3-D shape descriptors and distance metrics for content-based artefact retrieval

Simon Goodall, Paul Lewis and Kirk Martinez

School of Electronics and Computer Science,
University of Southampton,
SO17 1BJ,
United Kingdom

ABSTRACT

The growing number of large multimedia collections has led to an increased interest in content-based retrieval research. Applications of content-based techniques to image retrieval is an active research area but much less work has been reported on content-based retrieval of 3-D objects in a multimedia database context. Increasingly such objects are being captured and added to multimedia collections and the European project, SCULPTEUR, is developing a museum information system which includes the introduction of facilities for content-based retrieval of the 3-D representations.

This paper provides a comparison and evaluation of a range of 3-D shape descriptors and distance metrics which have been introduced into the SCULPTEUR project to demonstrate their use for content-based retrieval applications.

Results show that while particular descriptors and distance metrics provide good overall performance, it can be more appropriate to choose different descriptors for different search tasks.

Keywords: SCULPTUER, content-based retrieval, Princeton shape benchmark, 3-D, performance, descriptor, metric, museum

1. INTRODUCTION

The growing number of large multimedia collections has led to an increased interest in content-based retrieval research. Applications of content-based techniques to image retrieval is an active research area but much less work has been reported on content-based retrieval of 3-D objects in a multimedia database context. Increasingly such objects are being captured and added to multimedia collections and the European project, SCULPTEUR,[1] is developing a museum information system which includes the introduction of facilities for content-based retrieval of the 3-D representations. The project is also concerned with another rapidly developing area: the semantic web, which will aid the use of semantically described data both for machine processing and enhanced human interaction.

This paper provides a comparison and evaluation of a range of 3-D shape descriptors and distance metrics which have been introduced into the SCULPTEUR project to demonstrate their use for content-based retrieval applications.

Further author information: (Send correspondence to S. Goodall.)
S. Goodall: E-mail: sg02r@ecs.soton.ac.uk, Telephone: +44 (0)23 8059 3256

2. BACKGROUND & RELATED WORK

The project involves five major European galleries, the Uffizi in Florence, the National Gallery and the Victoria and Albert Museum in London, the Museé de Cherbourg and the Centre de Recherche et de Restauration des Museés de France (C2RMF) which is the Louvre related art restoration centre. Each of these galleries maintains a substantial digital archive of its collections. Other technical partners include Centrica in Italy and GET-ENST in Paris. The project builds on the work of an earlier project, ARTISTE,[2] involving many of the current partners in the design of distributed content and textual metadata-based image retrieval and navigation facilities.

Increasingly museums are recognising the value of 3-D visualisation of their artefacts not only for researchers, curators and historians but also potentially as an information source for a wider public and as a basis for e-learning and commercial activity.

The project is concerned with the perceived benefit of structuring and integrating the knowledge associated with museum artefacts, enabling users to more fully exploit the richness of the data, facilitating more versatile browsing, retrieval and navigation within collections, and also enabling cross collection searching and interoperability with external systems. Starting with the CIDOC conceptual reference model,[3] ontological descriptions of the museum collections are being developed. Metadata associated with the artefacts is being mapped to the ontology to form an integrated knowledge base and graphical tools are being developed to provide browsing of the concepts, relationships and instances within the collections. Integrated concept, metadata and content-based retrieval and navigation facilities are being implemented to explore the knowledge base.

A novel aspect of the project is the ability to search and retrieve objects from the system through a range of methods. In addition to the standard textual search interface, objects can also be retrieved by browsing the ontology (concept-based retrieval) or by providing an example to the system (content-based retrieval) or through a combination of these methods. For example a user can give an example model of a vase to the system and give a text based entry of "Greek" to retrieve Greek vases that are of a similar shape to the example.

Other goals of the project include the development of a web agent which automatically identifies missing instance information in the knowledge base and attempts to locate it on the web[4]; the development of an agent to assist with the classification of new acquisitions based on existing classifications within the collection and finally to demonstrate exploitation of the knowledge base by development of an e-learning product to make use of it.

