Compressed Domain Image Retrieval: A Comparative Study of Similarity Metrics

M. Hatzigiorgaki and A. N. Skodras

Electronics Laboratory, University of Patras, GR-26110 Patras, Greece
Research Academic Computer Technology Institute, GR-26221 Patras, Greece
School of Sciences & Technology, Hellenic Open University, GR-26222 Patras, Greece

ABSTRACT

Content based image retrieval has gained considerable attention in nowadays as a very useful tool in a plethora of applications. Web has become the most important application, because over 70% of it is devoted to images, and looking for a specific image is a really daunting task. The vast majority of these images are JPEG compressed. An extensive study of eighteen similarity measures used for image retrieval has been conducted and the corresponding results are reported in the present communication. The energy histograms of the low frequency DCT coefficients have been used as the feature space for similarity testing. Query-by-image-example was used in all tests.

Keywords: Compressed domain image retrieval, CBIR, JPEG, DCT

1. INTRODUCTION

Content-based image retrieval (CBIR) has gained considerable attention especially in the last decade. Image retrieval based on content is extremely useful in a plethora of applications, such as publishing and advertising, historical research, fashion and graphic design, architectural and engineering design, crime prevention, medicine, etc.1-3 The most important application, however, is the Web, as at least 73% of it is devoted to images, and searching for a specific image is indeed a daunting task.4,5 Numerous commercial and experimental CBIR systems are now available, as for example IBM’s QBIC (Query by Image Content), Virage’s VIR Image Engine, Excalibur’s Image RetrievalWare, or Columbia University’s WebSEEK. Also, many web search engines1-3 are now equipped with CBIR facilities, as for example Alta Vista and Yahoo!

Content-based image retrieval differs from classical information retrieval in that image databases are essentially unstructured, since digitized images consist purely of arrays of pixel intensities, with no inherent meaning. CBIR retrieves stored images from such a collection by comparing features extracted from the images themselves. The commonest features used are mathematical measures of color, texture and shape. These features belong to the primitive or literal or first level features. The higher level features that are closer to the user requirements are the logical or perceptual and abstract or semantic1-3. Logical features have to do with the identity of the objects shown on an image, while abstract attributes denote the significance of scenes depicted.

Most of the work on image retrieval done so far is devoted to the pixel- or spatial-domain manipulation of the images.6-10 This means that all processing and similarity tests are performed directly in the pixel-domain. However, the vast majority of images, especially those residing on the web, are JPEG compressed. Thus, searching for images directly in the compressed-domain (DCT – discrete cosine transform domain) becomes more beneficial compared to the spatial-
domain search. This is due to the following two reasons: (i) the time needed to fully decompress each image is substantially reduced, and (ii) the whole process (similarity measurement) is performed on a reduced number of components, as most of the DCT coefficients are zero after the quantization stage. These advantages have directed research towards compressed-domain image retrieval, and a lot of nice work has appeared in the open literature\textsuperscript{11-18}.

Content-based image retrieval systems very often work on \textit{query-by-image-example} basis. Namely, the query image is fed to the system, and the system is asked to retrieve further examples of similar kind. Image similarity, either in the spatial-domain or the compressed-domain, is typically defined using a metric on a feature space. This implies that, if the “right” features are chosen, proximity in feature space will correspond to perceptual similarity. However, there are several reasons to doubt this, the most fundamental being the metric assumption. There is evidence that human similarity judgments do not obey the requirements of a metric. Numerous similarity metrics have been proposed so far. An extensive comparative study about their effectiveness in compressed-domain image retrieval is attempted in the present communication. The feature space used is the histogram of the DCT coefficients and different similarity measures are applied on a variable number of coefficient histogram bins. The paper is structured in the following way. In Section 2 the DCT compressed-domain image retrieval is described. In Section 3 the eighteen different similarity metrics used in the present study, are presented. Comparative retrieval results are given in Section 4 and conclusions are drawn in the last section.

