DISSIMILARITY MEASURES FOR CONTENT-BASED IMAGE RETRIEVAL

Rui Hu, Stefan Rüger, Dawei Song, Haiming Liu and Zi Huang

The Open University,
Knowledge Media Institute
Milton Keynes, MK7 6AA, UK

ABSTRACT

Dissimilarity measurement plays a crucial role in content-based image retrieval. In this paper, sixteen core dissimilarity measures are introduced and evaluated. We carry out a systematic performance comparison on three image collections, including Corel, Getty and Trecvid2003, with seven different feature spaces. Two search scenarios are considered: single image queries based on Vector-Space-Model, and multi-image queries based on k-Nearest Neighbours search. A number of observations is drawn, which will lay a foundation for developing more effective image search technologies.

***Index Terms*—** dissimilarity measure, feature space, content-based image retrieval

1. INTRODUCTION

Content-based Image Retrieval (CBIR) provides users a way to browse or retrieve images from large image collections based on visual similarity. Visual feature extraction and dissimilarity measures are the key issues for a CBIR system. The combination of these two attributes determines the overall effectiveness of the system. Therefore, given the visual feature(s) generated in a CBIR system, it is crucial to choose the most appropriate dissimilarity measure to achieve the best possible mean average precision.

There have been some attempts in theoretically summarizing existing dissimilarity measures [1], evaluating dissimilarity measures for texture [2], and shape based image search [3]. Our previous work [4] gives a description of fourteen dissimilarity measures on six feature spaces, but only single-image queries is conducted on one image collection (Corel), which makes the conclusions difficult to generalize to other collections. There is still a lack of a systematic investigation of dissimilarity measures on different feature spaces, with large-scale real-world image collections.

In this paper, we conduct a systematic investigation on this issue. Firstly, based on [4] we introduce and categorize 16 typical dissimilarity measures theoretically. Then, experiments are carried out on three image collections, with seven different typical feature spaces, using both single image queries and multi-image queries. Our empirical evaluation provides evidence and insights on which dissimilarity measure works better on which feature spaces.

2. DISSIMILARITY MEASURES

Dissimilarity measures are classified into three categories according to their theoretical origins. The detailed information can be find in [4].

Geometric Measures Geometric Measures treat objects as vectors. Forwardly, let v and w be two vectors in a n-dimensional real vector space, i.e. $(v, w) \in \mathbb{R}^n$. Then, the distances between v and w are as following:

Minkowski Family: $(\sum_{i=1}^{n} |v_i - w_i|^p)^{\frac{1}{p}}$
Cosine Function Based: $1 - \frac{v \cdot w}{|v| \cdot |w|}$
Canberra: $\sum_{i=1}^{n} \frac{|v_i - w_i|}{|v_i| + |w_i|}$
Squared Chord: $\sum_{i=1}^{n} (\sqrt{v_i} - \sqrt{w_i})^2$
Partial-Histogram Intersection [5]: $1 - \frac{\sum_{i=1}^{n}(\min(v_i, w_i))}{\min(|v|, |w|)}$

The Minkowski distance is a general form of a series of distance measures, such as *Euclidean* (p=2), *City Block* (p=1), *Chebyshev* ($p = \infty$), and *Fractional* distances (i.e., $0 < p < 1$) [6]. In this paper, we conducted Fractional distances with three different parameters $p = 0.25, 0.5, 0.75$. Note that the fractional distances are not metric because it violates the triangle inequality.

Information Theoretic Measures are derivatived from the Shannon's entropy theory and treat objects as probabilistic distributions, i.e. $v_i \geq 0, \Sigma v_i = 1$.