The main contribution of this paper is to provide a comparison and evaluation of the 3-D shape descriptors and distance metrics which have been introduced into the architecture and to demonstrate their use for content-based retrieval applications. Recent work in the area of descriptor evaluation has used the Princeton Shape Benchmark (PSB)[5] to perform comparisons of a large range of shape descriptors. The benchmark provides a large dataset of classified objects (approx 1,800) from a wide range of classes and the tools to evaluate descriptor performance using a range of criteria (Precision-recall, nearest neighbour, first-tier, second-tier, E-Measure, F-Measure and DCG). Many other datasets have been used in the literature. However, they typically contain a small range of classified objects, or use a private dataset. The importance of classified objects is that manual classification is not required and more direct comparisons to other work using the dataset can be achieved.

The Viewpoint dataset[6] also has a large number of classified objects, however this is a commercial database and is considered too expensive for research purposes. Another commonly used dataset is the one by Osada et al..[7] However, it only contains 133 classified objects. Another larger dataset used by Chen[8] contains around 550 classified objects and these are limited to mostly vehicle and household objects.

In the next section the shape descriptors implemented within the project are introduced. This is followed by sections describing the distance metrics, the evaluation criteria and the model collections used. The methodology, results and future aims complete this paper.

3. 3-D MODELS & MATCHING

3-D object matching is a growing research area, and a wide range of differing techniques have been developed. 3-D content-based retrieval typically goes through three stages. The first stage is to normalise the object into a canonical coordinate frame. This stage is optional however, depending upon the requirements for a particular algorithm. Stage two is to generate the descriptor from the object, and stage three is to compare the descriptor with other descriptors using an appropriate distance metric.

As part of the ongoing work in the SCULPTEUR project, several 3-D matching algorithms have been implemented and integrated with our pre-existing retrieval facilities to provide 3-D content-based retrieval. These include the D2 shape distribution descriptors from the Princeton Shape Retrieval and Analysis Group[7] and the histogram descriptors from Paquet and Rioux developed as part of the Nefertiti system.[9] An area to volume ratio descriptor,[10] which is a single valued statistic giving the ratio of the surface area of the model to its enclosed volume is also introduced to provide a fast discriminator which can reduce the search space.

The Shape Distributions are a collection of descriptors that capture distributions of various features of the shape of an object. The work done by Osada et al.[7] determined that the D2 variant performed best overall and hence this variant is used in the project. The Shape D2 descriptor captures the distribution of distances between random pairs of points on the shape surface. It is rotation and translation invariant and robust to changes in mesh resolution. However, it is not scale invariant and so requires some pre-processing.

There are three versions of the histogram descriptors of Paquet and Rioux.[9] They define a cord as the vector between the centre of mass of an object and a point on its surface. Their first histogram records the distribution of the cord lengths for all points within the mesh (Cord Hist 1). The other two variations record the distribution of angles between cords and the first (Cord Hist 2) and second (Cord Hist 3) principal axis respectively. We have implemented the cord histograms as individual descriptors (one per type) and as a combined descriptor referred to as Cord Histogram from now on. The histograms are rotation and translation invariant but again normalisation for scale is required.

The reader is referred to Tangelder et al.[11] for a more comprehensive overview of 3-D retrieval methods.

3.1. Distance Metrics

In order to establish the similarity (closeness) of two feature vectors in some feature space, a wide range of distance metrics have been presented in the literature. The most commonly used metrics are the Minkowski norms, typically the L_1 norm (the city block distance) and the L_2 norm (the Euclidean distance). (See e.g.[7, 12]). The norms are particularly attractive as they are simple to calculate and generally produce good results. However other distance metrics may provide better results when used in combination with specific descriptors and types of object.

Osada et al.[7] suggest a range of distance metrics that could be used for comparison purposes. These are the Kolmogorov-Smirnov distance, Kullback-Leibler divergence distance, Match distances, Earth Mover's distance and the Bhattacharyya distance. Hetzel et al.[13] suggest the histogram intersection and the χ^2 distance, while Ankerst et al.[12] suggest the Quadratic distance.