2. COMPRESSED-DOMAIN IMAGE RETRIEVAL

The enormous majority of the imagery produced in nowadays is lossy compressed by means of the JPEG standard. This is especially true for the imagery stored on the web. Digital cameras and scanners that are mainly used for image capture, do directly produce JPEG compressed images. Of course, an application area for which this is not true, is the medical imagery one. The upcoming JPEG2000\textsuperscript{19} standard will gradually change this situation, but this will take some time until it becomes an everyday practice.

The JPEG\textsuperscript{20} decompression engine, as it is depicted in Fig.1, is well known for more than a decade now. Searching for an image in the compressed-domain means that only the inverse variable length coding and inverse quantization steps have to be performed for each 8x8 block, thus calculating the 64 DCT coefficients, i.e. the DC and the 63 AC coefficients. Depending on the compression factor (i.e. the picture quality), most of these coefficients are equal to zero, especially those corresponding to the high frequencies in the zig-zag pattern. The coefficients that carry most of the image energy are the DC and the first AC ones. Those are the ingredients of the feature space usually used for retrieval purposes\textsuperscript{11-18}. It has been shown in previous works that the DC alone, or the DC and the first three or eight AC can result in high retrieval efficiency. Recall that the DC coefficient corresponds to the average intensity of the image component block. These coefficients, which constitute the upper left 2x2 or 3x3 part of each 8x8 block, retain the low frequency information of the original image, i.e. a coarse or blurred version of it. This approach is the one adopted in the present work.

![Figure 1. Block diagram of the JPEG decompression engine.](image-url)
The histogram of the above mentioned coefficients is used as the feature space for image retrieval. This feature space is associated to the image itself. Searching the image database corresponds to searching the associated feature space repository, as shown in Fig. 2. In the case that a separate feature space repository does not exist, similarity comparisons could be performed on the fly by partially decompressing each stored image, as already described. After the similarity testing of the query feature space with the corresponding ones of the feature space repository, all similar images are retrieved. Usually a similarity threshold is employed and images are retrieved in descending order of similarity.

Various questions may arise during the implementation of the above described system. Given that retrieval is based on similarity comparisons of the coefficient histograms, the following two aspects are investigated here: (i) The number of histogram bins that should be used, as DCT coefficients can have up to 2048 integer values for 8-bit images, and (ii) the effectiveness of the similarity measure for the histograms. In order to answer these questions, an extensive number of comparisons has been conducted for four different cases of histogram bins (256, 512, 1024, 2048), eighteen different similarity metrics and three different coefficient sets (DC only, DC+3AC, DC+8AC). Comparative results are reported in Section 4.

3. MEASURING SIMILARITY

Similarity in the pixel-domain or coefficient-domain can be measured in a variety of ways. The most common similarity measures that are found in the open literature are reported below. \( x_i, y_i \) are the feature values (pixels, coefficients, histogram bins, other) of the query image and the image under test, respectively. In our case these represent the histogram bin values.

3.1. Euclidean Distance (L2)

One of the commonest distance metrics in image retrieval literature is the Euclidean distance. It corresponds to the Minkowski-form distance for \( r=2 \), and is defined as:

\[
D_2 = \sqrt{\sum (x_i - y_i)^2}
\]

3.2. City Block Distance (or Manhattan Distance)

The city block distance function is equivalent to the Minkowski-form distance for \( r=1 \). It requires less computations than many other distance metrics, and is defined as:

\[
D_1 = \sum |x_i - y_i|
\]
3.3. Canberra Metric

Canberra metric is very popular in CBIR applications. It has the advantage of a relatively low computational complexity and high retrieval efficiency.

\[ D_3 = \sum_i \frac{|x_i - y_i|}{x_i + y_i} \]

3.4. Histogram Intersection

The histogram intersection metric was first proposed for color image retrieval in the spatial domain by Swain and Ballard\(^{21}\). The intersection metric they used is not symmetric, but it can be modified in order to produce a true distance metric.

\[ D_4 = 1 - \frac{\sum_i \min(x_i, y_i)}{\min(\sum_i x_i, \sum_i y_i)} \]

