Comparing Dissimilarity Measures for Content-Based Image Retrieval

Haiming Liu, Dawei Song, Stefan Rüger, Rui Hu, and Victoria Uren

Knowledge Media Institute
The Open University, Walton Hall
Milton Keynes, MK7 6AA, UK
{h.liu,d.song,s.rueger,r.hu,v.s.uren}@open.ac.uk
http://kmi.open.ac.uk/mmis

Abstract. Dissimilarity measurement plays a crucial role in content-based image retrieval, where data objects and queries are represented as vectors in high-dimensional content feature spaces. Given the large number of dissimilarity measures that exist in many fields, a crucial research question arises: Is there a dependency, if yes, what is the dependency, of a dissimilarity measure's retrieval performance, on different feature spaces? In this paper, we summarize fourteen core dissimilarity measures and classify them into three categories. A systematic performance comparison is carried out to test the effectiveness of these dissimilarity measures with six different feature spaces and some of their combinations on the Corel image collection. From our experimental results, we have drawn a number of observations and insights on dissimilarity measurement in content-based image retrieval, which will lay a foundation for developing more effective image search technologies.

Keywords: dissimilarity measure, feature space, content-based image retrieval.

1 Introduction

Content-based image retrieval is normally performed by computing the dissimilarity between the data objects and queries based on their multidimensional representations in content feature spaces, for example, colour, texture and structure. There have been a large number of dissimilarity measures from computational geometry, statistics and information theory, which can be used in image search. However, only a limited number of them have been widely used in content-based image search. Moreover, the performance of a dissimilarity measure may largely depend on different feature spaces. Although there have been some attempts in theoretically summarizing existing dissimilarity measures [6], and some evaluation to find which dissimilarity measure for shape based image search [13], there is still lack of a systematic investigation into the applicability and performance of different dissimilarity measures in image retrieval field and the investigation into various dissimilarity measures on different feature spaces for large-scale image retrieval.

H. Li et al. (Eds.): AIRS 2008, LNCS 4993, pp. 44–50, 2008.

In this paper, we systematically investigate 14 typical dissimilarity measures from different fields. Firstly, we classify them into three categories based on their theoretical origins. Secondly, we experimentally evaluate these measures in content-based image retrieval, based on six different typical feature spaces from colour, texture and structure category and some of their combinations, on the standard Corel image collection. Our systematic empirical evaluation provides initial evidence and insights on which dissimilarity measure works better on which feature spaces.

2 Classification of Dissimilarity Measures

Based on McGill and others' studies on dissimilarity measures [6,4,12], we choose 14 typical measures that have been used in information retrieval.

2.1 Geometric Measures

Geometric measures treat objects as vectors in a multi-dimensional space and compute the distance between two objects based along pairwise comparisons on dimensions.

Minkowski Family Distances (d_p)

$$d_p(A,B) = (\sum_{i=1}^{n} |a_i - b_i|^p)^{\frac{1}{p}} \tag{1}$$

Here $A = (a_1, a_2, ..., a_n)$ and $B = (b_1, b_2, ..., b_n)$ are the query vector and test object vector respectively. The Minkowski distance is a general form of the **Euclidean** (p=2), **City Block** (p=1) and **Chebyshev** ($p = \infty$) distances. Recent research has also suggested the use of **fractional** dissimilarity (i.e., $0 < p < 1$) [3], which is not a metric because it violates the triangle inequality. Howarth and Rüger [3] have found that the retrieval performance would be increases in many circumstances when p=0.5.

Cosine Function Based Dissimilarity (d_{cos}). The cosine function computes the angle between the two vectors, irrespective of vector lengths [13]:

$$s_{cos}(A,B) = \cos\theta = \frac{A \cdot B}{|A| \cdot |B|}$$

$$d_{cos}(A,B) = 1 - \cos\theta = 1 - \frac{A \cdot B}{|A| \cdot |B|} \tag{2}$$

Canberra Metric (d_{can}) [4]

$$d_{can}(A,B) = \sum_{i=1}^{n} \frac{|a_i - b_i|}{|a_i| + |b_i|} \tag{3}$$

Squared Chord (d_{sc}) [4]

$$d_{sc}(A,B) = \sum_{i=1}^{n} (\sqrt{a_i} - \sqrt{b_i})^2 \tag{4}$$

Obviously, this measure is not applicable for feature spaces with negative values.