3.2. Evaluation Criteria

A wide range of criteria can be used to evaluate the quality of a descriptor for retrieval purposes. A common core of evaluation techniques can be established from Järvelin et al.,[14] Shilane et al.[5] and van Rijsbergen.[15]

The precision-recall graph is a commonly used method of evaluating the quality of a descriptor. Precision is defined as the proportion of relevant results, out of the results returned. Recall is defined as the proportion of relevant results returned out of all the possible relevant results. The basic precision-recall graph (showing precision against recall as the size of the returned set increases) is sometimes considered not to be enough. A common extension is to add the precision-recall curve for a random retrieval. Typically, one would expect that as recall increases, precision decreases. The E-Measure is one of several criteria that combine precision and recall into a single value. Other measures include the F-Measure and the Borko method. The nearest neighbour criterion is the percentage of objects for which the nearest object is of the same class. The first- and second-tier criteria are the percentage of the first K elements that are of the same class, where K, for the first-tier, is the size of the class. The second-tier uses K as twice the size of the class. The Discounted Cumulative Gain (DCG) is a measure that weights correct results returned earlier higher than those returned later within a ranked list. All these statistics are normalised to the range (0 to 1) for presentation purposes (Tables 2, 3, 4, and 5).

Class Name	Size
Bird Like	24
Head	24
Misc	38
Mold	34
Statue	54
Vases	108

Table 1. The Museum Dataset

4. METHODOLOGY

A quantitative comparison of shape descriptors and distance metrics in the SCULPTEUR system is presented. This is undertaken using a publicly available dataset, the Princeton Shape Benchmark (PSB), to provide a comparison with other work in this area. In particular the base training set is used. This contains 907 models in 90 non-empty classes. See Shilane et al.[5] for more details on the PSB dataset. For evaluation within a real setting, a dataset made up of real museum objects has been created. The museum data consists of 282 objects manually classified into 6 classes. Table 1 shows the classes and their size. This represents the majority of the objects supplied by museum partners. As can be seen in Table 1, the majority are either Vases or Statues. The Misc class contains objects that did not easily fit into any of the others. These objects are represented in VRML and range from containing a few thousand polygons, to many thousands of polygons.

The evaluation procedure will make use of the evaluation criteria, precision-recall graphs, nearest-neighbour, first- and second-tier, E-Measure, F-Measure and DCG statistics to give a broad view of the abilities of the descriptors. How the descriptors and distance metrics perform overall, and on a class basis are assessed, in order to identify on what types of object the descriptors work best, and on which ones they perform poorly.

In order to evaluate the per-class capabilities of each descriptor and metric, we present the percentage of classes for which each descriptor and metric performed “best”. To determine what was “best”, the sum of the different statistics (Nearest Neighbour, First-Tier, Second-Tier, E-Measure, F-Measure and DCG) is used, and the combination with the highest sum is taken. Alternative methods of combining may be more appropriate however as this method does not take into account the range in which each statistic typically operates, i.e. some statistics could typically return larger values that would dominate the statistics that returned smaller values.

The evaluation begins by comparing the descriptors side by side. The descriptors to be evaluated are the Shape D2, Cord Hist 1, Cord Hist 2, Cord Hist 3, Cord Histogram and Area Volume descriptors. The evaluation then continues to compare the distance metrics side by side. These are Minkowski L_1 and L_2 norms, Bhattacharyya, χ^2, Kullback-Leibler, histogram intersection and quadratic distances metrics. In addition, both these stages will compare the results from using both the PSB dataset and the Museum dataset.

5. RESULTS

The descriptors were all implemented with a bin count of 64, except in the case of the Area Volume descriptor which is only a single value. This value was chosen to try and match the implementation in Shilane et al.[5] to provide results that can be compared to those already published using the PSB. Experimentation on different bin sizes (8, 16, 32 and 64 bins) showed that there was little performance differences between them, however the lower bin sizes quite often showed marginal improvements in performance.

