Eurographics Workshop on 3D Object Retrieval (2015)
I. Pratikakis, M. Spagnuolo, T. Theoharis, L. Van Gool, and R. Veltkamp (Editors)

RETRIEVAL^{3D}: An On-line Content-Based Retrieval Performance Evaluation Tool

A. Koutsoudis[1], G. Ioannakis[2], I. Pratikakis[2] and Ch. Chamzas[2]

[1] Athena Research and Innovation Centre, Xanthi's Division, Multimedia Research Group, PO BOX 159, Xanthi 67100, Greece
[2] Democritus University of Thrace, Department of Electrical and Computer Engineering, Xanthi 67100, Greece

Abstract

Performance benchmarking is an absolute necessity when attempting to objectively quantify the performance of content-based retrieval methods. For many years now, a number of plot-based and scalar-based measures in combination with benchmark datasets have already been used in order to provide objective results. In this work, we present the first version of an integrated on-line content-based retrieval evaluation tool, named RETRIEVAL^{3D}, which can be used in order to quantify the performance of a retrieval method. The current version of the system offers a set of popular performance measures that can be accessed through a dynamic visualisation environment. The user is able to upload retrieval results using different input data structures (e.g. binary ranked lists, floating point ranked lists, dissimilarity matrices and groundtruth data) that are already encountered in the literature including the SHREC competition series. Moreover, the system is able to provide evaluation mechanisms for known within the retrieval research community benchmark datasets. It offers performance measures parameterisation that enables the user to determine specific aspects of the evaluated retrieval method. Performance reports archiving and downloading are some of the system's user-oriented functionalities.

Categories and Subject Descriptors (according to ACM CCS): H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval, I.3.5 [Computational Geometry and Object Modelling]: Geometric algorithms, languages, and systems.

1. Introduction

For more than fifteen years now we experience a continuous growth of multimedia datasets. This dictates the creation of efficient indexing, searching and retrieval mechanisms. The development of content-based retrieval (CBR) methods has become a very active research area as it provides answers to such needs. In fact, CBR has significant meaning when applied on large datasets were textual annotations are either not available or proved to be insufficient in describing the data content or to identify similarities on geometrical, structural and colour/texture levels.

The performance evaluation of CBR methods is an important task and according to literature it relies on procedures that offer unbiased performance measurements of their retrieval abilities in terms of precision, consistence, efficiency and robustness. As CBR can be used in a wide range of applications it is a fact that different requirements specifications in terms of shape matching accuracy and relevance discrimination are required. Still, a number of *plot*- and *scalar*-based measures are considered common while the assessment of their numeric values relies on the researcher and the requirements of each application domain. Furthermore, retrieval performance contests such as the SHape REtrieval Contest (SHREC) attempts for almost a decade now to appoint the best performing CBR methods on different retrieval challenges using a specific set of performance measures [SHREC06]. Moreover, the performance of different CBR methods relies on computing the performance measures over the same benchmark dataset using the same queries. Thus, a number of benchmarking datasets have been presented in the literature where some of them offer supplemental software tools (executable files) that implement performance measures to help researchers efficiently derive performance results of their CBR methods. An overview of the retrieval evaluation measures along with an extensive list of 3D benchmarking datasets are presented in [KPC13].

In this work, we present RETRIEVAL^{3D}; an online performance evaluation tool that provides a unified platform

DOI: 10.2312/3dor.20151062

to produce performance evaluation reports of CBR methods that have been applied on custom or known benchmarking datasets. We discuss on the functionalities of the current version of tool stating the various input data structures used along with the dynamic performance results visualisation environment and its reports export functionalities. We mention the performance measurements being implemented and the ability offered to the user to parameterise them according to his/her needs. We conclude by outlining the importance aspects of such a tool and the near future developments such as the installation of additional benchmark datasets found in the literature [KPC13].

2. RETRIEVAL[3D]: Current Tool Functionality

RETRIEVAL[3D] is a Web-based unified on-line platform that attempts to provide all the tools required for the evaluation process of a CBR method. The system has been implemented using open technologies such as PHP, JQuery, AJAX and MySQL.