Kullback-Leibler (K-L) Divergence [7]: $\sum_{i=1}^{n} v_i \log \frac{v_i}{w_i}$
Jeffrey Divergence: $\sum_{i=1}^{n} (v_i \log \frac{v_i}{m_i} + w_i \log \frac{w_i}{m_i})$, where $m_i = \frac{v_i + w_i}{2}$

Statistic Measures compare two objects in a distributed manner, and basically assume that the vector elements are samples.
χ^2 *Statistics* [8]: $\sum_{i=1}^{n} \frac{(v_i - m_i)^2}{m_i}$,

where $m_i = \frac{v_i + w_i}{2}$
Pearson's Correlation Coefficient: $1 - |p|$,
where $p = \frac{n\sum_{i=1}^{n} v_i w_i - (\sum_{i=1}^{n} v_i)(\sum_{i=1}^{n} w_i)}{\sqrt{[n\sum_{i=1}^{n} v_i^2 - (\sum_{i=1}^{n} v_i)^2][n\sum_{i=1}^{n} w_i^2 - (\sum_{i=1}^{n} w_i)^2]}}$
Kolmogorov-Smirnov [9]: $\max_{1 \le i \le n} |F_v(i) - F_w(i)|$
Cramer/von Mises Type: $\sum_{i=1}^{n}(F_v(i) - F_w(i))^2$

3. VISUAL FEATURES

Seven typical image features including HSV, margRGB-H, margRGB-M for color; Gabor, Tammura for texture; konvolution for structure and thumbnail are applied.

Colour: HSV is a three-dimensional joint colour histograms in the cylindrical colour-space; MargRGB-H does a one-dimensional histogram for each component individually; MargRGB-M records the first four central moments.

Texture: Gabor is a texture feature generated using Gabor wavelets; Tamura is a 3 dimensional texture feature composed by measures of image's coarseness, contrast and directionality [10].

Structure: Konvolution (Konv), discriminates between low level structures in an image, and is designed to recognize horizontal, vertical and diagonal edges.

Thumbnail: This is a feature created from the pixel intensity values of a scaled down image. Here we use a size of 40 by 30 resulting in a dense vector of length 1200.

4. RETRIEVAL METHODS

In the single-image-query model a database of images is searched to find images similar to a given query image. While in a multi-image-query model, more than one query examples are given, the examples can include positive examples and negative examples, the system aims to find images similar to the positive examples but not similar to the negative examples. In this papaer, we use vector space model for single-image queries, and use k-nearest neighbours for multi-image queries.

Vector Space Model (VSM). The images are represented as vectors in a multi-dimensional feature space and then ranked according to their distances to a query.

k-Nearest Neighbours (k-NN) [11] [6]. This is a variant of the distance-weighted k-Nearest Neighbours approach. Positive examples are supplied as the queries, and negative examples are randomly selected from the training set, excluding the categories that any positive query image belongs to. Test images are then classified according to their dissimilarity to these examples according to the equation below:

$$D(i) = \frac{\sum_{n \in N}(dist(i,n))^{-1}}{\sum_{p \in P}(dist(i,p))^{-1}}. \qquad (1)$$

where P and N are the sets of positive and negative examples respectively. p and n are the k nearest examples, $|p| + |n| = k$. $dist(i,n)$ is the distance between the test image i and the negative example n; $dist(i,p)$ is the distance between i and the positive example p. A value of $k = 40$ is used for our experiments.

5. EXPERIMENTS

An comprehensive empiricial performance study, using both Vector Space Model based single-image queries and k-Nearest Neighbor based multi-image queries, is conducted on three databases including Corel, Getty and Trecvid2003.

5.1. Data Sets

COREL. We use a subset of Corel dataset, which was created by Pickering and Rüger [11]. It consists of 6192 images, belonging to 63 categories. We randomly split the collection into 25% training data and 75% test data. For single image queries, we use every image in the training set as a query. Multi-image queries are conducted for each category with the number of positive examples varying from 1 to 6; 100 negative examples are randomly selected per query. As there are 63 categories, we generate 378 multi-image queries for each dissimilarity measure/feature space combination.

