A New Shape Benchmark for 3D Object Retrieval

Rui Fang[1], Afzal Godil[1], Xiaolan Li[1,2], and Asim Wagan[1]

[1] National Institute of Standards and Technology, Maryland, U.S.A
[2] Zhejiang Gongshang University, P. R. China
{rfang,godil,lixlan,wagan}@nist.gov

Abstract. Recently, content based 3D shape retrieval has been an active area of research. Benchmarking allows researchers to evaluate the quality of results of different 3D shape retrieval approaches. Here, we propose a new publicly available 3D shape benchmark to advance the state of art in 3D shape retrieval. We provide a review of previous and recent benchmarking efforts and then discuss some of the issues and problems involved in developing a benchmark. A detailed description of the new shape benchmark is provided including some of the salient features of this benchmark. In this benchmark, the 3D models are classified mainly according to visual shape similarity but in contrast to other benchmarks, the geometric structure of each model is modified and normalized, with each class in the benchmark sharing the equal number of models to reduce the possible bias in evaluation results. In the end we evaluate several representative algorithms for 3D shape searching on the new benchmark, and a comparison experiment between different shape benchmarks is also conducted to show the reliability of the new benchmark.

1 Introduction

With the increasing number of 3D models created and available on the Internet, many domains have their own 3D repositories such as the national design repository for CAD models [1], Protein Data Bank for biological macromolecules [2],CAESAR for Anthropometry [3], the AIM@SHAPE shape repository [4] and the INRIA-GAMMA 3D Database [5] for research purposes, etc.

Effectively searching a 3D repository for 3D shapes which are similar to a given 3D query model has become an important area of research. Traditional text based search engines are widely used in many 3D repositories to retrieve 3D shapes. Those text based search strategies only allow users to search 3D models by inputting keywords, file type, file size or by just browsing the thumbnail of 3D models, which may not meet all the demands. Because a text based method requires manually annotating the shapes which may introduce imprecision, it might not be able to properly represent the shape information. Furthermore, annotating a huge number of 3D models by hand is inefficient, inaccurate and impractical. In addition, the annotation of shapes is incomplete or not available in many cases. thus, a number of 3D shapes based search engines have been investigated to address this problem [6], [7], [8], [9], [10], etc.

G. Bebis et al. (Eds.): ISVC 2008, Part I, LNCS 5358, pp. 381–392, 2008.
© Springer-Verlag Berlin Heidelberg 2008

Shape based search methods do not require any manual annotation. The only thing they require is the shape descriptor which can automatically extract shape features and properly represent the shape information. It has been reported that the performances of shape based searching methods outperform text based methods [11]. For more about shape based searching algorithms, we refer readers to several surveys [12], [13], [14].

With a number of shape based retrieval methods appearing in the current literature, the question now is how to evaluate shape retrieval algorithms rationally with high confidence. Benchmarking is one answer to this question. Just like the methods used for evaluating text based [15], image based [16], [17], video based [18] and music based retrieval methods [19], [20], under a shape benchmark, different shape based retrieval algorithms are compared and evaluated in the same experimental environment by the same measurement tools from different aspects. Comparable results are then obtained and conclusions about their performance are drawn. By doing this, one then gets a good understanding of each algorithm and can judge which algorithm should be applied in a specific circumstance.

In section 2, the related work of previous benchmarks is briefly reviewed; in section 3, we discuss the construction of the benchmark; in section 4, we discuss the evaluation measures used; in section 5, the experiment results of different algorithms on the new benchmark are reported and analyzed; in section 6, the reliability of the new shape benchmark is discussed, and finally conclusions are drawn in section 7.

2 Related Work

The SHape REtrieval Contest (SHREC) [21] is organized every year since 2006 by Network of Excellence AIM@SHAPE to evaluate the effectiveness of 3D shape retrieval algorithms. Many tools are also provided to compare and evaluate 3D retrieval methods. In 2006, one track was conducted to retrieve 3D mesh models on the PSB [22]. From 2007, several tracks were conducted which focused on specialized problems: the watertight models track, the partial matching track, the CAD models track, the protein models track, the 3D face models track, the track on stability of watertight models, the track on the classification of watertight models and the generic models track [23], [21].