System access is performed through a user registration mechanism. The basic operational pipeline behind the system is organised into three major steps that are required so that the tool produces the performance evaluation reports.

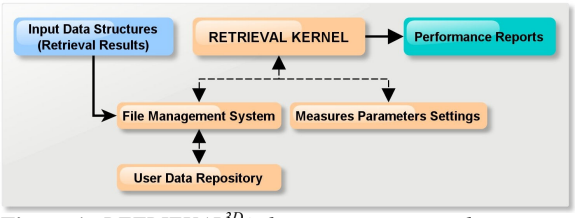

Figure 1: *RETRIEVAL[3D]'s basic operation pipeline.*

Initially, the system requires the uploading of retrieval results organised in some of the most commonly used, by the research community, data structures. These are: *i)* Binary Ranked List (BRL) *ii)* Floating Point Ranked List (FPRL) *iii)* Dissimilarity Matrix (DM). The latter two cases (FPRL, DM) should be coupled by the corresponding ground truth data. Once the input files are made available to the system, the user may alter the various parameters of the performance evaluation measures offered by the system in order to meet the needs of the used dataset and the requirements of the application domain. The system stores for each user the performance evaluation measures parameters in his/her profile. Additionally, the system offers detailed textual information about each measure and the role played by each of the parameters. Furthermore, the user may require to produce the performance evaluation reports that will be available either *on-line* for visualisation through dynamic plots and tables or *off-line* stored in various file types so that they can be further processed. The system stores all the uploaded data and through a file management system allows operations such as downloading, archiving and deletion. Through the same file management system the user can handle the produced performance reports. More specifically, the reports are available either as CSV files or archived into a ZIP file.

2.1 Input Data Structures

Performance evaluation of a CBR method requires retrieval results. These results should be organised into one of the data structures that can be currently handled by RETRIEVAL[3D].

i) The *Binary Ranked List* (BRL) is the simplest data structure supported by the system and describes the binary relevance classification of all objects in a ranked list. More specifically, the objects are ranked according to their similarity value in an ascending order and afterwards according to a classification file, each object is associated to *0 (non-relevant)* or *1(relevant)* in relation always to the query object. An example of the BRL data structure is shown in Table 1.

1.obj	*110010*	*6*	*3*
2.obj	*110100*	*6*	*3*
:	*:*	*:*	*:*
n.obj	*111000*	*6*	*3*

Table 1: *A Binary Ranked List (BRL) data structure.*

The first column refers to the query object, the second is the actual binary ranked list and the last two values represent the total number of objects in the dataset and the number of objects that are relevant to the query object (e.g. class cardinality). Thus, a complete class of a custom dataset can be included in a BRL file and each line can represent the results produced by a specific query file. The system will automatically produce averages of all performance measures based on the data found in a BRL file.

ii) The *Floating Point Ranked List* (FPRL) data structure contains the similarity distances of a query object in relation with all the objects in a dataset. The similarity distance depicts how similar the query object is in relation to another object in a given dataset, assuming the dataset consists of six objects the FPRL file will be as shown in Table 2.

1.obj	*0.010*	*4.430*	*2.130*	*3.124*	*6.345*	*11.223*

Table 2: *A FPRL data structure.*

The FPRL data structure is always combined with a *groundtruth* data structure. This groundtruth will provide the required classification information to the system in order to quantify the relevance of each object (binary classification). The system provides a detailed textual description of the groundtruth data structure as well as downloadable template files. Again, the system will produce averages of all performance measures based on the data found in a FPRL file.

iii) The *Dissimilarity Matrix (DM)* data structure expresses the similarity values pairwise all the possible combinations of objects that exist in a dataset. It is a square and symmet-

ric two-dimensional array with zeros on its diagonal. These zeroes depict the perfect similarity when comparing an object with itself. Again, each row depicts the similarity values of a query object in relation with the rest of a dataset (Table 3). Similar to the previous case, the DM data structure requires a groundtruth that defines the object classes. Downloadable sample template files are offered by the system.

