Efficient Descriptors of Hue Distributions from Kernel Density Estimators and Fourier Transforms

Linh V. Tran, Reiner Lenz
Department of Science and Technology, Linköping University
SE-60174 Norrköping, Sweden

Abstract

Many color-based image retrieval systems define the similarity (with regard to color) between two images as the similarity between the probability distributions of the color vectors in the images. These probability distributions are almost always estimated by histograms. Histograms have however the disadvantage that they are discontinuous and their form depends on the selection of the histogram bins. Results from probability theory and statistics show that kernel-based estimators are superior to the histogram in many respects. Previous studies in image retrieval have however shown that a naive application of kernel-based estimators provide no improvement in retrieval performance.

In this paper we first motivate why a combination of kernel-based estimators and Fourier transform theory provides good estimators of the similarity of hue-distributions. We then show that Fourier coefficients provide efficient descriptors of the probability distributions and that these Fourier coefficients can be directly used to compute the similarity between the hue distributions of images. Next we describe two methods to select the most relevant Fourier coefficients for image retrieval. We will argue that in image retrieval we should not select those Fourier coefficients that are most important for the description of the probability distributions themselves but that we should select those coefficients that are most important in the estimation of the difference between similar distributions. In the experimental part of the paper we describe the performance of these kernel-based methods when they are applied to image retrieval tasks involving the MPEG7 image database. We will show that the retrieval performance of the kernel based method is better than the performance of histogram methods and we will show that the retrieval performance is also relatively insensitive to the choice of the Kernel and the width of the Kernel.

Introduction

Color is widely used for content-based image retrieval. In these applications the color properties of an image are characterized by the probability distribution of the colors in the image. These probability distributions are very often estimated by histograms [3, 4]. Well-known problems of histogram based methods are: the sensitivity of the histogram to the placement of the bin edges, the discontinuity of the histogram as a step function and its inefficient use of the data in estimating the underlying distributions, compared to other estimators [2, 6, 5, 10]. These problems can be avoided by using other methods such as kernel density estimators.

To our best knowledge there are only a few papers [1, 7, 8] that use kernel density estimators in image retrieval. One reason is that the straightforward way of applying kernel density estimators gives very bad retrieval performance [7, 8]. Gaussian expansions were proposed in [7, 8] where an improvement of the kernel based method compared to the traditional histogram-based method was obtained. In this paper we use the Fourier Transform together with kernel based methods. We apply them to hue-based image retrieval and show that these estimates are superior to histogram based methods.

Non-parametric density estimators

Fast color based image retrieval methods usually describe the color properties of an image by the probability distribution of the colors in the image. The first decisions in the design of such a color based retrieval system are thus: (1) how to estimate the probability distribution and (2) how to represent the estimated distribution efficiently. Methods to estimate probability distributions come in two variations: parametric and non-parametric methods. Parametric methods choose a family of probability distributions first and then they estimate these parameters from the data. Typical examples are Gaussian mixture models. These methods are not appropriate for retrieval since the color distributions of interest can, and will, be of very different types. Non-parametric methods avoid these problems by describing the probabilities directly. The simplest non-parametric method is the histogram. It divides the color space into

different non-overlapping regions and then it describes a given distribution by counting how many pixels in an image are located in each region. Another slightly different approach is to estimate in each point x in color space the density f by

$$f(x) \approx \tilde{f}(x) = \frac{\text{Number of pixels in N}(x)}{\text{Number of samples} \cdot \text{Volume of N}(x)}$$

where $N(x)$ is a neighborhood of x. This gives a discrete, discontinuous estimate and therefore it is preferable to use kernel based methods. Such a kernel estimator $\hat{f}_K$ at point x of a distribution $f(x)$ is defined by

$$\hat{f}_K(x, h) = \frac{1}{Nh} \sum_{n=1}^{N} K_h(x, X_n) \qquad (1)$$

where N is the total number of samples (pixels), X_n are the data points (the color values of the pixels in the image), h denotes the window width of the estimator, also called the smoothing parameter or the bandwidth. The kernel $K_h(x, X_n)$ counts the data point X_n with a weight depending on its relation to the point x. Often $K_h(x, X_n)$ depends on the distance between x and X_n :

