he is looking for. In this context, we consider in our experiments the best s retrieval results, where we define s as the scope.

The problem is formulated as follows: Let $\mathcal{Q}_1, \ldots, \mathcal{Q}_n$ be the query images and for the ith query $\mathcal{Q}_i, \mathcal{I}_1^{(i)}, \ldots, \mathcal{I}_m^{(i)}$ 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 $|\{\mathcal{I}_j^{(i)} | \operatorname{rank}(\mathcal{I}_j^{(i)}) \leqslant s\}|/m$, and the precision p is defined as $|\{\mathcal{I}_j^{(i)} | \operatorname{rank}(\mathcal{I}_j^{(i)}) \leqslant s\}|/s$.

4.1. Experiments with Corel database

The first experiments were done using 8200 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 (Sebe and Lew, 1999). 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. Examples of the copy pairs are shown in Fig. 2. Note that we purposely chose a hard test set. The query image is typically very different from the target image. In the copy pair containing the child, the textures on the sleeve and on the hair are missing. Also, the cup and water jug are barely discernible. In the other copy pair, note the loss of details in the background mountainside and windows on the lower-right house wall.

Fig. 2. Two examples of copy pairs used: (a, c) the original image; (b, d) copy image.

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 (Fig. 3) 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. Fig. 4 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.

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.

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). Examples are shown in Fig. 5. For this experiment, we chose to implement a method

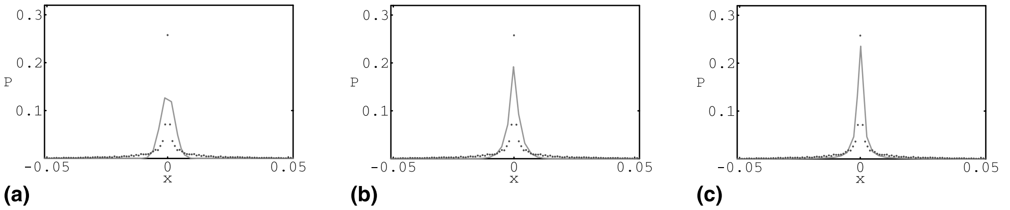

Fig. 3. Noise distribution in Corel database compared with: (a) the best fit Gaussian (approximation error is 0.106); (b) best fit Exponential (approximation error is 0.082); and (c) best fit Cauchy (approximation error is 0.068).

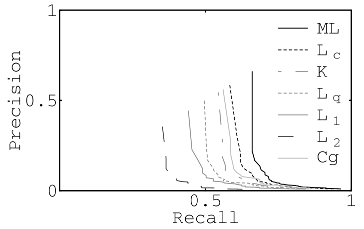

Fig. 4. Precision/recall in Corel database; for L_c, $\boldsymbol{a} = 1.32$.

Table 1
Retrieval accuracy (%) in the Corel database

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, $\boldsymbol{a}=1.32$	71.95	79.26	92.68
ML	75.60	82.92	96.34
Cg	71.09	79.63	88.17

proposed in (Gevers and Smeulders, 1999) 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 Exponential distribution was a better match than the Gaussian (Fig. 6). 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.

Fig. 7 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.

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 than L_1, L_2, L_q, and K.

Fig. 5. Example of images of one object taken from different camera viewpoints.

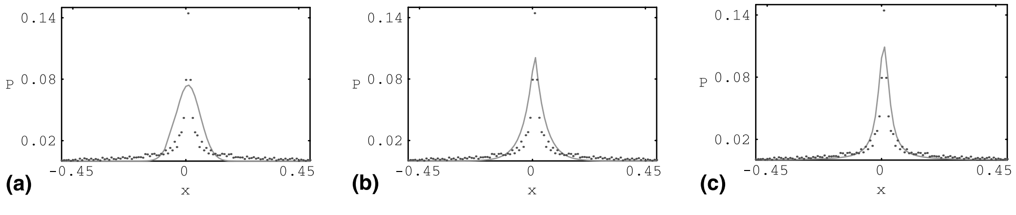

Fig. 6. Noise distribution in color objects database compared with: (a) the best fit Gaussian (approximation error is 0.123); (b) best fit Exponential (approximation error is 0.088); and (c) best fit Cauchy (approximation error is 0.077).

Table 2
Recall/precision vs scope; for $L_c, \boldsymbol{a} = 2.88$

Scope	Precision			Recall		
	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

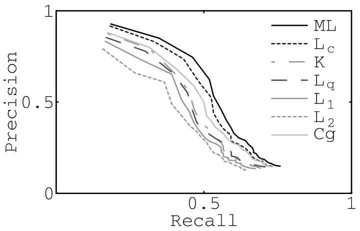

Fig. 7. Precision/recall for color objects database; for L_c, $\boldsymbol{a} = 2.88$.

5. Conclusions and discussion

In summary, we examined two applications from computer vision which involve distortions derived from changes in viewpoint and the process of printing and scaning. 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 finding all of the images of an object in a database where the images were taken from different viewpoints. Both the ground truth and the algorithm came from the work by Gevers and Smeulders (1999). Note that in their work, they used the L_1 metric. Furthermore, for both applications, we implemented Hafner's quadratic perceptual similarity measure (Hafner et al., 1995) and Huang's correlogram (Huang et al., 1997) as benchmarks.

For both applications in our experiments, the ML metric consistently outperformed all of the other metrics including the algorithms by Hafner et al. (1995) and Huang et al. (1997). 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, our technique is not applicable 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 show 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.

References

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

Gevers, T., Smeulders, A., 1999. Color-based object recognition. Pattern Recognition 32 (3), 453–464.

Hafner, J., Sawhney, H., Equitz, W., Flicker, M., Niblack, W., 1995. Efficient color histogram indexing for quadratic form distance functions. IEEE Trans. Pattern Anal. Machine Intell. 17 (7), 729–736.

Huang, J., Ravi Kumar, S., Mitra, M., Zhu, W., Zabih, R., 1997. Image indexing using color correlogram. IEEE Conf. Computer Vision Pattern Recognition, 762–768.

Huber, P.J., 1981. Robust Statistic. Wiley, New York.

Kullback, S., 1968. Information Theory and Statistics. Dover Publications, New York.

Sebe, N., Lew, M.S., 1999. Robust color indexing. ACM Multimedia, 239–242.

Smith, J.R., 1997. Integrated spatial and feature image systems: retrieval, compression and analysis. Ph.D. thesis, Columbia University, February 1997.

Swain, M.J., Ballard, D.H., 1991. Color indexing. Int. J. Computer Vision 7 (1), 11–32.