	1.obj	*2.obj*	*3.obj*	*n.obj*
1.obj	***0***	*0.15*	*7.75*	*2.78*
2.obj	*0.01*	***0***	*0.60*	*6.43*
3.obj	*7.00*	*0.65*	***0***	*1.14*
:	*:*	*:*	*:*	*:*
n.obj	*4.52*	*1.23*	*0.87*	***0***

Table 3: *Dissimilarity matrix data structure example.*

2.2 Benchmarking Measures

The system provides a set of plot and scalar based measures that have been used by the research community [KPC13][SHREC06]. The measures are considered complementary and attempt to capture important and distinct aspects of a CBR method's performance. All measures are used to quantify the performance of a CBR method over a given benchmark dataset using the same query set. RETRIEVAL^{3D} uses the FLOT library [FLOT07], in order to visualize the plot-based measures.

The proposed system is capable of calculating the following plots: (*i)* Precision-Recall, (*ii)* Precision, (*iii)* Recall, (*iv)* F-Measure, (*v)* E-Measure, (*vi)* Binary Relevance Ranked List Scatter Plot and (*vii)* Thumbnails-based Ranked List (where thumbnails are available). All of them are used to visualise the retrieval performance of a CBR method in all the available ranking positions. Figure 2 illustrates a screenshot of a precision-recall plot.

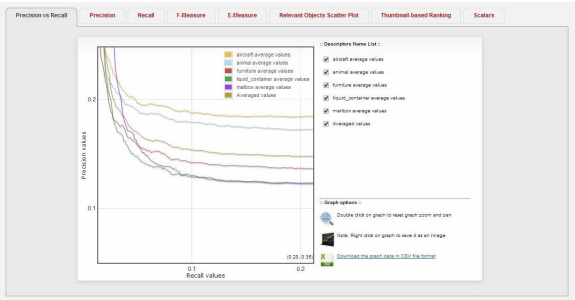

Figure 2: *A Precision-Recall plot example.*

The user can zoom in/out, hide/show an individual curve, and acquire values from specific points on a curve. Plots can also be saved as bitmap images. Additionally, the user can select among his uploaded dataset and produce combined performance plots. The system will perform the report generation even if this combination has is not reasonable. Hence, the user should select the appropriate data when requesting combined performance reports. On the other hand, the system also calculates scalars such as: (*i)* Nearest Neighbour (NN), (*ii)* Next to Nearest Neighbour (NN+1), (*iii)* 1st Tier, (*iv)* 2nd Tier, (*v)* Last Place Ranking, (*vi)* Average Precision (AP), (v*ii)* Average Dynamic Call, (v*iii)* Cumulative Gain (CG), (*ix)* Discount Cumulative Gain (DCG),(*ix)* Normalised Discount Cumulative Gain (NCG), (*x)* Fall-out rate, (*xi)* F-Measure, Weighted F-Measure and (*xii)* E-Measure at a given ranked list position. Figure 3 depicts a screenshot of a report with scalar measures along with maximum, minimum and averaged values.

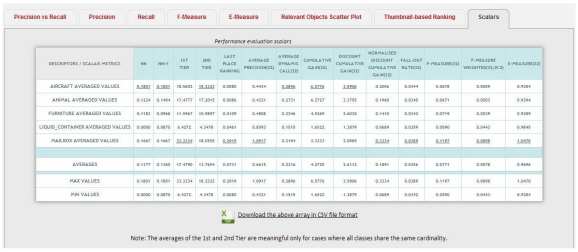

Figure 3: *Scalars performance table report.*

2.2 Benchmarking Datasets Management

In order to provide a unified performance evaluation tool it is important to integrate benchmarking datasets that are found in the literature.

As the system is still under development, there is currently only one benchmark dataset fully installed. We have selected the Princeton Shape Benchmark (PSB) [SMKF04] as the first to be installed due to the fact that it provides an extended (up to four levels) object classification scheme. All other benchmark datasets of similar or simpler classification schemes can now be installed efficiently using or altering parts of the existing source code. The system is designed in a way that each one of the predefined benchmark datasets will have its own handling kernel and front-end according to its specification and user input requirements (Figure 4). In the case of PSB, the system allows the user to upload a dissimilarity matrix (*907x907*) for the whole dataset. Then the user can select which of the classes (subclasses or all the classes) will be evaluated along with the classification level on which the binary relevance comparison will be performed. Then a batch of such jobs can be assigned to the system in order for it to create the performance reports based on these selections. A logging monitor in combination with a pop-up messaging mechanism informs the user about the progress of the assigned jobs. Currently, there is a limit of the total number of jobs (currently <=*5*) that can be assigned each time to the system. Nevertheless, in the case of PSB, the system offers an option for an overall performance report using the first classification level (e.g. 35 classes).