$$K_h(x, X_n) = K_h(\|x - X_n\|)$$

In almost all applications we define a kernel as a non-negative real function K with $\int K(x)dx = 1$ and we measure the influence between points as a function of their difference:

$$\hat{f}_K(x, h) = \frac{1}{Nh} \sum_{n=1}^{N} K(x - X_n)/h\}$$

$$= \frac{1}{N} \sum_{n=1}^{N} K_h(x - X_n) \qquad (2)$$

The scaled kernel is $K_h(u) = h^{-1}K(u/h)$.

From this form of the density estimate we make two observations and we will show that both of them will lead naturally to the Fourier Transform.

Symmetry-Based Compression

In the following discussion we consider only estimators of hue distributions which provide in a certain sense the simplest example. We also restrict us to the case of continuous distributions. The discrete case can be treated in the same way within the framework of the discrete Fourier transform. Most of the characteristic properties can be substantially generalized in the general framework of harmonic analysis.

Hue is in almost all color systems described as an angular variable. The values of x, X_n etc. are therefore all located on the unit circle. For a given, fixed, image ω the true (unknown) density f_ω is therefore a function defined on the unit circle. We now select a basis b_k for the Hilbert space of square integrable functions on the unit circle. In this system the density has an expansion

$$f_\omega = \sum_k \langle f_\omega, b_k \rangle b_k \qquad (3)$$

In the retrieval application we cannot keep all the coefficients $\langle f\omega, b_k \rangle$ but we have to select a finite number (say L), of them: $\langle f_\omega, b_1 \rangle, \ldots \langle f_\omega, b_L \rangle$. For each image, i.e. each f_ω such an approximation leads to an error $\delta(\omega, L)$ and it is natural to choose the $b_1, \ldots, b_L$ such that this approximation error is minimized. We have now formulated the problem as an minimum-least-squared error problem and it can be shown that for shift-invariant problems the optimal basis are the complex exponentials $b_k(x) = e^{ikx}$. Shift-invariance means in this case that for every distribution $f(x)$ and every constant ξ there is an equally likely shifted distribution $f(x + \xi)$. Here the assumption of shift-invariance of the hue-distributions leads to Fourier series as an optimal solution.

Computational Advantage

Next we note that the sum $\sum_{n=1}^{N} K_h(x - X_n)$ is the convolution of the kernel function K_h with the sum of Dirac delta functions located at the positions of the data points $X_n : \hat{f}_K(x, h) = (K_h \star F)(x)$ where F puts weight $1/N$ at each data point. This convolution form of the estimate suggests that the Fourier Transform can lead to substantial simplifications. We therefore compute the Fourier transform $\mathfrak{F}(f)_K(y, h)$ of the distribution in Eq. 1 as follows:

$$\mathfrak{F}(f)_K(y, h) = \int \hat{f}_K(x, h) \cdot e^{-ixy} dx$$

$$= \frac{1}{Nh} \int \sum_{n=1}^{N} K\{(x - X_n)/h\} \cdot e^{-iyx} dx$$

$$= \frac{1}{N} \sum_{n=1}^{N} \int K(t) \cdot e^{-iy(ht + X_n)} dt$$

$$= \frac{1}{N} \left\{ \sum_{n=1}^{N} e^{-iyX_n} \right\} \int K(t) \cdot e^{-iyht} dt$$

$$= \frac{1}{N} \left\{ \sum_{n=1}^{N} e^{-iyX_n} \right\} \mathfrak{F}(K)(yh) \qquad (4)$$

where $\mathfrak{F}(K)$ is the Fourier transform of the kernel K. We see that the factor $\sum_{n=1}^{N} e^{-iyx_n}$ of the Fourier transform $\mathfrak{F}(f)_K(y, h)$ in Eq. 4 is independent of the kernel and the smoothing parameter h. It can thus be computed from the

data once and then new estimates with different kernels and smoothing parameters can be computed without accessing the data again.