The Princeton Shape Benchmark (PSB) is a publicly available database of 3D polygonal models with a set of software tools that are widely used by researchers to report their shape matching results and compare them to the results of other algorithms [24].The Purdue engineering shape benchmark (ESB) [25] is a public 3D shape database for evaluating shape retrieval algorithms mainly in the mechanical engineering domain. The McGill 3D shape benchmark [26] provides a 3D shape repository which includes a number of models with articulating parts. Other current shape benchmarks were introduced and analyzed in [24].

Although previous shape benchmarks have provided valuable contributions to the evaluation of shape matching algorithms, there are some limitations to evaluate general purpose shape matching algorithms.

- Domain dependent benchmarks like the ESB and the McGill benchmark can only be used to evaluate retrieval algorithms in their respective domains.
- A number of classes in the PSB (basic classification) contain too few models. Actually, a benchmark database must have a reasonable number of models in each of its classes; five or ten is too few to get a statistically sound evaluation [27]. Take the base classification in the PSB as an example, in the training set there are 90 classes, 55 of which have no more than 10 models; 27 of which have no more than 5 models; in the testing set there are 92 classes, 59 of which have no more then 10 models; 28 of which have no more then 5 models.
- Unequal number of 3D models in each class could cause bias when using the benchmark to evaluate different shape retrieval algorithms. Some authors [28] reported their results on the PSB and concluded that the quality of the retrieval results depends on the size of the query model's class, and the higher the cardinality of a class the more significant the results will be (table 4 in [28]). Similar results were also reported in table 5 of [24], table 3 of [29]. Sometimes it is hard to decide which factor affects the results, the discriminative power of the algorithm on different model classes or the number of models in these classes. For example, suppose there are two algorithms A and B, and there are two kinds of 3D models: 5 birds and 30 cars; we want to evaluate the discrimination ability of the two algorithms; and further assume that algorithm A is good at discriminating birds and algorithm B is good at discriminating cars. Then the result of evaluation based on the given 3D models would favor B due to the unequal number of models in the two classes. Due to this reason, in SHREC 2007 watertight models track, all classes in the database were made up of the same number of objects to keep the generality of each query [23].
- In addition, models in some benchmarks have mesh errors like inconsistent normal, degenerated surfaces, duplicate surfaces, intersecting surfaces that do not enclose a volume, etc.

To overcome some limitations of the current benchmarks, we propose a new shape benchmark with a set of tools for direct comparison and analysis of general purpose shape matching algorithms.

3 Benchmark Design Principle

There are two main steps to benchmark a shape database, the first of which is to get enough 3D shape models. All the 3D models in the new shape benchmark were acquired by the web crawler Methabot [30] from major 3D repositories on the Internet. We have obtained permission to freely redistribute all these models only for research purposes. The other step is to classify the 3D shape models into a ground truth database; we discuss it below in detail.

As in the shape retrieval problem, we retrieve 3D objects solely according to their shape information other than their color, texture, etc., so in the benchmark, we use ASCII Object File Format (*.OFF) to represent the shape information of

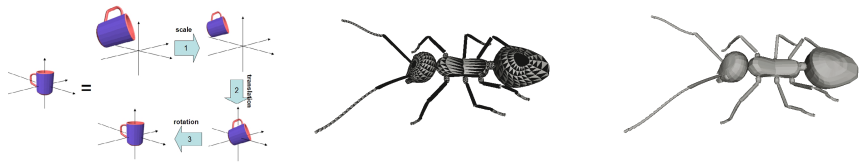


Fig. 1. Normalization procedure (left), a model with inconsistent normal orientation in some benchmarks (middle), and the same model with consistent normal orientation in the new benchmark (right)

each model, which consists of polygons, that is, the coordinates of points and the edges that connect these points. This shape format has benefits of simplicity and it only contains shape information, which allows us to concentrate on shape itself.

3D models downloaded from websites are in arbitrary position, scale and orientation, and some of them have many types of mesh errors [31]. Shapes should be invariant to rotation, translation and scaling, which require the process of pose normalization before many shape descriptors can be applied to extract shape features. Unfortunately, few previous shape benchmarks have done this simple but important step. For this purpose, in the benchmark database, every model is normalized: all the polygons are triangulated, scaled to the same size, translated to the center of mass, and rotated to the principle axes. Figure 1 (left side) gives an example of the normalization procedure. We partially solve some mesh errors like the inconsistent normal orientation problem. In figure 1, the model in the middle was taken from some benchmarks. It has inconsistent orientation of polygons on the same surface, (dark area and white area appear on the same surface). The model on the right side of figure 1 shows the model in our new shape benchmark with consistent orientation of polygons on the same surface.

