EVALUATION OF SIMILARITY MEASUREMENT FOR IMAGE RETRIEVAL

Dengsheng Zhang and Guojun Lu

Gippsland School of Computing and Info Tech
Monash University
Churchill, Victoria 3842
dengsheng.zhang, guojun.lu@infotech.monash.edu.au

ABSTRACT

Similarity measurement is one of the key issues in content based image retrieval (CBIR). In CBIR, images are represented as features in the database. Once the features are extracted from the indexed images, the retrieval becomes the measurement of similarity between the features. Many similarity measurements exist. A number of commonly used similarity measurements are described and evaluated in this paper. They are evaluated in a standard shape image database. Results show that *city block distance* and χ^2 *Statistics measure* outperform other distance measure in terms of both retrieval accuracy and retrieval efficiency.
Keywords: image retrieval, distance measure, CBIR, shape.

1. INTRODUCTION

Image description and retrieval is an important multimedia application. Due to tremendous increase of digital images and the limitations associated with traditional image indexing, recent research on image indexing focuses on Content Based Image Retrieval (CBIR). CBIR employs low level image features such as color, shape or texture to achieve *objective* and *automatic* indexing, in contrast to *subjective* and *manual* indexing in traditional image indexing. For contend based image retrieval, the image feature extracted is usually an *N*-dimensional feature vector which can be regarded as a point in a *N*-dimensional space. Once images are indexed into the database using the extracted feature vectors, the retrieval of images is essentially the determination of similarity between the features of query image and the features of target images in database, which is essentially the determination of distance between the feature vectors representing the images. The desirable distance measure should reflect human perception. That is to say, perceptually similar images should have smaller distance between them and perceptually different images should have larger distance between them. Therefore, for a given shape feature, the higher the retrieval accuracy, the better the distance measure. Of course, for online retrieval, computation efficiency should also be considered when choosing a distance measure. Various distance measures have been exploited in image retrieval, they include city block distance [6, 7], Euclidean distance [9], cosine distance [5], histogram intersection distance [5, 7], χ^2 statistics distance [3], quadratic distance [1, 2, 4, 10], and Mahalanobis distance [5, 8]. In this paper, several commonly used distance measurements will be described and evaluated. The purpose of this evaluation is to find a desirable similarity measure for shape based image retrieval.

The rest of the paper is organized as following. In Section 2, different similarity measurements are described in details. Section 3 presents the experiment results and discussions. The paper is concluded in Section 4.

2. SIMILARITY MEASUREMENTS

A similarity measurement is normally defined as a metric distance. In this section different similarity measurements are described in details.

2.1 Minkowski-form Distance

The Minkowski-form distance is defined based on the L_p norm:

$$d_p(\mathbf{Q},\mathbf{T}) = \left(\sum_{i=0}^{N-1}(Q_i - T_i)^p\right)^{\frac{1}{p}}$$

where $\mathbf{Q} = \{Q_0, Q_1, \ldots Q_{N-1}\}$ and $\mathbf{T} = \{T_0, T_1, \ldots, T_{N-1}\}$ are the query and target feature vectors respectively. When $p = 1$, $d_1(\mathbf{Q}, \mathbf{T})$ is the *city block distance* or *Manhattan distance* (L_1)

$$d_1(\mathbf{Q},\mathbf{T}) = \sum_{i=0}^{N-1}|Q_i - T_i|$$

When $p = 2$, $d_2(\mathbf{Q}, \mathbf{T})$ is the *Euclidean distance* (L_2)

$$d_2(\mathbf{Q},\mathbf{T}) = \left(\sum_{i=0}^{N-1}(Q_i - T_i)^2\right)^{\frac{1}{2}}$$

When $p \to \infty$, we get L_∞,

$$L_\infty(\mathbf{Q},\mathbf{T}) = \max_{0 \le i < N}\{|Q_i - T_i|\}$$

2.2 Cosine Distance

The cosine distance computes the difference in direction, irrespective of vector lengths. The distance is given by the angle between the two vectors. By the rule of dot product

$\mathbf{Q} \cdot \mathbf{T} = \mathbf{Q}^t\mathbf{T} = |\mathbf{Q}| \cdot |\mathbf{T}| \cos\theta$

$$d_{\cos}(\mathbf{Q}, \mathbf{T}) = 1 - \cos\theta = 1 - \frac{\mathbf{Q}^t\mathbf{T}}{|\mathbf{Q}| \cdot |\mathbf{T}|}$$

2.3 χ^2 Statistics

The χ^2 statistics is defined as

$$d_{\chi^2}(\mathbf{Q}, \mathbf{T}) = \sum_{i=0}^{N-1} \frac{(Q_i - m_i)^2}{m_i}$$

where $m_i = \frac{Q_i + T_i}{2}$. This quantity measures how unlikely it is that one distribution was drawn from the population represented by the other [3].