Partial-Histogram Intersection (d_{p-hi}): This measure is able to handle partial matches when the sizes of the two object vectors are different [13]. When A and B are non-negative and have the same size, in terms of the City Block metric ($|x| = \sum_i |x_i|$), it is equivalent to the City Block measure. [12, 9]

$$d_{p-hi}(A,B) = 1 - \frac{\sum_{i=1}^{n}(\min(a_i, b_i))}{\min(|A|, |B|)} \tag{5}$$

2.2 Information Theoretic Measures

Information-theoretic measures are various conceptual derivatives from the Shannon's entropy theory and treat objects as probabilistic distributions. Therefore, again, they are not applicable to features with negative values.

Kullback-Leibler (K-L) Divergence (d_{kld}). From the information theory point of view, the K-L divergence measures how one probabilistic distribution diverges from the other. However, it is non-symmetric. [7]

$$d_{kld}(A,B) = \sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \tag{6}$$

Jeffrey Divergence (d_{jd})

$$d_{jd}(A,B) = \sum_{i=1}^{n} (a_i \log \frac{a_i}{m_i} + b_i \log \frac{b_i}{m_i}), \tag{7}$$

where $m_i = \frac{a_i + b_i}{2}$, Jeffrey divergence, in contrast to the Kullback-Leibler divergence, is numerically stable and symmetric. [10]

2.3 Statistic Measures

Unlike geometric measures, statistical measures compare two objects in a distributed manner rather than simple pair wise distance.

χ^2 Statistics (d_{χ^2})

$$d_{\chi^2}(A,B) = \sum_{i=1}^{n} \frac{(a_i - m_i)^2}{m_i}, \tag{8}$$

where $m_i = \frac{a_i+b_i}{2}$. It measures the difference of query vector (observed distribution) from the mean of both vectors (expected distribution). [13]

Pearson's Correlation Coefficient (d_{pcc}). A distance measurement derived from Pearson correlation coefficient [5] is defined as

$$d_{pcc}(A, B) = 1 - |p|, \tag{9}$$

where

$$p = \frac{n\sum_{i=1}^{n} a_i b_i - (\sum_{i=1}^{n} a_i)(\sum_{i=1}^{n} b_i)}{\sqrt{[n\sum_{i=1}^{n} a_i^2 - (\sum_{i=1}^{n} a_i)^2][n\sum_{i=1}^{n} b_i^2 - (\sum_{i=1}^{n} b_i)^2]}}$$

Note the larger $|p|$ is the more correlated the vectors A and B. [1]

Kolmogorov-Smirnov (d_{ks}). Kolmogorov-Smirnov distance is a measure of dissimilarity between two probability distributions [2]. Like K-L divergence and Jeffrey divergence, it is defined only for one-dimensional histograms [12]:

$$d_{ks}(A, B) = \max_{1 \le i \le n} |F_A(i) - F_B(i)| \tag{10}$$

$F_A(i)$ and $F_B(i)$ are the simple probability distribution function (PDF) of the object vectors, which are interpreted as probability vectors of one-dimensional histogram.

Cramer/von Mises Type (CvM) (d_{cvm}). A statistics of the Cramer/von Mises Type(CvM) is also defined based on probability distribution function (PDF) [11]:

$$d_{cvm}(A, B) = \sum_{i=1}^{n} (F_A(i) - F_B(i))^2 \tag{11}$$

3 Empirical Performance Study

Our experiment aims to address the performance of 14 dissimilarity measures on different feature spaces. We use mean average precision as the performance indicator.

3.1 Experimental Setup

Data Set. In this experiment, we use a subset of the Corel collection, developed by [8]. There are 63 categories and 6192 images in the collection, which is randomly split into 25% training data, and 75% test data. We take the training set as queries to retrieve similar images from the test set.

Features. Six typical image feature spaces are applied in the experiment.

- Colour feature spaces: RGB is three-dimensional joint colour histogram, which contains a different proportion of red, green and blue; MargRGB-H does a one-dimensional histogram for each component individually; MargRGB-M only records the first several central moments; HSV is similar to RGB, which are hue, saturation and value of colour-space.
- Texture feature spaces: Gabor, is a texture feature generated using Gabor wavelets; Tamura is a texture feature generated by statistical processing points of view.
- Structure feature space: Konvolution (Konv), discriminates between low level structures in an image, which designed to recognize horizontal, vertical and diagonal edges.

