Histograms analysis for image retrieval

ITC-irst Tech. Rep. 9812-03

To appear on Pattern Recognition

R. Brunelli and O. Mich
Istituto per la Ricerca Scientifica e Tecnologica
I-38050 Povo, Trento, ITALY

Abstract— **This paper analyzes the use of histograms of low level image features, such as color and luminance, as descriptors for image retrieval purposes. A novel definition of histogram capacity curve taking into account the density distribution of histograms in the corresponding spaces is proposed and used to quantify the effectiveness of image descriptors and histogram dissimilarities in image retrieval applications. The results permit the design of scalable image retrieval systems which make optimal use of computational and storage resources.**

Keywords: **image retrieval, histograms, density estimation, distribution comparison.**

1. Introduction

A currently active line of research and development in the Computer Vision community is the design and development of efficient tools for accessing multimedia material, such as video and still images, using their media specific features. In particular, several research papers and tools have been presented for image retrieval based on low level visual features, such as color and luminance, as image descriptors ([3], [5], [6], [7], [8], [9], [10], [15], [16], [17], [19], [22]).

This paper considers how an efficient and effective system for image retrieval can be based on the statistics of such low level features. A novel definition of histogram capacity curve taking into account the density distribution of histograms in the corresponding spaces is proposed and used to quantify the effectiveness of image descriptors and histogram dissimilarities in image retrieval applications.

In the light of this definition, the following problems have been considered:

- how descriptor effectiveness should be assessed,
- how histograms should be compared,
- how many bins are necessary,
- how the low level features should be mapped before being used to compute the corresponding histograms,
- the effect of the introduction of spatial information by using multiple regions of interest.

The results of histograms analysis can be used for the development of an efficient image retrieval system that minimizes the size of the image descriptors while maintaining good discriminating ability. The next section discusses how images can be described by the probability distributions of low level features. The notion of *histogram capacity*, upon which all the analyses performed in this paper are based, is then introduced. Methods for the comparison of histograms and for choosing the number of bins are presented and discussed in the two following sections. Finally, some applications are considered and the last section of the paper summarizes the results.

2. Histogram capacity

Digital images are usually represented as a set of elements, called pixels, arranged in a regular structure, e.g. a square grid. A small set of numbers is associated to each pixel (its luminance, color components, etc). It is then possible to represent a digital image $\mathcal{I}$ with the following notation:

$$\mathcal{I} = \{(x, y, \boldsymbol{v}(x, y))\}_{(x,y)\in \boldsymbol{S}, \boldsymbol{v}\in \boldsymbol{V}} \tag{1}$$

where $\boldsymbol{S}$ represents the set of possible pixel locations and $\boldsymbol{V}$ the set of values associated to the pixel locations. By quantizing $\boldsymbol{S}$ and $\boldsymbol{V}$, a multivariate frequency distribution can be derived from the population data (i.e. the pixels). The resulting distribution can be represented by a multidimensional histogram, giving for each of the cells of the quantized spaces, the fraction of pixels whose description falls within the cell. The data can be further summarized by considering marginal distributions of the pixel descriptors, that is, by integrating over some of the data dimensions. Several commonly used descriptors are then easily obtained:

- Global image histograms: they are obtained by integration over the spatial coordinates. All spatial information is lost and only the population of $\boldsymbol{V}$ is considered.
- Horizontal/vertical projections: only one of the spatial coordinates is integrated over.
- Multiple region histograms: none of the spatial coordinates are integrated over, but they are heavily quantized (approaching the state of categorical data).

The resulting descriptors are density distributions that can be compared for similarity using the values of the statistics upon which common tests, such as the χ^2, are based. As the densities are considered in a binned form (histograms), vector distances, such as the Euclidean and the L_1 norms, can also be used to quantify the similarity of two descriptors as they are represented by numerical vectors.

Binned representations of the densities of low level image features can be used in an image retrieval task to sort the images stored in a database by decreasing similarity

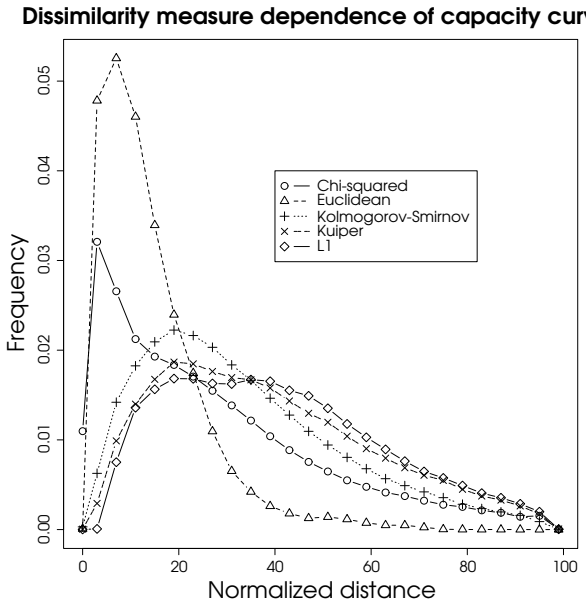

Fig. 1. The figure reports the capacity curves for several dissimilarity measures. The curves are built by averaging the curves derived by the three descriptors I, H, and E. Note how the L_1 norm provides the best results. All curves are computed on the VIDEO database.

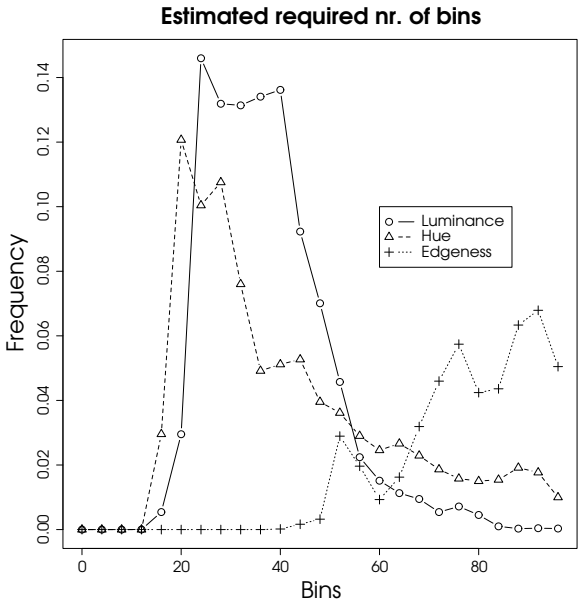

Fig. 2. The curves report the distribution of the estimated required number of bins for some of the descriptors used in the tests over the VIDEO database according to Scott's rule.