3.1 Building a Ground Truth for the Benchmark

The purpose of benchmarking is to establish a known and validated ground truth to compare different shape matching algorithms and evaluate new methods by standard tools in a standard way. Building a ground truth database is an important step of establishing a benchmark. A good ground truth database should meet several criteria [32], like having a reasonable number of models, being stable in order to evaluate different methods with relatively high confidence, and having certain generalization ability to evaluate new methods.

To get a ground truth dataset, in text retrieval research, TREC [15] uses pooling assessment [32] . In image retrieval research, as there is no automatic way to determine the relevance of an image in the database for a given query image [33], the IAPR benchmark [16] was established by manually classifying images into categories. In image processing research, the Berkeley segmentation dataset and benchmark [34] assumes that the human segmented images provide valid ground truth boundaries, and all images are segmented and evaluated by a group of people. In shape retrieval research, the PSB manually partitioned the

Table 1. 40 classes used in the new Shape benchmark

1 Bird	2 Fish	3 NonFlyingInsect	4 FlyingInsect
5 Biped	6 Quadruped	7 ApartmentHouse	8 Skyscraper
9 SingleHouse	10 Bottle	11 Cup	12 Glasses
13 HandGun	14 SubmachineGun	15 MusicalInstrument	16 Mug
17 FloorLamp	18 DeskLamp	19 Sword	20 Cellphone
21 DeskPhone	22 Monitor	23 Bed	24 NonWheelChair
25 WheelChair	26 Sofa	27 RectangleTable	28 RoundTable
29 Bookshelf	30 HomePlant	31 Tree	32 Biplane
33 Helicopter	34 Monoplane	35 Rocket	36 Ship
37 Motorcycle	38 Car	39 MilitaryVehicle	40 Bicycle

models of the benchmark database primarily based on semantic and functional concepts and secondarily based on shape attributes.

As there is no standard measure of difference or similarity between two shapes, in the new shape benchmark, two researchers were assigned as assessors to manually classify objects into ground truth categories. When there are disagreements on which category some objects should belong, another researcher was assigned as the third assessor to make the final decision. This classification work is purely according to shape similarity, that is, geometric similarity and topology similarity. Each model was input to a 3D viewer, the assessor rendered it in several viewpoints to make a final judgment towards shape similarity.

The assessors were told to assume that, within the same class, the shape of the objects should be with high similarity to each other, while, between classes, the objects should have distinctive shape differences. In this benchmark, we equalize each class and make them contain the same numbers of 3D models (20 models). Here are two main reasons for this: to avoid possible bias of evaluation, as it is discussed in section 2; and some evaluation measures are unstable due to few models in a class [35].

Table 1 shows the 40 classes used in the new shape benchmark and each class contains 20 models which are common in daily life.

4 Evaluation Measures

The procedure of information retrieval evaluation is straightforward. In response to a given set of users' queries, an algorithm searches the benchmark database and returns an ordered list of responses called the ranked list(s). The evaluation of the algorithm then is transformed to the evaluation of the quality of the ranked list(s). As different evaluation metrics measure different aspects of shape retrieval behavior, in order to make a thorough evaluation of a 3D shape retrieval algorithm with high confidence, we employ a number of common evaluation measures used in the information retrieval community: Precision-Recall curve [36]; Average Precision(AP) and Mean Average Precision(MAP) [37]; E-Measures [36]; Cumulated gain based measurements [38]; Nearest Neighbor (NN), First-Tier (Tier1) and Second-Tier (Tier2) [24].

5 Comparison Experiments and Discussion

In this section, in order to examine how the benchmark reveals features of different shape matching algorithms, we include several kinds of algorithms to compare on the new benchmark. Moreover, comparison experiments are conducted on both the entire benchmark and a specific class of the benchmark to get comprehensive understanding of shape matching algorithms.

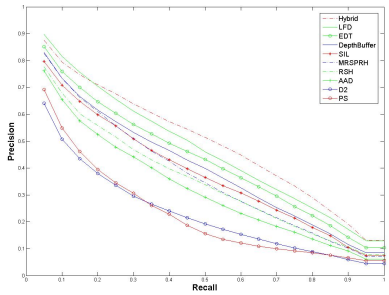
First we classify algorithms into several categories according to the kind of features the algorithms need and the way how these algorithms extract shape features.