Approach. Here, we use the vector space model approach for image retrieval. The difference from [8] is that we aim to test various dissimilarity measures instead of using traditional cosine based or city block measures.

3.2 Single Feature Spaces

We investigate the performance of the 14 dissimilarity measures on 6 single image feature spaces as described above.

3.3 Combined Feature Spaces

In a further experiment, we picked up three typical features from colour, texture and structure, respectively. This experiment we use the same set up on the three and their combined feature spaces, HSV and Gabor, HSV and Konv, Gabor and Konv, and HSV, Gabor and Konv.

3.4 Results

Table 1 and Table 2 show the experimental results, from which the following observations can be made. Firstly, most of the dissimilarity measures from the geometric category have better performance than other two categories; Secondly, the performance of most of the dissimilarity measures in the color feature spaces outperform the other feature space; Finally, after identifying the top five performing dissimilarity measures on every feature space, we find Canberra metric, Squared Chord from the geometric measures category, Jeffrey Divergence from the information-theoretic measures category, and χ^2 from the statistical measures category have better performance than Euclidean and City Block dissimilarity measures, which have been most widely used in image retrieval field. Significance tests, using the paired Student's t-test (parametric test), the sign test and the paired Wilcoxon signed-rank test (non-parametric test), have shown that the improvements over the city-block measure are statistically significant (p-value less than 0.05). Therefore we would recommend them for image retrieval applications.

Table 1. Mean Average Precision on Single Feature Spaces

	HSV	margRGB-H	margRGB-M	gabor	tamura	konv
Geometric Measures						
Fractional(p=0.5)	0.1506	0.1269	0.0871	0.1490	0.1286	0.0731
City Block(p=1)	0.1682	0.1207	0.0912	0.1350	0.0949	0.0951
Euclidean(p=2)	0.1289	0.1128	0.0917	0.1161	0.0678	0.0761
Chebyshev(p=∞)	0.1094	0.1013	0.0886	0.0615	0.0358	0.0555
Cosine Similarity	0.1345	0.1204	0.0778	0.1057	0.0671	0.0716
Canberra Metric	0.1568	0.1333	0.0824	0.1496	0.1267	0.0709
Squared Chord	0.1876	0.1294	0.0967	0.1259	0.0880	0.0984
Partial-Histogram	0.1682	0.1207	0.0566	0.0320	0.0209	0.0301
Information-Theoretic Measures						
Kullback-Leibler Divergence	0.1779	0.1113	0.0893	0.1019	0.0528	0.0948
Jeffrey Divergence	0.1555	0.1185	0.0902	0.1353	0.0960	0.0950
Statistic Measures						
χ^2 Statistics	0.1810	0.1282	0.0832	0.1303	0.0897	0.0984
Pearson's Correlation	0.1307	0.1182	0.0818	0.1035	0.0692	0.0763
Kolmogorov-Smirnov	0.0967	0.1041	0.0750	0.0575	0.0426	0.0598
Cramer/von Mises Type	0.0842	0.1077	0.0724	0.0529	0.0406	0.0516

Table 2. Mean Average Precision on Combined Feature Spaces

	HSV	Gabor	Konv	HSV +Gabor	HSV +Konv	Gabor +Konv	HSV+Gabor +Konv
Geometric Measures							
Fractional(p=0.5)	0.1506	0.1490	0.0731	0.0693	0.0733	0.0686	0.0686
City Block(p=1)	0.1682	0.1350	0.0951	0.1350	0.0964	0.1396	0.1397
Euclidean(p=2)	0.1289	0.1161	0.0761	0.1163	0.0782	0.1198	0.1199
Chebyshev(p=∞)	0.1094	0.0615	0.0555	0.0623	0.0576	0.0721	0.0727
Cosine Similarity	0.1345	0.1057	0.0716	0.1542	0.1435	0.1164	0.1617
Canberra Metric	0.1568	0.1496	0.0709	0.1573	0.0765	0.1617	0.1627
Squared Chord	0.1876	0.1259	0.0984	0.1261	0.1116	0.1304	0.1306
Partial-Histogram	0.1682	0.0320	0.0301	0.0301	0.0320	0.0209	0.0205
Information-Theoretic Measures							
Kullback-Leibler Divergence	0.1779	0.1019	0.0948	0.0411	0.0306	0.0414	0.0414
Jeffrey Divergence	0.1555	0.1353	0.0950	0.1283	0.1085	0.1329	0.1330
Statistic Measures							
χ^2 Statistics	0.1810	0.1303	0.0984	0.1304	0.1062	0.1351	0.1352
Pearson's Correlation	0.1307	0.1035	0.0763	0.0529	0.1083	0.0316	0.0528
Kolmogorov-Smirnov	0.0967	0.0575	0.0598	0.1099	0.1155	0.0438	0.1163
Cramer/von Mises Type	0.0842	0.0529	0.0516	0.1291	0.1420	0.0529	0.1422