Dissimilarity measure	$\mathcal{E}_{\texttt{VIDEO}}$	$\mathcal{E}_{\texttt{STILLS}}$
L_1	41	50
Kuiper	39	48
Kolmogorov-Smirnov	34	42
Chi-square	27	36
Euclidean	15	20

TABLE I

THE EFFECTIVENESS $\mathcal{E}$ OF THE DIFFERENT DISSIMILARITY MEASURES FOR THE TWO DATABASES VIDEO AND STILLS, SCALED SO THAT THE MAXIMUM POSSIBLE VALUE CORRESPONDS TO 100. THE CAPACITY CURVES ARE AVERAGED OVER THREE DESCRIPTORS: HUE, LUMINANCE, AND EDGENESS.

with a query image. All the database images are indexed by their descriptors histograms which are then used as access keys. Let us assume, for the moment, that some design issues, such as histogram resolution and comparison, have been solved (this issues are discussed at length in the following sections). The notion of *histogram capacity* has been introduced in [20] as a measure of the effectiveness of histogram indexing in image retrieval tasks. The capacity of an n-bins histogram space $\mathcal{H}$ is defined as the maximal number of spheres with a given radius that can be packed in the embedding R^n space so that the sphere centers lie within $\mathcal{H}$. Starting from this definition, a new one can be introduced that does not rely on the geometrical structure of space $\mathcal{H}$ but takes into account the distribution of histograms within it. This is particularly important as the distribution of histograms within the space cannot be assumed uniform. The following two definitions formalize the new notion of capacity:

Definition 1: Given an n-dimensional histogram space $\mathcal{H}$, a dissimilarity measure[1] d on $\mathcal{H}$, the *capacity curve* C of $\mathcal{H}$ is defined as the density distribution of the dissimilarity between the two elements of all possible histogram couples within $\mathcal{H}$.

Definition 2: The *capacity* $\mathcal{C}(t)$ of an histogram space $\mathcal{H}$ is given by

$$\mathcal{C}(t) = \int_{y>t} C(y)\, dy \qquad (2)$$

where C is the capacity curve of $\mathcal{H}$ with respect to a given dissimilarity measure.

There are two major differences from the definition introduced in [20]:

1. there is no need for a distance function: the triangular inequality of distances is not necessary for image retrieval applications, and requiring it could be a limiting factor in the design of a retrieval system;
2. the difficult task of estimating a 'maximal number' has been transformed into the easier estimation of an average value using the empirical capacity curve computed by considering all image couples within the database.

Histogram capacity curves provide a basis on which the effectiveness, i.e. the discrimination ability, of different image descriptors can be compared. The shape of $C(t)$ is an indicator of the distribution of histograms in $\mathcal{H}$ with the

[1] In this context, a dissimilarity measure is a bounded, positive, and symmetric function defined over a subset of $\boldsymbol{R}^n \times \boldsymbol{R}^n$.

topology induced by the selected comparison dissimilarity measure. If the average value of dissimilarity is low, histograms are not sparse enough in $\mathcal{H}$ and histogram indexing is not effective. This can be formalized by the following definition:

Definition 3: The *indexing effectiveness* $\mathcal{E}$ of an histogram space $\mathcal{H}$ is given by the average dissimilarity value:

$$\mathcal{E} = \int y\, C(y)\, dy \tag{3}$$

The indexing effectiveness $\mathcal{E}$ can be used to assess several descriptor-dissimilarity combinations for image retrieval applications.

3. Histogram comparison

As histogram capacity depends on the dissimilarity measure d chosen for the comparison, it is useful to compare the capacity over a given space $\mathcal{H}$ for different choices of d to see which one gives the best indexing effectiveness. Binned densities can be compared using statistical tests as well as using vector norms. Non-parametric tests are required as *a-priori* knowledge about the shape of the densities of low level image descriptors is generally not available. Note however that in an image comparison task we do not need the significance associated to the test statistics, but only the statistics values themselves, e.g. normalized to the interval $[0, 100]$, low values representing similar densities and high values different ones. The following tests are considered in this paper:

- **Chi-square**: the χ^2 test can be applied to binned distributions P, Q and the corresponding statistic is:

$$\chi^2(P,Q) = \sum_i \frac{(P_i - Q_i)^2}{P_i + Q_i} \tag{4}$$

  where subscript i represents the bins.
- **Kolmogorov-Smirnov**: it is applicable to unbinned (cumulative) distributions $S(x), R(x)$ of a scalar variable and is defined by the following statistic:

$$D_{KS}(S,R) = \max_{-\infty<x<\infty} |S(x) - R(x)| \tag{5}$$

  An important characteristic of D_{KS} is its invariance to reparametrization of the variable x. The test can also be applied to binned distributions but the resulting statistic underestimates the true value. Generalizations of the Kolmogorov-Smirnov statistic to multi-dimensional data exist but are much more computationally demanding (see [4]).
- **Kuiper**: it is applicable to unbinned distributions of a scalar variable and it is based on a statistic similar to the one used by the Kolmogorov-Smirnov test:

$$\begin{aligned} D_{Ku}(S,R) \quad = \quad & \max_{-\infty<x<\infty} [S(x) - R(x)] + \\ & \max_{-\infty<x<\infty} [R(x) - S(x)] \end{aligned} \tag{6}$$

  It can also be applied to binned distributions and presents some advantages over the Kolmogorov-Smirnov statistic: it is appropriate for circular data, and it is more sensitive to the tails of the distributions.