Evaluation begins by comparing the descriptors against each other using the Euclidean distance. Figure 1 shows the precision-recall graphs for the different descriptors and the Euclidean distance using the PSB dataset. As can be seen here, there is little difference between the descriptors, although two groupings can be distinguished. The Shape D2, Cord Hist 2, Cord Hist 3 and Cord Histogram descriptors perform quite similarly to each other, as does the Area Volume descriptor and the Cord Hist 1 descriptor, although there is noticeably

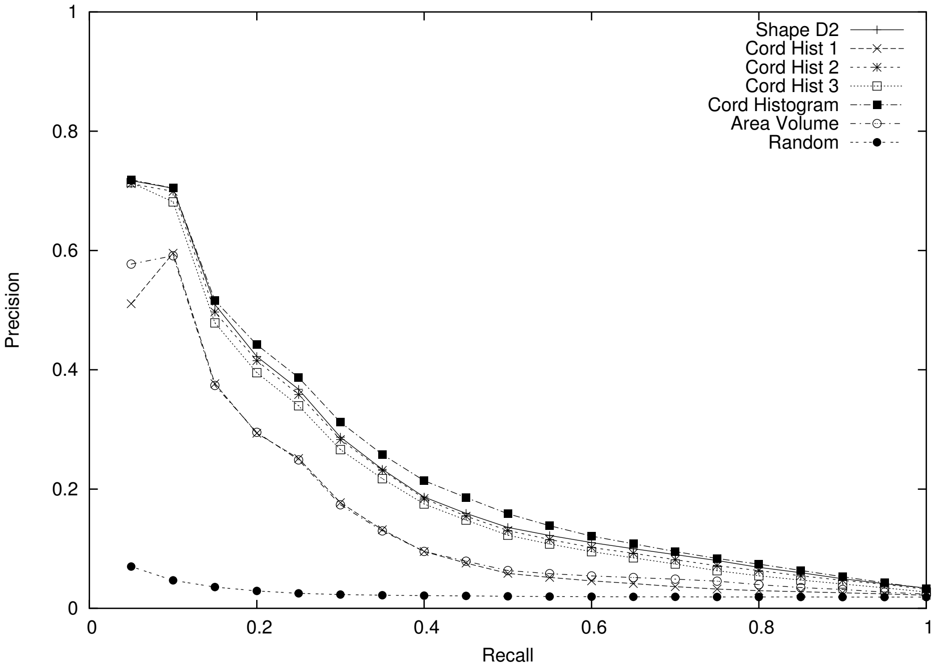

Figure 1. Overall Performance (PSB)

Descriptor	Nearest Neighbour	First Tier	Second Tier	E-Measure	F-Measure	DCG	Percent of Classes
Area Volume	0.074	0.143	0.183	0.080	0.094	0.392	2.2%
Cord Hist 1	0.104	0.143	0.174	0.074	0.089	0.380	6.7%
Cord Hist 2	0.246	0.202	0.258	0.121	0.140	0.446	15.6%
Cord Hist 3	0.212	0.197	0.251	0.116	0.135	0.440	12.2%
Cord Histogram	0.258	0.216	0.282	0.133	0.154	0.463	24.4%
Shape D2	0.245	0.207	0.274	0.126	0.146	0.455	38.9%
Random	0.018	0.015	0.033	0.017	0.019	0.306	N/A

Table 2. Overall Performance (PSB)