2.4 Histogram Intersection

The histogram intersection was proposed by Swain and Ballard [7]. Their objective was to find known objects within images using color histograms. It is able to handle partial matches when the object (with feature **Q**) size is less than the image (with feature **T**) size. The original definition of histogram distance is given as

$$d_{hi}(\mathbf{Q}, \mathbf{T}) = 1 - \frac{\sum_{i=0}^{N-1} \min(Q_i, T_i)}{|\mathbf{Q}|}$$

It has been extended into a metric distance as [5]

$$d_{hi}(\mathbf{Q}, \mathbf{T}) = 1 - \frac{\sum_{i=0}^{N-1} \min(Q_i, T_i)}{\min(|\mathbf{Q}|, |\mathbf{T}|)}$$

2.5 Quadratic Distance

The distances calculated from above described distance measures only take account for the correspondence between each dimension, and do not make use of information across dimensions. This issue has been recognized in histogram matching. As a result, quadratic distance is proposed to take similarity across dimensions into accounted [2, 5]. It has been reported to provide more desirable result than "like-bin" only matching between color histograms. The quadratic-form distance between two feature vectors **Q** and **T** is given by:

$$d_{qad}(\mathbf{Q}, \mathbf{T}) = [(\mathbf{Q}-\mathbf{T})^t\mathbf{A}(\mathbf{Q}-\mathbf{T})]^{\frac{1}{2}}$$

where $\mathbf{A} = [a_{ij}]$ is an $N \times N$ matrix, and a_{ij} is the similarity coefficient between indexes (dimensions) i and j. a_{ij} is given by

$a_{ij} = 1 - d_{ij} / d_{max}$ and $d_{ij} = |Q_i - T_j|$

For calculation, the quadratic-form distance is rewritten as [1]

$$d_{qad}(Q, T) = \left(\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} a_{ij} Q_i Q_j + \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} a_{ij} T_i T_j + \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} Q_i T_j\right)^{\frac{1}{2}}$$

2.6 Mahalanobis Distance

The Mahalanobis distance is a special case of the quadratic-form distance metric in which the transform matrix is given by the covariance matrix obtained from a training set of feature vectors, that is $\mathbf{A} = \Sigma^{-1}$. In order to apply the Mahalanobis distance, the feature vectors are treated as random variables $\mathbf{X}=[x_0, x_1, \ldots, x_{N-1}]$, where x_i is the random variable of ith dimension of the feature vector. Then, the correlation matrix is given by **R** where $\mathbf{R}=[r_{ij}]$ and $r_{ij} = E\{x_i x_j\}$. $E\{x\}$ gives the mean of the random variable x. Then, the covariance matrix is given by Σ, where $\Sigma=[\sigma_{ij}^2]$ and

$\sigma_{ij}^2 = r_{ij} - E\{x_i\}E\{x_j\}$

The Mahalanobis distance between two feature vectors **Q** and **T** is obtained by letting $\mathbf{X_Q} = \mathbf{Q}$ and $\mathbf{X_T} = \mathbf{T}$, which gives:

$$d_{mah} = [(\mathbf{X_Q} - \mathbf{X_T})\, \Sigma^{-1}\, (\mathbf{X_Q} - \mathbf{X_T})]^{\frac{1}{2}}$$

In the special case where x_i are statistically independent, but have unequal variances, Σ is diagonal matrix as following

$$\Sigma = \begin{bmatrix} \sigma_0^2 & & & 0 \\ & \sigma_1^2 & & \\ & & \ddots & \\ 0 & & & \sigma_{N-1}^2 \end{bmatrix}$$

In this case, the Mahalanobis distance is reduced to a simpler form [5]:

$$d_{mah}(\mathbf{Q}, \mathbf{T}) = \sum_{i=0}^{N-1} \frac{(Q_i - T_i)^2}{\sigma_i^2}$$

It is a weighted L_2 distance. It gives more weight to dimension with smaller variance and gives less weight to dimension with larger variance.

3. EVALUATION OF SIMILARITY MEASUREMENTS

In the above, different distance measurements have been described and discussed. In order to determine which distance measurement is most suitable for image retrieval, the retrieval accuracy and retrieval efficiency of different distance measurements are tested using two general shape features and two standard shape databases.