Binned densities can be represented as vectors and their difference can be quantified by any metric defined on the corresponding vector spaces. A widely used family of metrics is the L_p family defined by:

$$L_p(\boldsymbol{x}, \boldsymbol{y}) = \left(\sum_i |x_i - y_i|^p \right)^{1/p}, \quad p \geq 1 \tag{7}$$

Among the most used metrics we find the Euclidean norm (L_2) and the Manhattan norm (L_1).

In order to assess the relative merits of the different dissimilarity measures for image retrieval, some experiments have been performed on two databases (`VIDEO` and `STILLS`) using the following low level image descriptors:

- hue, H: a scalar descriptor which associates to an (r, g, b) triple representing the pixel color its tint; the resulting density represents a circular variable;
- luminance, I: a scalar descriptor which associates to an (r, g, b) triple representing the pixel color the normalized sum of its components;
- edgeness, E: the magnitude of the gradient $\sqrt{(\partial_x I)^2 + (\partial_y I)^2}$ where I represents the image luminance;
- hue co-occurrence: space $\boldsymbol{S}$ is partitioned into couples of pixels by means of a binary spatial relation: a pixel located at (x, y) is associated to a pixel at $(x + \Delta x, y + \Delta y)$ and the tints of the two pixels are used as indices in a 2-dimensional histogram;
- luminance co-occurrence: the same as hue co-occurrence using pixel luminances.

The two image databases used for the computation of the descriptors capacity curves present distinct characteristics:

- `VIDEO`: a set of 40000 frames from nine different video clips. Each video clip, approximately five minutes long, was sampled at 25 frames per second. The video material was varied, ranging from comics, news, to documentaries and action movies. This database is important for two reasons:
  - it covers a wide variety of images over which image searches could be useful (e.g. for video on demand systems, news production, etc);
  - for each given image (frame) it provides a set of similar images, the neighboring frames: this is relevant when the database is used for user-validation of an image retrieval system. It is also important as it provides an adequate coverage of the neighborhood of each query density for statistical analysis.
- `STILLS`: a set of 3500 still images from a commercial collection, providing more colorful and high quality images than the *average* video material of the above database.

For the reasons detailed above, the `VIDEO` database has been chosen as the reference database for the analyses reported in the following sections while the `STILLS` database is used to cross validate the results.

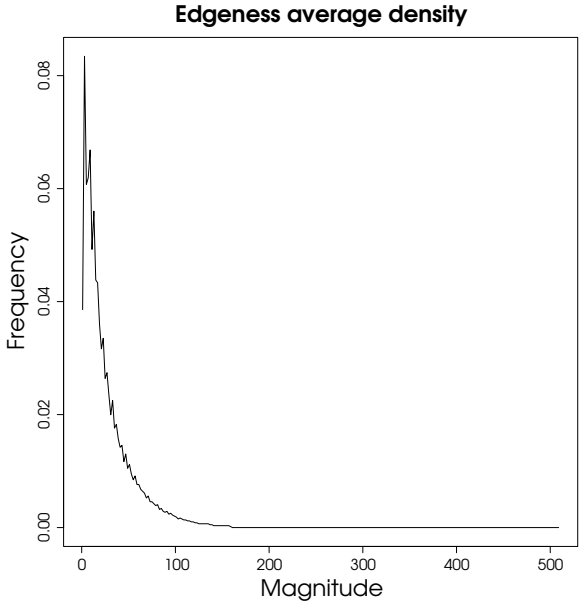

Fig. 3. The average density of the edgeness descriptor.

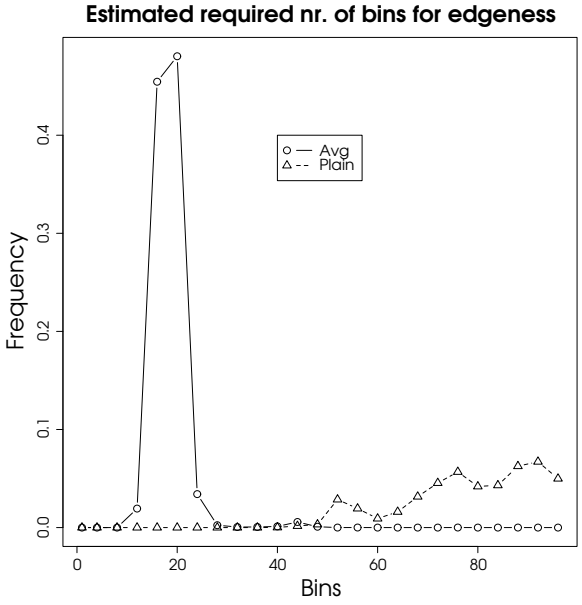

Fig. 4. The curves report the estimated required number of bins using Scott's rule for the edgeness descriptor without any mapping, and using the average cumulative distribution for equalizing the descriptor. Note that the number of bins estimated according to Sturges' rule is 14 as the images considered have 6336 pixels and the descriptor was normalized to a range $\Delta = 256$. This value grossly underestimates the required number of bins, especially when no mapping is applied to the edgeness descriptor.

The capacity curves computed using the previously described dissimilarity measures are reported in Figure 1 for the VIDEO database. In order to compare the results from different dissimilarity measures, characterized by different ranges of values, the measures have been rescaled so that their maximum value is 100. Inspection of the plots shows that the L_1 norm provides the most effective curve, with an even distribution over the available range. It is also interesting to note that the L_1 and Kuiper curves are close. The left tail of the different capacity curves on the VIDEO database deserves some attention. For each image, a subset can be found of very similar images: the neighboring frames. The χ^2 and Euclidean curves are very steep at the left tail: they do not make optimal use of the available range and are possibly much more sensitive to small differences between images. The indexing effectiveness $\mathcal{E}$ is reported in Table I for the two databases. These values support the preceding considerations on which dissimilarity measure is best suited to image retrieval tasks.