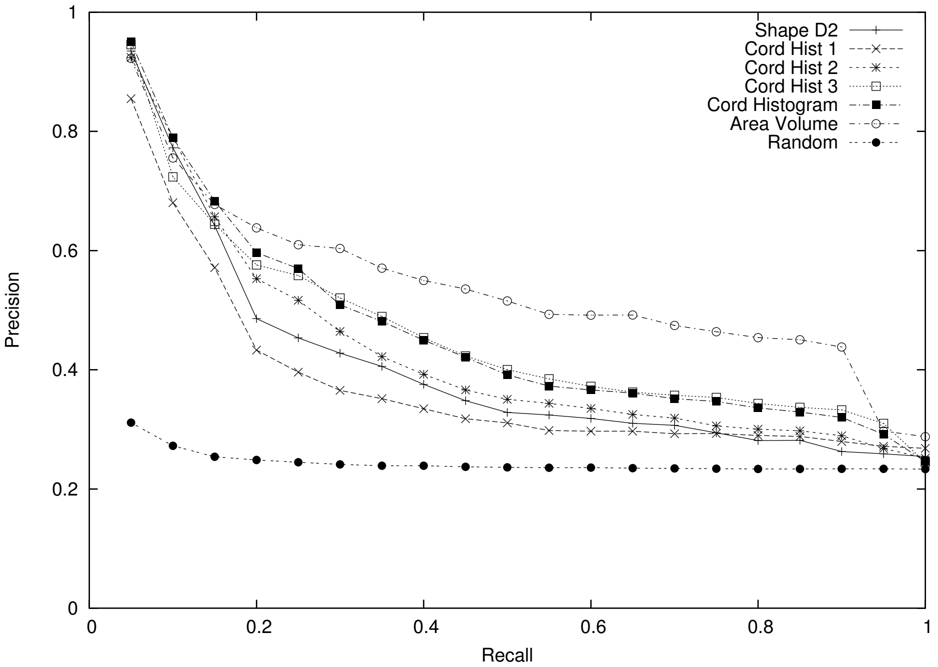

Figure 2. Overall Performance (Museum)

Descriptor	Nearest Neighbour	First Tier	Second Tier	E-Measure	F-Measure	DCG	Percent of Classes
Area Volume	0.986	0.518	0.646	0.438	0.359	0.798	33.3%
Cord Hist 1	0.986	0.310	0.565	0.307	0.262	0.753	33.3%
Cord Hist 2	0.986	0.378	0.610	0.354	0.287	0.766	16.7%
Cord Hist 3	0.986	0.406	0.585	0.373	0.291	0.761	0.0%
Cord Histogram	0.986	0.410	0.621	0.380	0.301	0771	16.7%
Shape D2	1.000	0.350	0.587	0.322	0.267	0.764	0.0%
Random	0.241	0.233	0.460	0.182	0.144	0.641	N/A

Table 3. Overall Performance (Museum)

lower performance results from them. Table 2 shows the results from the statistic evaluators, which also shows similar groupings in performance for the descriptors. It is worth noting that the Shape D2 performed better on an individual class basis then the Cord Histogram which has the better overall statistics.

Figure 2 and Table 3 shows the equivalent results for the Museum dataset. It is plain to see that the descriptors and metrics perform significantly differently against this dataset than against the PSB. The Museum dataset shows much higher performance values for each descriptor, and even for the random retrieval due to the small number of classes. One surprise is the Area Volume descriptor achieving the best performance. In this case the Shape D2 performs quite badly compared to most of the other descriptors. However, overall, the Cord Hist 1 descriptor performs the worst compared to the PSB dataset. However the Cord Hist 1 descriptor performs equally well on the percentage of classes as the Area Volume descriptor.