3.1 Test Setup

In order to determine which distance measure best conforms to human perception, the retrieval effectiveness using different distance measures is tested. Efficiency of the different similarity measurements is also tested. The similarity measurements are tested on two datasets. The first dataset is MPEG-7 contour shape database Set B, which consists of 1400 shapes of natural objects and their variants. The second dataset is MPEG-7 region shape database, which consists of 3621 mainly trademark shapes and their variants. The first dataset has been subjectively classified into 70 classes by MPEG-7, with each class having 20 similar shapes. All the 70 classes of shapes are used as queries to test the retrieval result. 651 of the region shapes in the second dataset have been classified into 31 classes by MPEG-7, with each class having 21 similar shapes. All the 651 shapes are used as queries to test the retrieval result.

The feature describing the shape images in the first dataset is the Fourier Descriptor (FD), which is a normalized 10-dimension feature vector. The feature describing the shape images in the second database is the Generic Fourier Descriptor (GFD), which is a normalized 36-dimension feature vector. Detailed information on the testing datasets and the image features can be found in [11]. The retrieval performance measure used in the evaluation is the *precision* and *recall* pair. For each query, the precision of the retrieval at each level of the recall is obtained. The result precision of retrieval is the average precision of all the queries.

3.2 Results

The average precision-recall obtained from the first dataset using the 8 similarity measurements is shown in Figure 1. The precision-recall obtained from the second dataset using the 8 similarity measurements is shown in Figure 2. The time taken for the retrieval in each dataset is given in Table 1. The time reported is obtained on Windows platform of a PC with PIII-866 processor and 256M memory. The time shown in the table is the average time of three processes of running.

Since the datasets have been subjectively tested, the higher the retrieval precision, the better the distance measure conforms to human perception.

3.3 Discussions

From the retrieval effectiveness, it is found that histogram intersection distance and Mahalanobis distance perform significantly lower than other distance measure. The histogram intersection distance filters out the irrelevant elements in the matching of two feature vectors. This is appropriate for object-to- image matching because it filters out the part of features irrelevant to the object. However, it causes large error in object-to-object and image-to-image matching in general. Mahalanobis distance does not work as well as Euclidean distance. This is because the weight assigned to each dimension of the feature vector may not reflect the true significance of that dimension. Although quadratic distance measure considers the relationship between each feature element with all other feature elements, it gives equal weight to each feature element in the relationship establishment. In fact, a feature element normally only has close relationship with its neighbor feature elements. This causes database dependant performance of quadratic distance measure, as can be seen in Figure 1 and 2. In general, Euclidean distance, city block distance and χ^2 statistics are the more desirable distance measures in terms of both retrieval effectiveness and retrieval efficiency. City block distance and χ^2 statistics give slight better results than Euclidean distance measure. In terms of retrieval efficiency, except the quadratic distance measure which involves computation of $O(N^2)$, all the other 7 distance measure involves $O(N)$ computation (where N is the dimension of the feature vector).

4. CONCLUSIONS

In this paper, different similarity measurements commonly used in image retrieval have been described and evaluated. They are evaluated using shape features and standard shape datasets. Experiment results show that in terms of both retrieval effectiveness and retrieval efficiency, *city block distance* and χ^2 *statistics distance* are more desirable than other distance measurements for determining image similarity. City block distance is simpler than χ^2 statistics distance, it is more desirable for online retrieval. Although the evaluation is conducted using shape features and on shape databases, it should have similar implications to other image features and image databases.

5. REFERENCES

[1] Y. Deng. *A Region Based Representation for Image and Video Retrieval*. PhD thesis, University of California, 1999.

[2] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos and G. Taubin. *The QBIC Project: Querying Image By Content Using Color, Texture and Shape*. In Proc. SPIE Storage and Retrieval for Image and Video Databases, vol.1908, pp.173-187, 1993.

[3] Y. Rubner. *Perceptual Metrics for Image Database Navigation*. PhD thesis, Stanford University, 1999.

[4] T. Seidl and H. Kriegel. *Efficient User-adaptable Similarity Search in Large Multimedia Databases*. In Proc. of the 23rd International Conf. On Very Large Data Bases (VLDB'97), pp.506-515, Athens, Greece, August, 1997.

[5] J. R. Smith. *Integrated Spatial and Feature Image System: Retrieval, Analysis and Compression*. PhD thesis, Columbia University, 1997.

[6] M. Stricker and M. Orengo. *Similarity of Color Images*. In Proc. SPIE: Storage and Retrieval for Image and Video Databases, Vol.2420, pp.381-392, 1995.

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

[8] H. L. Van Trees. *Detection, Estimation, and Modulation Theory*. New York, Wiley, 1971.

[9] H. Voorhees and T. Poggio. *Computing Texture Boundaries from Images*. Nature, 333:364-367, 1988.

[10] E. Wold, T. Blum, D. Keislar and J. Wheaton. *Content-based Classification, Search, and Retrieval of Audio*. IEEE Multimedia, 3(3):27-36, 1996.