3.5. Jeffrey Divergence

The Jeffrey divergence metric is defined as:

\[ D_5 = \sum_i [x_i \log \left(\frac{x_i}{m_i}\right) + y_i \log \left(\frac{y_i}{m_i}\right)] \]

where \( m_i = \frac{x_i + y_i}{2} \)

3.6. Bhattacharyya Distance

This distance metric is not frequently used in image retrieval applications mainly because of its complexity in computation. Bhattacharyya distance is given by:

\[ D_6 = \log_{10} \left( \frac{\sum_i \sqrt{|x_i - y_i|}}{\sum_i x_i} \right) \]

3.7. Chi-Square

The Chi-square statistic is applicable to unbinned distributions, but it can also be used for comparison between binned distributions such as histograms. Chi-square distance is one of the most popular metrics and is given by the formula:

\[ D_7 = \sum_i \frac{(x_i - y_i)^2}{x_i + y_i} \]

3.8. Bray Curtis Distance

Bray Curtis distance is quite similar to Canberra metric. It is defined as:

\[ D_8 = \frac{\sum_i |x_i - y_i|}{\sum_i (x_i + y_i)} \]
3.9. Angular Separation Distance
The angular separation metric is defined as:

\[ D_9 = 1 - \frac{\sum_i x_i \cdot y_i}{\sqrt{\sum_i x_i^2 \cdot \sum_i y_i^2}} \]

3.10. Chord Distance
Chord distance measures the distance between the points where vectors cross a unit sphere. It also emphasizes qualitative aspects of data sets. It is defined as:

\[ D_{10} = \sqrt{2 - 2 \cdot \frac{\sum_i x_i \cdot y_i}{\sqrt{\sum_i x_i^2 \cdot \sum_i y_i^2}}} \]

3.11. Non-Correlation
The non-correlation metric is defined as:

\[ D_{11} = 1 - \sqrt{\frac{\sum_i |x_i - \bar{x}| \cdot |y_i - \bar{y}|}{\sqrt{\sum_i (x_i - \bar{x})^2 \cdot \sum_i (y_i - \bar{y})^2}}} \]

3.12. Matusita Distance
The Matusita distance is defined as:

\[ D_{12} = \sqrt{\sum_i \left(\sqrt{x_i} - \sqrt{y_i}\right)^2} \]

3.13. Soergel
The Soergel metric is defined as:

\[ D_{13} = \frac{\sum_i |x_i - y_i|}{\sum_i \max(x_i, y_i)} \]

3.14. Wave - Hedges
The Wave-Hedges metric is defined as:

\[ D_{14} = \sum_i \left(1 - \frac{\min(x_i, y_i)}{\max(x_i, y_i)}\right) \]

3.15. WED Distance
The WED (Weighted Euclidean Distance) has the same form as Euclidean distance except that the square difference of the two distributions for every \( i \) is multiplied by a weight \( w_i \) depending on the value of distribution \( x \).

\[ D_{15} = \sqrt{\sum_i w_i \cdot (x_i - y_i)^2} \]

where \( w_i = x_i \) if \( x_i \neq 0 \), and \( w_i = 1 \) otherwise.
3.16. Kolmogorov-Smirnov Statistic

Kolmogorov-Smirnov test (or K-S statistic) is a particularly simple measure. It is defined as the maximum value of the absolute difference between two cumulative distribution functions. For binned distributions it is actually the maximum distance or Chebychev distance. For the two different cumulative distribution functions \( x_i \) and \( y_i \), the K-S statistic has the form:

\[
D_{16} = \max_i |x_i - y_i|
\]

3.17. Kuiper

Kuiper statistic is quite similar to the Kolmogorov-Smirnov test. Although it is applicable to unbinned distributions its discrete version can be used to measure distance between histograms.

\[
D_{17} = \max \ (x_i - y_i) + \max \ (y_i - x_i)
\]

3.18. Mean Distance

Mean distance is one of the metrics used in cluster analysis for measuring distances between clusters. It can also be used to measure distances between histograms. It has the simple form:

\[
D_{18} = |m_x - m_y|
\]

where \( m_x \), \( m_y \) correspond to the mean value of each distribution, namely \( m_x = \sum_i x_i \ p(x_i) \) and \( m_y = \sum_i y_i \ p(y_i) \).