4 Conclusion and Future Work

We have reviewed fourteen dissimilarity measures, and divided them into three categories: geometry, information theory and statistics, in terms of their theoretical characteristic and functionality. In addition, these dissimilarity measures have been empirically compared on six typical content based image feature spaces, and their combinations on the standard Corel image collection.

Interesting conclusions are drawn from the experimental results, based on which we recommend Canberra metric, Squared Chord, Jeffrey Divergence, and χ^2 for future use in the Content based Image Retrieval.

This work will be a foundation for developing more effective content-based image information retrieval systems. In the future, we are going to test how the dissimilarity measures work on multi-image queries, and what their performances are on different data collections.

Acknowledgments

This work was funded in part the European Union Sixth Framework Programme (FP6) through the integrated project Pharos (IST-2006-045035). In addition, we would like to thank Peter Howarth for helping construct the feature spaces and setting up the experiments.

References

1. Chen, C.-C., Chu, H.-T.: Similarity measurement between images. In: Proceedings of the 29th Annual International Computer Software and Applications Conference (COMPSAC 2005), IEEE, Los Alamitos (2005)
2. Geman, D., Geman, S., Graffigne, C., Dong, P.: Boundary Detection by Constrained Optimization. IEEE Transactions on Pattern Analysis and Machine Intelligence 12(7), 609–628 (1990)
3. Howarth, P., Rüger, S.: Fractional distance measures for content-based image retrieval. In: Losada, D.E., Fernández-Luna, J.M. (eds.) ECIR 2005. LNCS, vol. 3408, Springer, Heidelberg (2005)
4. Kokare, M., Chatterji, B., Biswas, P.: Comparison of similarity metrics for texture image retrieval. In: Proceeding of IEEE Conf. on Convergent Technologies for Asia-Pacific Region, vol. 2, pp. 571–575 (2003)
5. Luke, B.T.: Pearson's correlation coefficient. Online (1995)
6. Noreault, T., McGill, M., Koll, M.B.: A performance evaluation of similarity measures, document term weighting schemes and representations in a Boolean environment. In: Proceeding of the 3rd annual ACM Conference on Research and development in inforamtion retreval, SIGIR 1980, Kent, UK, pp. 57–76. ACM, Butterworth Co. (1980)
7. Ojala, T., Pietikainen, M., Harwood, D.: Comparative study of texture measures with classification based on feature distributions. Pattern Recognition 29(1), 51–59 (1996)
8. Pickering, M.J., Rüger, S.: Evaluation of key frame-based retrieval techniques for video. Computer Vision and Image Understanding 92(2-3), 217–235 (2003)
9. Puzicha, J.: Distribution-Based Image Similarity, ch. 7, pp. 143–164. Kluwer Academic Publishers, Dordrecht (2001)
10. Puzicha, J., Hofmann, T., Buhmann, J.M.: Non-parametric similarity measures for unsupervised texture segmentation and image retrieval. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, San Juan (1997)
11. Puzicha, J., Rubner, Y., Tomasi, C., Buhmann, J.M.: Empirical evaluation of dissimilarity measures for color and texture. In: Proceeding of the international conference on computer vision, vol. 2, pp. 1165–1172 (September 1999)
12. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover's distance as a metric for image retrieval. International Journal of Computer Vision 40(2), 99–121 (2004)
13. Zhang, D., Lu, G.: Evaluation of similarity measurement for image retrieval. In: Procedding of IEEE International Conference on Neural Networks Signal, Nanjing, December 2003, pp. 928–931. IEEE, Los Alamitos (2003)