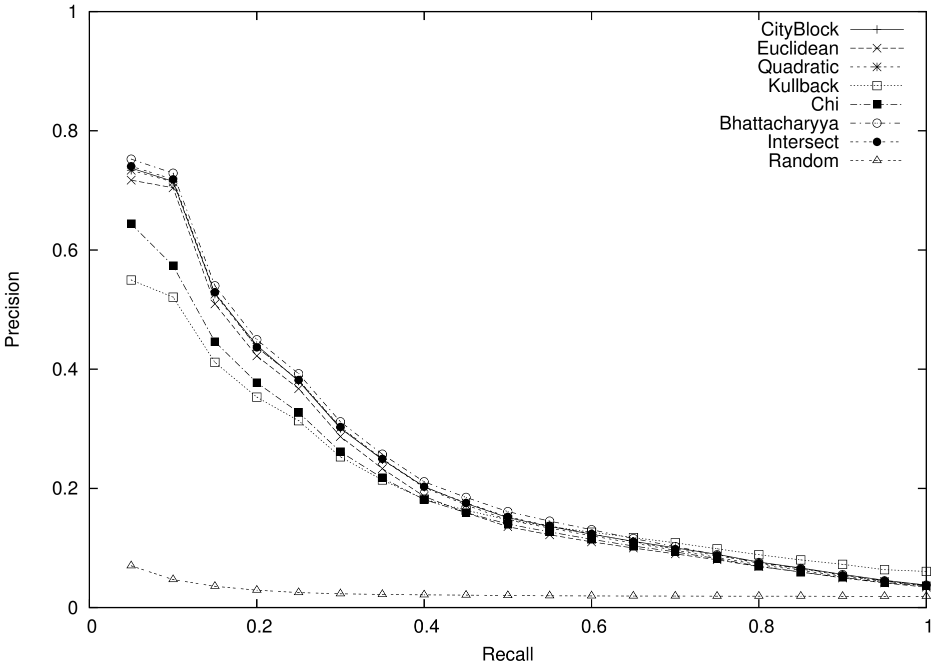

Figure 3. Distance Metric Performance (PSB)

Distance Metric	Nearest Neighbour	First Tier	Second Tier	E-Measure	F-Measure	DCG	Percent of Classes
Bhattacharyya	0.286	0.224	0.295	0.138	0.159	0.471	25.6%
Chi	0.241	0.191	0.258	0.121	0.140	0.446	3.3%
City Block	0.267	0.218	0.288	0.131	0.153	0.466	11.1%
Euclidean	0.245	0.207	0.274	0.126	0.146	0.455	12.2%
Intersection	0.269	0.218	0.286	0.131	0.152	0.468	13.3%
Kullback	0.222	0.188	0.255	0.108	0.124	0.432	18.9%
Quadratic	0.272	0.214	0.282	0.131	0.153	0.464	15.6%
Random	0.018	0.015	0.033	0.017	0.019	0.306	N/A

Table 4. Distance Metric Performance (PSB)

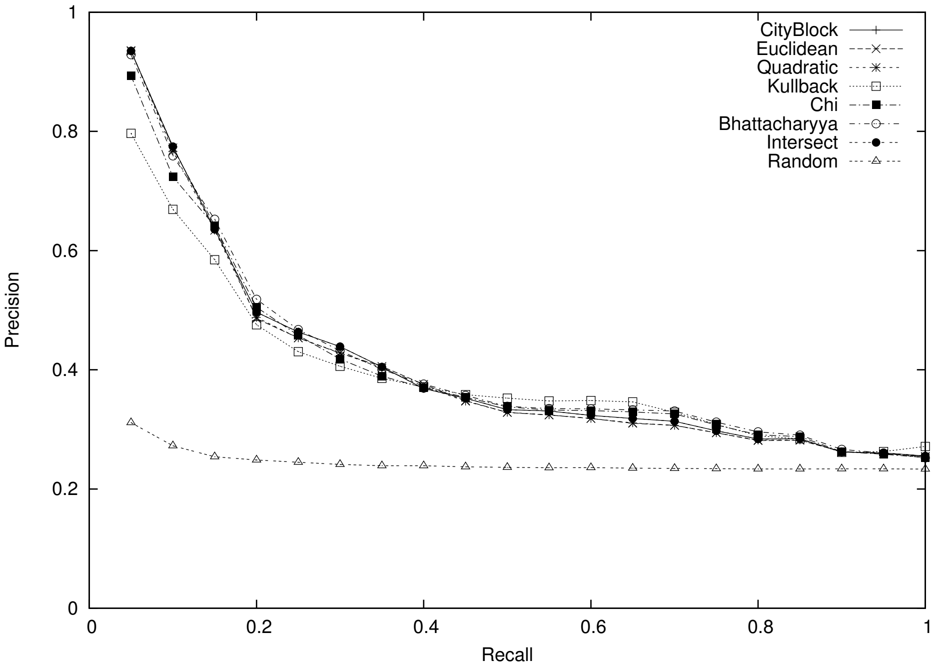

Figure 4. Distance Metric Performance (Museum)