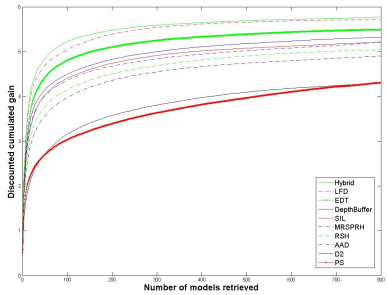
- View-based algorithm: Light Field Descriptor(LFD) [39]; Depth Buffer-Based (DepthBuffer) [40]; Silhouette-Based(SIL) [40].
- Statistic-based algorithm: D2 Shape Distributions(D2) [41]; AAD Shape Distributions (AAD) [42]; Linearly Parameterized Statistics(PS) [43].
- Ray-based algorithm: Ray-based with Spherical Harmonic representation (RSH) [40].
- Multiple descriptors: Hybrid Descriptors(Hybrid) [44]; Multiple Resolution Surflet-Pair-Relation Histograms(RSPRH) [45]; Exponentially decaying Euclidean distance transform(EDT) [40].

Now we perform comparison experiments on the whole benchmark using the evaluation measurements described in Section 5. Figure 2 shows the precision recall curves, the mean DCG curves and other measure score results of the 10 algorithms, respectively. Our main findings from the experiments are as follows:

- View-based methods have obtained considerable results on the whole benchmark. We conjecture the reasons for this are that: they extract high level features of a 3D shape and the ground truth database was established mainly according to visual similarity of shapes. The idea of view-based shape descriptors is that two similar shapes should look similar in different viewpoints. This corresponds directly with the way in which human beings judge shape similarity.
- Multiple descriptors also get very good results. In figure 2 (left), with the increasing recall value (bigger than 0.2), the performance of the hybrid descriptor outperforms the light field descriptor. We think the reason might be that the descriptor takes strength of each single shape descriptor, extracts and integrates the most important features at both high and low level.
- The performance of statistic-based methods is not so good compared to other kinds of algorithms. One of the possible reasons might be that, statistical features alone are not capable enough to discriminate shapes with high accuracy, and that these statistical features are relatively low level features. Therefore, new and more powerful statistic features should be explored and created to get better retrieval performance.

Comparison results on individual classes of the benchmark. In this subsection, we perform comparison experiments on every individual class given





	NN	Tier1	Tier2	E-Measure	DCG	MAP
1.Hybrid	81.13%	45.46%	57.72%	36.65%	76.09%	48.54%
2.LFD	84.50%	42.58%	54.03%	35.05%	75.40%	49.79%
3.EDT	77.50%	39.78%	52.09%	33.03%	72.37%	42.20%
4.DepthBuffer	75.88%	37.37%	47.96%	31.15%	70.08%	38.46%
5.SIL	71.00%	35.54%	47.98%	30.06%	68.57%	37.16%
6.MRSPRH	70.00%	35.15%	45.99%	29.98%	68.77%	36.38%
7.RSH	69.13%	32.83%	44.11%	28.18%	66.43%	34.53%
8.AAD	65.00%	30.61%	41.69%	26.67%	64.58%	31.59%
9.D2	49.38%	22.33%	32.08%	20.56%	56.58%	22.53%
10.PS	50.00%	21.79%	29.40%	20.51%	56.73%	22.38%

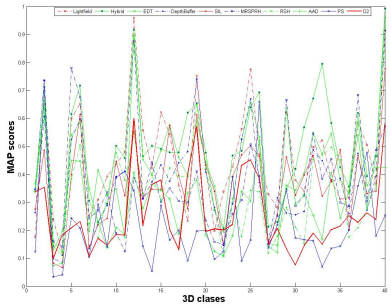


Fig. 2. The precision-recall curve on whole benchmark (top left), the mean DCG curve on whole benchmark (top right), other comparison results on the whole benchmark (bottom left) and the MAP scores of the evaluated algorithms on each individual class of the proposed benchmark (bottom right)

in Table 1, and explore how different algorithms perform on a specific kind of 3D objects. The Mean Average Precision (MAP) is used to evaluate the performance of each algorithm on the 40 individual classes of the benchmark. Figure 2(bottom right) shows the MAP evaluation results of the ten algorithms on each individual shape class of the new benchmark. Our main findings from the experiments on individual classes are as follows:

- Most algorithms do especially well in geometrically simple structure models, for example, Fish, Glasses, and Sword.
- The performance of different shape matching algorithms on a specific class can vary a lot. Also, a shape matching algorithm will get quite different evaluation scores on different classes. The reason might be that, due to the characteristics of a shape matching algorithm, it is easy for the shape matching algorithm to extract the feature for certain classes of objects, while this is not easy for some other classes of objects.