Similarity and Compression

Motivated by the previous observations we choose to estimate the probability density by its Fourier transform Eq. (4) and we show how to compute the similarity between images by these Fourier transforms.

We define the distance between two images I_1, I_2 as the distance between the two corresponding hue distributions $f_1(x)$ and $f_2(x)$ and compute the similarity between them using the Parsevals relation:

$$
\begin{aligned}
\text{similarity}(I_1, I_2) \quad &= \quad \text{similarity}(f_1(x), f_2(x)) \\
&= \langle f_1(x), f_2(x) \rangle \\
&= \frac{1}{2\pi} \langle \mathfrak{F}(f_1)_K(y,h), \mathfrak{F}(f_2)_K(y,h) \rangle \quad (5)
\end{aligned}
$$

For hue-distributions the Fourier transform is actually a Fourier series since the functions are all defined on the circle. The scalar product $\langle \mathfrak{F}(f_1)_K(y,h), \mathfrak{F}(f_2)_K(y,h) \rangle$ is thus equal to the scalar product of the Fourier coefficients of the hue probability distributions and we can regard the Fourier coefficients as descriptors of the hue properties of an image. Since it is desirable to minimize the number of descriptors we have to choose a method to select the most important Fourier coefficients. Arguing as above we can select those coefficients that contribute most to the average reduction of the reconstruction error. We call this the PCA selection method and denote it by the subscript D (**D**irect estimation based on the reduction of the L^2 reconstruction error). This may not be an optimal solution for image retrieval since in retrieval we are not primarily interested in the description of the distributions but we are mainly interested in estimating the similarity between similar images. For an image we thus want to have a good estimation of the similarity to other images that have very similar properties. The approximation for very different images is of no importance since they are of no interest in retrieval. In our experiments we implemented this, difference-based, feature selection method as follows:

1. We select from the image database 100 random images and use each of them as query image

2. For each of these 100 images we select the 50 most similar (using all Fourier coefficients)

3. For each of the query images and each of its neighboring images we compute a difference image. This results in 5000 images

4. From this database of 5000 images we compute those Fourier coefficients with the maximal mean value (best approximation) of the differences.

We call this the local estimation method and denote it by the subscript L (i.e. the **L**ocal estimation method).

In both cases we describe the two Fourier transforms (or two hue distributions of the two images) by selecting the most important Fourier coefficients $\eta_{(1,m)}, \eta_{(2,m)}$ of $\mathfrak{F}_1$ and $\mathfrak{F}_2$ with $m = 0, \ldots M$ and estimate the similarity between the two images by the inner product of two low dimensional vectors:

$$
\text{similarity}(I_1, I_2) \approx \frac{1}{2\pi} \sum_m \eta_{(1,m)} \cdot \overline{\eta_{(2,m)}} \quad (6)
$$

Experiments

In our experiments we evaluated the retrieval performance of the above new descriptors in our image retrieval engine using the MPEG7 database of 5466 images (see [11] for a description).

Evaluation and comparison of different retrieval methods is a difficult problem. In our experiments we used the ANMRR (average normalized modified retrieval rank) method. The definition of ANMRR is quite complex and here we only need to know that a lower ANMRR value means better retrieval performance (for a complete description of ANMRR see [11]).

We first compare the retrieval performance of the histogram and the kernel based method (see Figure 1). In this experiment the triangular kernel and the smoothing parameter $h = 0.0056$ was used. The experiment shows that the kernel based retrieval is always better than the histogram based method. For example, using 10 Fourier coefficients gives a retrieval performance comparable to the retrieval based on 23 coefficients from a histogram. The improvement is largest for a low number of coefficients, which is the most relevant case for large image databases.