4. Optimal histogram resolution

In a typical image retrieval application, many items are usually involved: it is then important for the descriptive data to require only a fraction of the image size for storage. Furthermore, the complexity of the search depends on the descriptors size: the shorter the descriptors, the faster the search. Binning a probability distribution is a way to summarize it, and the applied quantization (bin size) controls how much the distribution is summarized. The bin width (or widths if a non uniform quantization is employed) is therefore the most important parameter of a histogram. The chosen quantization may result in a over-smoothed, correct, or under-smoothed representation of the corresponding distribution. A very simple rule proposed by Sturges [21] relates the bin width $\hat{h}$ to the range of data Δ and the sample size n:

$$\hat{h} = \frac{\Delta}{1 + \log_2(n)} \tag{8}$$

Theoretical analysis [13] shows that the resulting bin width provides an over-smoothed histogram (especially for large samples). The optimal rate of decay of the bin width, with respect to L_p norms, is $n^{-1/3}$ and rules have been proposed of the form:

$$\hat{h} = \hat{C} n^{-1/3} \tag{9}$$

where $\hat{C}$ is an appropriate statistic. Scott [14] proposed a rule of this type, based on calibration with a normal distribution, with the following form:

$$\hat{h} = 3.49 \; \hat{\sigma} \; n^{-1/3} \tag{10}$$

where $\hat{\sigma}$ is an estimate of the standard deviation. As noted in [23], this rule is actually the simplest member of a family of rules exhibiting good theoretical properties and practical performance, but requiring, with the exception of eq. (10), extensive computations. Scott's rule is appropriate whenever the distribution has a roughly unimodal shape. The distribution of the number of bins estimated according to Scott's rule is reported in Figure 2 for some of the image descriptors.

By looking at the plots of Figure 2, it is apparent that the edgeness descriptor requires a very high number of bins for accurate representation. This is due to the shape of its average density, which is reported in Figure 3. The density

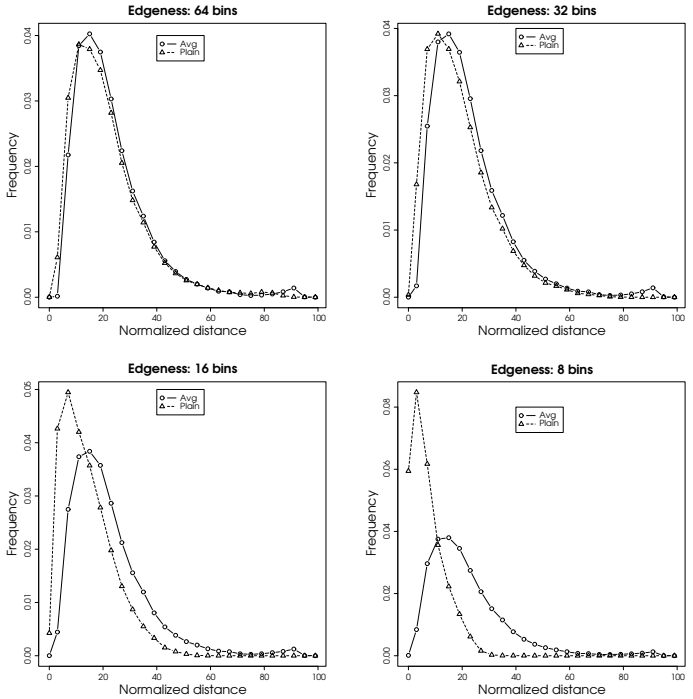

Fig. 5. The plots report the effect of using the cumulative distribution for increasing the uniformity of bin populations. Note that the effect is marked only when a reduced number of bins is used.

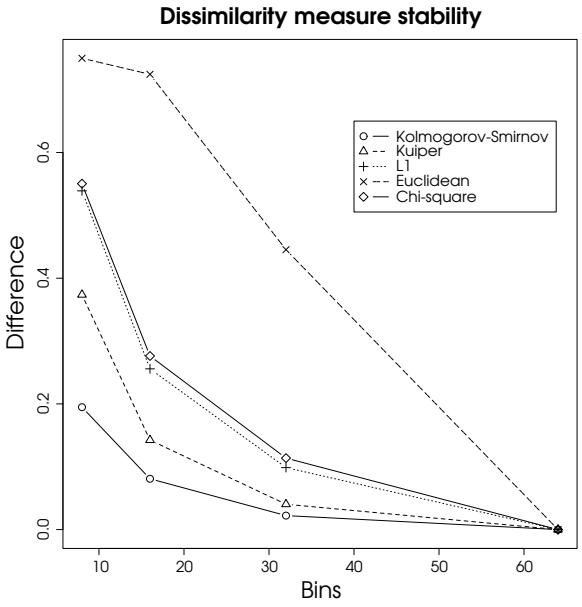

Fig. 6. The plots highlight the dependency of the capacity curves on the number of bins for the dissimilarity measures on the hue descriptor. The capacity curves at 64 bins are taken as reference, and the area (i.e. the L_1 distance) between the reference capacity curve and the corresponding curve at a lower histogram resolution is reported in the graph.

is biased towards low values, with the bins corresponding to the highest values being scarcely populated. This has two negative effects:

- many bins corresponding to the right tail are empty: they do not provide (on average) much information, yet they must be stored and compared;
- if the comparison of densities is done using tests such as the χ^2, high fluctuations in the population of the right tail bins may invalidate the significance of the comparison.

A partial solution to this problem is to re-map the descriptor values using the probability distribution $\mathcal{P}(x)$ (i.e. the integral of the density):

$$x' = x_{\min} + (x_{\max} - x_{\min})\mathcal{P}(x) \qquad (11)$$

effectively equalizing the distribution of the values: the resulting density more closely resembles a uniform density [11].