GETTY. We use a subset of Getty dataset, which was created by Yavlinsky and Rüger [12]. We randomly split the dataset into 2560 training and 5000 test images. We use each image in training set as a query. The groundtruth is generated by considering the images in the test set, sharing at least one common keyword (the same 184 keywords as in [12]) with a query, as relevant to the query. For multi-image queries, we use each image in training set as a query; 100 negative images are randomly selected per query. There are 2560 multi-image queries for each dissimilarity measure/feature space combination.

TRECVID2003. It is comprised of 32,318 key-frames from Trecvid 2003 video collection. The search task consists of 25 topics [13] as query images. For multi-image queries, the number of positive examples per query ranges from 1 to 3, and 100 negative images per query. That is 75 multi-image queries for each dissimilarity measure/feature space combination.

Table 4. Recommended Dissimilarity Measures

	VSM	KNN
HSV	Squared Chord, χ^2, Histogram, City Block	Squared Chord, χ^2, Fractional(p=0.75)
margRGB-H	Fractional(p=0.5)	Squared Chord, χ^2
margRGB-M	Euclidean, City Block	Squared Chord, City Block, Euclidean
konv	Squared Chord, χ^2, City Block, Jeffrey	Squared Chord, χ^2, City Block
gabor	Fractional(p=0.25), Fractional(p=0.5)	Fractional(p=0.5), Canberra, χ^2, Squared Chord
tamura	Fractional(p=0.5), Fractional(p=0.75)	Canberra, Fractional(p=0.75)
thumbnail	City Block, Jeffrey	Canberra, Fractional(p=0.5)

5.2. Experimental Results and Analysis

For each dissimilarity measure, single-image queries and multi-image queries are performed on seven feature spaces. We use mean average precision (MAP), which has been extensively used by the Text REtrieval Conference (TREC) community [14], as the performance measure.

Results on the three datasets are listed in Table 1- 3. Each number in the tables is the MAP resulted by applying one of the sixteen dissimilarity measures on one of the seven feature spaces. The MAP for single-image and multi-image queries are shown respectively at the left hand side and right hand side of each cell.

We can observe that for each feature space, the effects of different dissimilarity measures follow a similar trend on different databases. The overall results show that the Squared Chord, Minkowski(p=0.5), χ^2 and Cityblock usually get a better performance than the other measures. For each feature space, dissimilarity measures which give top five MAP values for all the three databases are listed in Table 4, which we would recommend for future use.

6. CONCLUSION

A comprehensive performance study has been conducted for sixteen dissimilarity measures, on seven typical feature spaces, using two search methods. In order to make a more reliable conclusion, three typical databases are used.

For each feature space we list dissimilarity measures which give top five mean average precisions for all the three databases, for both vector space model based single image queries and k-nearest neighbours method based multi-image queries. This conclusion can be a foundation for developing more effective content based image retrieval systems.

7. REFERENCES

[1] T. Noreault, M. McGill, and M. B. Koll, "A performance evaluation of similarity measures, document term weighting schemes and representations in a boolean environment," in *3rd ACM Conf. on Research and development in inforamtion retreval, SIGIR*, 1980, pp. 57–76.

[2] M. Kokare, B.N. Chatterji, and P.K. Biswas, "Comparison of similarity metrics for texture image retrieval," in *IEEE Conf. on Convergent Technologies for Asia-Pacific Region*, 2003, vol. 2, pp. 571–575.

[3] D. Zhang and G. Lu, "Evaluation of similarity measurement for image retrieval," in *IEEE International Conf. on Neural Networks Signal*, 2003, pp. 928–931.

[4] H. Liu, D. Song, S. Rüger, R. Hu, and V. Uren, "Comparing dissimilarity measures for content-based image retrieval," *to be appear*.

[5] M. J. Swain and D. H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, pp. 11–32, 1991.

[6] P. Howarth and S. Rüger, "Fractional distance measures for content-based image retrieval," in *European Conf. on IR research*, 2005.