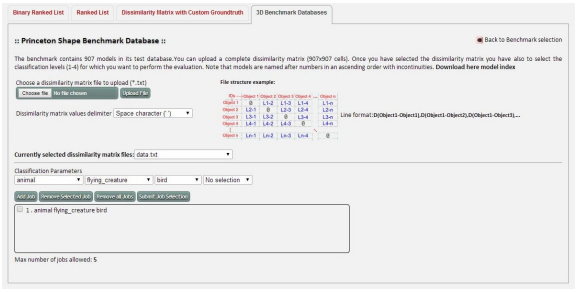

Figure 4: *The front-end of the PSB benchmark.*

When using PSB or any future installed benchmark datasets, the system offers the Thumbnails-based Ranked List visualisation approach in order for the user to monitor the performance of his/her CBR method and to determine any outliers or other performance issues related to specific objects. Figure 6 shows a thumbnails-based ranked list visualisation of the PSB mailbox class along with the retrieval results produce by querying with a specific object. The green and red framing around each thumbnail indicate the positive and negative relevance of each object in relation to the query object.

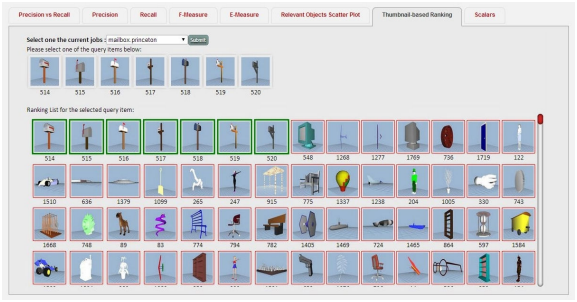

Figure 5: *The retrieved items ranking list for a specific query item. Border colour indicates each object's relevance (Green is relevant and Red is non- relevant).*

3. Conclusions

In this work, we have presented the first version of an on-line performance evaluation tool for content-based retrieval methods. This is an attempt to provide a common evaluation platform that can be used by the research community. The development of the system is based on the idea to provide performance evaluation functionalities for both custom and known benchmark datasets. The system is temporarily hosted in a server that can be accessed from http://retrieval.ceti.gr.

Currently, we are gathering test data in order to validate the system's calculations black-box testing. Additionally, we are working on enriching the functionalities of the system by integrating other benchmark datasets found in the literature while enriching the performance reports with additional sections. Also, we are working towards the optimisation of the recursive and computationally expensive functions such plot and scalar calculations in order to improve the system's responsiveness for the performance reports generation. Currently, large jobs (e.g. complete PSB dissimilarity matrices *907x907* objects) require long computational times (approx. *15* minutes on this server) that are not common for Web-based systems where the user expects almost instant interaction.

Acknowledgements

This research has been co-financed by the European Union (European Social Fund - ESF) and Greek National Funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALES (MIS 379516) Investing in knowledge society through the European Social Fund.

References

[FLOT07] FLOT Javascript plotting library for jQuery. http://www.flotcharts.org, Accessed 19 February 2015.

[KPC13] KOUTSOUDIS A., PRATIKAKIS I., CHAMZAS C.: On 3D Object Retrieval Benchmarking. *3D Research* 4, 3 (Oct. 2013), 2-11. doi: 10.1007/3DRes.04(2013)3.

[SHREC06] Shape Retrieval Contest (SHREC) (2006). http://www.aimatshape.net, Accessed 19 February 2015.

[SMKF04] SHILANE P., MIN P., KAZHDAN M., FUNKHOUSER T.: The Princeton Shape Benchmark. *Shape Modeling International*, (Jun. 2004), C167-C178, http://shape.cs.princeton.edu/benchmark/benchmark.pdf. doi: 10.1109/SMI.2004.1314504.