4. COMPARATIVE RESULTS AND DISCUSSION

The study has been based on a grayscale database of approximately 200 images of size 256x256 each. A sample collection of these images is shown in Fig. 3. Many of the images were similar, as they were produced out of the same image by translation, rotation, cropping or filtering. Each of these sets of similar images constituted the ground truth set for that particular image. The ground truth sets for the test images ‘boat’ and ‘vehicles’ are illustrated in Fig. 4. This means that a successful retrieval system would have been able to retrieve all versions of the same image (ground truth set). There are many ways for the evaluation of the retrieval efficiency of each method. The recall or retrieval rate has been employed in the present study as the ground truth set for each image was known and the recall rate could be precisely calculated. Query-by-internal-image-example was used in all experiments. This means that the query image was always one of the images residing in the image repository. This explains why the performance of the system was never zero, as it could always retrieve the image used for the query. This was giving an additional assurance that the similarity metrics were correctly implemented. The recall rate, \( R_a^q \), for a particular query image \( q \) is defined as

\[
R_a^q = \frac{N_a}{N_G}
\]

where \( N_a \) is the number of ground truth images retrieved within the first \( aN_G \) retrievals, and \( N_G \) is the size of ground truth set for query image \( q \). The recall rate \( R_a^q \) takes values between 0 and 1, where 0 stands for “no image found” and 1 for “all (ground truth) images found”. In our case, as mentioned above, the recall rate was always greater than 0 because of the query-by-internal-image-example. Parameter \( a \) should be at least 1. As \( a \) increases, tolerance also increases. In Fig. 5 two query examples are shown along with the first eight best retrieval results for three different similarity metrics (Euclidean, Canberra, Wave-Hedges) for the DC and 3 AC coefficients and 256 histogram bins. From these and similar tests it is demonstrated that the Canberra and Wave-Hedges metrics perform constantly better than the others. The Euclidean distance metric has a poor retrieval rate in most of the cases.
The retrieval efficiency of all similarity metrics for 256 and 1024 DCT coefficient histogram bins is depicted in Fig.6 for the query images ‘boat’ and ‘vehicles’. Parameter $a$ was set to 1. The retrieval results for other query images were very similar to that. From these graphs some useful conclusions can be drawn:

(i) Retrieval efficiency is almost independent of the number of coefficients histogram bins used.
(ii) Retrieval efficiency does not necessarily improve as the number of AC coefficients increases.
Some similarity metrics perform constantly better than other, independently of the number of AC coefficients or histogram bins used. Such metrics are the Canberra, the Wave-Hedges, the Jeffrey Divergence, the Bhattacharyya and the Matusita. By taking into account the computational complexity of each similarity metric, it can be concluded that the Canberra and Wave-Hedges metrics are preferable than the others.

Figure 5. Retrieval results for the Euclidean, Canberra and Wave-Hedges metrics (from top to bottom) for the DC and the first 3 AC coefficients (in zig-zag order) and 256 histogram bins.
5. CONCLUSIONS

The effect of the similarity metric on the image retrieval by content has been approached in the present work. The histogram of the DC and the first three and eight AC DCT coefficients was used as the feature space for the retrieval. It has been realized that the number of histogram bins does not affect profoundly retrieval efficiency. Some commonly used distance measures, as the Euclidean distance, do not result in good retrieval performance, while others perform well in the vast majority of the tests. The Canberra and Wave-Hedges metrics seem to be beneficial from the retrieval efficiency and the computational complexity points of view. This study should be extended to cover the cases of color images of various sizes. Also, the difference in the statistical properties of the DC and the AC coefficients should be taken into account. This would lead to applying different similarity metrics for the DC and the AC coefficients. Such a combination should improve retrieval performance considerably, as some initial experimental results have already shown.

Figure 6. Retrieval rate versus similarity metric for 256 and 1024 coefficient histogram bins for the query images ‘boat’ (top) and ‘vehicles’ (bottom).
REFERENCES