[7] T. Ojala, M. Pietikainen, and D. Harwood, "Comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[8] J. Puzicha, T. Hofmann, and J. M. Buhmann, "Non-parametric similarity measures for unsupervised texture segmentation and image retrieval," *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 267–272, 1997.

[9] D. Geman, S. Geman, C. Grafgne, and P. Dong, "Boundary detection by constrained optimization," *IEEE Trans. Pattern Anal. Machine Intell*, vol. 12, pp. 609–628, 1990.

[10] P. Howarth and S. Rueger, "Robust texture features for still-image retrieval," *IEE Proceedings on Vision, Image and Signal*, vol. 152, pp. 868–874, 2005.

[11] M. J. Pickering and S. Rüger, "Evaluation of key frame-based retrieval techniques for video," *Computer Vision and Image Understanding*, vol. 92, pp. 217–235, 2003.

[12] A. Yavlinsky, E. Schofield, and S. Rueger, "Automated image annotation using global features and robust non-parametric density estimation," *CIVR*, vol. 3568, pp. 507–517, 2005.

[13] A. Smeaton, W. Kraaij, and P. Over, "Trecvid2003-an introduction," *TRECVID workshop*, pp. 1–10, 2003.

[14] E. M. Voorhees and D. Harman, "Overview of the eighth text retrieval conference (trec-8)," *Proceedings of TREC*, pp. 1–24, 2000.

Table 1. Mean Average Precision for Corel dataset

	HSV		margRGB-H		margRGB-M		konv		gabor		tamura		thumbnail	
Geometric Measures														
Fractional(p=0.25)	0.1059	0.1807	0.1294	0.1912	0.0823	0.1339	0.0677	0.0801	0.1566	0.1605	0.1437	0.1448	0.1329	0.1375
Fractional(p=0.5)	0.1506	0.2953	0.1269	0.1964	0.0871	0.1461	0.0731	0.1086	0.1490	0.1882	0.1286	0.01773	0.1289	0.1503
Fractional(p=0.75)	0.1733	0.2747	0.1236	0.1911	0.0898	0.1489	0.0850	0.1383	0.1416	0.1811	0.1097	0.1626	0.1238	0.1445
City Block(p=1)	0.1682	0.2532	0.1207	0.1877	0.0912	0.1495	0.0951	0.1481	0.1350	0.1791	0.0949	0.1538	0.1176	0.1398
Euclidean(p=2)	0.1289	0.1969	0.1128	0.1855	0.0917	0.1476	0.0761	0.1043	0.1161	0.1789	0.0678	0.1024	0.0929	0.1293
Chebyshev(p=∞)	0.1094	0.1559	0.1013	0.1591	0.0886	0.1412	0.0555	0.0772	0.0615	0.1205	0.0358	0.0536	0.0332	0.0592
Cosine	0.1345	0.1559	0.1204	0.1591	0.0778	0.1412	0.0716	0.0772	0.1057	0.1205	0.0671	0.0536	0.0756	0.0592
Canberra	0.1568	0.2779	0.1333	0.2016	0.0824	0.1396	0.0709	0.1104	0.1496	0.2296	0.1267	0.1880	0.1211	0.1593
Squared Chord	0.1876	0.2894	0.1294	0.2044	0.0967	0.1607	0.0984	0.1597	0.1259	0.1898	0.0880	0.1507	0.0904	0.1170
Histogram	0.1682	0.1559	0.1207	0.1591	0.0720	0.1412	0.0551	0.0772	0.0680	0.1205	0.0319	0.0536	0.0486	0.0592
Information-Theoretic Measures														
Kullback-Leibler	0.1779	0.1052	0.1113	0.1888	0.0893	0.1443	0.0528	0.1444	0.1019	0.1205	0.0948	0.0672	0.0467	0.0828
Jeffrey	0.1555	0.2345	0.1185	0.1808	0.0902	0.1470	0.0960	0.1473	0.1353	0.1782	0.0950	0.1562	0.1196	0.1404
Statistic Measures														
χ^2 Statistics	0.1810	0.2754	0.1282	0.2010	0.0832	0.1352	0.0897	0.1597	0.1303	0.1966	0.0984	0.1573	0.0940	0.1198
Pearson	0.1307	0.1825	0.1182	0.1832	0.0818	0.1417	0.0692	0.1240	0.1035	0.1663	0.0763	0.1010	0.0665	0.0933
Kolmogorov	0.0967	0.1477	0.1041	0.1687	0.0750	0.1132	0.0426	0.0878	0.0575	0.0383	0.0598	0.0583	0.0618	0.0769
Cramer	0.0842	0.1352	0.1077	0.1699	0.0724	0.1088	0.0406	0.0675	0.0529	0.0766	0.0516	0.0439	0.0564	0.0513