It is not necessary to use eqn. (11) to get an improvement: in the specific edgeness case, a good result can also be obtained by log-mapping the magnitude value. The difference in the estimated required number of bins is reported in Figure 4. The impact of this mapping on the effectiveness of the descriptor is more marked when the number of bins is very small, as can be seen in Figure 5.

The capacity curves exhibit a dependence on histogram resolution which is specific to the chosen dissimilarity measure. This is related to the fact that the effectiveness of histogram indexing depends on the number of bins used (if the representation is too coarse, discrimination is severely impaired) and to the different sensitivity of the dissimilarity measures to histogram resolution. The dependence of the capacity $\mathcal{H}$ on histogram resolution is a very important issue as the higher the number of bins, the higher the computational cost incurred for image retrieval. For the development of an efficient retrieval system, histogram resolution should be kept as low as possible without affecting retrieval effectiveness.

For each dissimilarity function d, it is possible to quantify the impact of the number of bins on the effectiveness of histogram indexing by computing the distance between the capacity curves at a reference resolution (in our case 64 bins) and the corresponding curves at a reduced number of bins (in our case 32, 16, 8). The plot of these distances computed using the L_1 norm, which correspond to the area between the curves, is reported in Figure 6 for the hue descriptor and for the dissimilarities considered in this paper. The plot suggests that the dissimilarities can be divided into three classes: the Euclidean norm being the less stable, the χ^2 and the L_1 measures with approximately the same, intermediate stability, the Kolmogorov-Smirnov and the Kuiper statistics being the most stable of the group. The minimum resolution at which the discrimination ability is close to the one at the reference resolution is 16 bins, in accordance with Sturges' rule. A resolution of 32 bins, in accordance with Scott's rule, provides essentially the same effectiveness as the reference resolution, at least for the most stable dissimilarities.

The histograms considered so far were obtained by marginalizing over the spatial components of the corresponding distributions. While resulting data are compact,

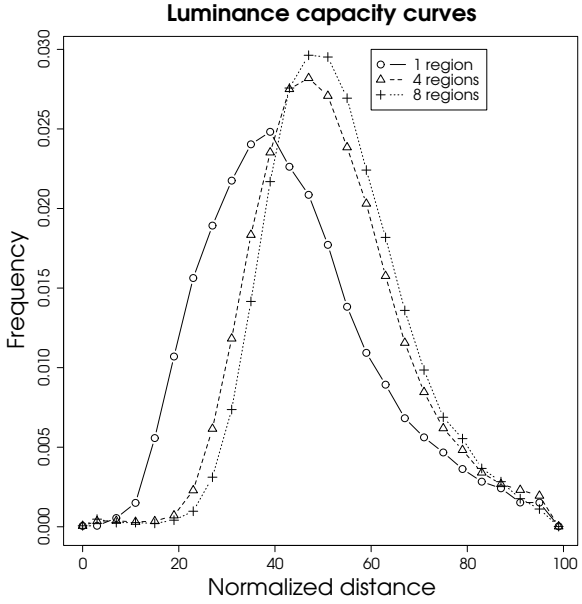

Fig. 7. The plots show the effect of incomplete marginalization over the spatial variables when keeping the size of the descriptors fixed (in this case 64 bins). Note how the use of multiple regions increases the effectiveness of the luminance descriptor. The L_1 metric is used for the computation of the curves.

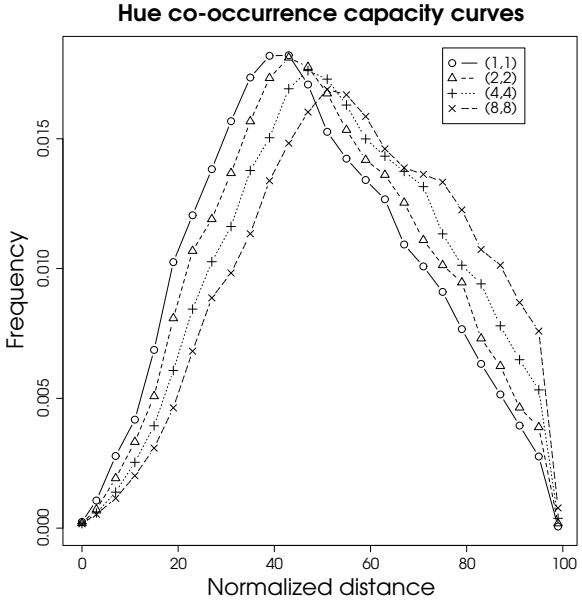

Fig. 8. The plots show the effect on the capacity curves of varying the spatial relation $(\Delta x, \Delta y)$ used in the computation of hue co-occurrence. Note that by increasing the distance between the pixels of the couple, the effectiveness increases. The curves are computed using the L_1 metric.

they represent only a rough summary of the information originally available in the image. In particular, all the information regarding the spatial distribution of the related descriptors has been lost. A compromise to complete marginalization over the spatial components is to quantize the coordinate space into a limited number of cells. By analyzing the capacity curves it is possible to check for increased effectiveness of the resulting descriptors when this quantization is introduced, and if there are any advantages in using spatial information while constraining the overall number of bins used (see Figure 7). Another way spatial information can be taken into account is to partition the set of pixels by imposing some relation other than spatial proximity. For example, a binary relation could be imposed, associating each pixel with the pixel whose position is obtained by a fixed translation. The hue (and luminance) co-occurrence descriptors are obtained in this manner. The imposed spatial relation influences the effectiveness of the descriptor as detailed by Figure 8.

The capacity curves for the descriptors at 64 bins (using the appropriate mapping for edgeness) are reported in Figure 9. The resulting ranking according to effectiveness is further confirmed by the principal component analysis of the data as reported in Figure 10. The rating of the descriptors, given by their indexing effectiveness $\mathcal{E}$, is consistent across the two databases even if the curves show some differences, and is reported in Table II.