[11] D. S. Zhang and G. Lu. *Shape Based Image Retrieval Using Generic Fourier Descriptors*. Signal Processing: Image Communication, 17:(10), pp. 825-848, 2002.

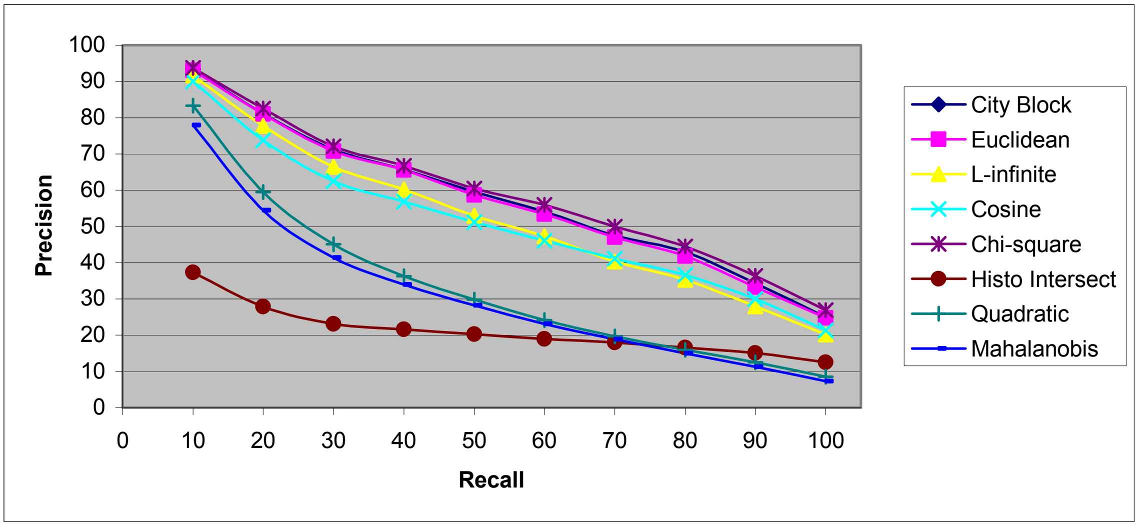

Figure 1. Retrieval performance of different distance measurements on Set B of MPEG-7 contour shape database.

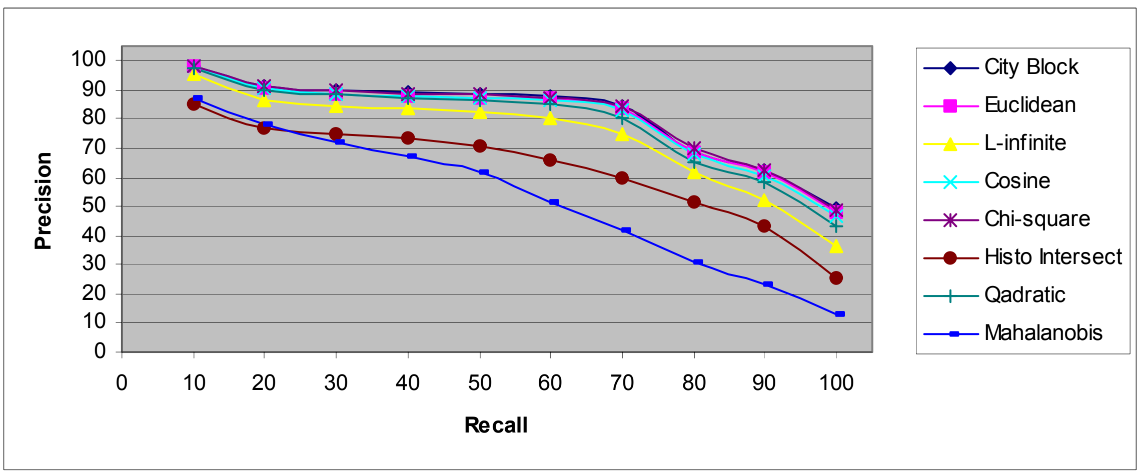

Figure 2. Retrieval performance of different distance measurements on MPEG-7 region shape database.

Table 1. Retrieval efficiency of different distance measurements

Distance Measure	Average time taken for retrieving each shape in MPEG-7 contour shape database (*ms*)	Average time taken for retrieving each shape in MPEG-7 region shape database (*ms*)
City Block (L_1)	23	33
Euclidean (L_2)	25	59
L_∞	21	34
Cosine	15	35
χ^2 statistics	23	35
Histogram Intersection	34	41
Quadratic	845	2218
Mahalanobis	27	43