Table 2. Mean Average Precision for Getty dataset

	HSV		margRGB-H		margRGB-M		konv		gabor		tamura		thumbnail	
Geometric Measures														
Fractional(p=0.25)	0.1408	0.1546	0.1505	0.1501	0.1441	0.1454	0.1414	0.1431	0.1527	0.1526	0.1544	0.1531	0.1458	0.1526
Fractional(p=0.5)	0.1482	0.1724	0.1499	0.1555	0.1465	0.1518	0.1427	0.1509	0.1509	0.1584	0.1502	0.1582	0.1459	0.1539
Fractional(p=0.75)	0.1575	0.1743	0.1487	0.1559	0.1484	0.1541	0.1448	0.1531	0.1492	0.1571	0.1469	0.1551	0.1458	0.1536
City Block(p=1)	0.1628	0.1740	0.1475	0.1557	0.1497	0.1557	0.1472	0.1554	0.1479	0.1561	0.1445	0.1531	0.1455	0.1534
Euclidean(p=2)	0.1503	0.1586	0.1449	0.1551	0.1523	0.1581	0.1442	0.1518	0.1445	0.1541	0.1396	0.1485	0.1445	0.1528
Chebyshev(p=∞)	0.1510	0.1514	0.1426	0.1531	0.1520	0.1575	0.1396	0.1474	0.1392	0.1486	0.1311	0.1408	0.1391	0.1462
Cosine	0.1561	0.1565	0.1498	0.1512	0.1507	0.1553	0.1420	0.1473	0.1341	0.1442	0.1298	0.1412	0.1324	0.1409
Canberra	0.1484	0.1629	0.1421	0.1501	0.1451	0.1506	0.1420	0.1503	0.1445	0.1599	0.1434	0.1572	0.1408	0.1521
Squared Chord	0.1657	0.1788	0.1484	0.1586	0.1489	0.1577	0.1480	0.1563	0.1470	0.1574	0.1408	0.1519	0.1435	0.1524
Histogram	0.1628	0.1661	0.1475	0.1504	0.1432	0.1494	0.1319	0.1502	0.1253	0.1420	0.1218	0.1385	0.1222	0.1364
Information-Theoretic Measures														
Kullback-Leibler	0.1140	0.1243	0.1391	0.1525	0.1422	0.1428	0.1448	0.1419	0.1329	0.1388	0.1285	0.1390	0.1351	0.1398
Jeffrey	0.1582	0.1772	0.1466	0.1584	0.1493	0.1499	0.1472	0.1563	0.1480	0.1575	0.1454	0.1519	0.1458	0.1525
Statistic Measures														
χ^2 Statistics	0.1640	0.1760	0.1482	0.1579	0.1453	0.1500	0.1479	0.1563	0.1471	0.1574	0.1415	0.1520	0.1438	0.1526
Pearson	0.1517	0.1614	0.1447	0.1501	0.1500	0.1602	0.1433	0.1525	0.1339	0.1493	0.1296	0.1455	0.1337	0.1404
Kolmogorov	0.1433	0.1452	0.1513	0.1612	0.1386	0.1436	0.1391	0.1479	0.1398	0.1478	0.1369	0.1450	0.1389	0.1368
Cramer	0.1415	0.1434	0.1552	0.1629	0.1381	0.1431	0.1378	0.1459	0.1391	0.1436	0.1372	0.1448	0.1381	0.1427