The characterization of the image descriptors and the assessment of the comparison functions rely on the use of large image sets. This would be avoided if the histograms could be generated randomly using a realistic model preserving the characteristic distribution in the histogram space [20]. In this case, the results should be interpreted with care. We have generated a set of random histograms to simulate the luminance histograms. They were obtained by combining two Gaussians of random normalization, σ, and centers (one in the left half of the luminance range, the other one in the right) modulated by a multiplicative noise. The resulting capacity curve is compared to the curve obtained on the database of still images for the luminance descriptor using the L_1 metric in Figure 11: the resulting curves are very similar. However, if we use a different dissimilarity measure for computing the capacity curves, their difference is more marked, suggesting that the random histograms do not accurately reflect the real distribution.

The dependence of the capacity curves on the database considered is clearly reflected in Figure 12. While the distribution of the luminance for the two databases is very similar, the distributions of hue are different: the database of still images is much more colorful than the video database and this is reflected by a shift of the histogram distance from lower values (`VIDEO`) to higher values (`STILLS`).

5. Applications

The analysis of histogram effectiveness in image discrimination is mainly useful for the design of image retrieval systems. However, it is not limited to image retrieval and can also be used to derive some global descriptive information from a collection of images such as the frames of a complete video clip. Another application is that of *template matching*, i.e. the search for image regions that are

Descriptor	$\mathcal{E}_{\texttt{VIDEO}}$	$\mathcal{E}_{\texttt{STILLS}}$
Co-occurrence (hue)	57	68
Hue	55	70
Co-occurrence (lum)	52	50
Luminance	43	46
Edgeness	22	32

TABLE II

THE EFFECTIVENESS $\mathcal{E}$ OF THE DIFFERENT DESCRIPTORS. NOTE THAT THE EFFECTIVENESS OF HUE AND HUE CO-OCCURRENCE ARE NEARLY THE SAME.

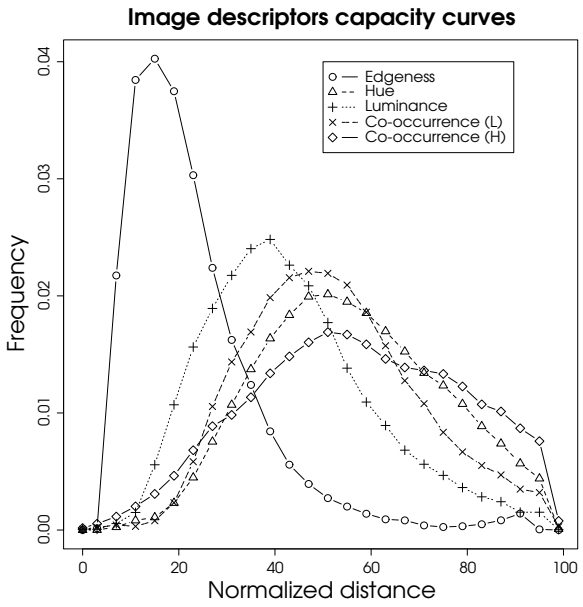

Fig. 9. The plots report the capacity curves, computed using the L_1 distance and 64 bins per histogram, for the image descriptors considered in this paper on the VIDEO database.

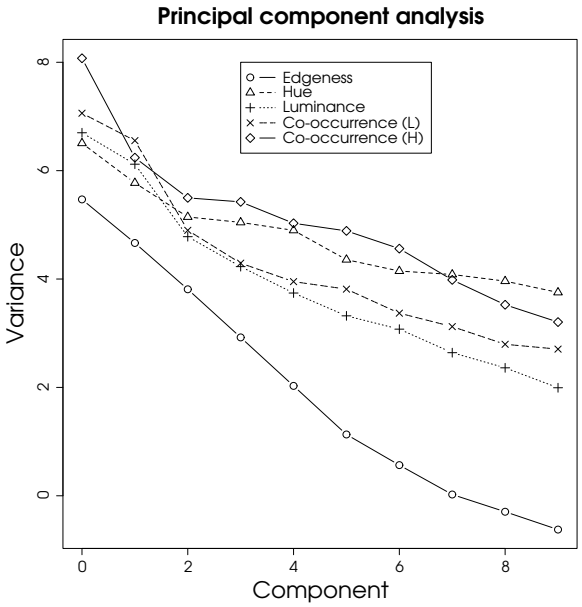

Fig. 10. The plots show the (logarithmic) values of the variance described by the *principal components* of the descriptors (VIDEO database, 64 bins). The curves would be flat in the case of a multidimensional Gaussian distribution of points characterized by the same variance along the different axes.

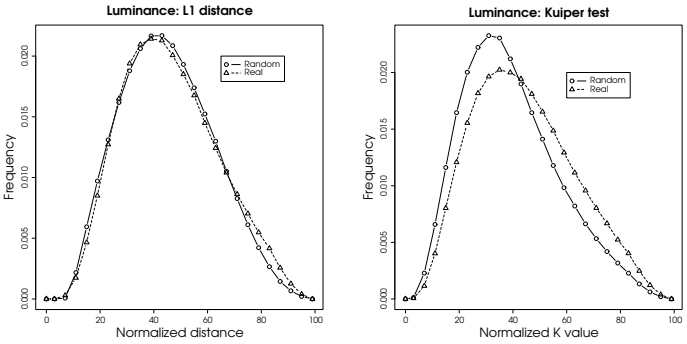

Fig. 11. LEFT: the capacity curves for a real database and for a randomly generated one using the L_1 distance; RIGHT: the capacity curves when the Kuiper statistic is used.

similar to one or more reference templates.