So, besides comparing the results on the whole benchmark, it is helpful to further discover the nature of a shape matching algorithm on each single class.

6 The Reliability of the New Shape Benchmark

In this section, we explore the reliability of the new proposed benchmark by testing the effect of class set size on retrieval error. Voorhees and Buckley [46] proposed a method to estimate the reliability of retrieval experiments by computing the probability of making wrong decisions between two retrieval systems over two retrieval experiments. They also showed how the topic set sizes affect the reliability of retrieval experiments. A theoretical justification of this method has been discussed by Lin and Hauptmann [47].

We use this method to conduct an experiment and test the reliability of the new shape benchmark. The procedure of computing error rates was described in [46]. Here we summarize the procedure as follows:

- Randomly split the benchmark database into two equal sets, each set containing 20 classes.
- Use the MAP score of the 10 algorithms described in the last section, so that the number of algorithm pairs run in the benchmark is $C_{10}^2 = 45$. Randomly take several classes (from 2 classes up to 20) each time from one set and compute the MAP of every algorithm on these chosen classes. The same procedure is done on the other set.
- The difference between two algorithms' MAP is categorized into one of 19 bins ranging from 0.00 to 0.20 with 0.01 increment (i.e. $(0.00, 0.01]$, $(0.00, 0.01]$,...,$(0.19, 0.20]$).
- Based on step 2, count the number of swaps for all pairs of algorithms and the number of MAP differences on that bin. For example, a swap occurs when in one set the MAP of algorithm A is bigger than that of B, while in the other set, the MAP of A is smaller than that of B. Count it when the size of differences between two algorithms' MAP is in a certain bin.
- Compute error rate at each bin by dividing the number of swaps by the total number of MAP difference on that bin.
- For each bin of MAP difference, draw the error rates for the number of classes from 1 to 20, go over every bin and repeat the procedure.
- Estimate the error rates of the whole benchmark by extrapolating the error rates using the number of classes up to half of the benchmark.

Figure 3 (top left) shows the estimated error rates with the number of classes up to 40. From this figure, we can see that: for each curve, the error rate decreases with the increasing class set size, and for each class set size, the error rate decreases with the increasing size of difference of MAP scores. By looking at curves whose error rates do not exceed a certain value (e.g. 5%), we can estimate how much MAP difference is needed to conclude that retrieval algorithm A is better than retrieval algorithm B with 95% confidence. When the class set size increases from 15 to 40, the error rates are stably converging to zero. Especially, when the number of class is 40, it needs less than 0.03 difference in MAP scores to guarantee that the error rate is below 5%.