Table 3. Mean Average Precision for Trecvid2003 dataset

	HSV		margRGB-H		margRGB-M		konv		gabor		tamura		thumbnail	
Geometric Measures														
Fractional(p=0.25)	0.0105	0.0126	0.0090	0.0140	0.0069	0.0132	0.0115	0.0264	0.0263	0.0290	0.0187	0.0210	0.0192	0.0280
Fractional(p=0.5)	0.0137	0.0168	0.0097	0.0142	0.0077	0.0132	0.0120	0.0172	0.0259	0.0290	0.0208	0.0210	0.0204	0.0260
Fractional(p=0.75)	0.0161	0.0180	0.0100	0.0143	0.0081	0.0136	0.0133	0.0172	0.0254	0.0262	0.0210	0.0222	0.0215	0.0240
City Block(p=1)	0.0149	0.0176	0.0101	0.0136	0.0084	0.0140	0.0139	0.0176	0.0249	0.0262	0.0209	0.0238	0.0223	0.0228
Euclidean(p=2)	0.0106	0.0164	0.0101	0.0139	0.0090	0.0140	0.0115	0.0168	0.0233	0.0250	0.0189	0.0230	0.0229	0.0236
Chebyshev(p=∞)	0.0086	0.0144	0.0088	0.0137	0.0084	0.0144	0.0107	0.0136	0.0169	0.0238	0.0093	0.0170	0.0079	0.0168
Cosine	0.0120	0.0121	0.0104	0.0132	0.0101	0.0116	0.0135	0.0116	0.0255	0.0154	0.0177	0.0162	0.0219	0.0152
Canberra	0.0118	0.0132	0.0087	0.0136	0.0083	0.0136	0.0118	0.0180	0.0257	0.0274	0.0165	0.0242	0.0207	0.0232
Squared Chord	0.0160	0.0180	0.0104	0.0145	0.0096	0.0140	0.0143	0.0176	0.0264	0.0278	0.0183	0.0242	0.0221	0.0272
Histogram	0.0149	0.0127	0.0101	0.0129	0.0062	0.0116	0.0072	0.0116	0.0059	0.0150	0.0067	0.0182	0.0059	0.0140
Information-Theoretic Measures														
Kullback-Leibler	0.0058	0.0120	0.0076	0.0140	0.0071	0.0132	0.0105	0.0128	0.0155	0.0278	0.0097	0.0174	0.0139	0.0136
Jeffrey	0.0133	0.0178	0.0101	0.0146	0.0083	0.0132	0.0138	0.0128	0.0246	0.0274	0.0209	0.0174	0.0219	0.0136
Statistic Measures														
χ^2 Statistics	0.0157	0.0181	0.0104	0.0145	0.0091	0.0152	0.0143	0.0176	0.0265	0.0274	0.0190	0.0234	0.0223	0.0272
Pearson	0.0119	0.0145	0.0105	0.0140	0.0091	0.0136	0.0120	0.0192	0.0201	0.0266	0.0176	0.0242	0.0166	0.0276
Kolmogorov	0.0065	0.0.131	0.0077	0.0128	0.0058	0.0124	0.0078	0.0132	0.0056	0.0166	0.0060	0.0174	0.0074	0.0156
Cramer	0.0064	0.0124	0.0089	0.0146	0.0057	0.0124	0.0065	0.0128	0.0052	0.0158	0.0064	0.0174	0.0068	0.0252