The capacity of the visual descriptors can be used for the design of a scalable image retrieval system which makes optimal use of the available storage and computation resources. Low level image descriptors can be introduced into the system according to their indexing effectiveness $\mathcal{E}$, the most discriminant being introduced first (see Figure 13 for the capacity curves when multiple descriptors are used). Sorting the descriptors according to their effectiveness may also improve the efficiency of the computation of histogram differences. In image retrieval tasks, a threshold on the minimum acceptable similarity is usually imposed to limit the number of retrieved items: the computation of d can be stopped as soon as its monotonically increasing value exceeds the retrieval threshold. Comparing the descriptors sorted by decreasing effectiveness is expected to increase the computational savings. Furthermore, for each of the descriptors an appropriate resolution can be determined by the analysis of the capacity curves. The effectiveness of the introduction of multiple regions in the descriptions can also be monitored, adding another dimension for scalability of the resulting system. The results of the analysis on the dissimilarity functions and histogram resolution, relatively to the two databases considered, support the following conclusions:

- the L_1 norm provides the best qualitative results (see Figure 1) being at the same time the less computationally demanding;
- histogram resolution can be as low as 16 bins (see Figure 6).

The latter result is important as the computational requirements scale linearly with the number of bins. Furthermore, using a reduced number of bins makes the representation of data using few digits in fixed point format feasible. A very performing system can then be built where images can be compared at a rate of over one million images per second

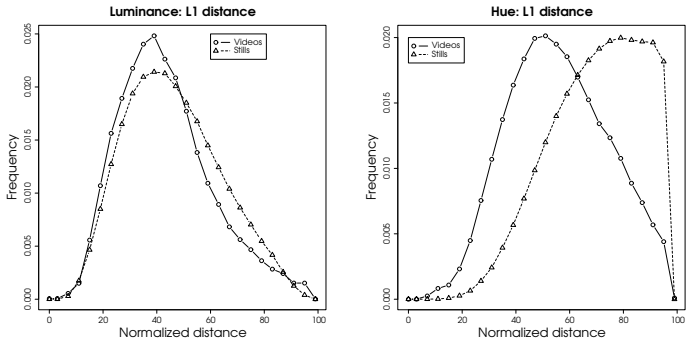

Fig. 12. The plots highlight the descriptor specific database dependency by comparing the capacity curves over the VIDEO and STILLS databases when the L_1 metric is used.

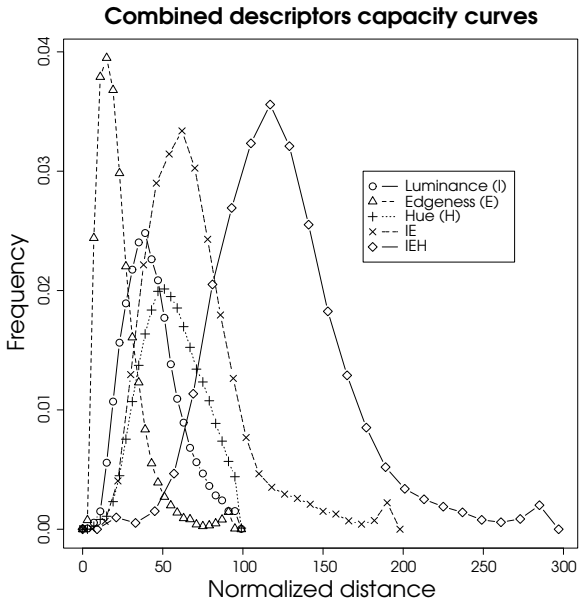

Fig. 13. The plots show the effect of using multiple descriptors (VIDEO database, 64 bins).

on a standard PC (see [2]).

The capacity curves can also be used to infer some global descriptive information which can be used to characterize a whole set of images, such as the frames of a complete feature movie. As an example, we show how to obtain an estimate of the number of key-frames required to represent a video clip. Let us assume that the video is composed by m different shots, each shot being built by similar frames. The point at which the capacity curve slope abruptly increases determines the threshold θ_0 to be used for judging two images as similar. Let $[0, \theta_0]$ represent the region to the left of the threshold. If the shots have the same length, $k_i = k, \forall i$, we have:

$$b_0 = \sum_i \frac{k_i(k_i - 1)}{2} = m\frac{k(k-1)}{2} \tag{12}$$

where b_0 represents the area to the left of θ_0 and the sum is over the shots. If the shots have different lengths:

$$\begin{aligned} b_0 &= \sum_i \frac{k_i(k_i-1)}{2} \\ &\approx \sum_i \frac{k_i^2}{2} \\ &\approx \frac{m(\tilde{k}^2 + \sigma^2)}{2} \end{aligned} \tag{13}$$

where $\tilde{h}$ and σ are respectively the arithmetic average and the standard deviation of the set $\{k_i\}$, and the relation $\sum k_i^2 - m\tilde{k}^2 = m\sigma^2$ has been used. The overall area of the curve is given by:

$$b \approx \frac{m^2\ \tilde{k}^2}{2} \tag{14}$$

using the fact that the overall number of frames is $n = m\tilde{k}$. By taking the ratio:

$$\frac{b}{b_0} \approx m\frac{1}{1 + \left(\frac{\sigma}{\tilde{k}}\right)^2} \tag{15}$$

an approximate value for the number of shots m can be derived:

$$m \approx \frac{b}{b_0}\ \left[1 + \left(\frac{\sigma}{\tilde{k}}\right)^2\right] \tag{16}$$

An example is reported in Figure 14. The plot shows the characteristic shape of the left tail of the capacity curve for a 5 minutes long video clip. The point at which the slope abruptly increases ($\theta_0 = 15$) is clear and the estimated number of key frames ($m = 138$), each one representing an homogeneous portion of video, is in good accordance with the available ground truth ($m = 141$).

Template matching tasks are very similar to image retrieval applications. The major difference is that each single image represents the set of all possible overlapping subregions whose size corresponds to that of the reference template. Matching by histogram comparison is an efficient strategy, as the computation of histograms for nearby regions requires only minimal updating. The increased efficiency and effectiveness obtained for image retrieval applications through appropriate choices of the comparison dissimilarity measure, low level image descriptors, and histogram resolution are automatically transferred to template matching tasks. Future work will address two related issues: the extension the functionalities of an image retrieval system by enabling the search for image details (i.e. looking for a query image in an arbitrary position within larger images) and the analysis of the *robustness* of histogram matching (see [1] for a simple introduction of the concept *robustness* in template matching tasks).