Distance Metric	Nearest Neighbour	First Tier	Second Tier	E-Measure	F-Measure	DCG	Percent of Classes
Bhattacharyya	1.000	0.355	0.594	0.338	0.278	0.765	33.3%
Chi	0.965	0.349	0.590	0.330	0.271	0.756	0.0%
City Block	1.000	0.353	0.584	0.324	0.268	0.764	0.0%
Euclidean	1.000	0.350	0.587	0.322	0.267	0.764	33.3%
Intersection	1.000	0.353	0.584	0.324	0.268	0.764	0.0%
Kullback	0.858	0.350	0.562	0.306	0.248	0.736	33.3%
Quadratic	1.000	0.352	0.584	0.322	0.267	0.763	0.0%
Random	0.241	0.233	0.460	0.182	0.144	0.641	N/A

Table 5. Distance Metric Performance(Museum)

The next section of results evaluates distance metric performance using the Shape D2 descriptor. Figure 3 shows the precision-recall curves for the PSB dataset with different distance metrics while Table 4 shows the corresponding statistic evaluators. This shows a similar story to that of the descriptors; some close groupings of distance metrics. The precision-recall curves suggests that all metrics except the χ^2 and Kullback distances performed approximately equally well with the Bhattacharyya just slightly in the lead. This is supported by the data in Table 4 and the Bhattacharyya performs significantly better than the others in the class by class analysis. The Kullback metric performs worst in the overall tests but is second to the Bhattacharyya in the class by class tests.

Figure 4 and Table 5 show the results for the museum dataset. A similar overall improvement in performance

Figure 5. Top 10 Results (Museum)

compared with that of the PSB set can be seen with the distance metrics, as it was with the descriptors. However, there is little overall change in the relative performance of distance metrics between the datasets, unlike the situation for the descriptors. On a class basis, the Bhattacharyya, Euclidean and Kullback have equal shares of the classes, whereas the others have none. Again, overall the Bhattacharyya distance shows slightly better statistics than the others, however for this dataset, there is very little difference between them.

Figure 5 shows an example query from the Museum dataset. The top 10 results using the Cord Histogram at 64 bins with the Euclidean distance is shown. The top left object is the closest match and it is also the query object as would be expected. This figure also shows some of the duplication of objects within the dataset.

6. DISCUSSION

Comparisons between the different shape descriptors and distance metrics have been performed using the Princeton Shape Benchmark dataset. This has shown that the Cord Histogram gives both good and consistent results across both datasets. While the Shape D2 and the Area Volume descriptors preform best in the PSB and Museum dataset respectively, their performance is poor against the other dataset. It is surprising that the performance of the Area Volume descriptor was so good. As the Area Volume ratio is just a single value, intuition would suggest that it would not be capable of capturing as much useful data as the multi-valued representations.

Typically, it seems that the choice of distance metric is less important than the choice of descriptor, as would be expected. However in some cases the choice of metric is enough to improve the retrieval performance beyond those of other descriptors. The Bhattacharyya distance metric seems to perform best overall, however, in computational terms, it is quite expensive. Good results were seen with the city block metric. This produced similar results to that of the Bhattacharyya distance and is much cheaper computationally.

The differences in the two datasets account for some of the differences in performance. The improved retrieval rates of the Museum dataset can in part be attributed to the small number of classes compared to the PSB dataset. The small number of classes helps improve the random retrieval results as it improves from an approximate 1 in 90 chance of selecting the right class to a 1 in 6 chance, and this can be applied to the descriptors themselves. The Museum dataset also has a less diverse range of objects which allows a clearer distinction between classes. However this is the supplied data from the museums and this is the kind of situation in which these descriptors need to work.

7. CONCLUSIONS & FUTURE WORK

It has been shown that there are large differences in the performance of the shape descriptors and metrics depending on the dataset used. Overall it would seem that the best combination is the Cord Histogram and the city block distance metric. However other combinations have been shown to work better under particular circumstances, unfortunately benefiting from this requires prior knowledge of what you are looking for, and what is the best way to find it.

Further investigation is required as to what causes the Area Volume descriptor to perform so well against the Museum dataset.

Current and future work includes the development of more versatile shape descriptors, for example to facilitate sub-object matching, the use of the texture map and the possibility of matching between 2-D and 3-D representations. The findings indicate that for certain types of object, the best descriptor is not necessarily the one that performs best on average across all types. One area of further work is determining what descriptor and metric is best for which particular types of object.