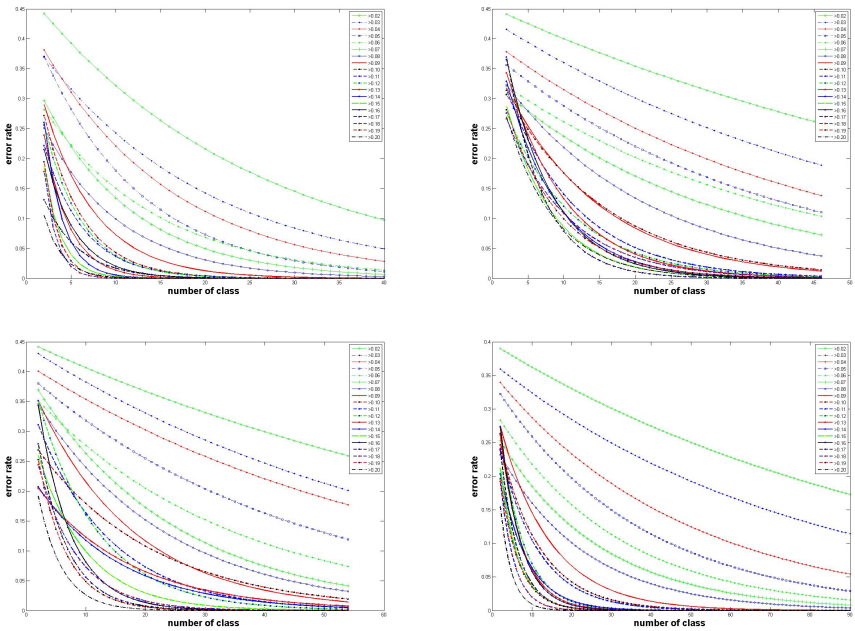


Fig. 3. Extrapolated error rates vs. class set sizes up to the whole benchmarks

We also conduct the same experiments on other different shape benchmarks: the base training database in PSB(90 classes) (Figure 3 (bottom right)) [6], the CCCC shape benchmark(55 classes) (Figure 3 (bottom left)) [7], and the National Taiwan University(NTU)'s shape benchmark (47 classified classes) (Figure 3 (top right)) [8]. We do not include the Utrecht University's shape benchmark because it contains too few (five) classified classes in the databases.

From the experiment results on other shape benchmarks, we can see that: to guarantee that the error rate is below 5% on the whole database, the NTU shape benchmark needs at least 0.07 difference in MAP scores, the CCCC shape benchmark needs at least 0.06 difference in MAP scores, and the Princeton shape benchmark needs at least 0.04 difference in MAP scores. At class set size of 40, which is also class set size of the new proposed shape benchmark, to guarantee that the error rate is below 5%, the NTU shape benchmark needs at least 0.07 difference in MAP scores, the CCCC shape benchmark needs at least 0.08 difference in MAP scores, and the PSB needs at least 0.07 difference in MAP scores. This indicates that the number of classes of the new proposed shape benchmark is large enough to get a sufficiently low error rate compared to other shape benchmarks.

From the above discussion, the experiment can be considered as a strong evidence to support the reliability of the new benchmark to evaluate shape retrieval algorithms.

7 Conclusion

We have established a new publicly available 3D shape benchmark with a suite of standard tools for evaluating generic purpose shape retrieval algorithms (the benchmark website is http://www.itl.nist.gov/iad/vug/sharp/benchmark). Several retrieval algorithms are evaluated from several aspects on this new benchmark by various measurements, and the reliability of the new shape benchmark is discussed. This reliable shape benchmark provides a new perspective to evaluate shape retrieval algorithms. It has several merits: high reliability (in terms of error rate) to evaluate 3D shape retrieval algorithms, sufficient number of good quality models as the basis of the shape benchmark, equal size of classes to minimize the bias of evaluation.

References

1. http://www.designrepository.org/
2. http://www.rcsb.org/pdb/home
3. http://store.sae.org/caesar/
4. http://shapes.aim-at-shape.net/
5. http://www-c.inria.fr/gamma/gamma.php
6. http://shape.cs.princeton.edu/search.html
7. http://merkur01.inf.uni-konstanz.de/CCCC/
8. http://3d.csie.ntu.edu.tw/
9. http://www.cs.uu.nl/centers/give/multimedia/3Drecog/3Dmatching.html
10. http://3d-search.iti.gr/3DSearch
11. Min, P., Kazhdan, M., Funkhouser, T.: A comparison of text and shape matching for retrieval of online 3d models. In: Proc. European conference on digital libraries, pp. 209–220 (2004)
12. Tangelder, J.W.H., Veltkamp, R.C.: A survey of content based 3d shape retrieval methods. In: Proceedings of the Shape Modeling International, pp. 145–156 (2004)
13. Bustos, B., Keim, A.D., Saupe, D., Schreck, T., Vranic, V.D.: Feature-based similarity search in 3d object databases. ACM Computing Surveys 37(4), 345–387 (2005)
14. Iyer, N., Jayanti, S., Lou, K., Kalyanaraman, Y., Ramani, K.: Three-dimensional shape searching: state-of-the-art review and future trends. Computer-Aided Design 37(5), 509–530 (2005)
15. http://trec.nist.gov/
16. http://eureka.vu.edu.au/~grubinger/IAPR/TC12_Benchmark.html
17. http://www.imageclef.org/
18. Smeaton, A.F., Over, P., Kraaij, W.: Evaluation campaigns and trecvid. In: Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval, pp. 321–330 (2006)
19. Orio, N.: Music retrieval: A tutorial and review. Foundations and Trends in Information Retrieval 1, 1–90 (2006)
20. Typke, R., Wiering, F., Veltkamp, R.C.: A survey of music information retrieval systems. In: Proceedings of the 6th International Conference on Music Information Retrieval, pp. 153–160 (2005)
21. http://www.aimatshape.net/event/SHREC