6. Conclusions

In this paper a general method for comparing the effectiveness of histograms in image comparison tasks has been introduced and used to solve some important design issues in the development of a scalable, efficient (i.e. fast), and

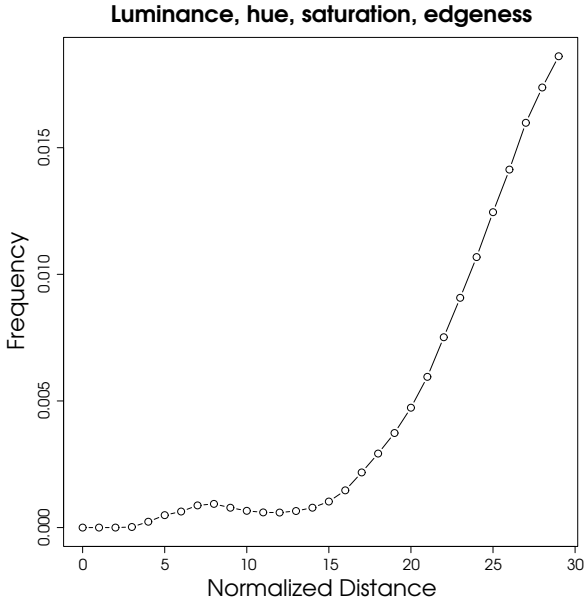

Fig. 14. The plot reports the left tail of the capacity curve, computed with the L_1 metric, for a video containing commercials. The video is approximately 5000 frames long and contains 141 shots. Choosing $\theta_0 = 15$ an estimate of 138 shots is obtained.

effective (i.e. discriminant) image retrieval system based on the query-by-example paradigm. The results, based on the analysis of the histogram capacity curves, show that, among the most commonly used dissimilarity measures, the L_1 norm is the most effective for histogram indexing. Furthermore, an image retrieval system can rely on low resolution histograms without severe degradation of retrieval performance. Several low level image descriptors (hue, luminance, etc.) have been compared and their retrieval effectiveness assessed through their capacity curves. This result permits the design of scalable image retrieval systems which make optimal use of computational and storage resources.

References

[1] R. Brunelli and S. Messelodi. Robust Estimation of Correlation with Applications to Computer Vision. *Pattern Recognition*, 28:833–841, 1995.

[2] R. Brunelli and O. Mich. COMPASS: a Distributed Image Retrieval System. Technical report, ITC-irst, 1998. In preparation.

[3] C. Faloutsos, M. Flickner, W. Niblack, D. Petkovic, W. Equitz, and R. Barber. Efficient and Effective Querying by Image Content. Technical Report RJ 9453 (83074), IBM Research Division, Almaden Research Center, 1993.

[4] G. Fasano and A. Franceschini. A Multidimensional Version of the Kolmogorov-Smirnov Test. *Monthly Notices of the Royal Astronomical Society*, 225:155–170, 1987.

[5] M. Flickner, 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. *Computer*, pages 23–32, September 1995.

[6] B. V. Funt and G. D. Finlayson. Constant Color Indexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5):522–529, May 1995.

[7] Y. Gong and M. Sakauchi. Detection of Regions Matching Specified Chromatic Features. *Computer Vision and Image Understanding*, 61(2):263–269, 1995.

[8] Y. Gong, H. Zhang, H. C. Chuan, and M. Sakauchi. An Image Database System with Content Capturing and Fast Image Indexing Abilities. In *Proceedings of IEEE International Conference on Multimedia Computing and Systems*, pages 121–130, Boston, May 1994.

[9] J. Hafner, H. S. Sawhney, W. Equitz, M. Flickner, and W. Niblack. Efficient Color Histogram Indexing for Quadratic Form Distance Functions. *IEEE Trans. on PAMI*, 17(7):729–735, July 1995.

[10] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin. The QBIC Project: Querying Images by Content Using Color, Texture, and Shape. In *Proceedings SPIE Storage and Retrieval for Image and Video Databases*, pages 173–181, 1993.

[11] G. Pass and R. Zabih. Comparing Images Using Joint Histograms. Technical report, Cornell University, 1997. Submitted to ACM Journal of Multimedia Systems.

[12] J. Rissanen, T. P. Speed, and B. Yu. Density Estimation by Stochastic Complexity. *IEEE Trans. on Information Theory*, 38(2):315–323, March 1992.

[13] D. W. Scott. *Multivariate Density Estimation*. New York: John Wiley, 1992.

[14] D.W. Scott. On Optimal and Data-Based Histograms. *Biometrika*, 66:605–610, 1979.

[15] J. R. Smith and S.-F. Chang. Automated Image Retrieval Using Color and Texture. Technical Report CU/CTR 408-95-14, Columbia University, July 1995.

[16] J. R. Smith and S.-F. Chang. Single Color Extraction and Image Query. In *Proceedings, IEEE International Conference on Image Processing (ICIP-95), Washington, DC*, October 1995.

[17] J. R. Smith and S.-F. Chang. Tools and Techniques for Color Image Retrieval. In *Storage & Retrieval for Image and Video Databases IV*, Proceedings of IS&T/SPIE Vol. 2670, 1996.

[18] M. A. Stephens. The Goodness-of-Fit Statistic V_n: Distribution and Significance Points. *Biometrika*, 52(3,4):309–321, 1965.

[19] M. Stricker and M. Swain. The Capacity and the Sensitivity of Color Histogram Indexing. Technical Report 94-05, Communications Technology Lab, ETH-Zentrum, 1994.

[20] M. Stricker and M. Swain. The Capacity of Color Histogram Indexing. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 704–708, 1994.

[21] H.A. Sturges. The Choice of a Class Interval. *Journal of the American Statistical Association*, 21:65–66, 1926.

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

[23] M. P. Wand. Data-Based Choice of Histogram Bin Width. *Statistical Computing and Graphics*, 51(1):59–64, February 1997.