ACKNOWLEDGEMENTS

The authors wish to thank: the European Commission for support through the SCULPTEUR project under grant IST-2001-35372. We would also like to thank our collaborators on the project, including Francis Schmitt and Tony Tung of ENST, Paris, Christian Lahanier of C2RMF, James Stevenson and Rachel Coates of the V&A museum, Joseph Padfield of the National Gallery, Raffaela Rimaboschi of the Uffizi and Jean-Pierre of the Musée de Cherbourg for many useful discussions, use of data and valuable help and advice; and Hewlett Packard's Art & Science programme for the donation of server equipment.

REFERENCES

1. S. Goodall, P. H. Lewis, K. Matrinez, P. A. S. Sinclair, F. Giorgini, M. J. Addis, M. J. Boniface, C. Lahanier, and J. Stevenson, "SCULPTEUR: Multimedia Retrieval for Museums," in *Image and Video Retrieval: Third International Conference, CIVR 2004)*, pp. 638–646, (Dublin, Ireland), July 2004.
2. P. H. Lewis, K. Martinez, F. S. Abas, M. F. A. Fauzi, S. C. Y. Chan, M. J. Addis, M. J. Boniface, P. Grimwood, A. Stevenson, C. Lahanier, and J. Stevenson, "An Integrated Content and Metadata Based Retrieval System for Art," *IEEE Transactions on Image Processing* **13**, pp. 302–313, March 2003.
3. N. Crofts, I. Dionissiadou, M. Doerr, and M. Stiff, "Definition of the CIDOC Object-Orientated Conceptual Reference Model, v.3.1," July 2001.
4. S. Kim, P. Lewis, K. Martinez, and S. Goodall, "Question Answering Towards Automatic Augmentations of Ontology Instances," in *Proceedings of the First European Semantic Web Symposium*, pp. 152–166, Springer, (Heraklion, Crete), May 2004.
5. P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser, "The Princeton Shape Benchmark," in *Shape Modeling International (SMI04)*, pp. 167–178, (Genova, Italy), June 2004.
6. T. Funkhouser, P. Min, M. Kazhdan, J. Chen, A. Halderman, D. Dobkin, and D. Jacobs, "A search engine for 3D models," *ACM Transactions on Graphics* **22**, pp. 83–105, Janurary 2003.
7. R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, "Matching 3D Models with Shape Distributions," in *Shape Modeling International*, pp. 154–166, (Genova, Italy), May 2001.
8. D.-Y. Chen, M. Ouhyoung, X.-P. Tian, and Y.-T. Shen, "On visual similarity based 3D model retrieval," *Computer Graphics Forum* , pp. 223–232, 2003.
9. E. Paquet and M. Rioux, "Nefertiti: a query by content system for three-dimensional model and image databases management," *Image and Vision Computing* **17**, pp. 157–166, February 1999.
10. T. Tung and F. Schmitt, "Augmented Reeb Graphs for Content-Based Retrieval of 3D Mesh Models," in *International COnference on Shape Modeling and Applications 2004*, pp. 157–166, (Genova, Italy), June 2004.
11. J. W. H. Tangelder and R. C. Veltkamp, "A Survey of Content Based 3D Shape Retrieval Methods," in *International Conference on Shape Modeling and Applications 2004*, pp. 145–156, (Genova, Italy), June 2004.

12. M. Ankerst, G. Kastenmüller, H.-P. Kriegel, and T. Seidl, "Nearest Neighbor classification in 3D protein databases.," in *7th International Conference on Intelligent Systems for Molecular Biology (ISMB'99)*, pp. 34–43, AAAI Press, (Heidelberg, Germany), 1999.
13. G. Hetzel, B. Leibe, P. Levi, and B. Schiele, "3D Object Recognition from Range Images using Local Feature Histograms," in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'01)*, **2**, pp. 394–399, (Kauai Island, Hawaii), December 2001.
14. K. Järvelin and J. Kekäläinen, "IR evaluation methods for retrieving highly relevant documents," in *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 41–48, ACM Press, (Athens, Greece), 2000.
15. C. J. van Rijsbergen, *Information Retrieval*, Butterworths, 1975.