22. Veltkamp, R.C., Ruijsenaars, R., Spagnuolo, M., van Zwol, R., ter Haar, F.: Shrec2006 3d shape retrieval contest, technical report uu-cs-2006-030. Technical report, Department of Information and Computing Science, Utrecht University (2006)
23. Veltkamp, R.C., ter Haar, F.B.: Shrec2007 3d shape retrieval contest, technical report uu-cs-2007-015. Technical report, Department of Information and Computing Science, Utrecht University (2007)
24. Shilane, P., Min, P., Kazhdan, M., Funkhouser, T.: The princeton shape benchmark. In: Proceedings of the Shape Modeling International, pp. 167–178 (2004)
25. Iyer, N., Jayanti, S., Lou, K., Kalyanaraman, Y., Ramani, K.: Developing an engineering shape benchmark for cad models. Computer-Aided Design 38(9), 939–953 (2006)
26. Zhang, J., Siddiqi, K., Macrini, D., Shokouf, A., Dickinson, S.: Retrieving articulated 3-d models using medial surfaces and their graph spectra. In: International Workshop On Energy Minimization Methods in Computer Vision and Pattern Recognition, pp. 285–300 (2005)
27. Voorhees, E.M.: Variations in relevance judgments and the measurement of retrieval effectiveness. In: Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval, pp. 315–323 (1998)
28. Papadakis, P., Pratikakis, I., Perantonis, S., Theoharis, T.: Efficient 3d shape matching and retrieval using a concrete radialized spherical projection representation. Pattern Recognition 40, 2437–2452 (2007)
29. Laga, H., Takahashi, H., Nakajima, M.: Spherical wavelet descriptors for content-based 3d model retrieval. In: Proceedings of the IEEE International Conference on Shape Modeling and Applications, pp. 15 (2006)
30. `http://bithack.se/methabot/start`
31. Veleba, D., Felkel, P.: Detection and correction of errors in surface representation. In: The 15th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (2007)
32. Jones, S., Van Rijsbergen, C.J.: Information retrieval test collections. Journal of Documentation 32, 59–75 (1976)
33. Leung, C.: Benchmarking for content-based visual information search. In: Laurini, R. (ed.) VISUAL 2000. LNCS, vol. 1929, pp. 442–456. Springer, Heidelberg (2000)
34. `http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/`
35. Buckley, C., Voorhees, E.M.: Evaluating evaluation measure stability. In: Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 33–40 (2000)
36. Rijsbergen, C.J.V.: Information Retrieval, 2nd edn. Butterworths (1979)
37. Salton, G., McGill, M.J.: Introduction to Modern Information Retrieval. McGraw Hill Book Co., New York (1983)
38. Jarvelin, K., Kekalainen, J.: 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 (2000)
39. Chen, D., Tian, X., Shen, Y., Ouhyoung, M.: On Visual Similarity Based 3 D Model Retrieval. Computer Graphics Forum 22, 223–232 (2003)
40. Vranic, D.V.: 3d model retrieval. PH.D thesis, University of Leipzig, German (2004)
41. Osada, R., Funkhouser, T., Chazelle, B., Dobkin, D.: Shape distributions. ACM Transactions on Graphics (TOG) 21, 807–832 (2002)

42. Ohbuchi, R., Minamitani, T., Takei, T.: Shape-similarity search of 3d models by using enhanced shape functions. International Journal of Computer Applications in Technology 23, 70–85 (2005)
43. Ohbuchi, R., Otagiri, T., Ibato, M., Takei, T.: Shape-similarity search of three-dimensional models using parameterized statistics. In: Proceedings of the 10th Pacific Conference on Computer Graphics and Applications, pp. 265–274 (2002)
44. Vranic, D.V.: Desire: a composite 3d-shape descriptor. In: Proceedings of the IEEE International Conference on Multimedia and Expo. (2005)
45. Ohbuchi, R., Hata, Y.: Combining multiresolution shape descriptors for 3d model retrieval. In: Proc. WSCG 2006 (2006)
46. Voorhees, E.M., Buckley, C.: The effect of topic set size on retrieval experiment error. In: Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 316–323 (2002)
47. Lin, W.H., Hauptmann, A.: Revisiting the effect of topic set size on retrieval error. In: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 637–638 (2005)