The application of kernel based methods requires the selection of the kernel to be used. It is therefore of interest to see how retrieval performance depends on the choice of the kernel. The results of our experiments with different kernels is summarized in Figure 2. In general we found that using different kernels gives comparable retrieval performance when the kernel is not over-smoothed. When $h < 0.01$ all kernels had identical retrieval properties. We tested seven different kernels (Epanechnikov, Biweight, Triweight, Normal, Triangular, Laplace, Logistic, detailed definition of kernels can be found in [10]). In the experiment shown in Figure 2 an over-smoothed kernel with $h = 0.05$ is used.
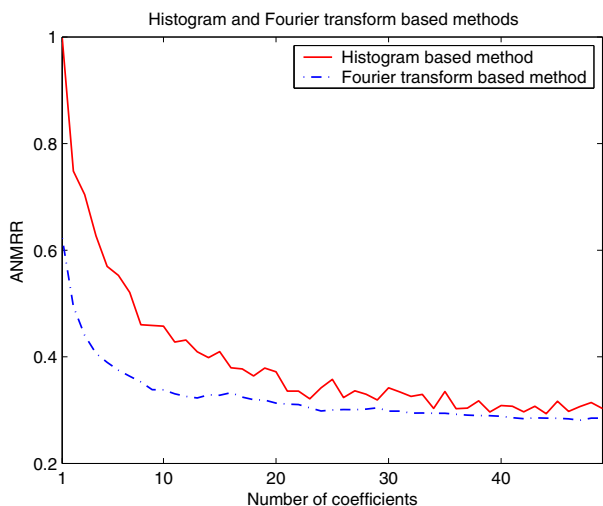
Figure 1: Retrieval performance of histogram and Fourier transform based method using triangular kernel, the smoothing parameter $h = 0.0056$
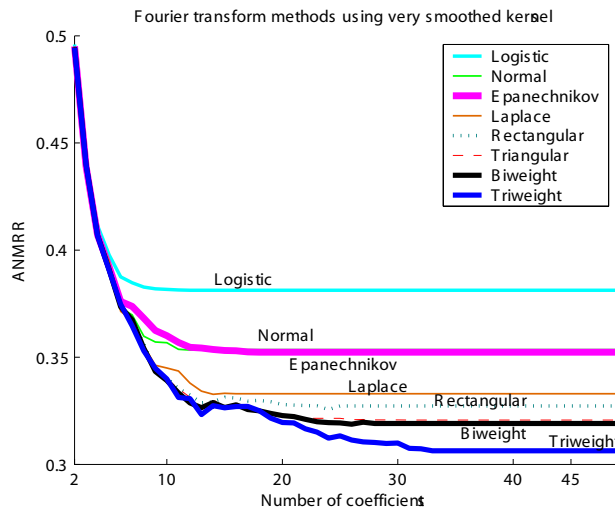

Figure 2: Retrieval performance of Fourier transform based method using different kernels with smoothing parameter $h = 0.05$

Once the kernel has been selected the next choice is the selection of the smoothing parameter h. The value of the bin-size is a very critical parameter in histogram methods and therefore it is important to test how critical the corresponding parameter h is in the kernel based methods. We used the MPEG database and computed the retrieval performance for a wide selection of smoothing parameters and number of Fourier coefficients. The result is illustrated in Figure 3. It shows that the retrieval performance is almost independent of the value smoothing parameter if it has been chosen in a reasonable region. From our experiments we conclude that using a larger smoothing parameter h gives a better retrieval performance. However the performance does not change for h below 0.005. We tested 30 different smoothing parameters ranging from 0.0001 to 0.2.

We also compared the performance of the direct and the local coefficient selection method. We used different kernels and different smoothing values. The results, collected in Table 1 shows that the local feature selection method is slightly more efficient than the direct method.

Conclusions

From the results of our experiments we found that applications of kernel methods in which the estimated density distribution is sampled and the sample values are used as descriptors are comparable to histogram methods of the same complexity. The additional computational cost is thus not justified and this may be one of the reasons why kernel methods are not widely used in image database retrieval applications.

In this paper we replaced the sample based description by a description in which the estimated probability distribution is expanded in a basis and the expansion coefficients are used as descriptors. The application of this strategy requires three major decisions:

1. Selection of the kernel function

2. Selection of the smoothing parameter

3. Selection of the basis used to describe the estimated distribution

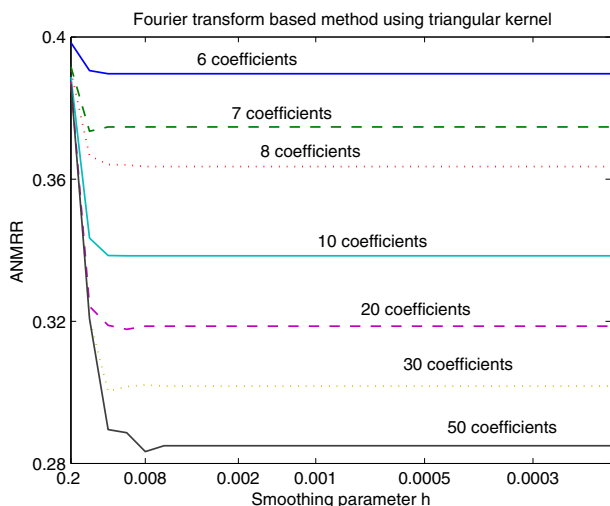Our experiments show that the selection of the kernel function and the smoothing parameter are not very critical. In

Figure 3: *Retrieval performance of Fourier transform-based method using triangular kernel with different smoothing parameters.*

Method	M_L	M_D
Biweight kernel, $h = 0.2$	0.4786	0.4954
Biweight kernel, $h = 0.05$	0.4749	0.4946
Biweight kernel, $h = 0.008$	0.4748	0.4945
Biweight kernel, $h = 0.0056$	0.4748	0.4945
Biweight kernel, $h = 0.001$	0.4748	0.4945
Biweight kernel, $h = 0.0002$	0.4748	0.4945
Triangular kernel, $h = 0.001$	0.4748	0.4945
Normal kernel, $h = 0.001$	0.4748	0.4945
Epenechnikov kernel, $h = 0.001$	0.4748	0.4945

Table 1: *The retrieval performance improvement of the M_L method over M_D method of selecting the coefficients of the most three important frequencies for CBIR.*

this paper we used only hue distributions and in this case we argued that the Fourier basis is optimal under very general conditions. When other types of distributions are used, other basis systems are certainly preferable.

Acknowledgements

References

[1] T. Gevers. Robust histogram construction from color invariants. In *Proc. of Intl. Conf. on Computer Vision*, 2001.

[2] T. Hastie, R. Tibshirani and J. Friedman. *The Elements of Statistical Learning*. Springer, New York, 2001.

[3] Y. Rui, T.S. Huang, and S.-F. Chang. Image retrieval: Current techniques, promising directions, and open issues. *J. of Visual Communication and Image Representation*, 10:39–62, 1999.

[4] R. Schettini, G. Ciocca, and S. Zuffi. Color in databases: Indexation and similarity. In *Proc. Int'l Conf. on Color in Graphics and Image Processing*, pages 244–249, 2000.

[5] D.W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, New York, 1992.

[6] B.W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, 1986.

[7] L. V. Tran. *Efficient Image Retrieval with Statistical Color Descriptors*. PhD thesis, Linköping University, 2003.

[8] L. V. Tran and R. Lenz. Characterization of color distributions with histograms and kernel density estimators. In *Proc. of SPIE Internet Imaging IV Conf., Electronics Imaging, Santa Clara, USA*, Jan. 2003.

[9] L.V. Tran and R. Lenz. PCA based representation for color based image retrieval. In *IEEE Proc. of International Conference on Image Processing*, 2001.

[10] M.P. Wand and M.C. Jones. *Kernel Smoothing*. Chapman and Hall, 1995.

[11] JTC1/SC29/WG11, "Common datasets and queries in MPEG-7 color core experiments," ISO/IEC, Tech. Rep. Doc. MPEG99/M5060, October 1999.