Eurographics Workshop on 3D Object Retrieval (2008)
I. Pratikakis and T. Theoharis (Editors)

On-line and open platform for 3D object retrieval

B. Le Bonhomme[1], B. Mustafa[2], S. Celakovsky[2], M. Preda[1], F. Prêteux[1] and D. Davcev[2]

[1]ARTEMIS Department, Institut TELECOM, France
[2]Faculty of Electrical Engineering, University Sts Cyril and Metodius, Karpos II bb, the former Yugoslav Republic of Macedonia

Abstract

In this paper we present the MyMultimediaWorld Internet-based platform designed to benchmark descriptors and description scheme for 3D object retrieval purpose. Relying on the MPEG-4 and MPEG-7 multimedia standards for data representation and description respectively, this open platform is designed to host multiple datasets, descriptors, descriptor extraction algorithms and similarity measures. We implemented an easy-to-use API designed to make the integration of the 3D object retrieval technology of third-party researchers agnostic to and independent of the global system complexity. Benchmarking results are automatically updated accordingly and presented qualitatively by displaying the 3D retrieved objects and quantitatively by providing the estimates of the state-of-the art performance criteria.

1. Introduction

Developments of graphics hardware and their adoption in personal computers and, now in mobile devices, together with the massive deployment of games, pushed the 3D graphics on similar levels as the traditional video and audio media. Contributors from academia and industry continuously provide novel solutions addressing all the elements in the media chain, including production (3D scanners and authoring tools), transmission (compression and streaming tools) and consumption (dedicated 3D rendering hardware and software solution). The increasing availability of 3D graphics contents originates nowadays a challenging interest in terms of automated indexing and retrieval for large 3D databases.

In the area of the Multimedia Information Retrieval (MIR), the main goals are to retrieve the content, to browse and to summarize the multimedia set with a human-centred approach. In [LSDJ06], different issues of MIR are reported: (1) feature extraction - such as the ones specified by MPEG-7 [MSS02] - and similarity measurements; (2) the ability of systems to learn and classify by exploiting relevance feedback mechanisms; (3) multimedia information browsing and summarization, exemplified by such systems as Muvis [KCG*03], VideoQ [CCS98], Photobook [PPS96], Visualseek [SC96], Calif & Emir [LBK03], SCHEMA [MDH*05]; (4) performance assessment, including the creation of large and representative datasets.

In this paper, we propose a MIR platform able to support issues (1) and (3) and bring an original contribution for addressing (4).

Currently, several benchmarking strategies are used by the research community. The most powerful (but also the most time-and resources-consuming) one is the organisation of dedicated conferences, such as TRECVID [SOK06], or of Core Experiments as in the MPEG consortium. Another typical example, dedicated to 3D graphics description, is SHREC [VH07] where participants evaluate their 3D object retrieval system using the same database and agreeing to the set of tasks.

The second strategy is to use a common database and test set. For 3D object description, the reference database is Princeton Shape Benchmark [SKMF04]. Another example related to this approach is the Berkeley Segmentation Dataset and Benchmark (`http://www.cs.berkeley.edu/projects/vision/grouping/segbench/`) for evaluating image segmentation algorithms. In [LE05], a rich review of benchmarks for Content Base Image Retrieval (CBIR) on large databases is provided. For video indexing, the reference data set is TRECVID and for text retrieval the TREC [GG07].

Another approach is a web-based benchmark such as the VIPER network [MMMMP00] for image retrieval evaluation. In this approach, MIR developers used the Multimedia Retrieval Mark-up Language (MRML) [MMMM*00] to communicate with the benchmark server. Once the MIR client downloads the image, the benchmark server sends queries for performance assessment.

The above-mentioned benchmark approaches are confronted with various limitations. In the off-line scenarios, it is almost impossible to allow the update of the database. Any update would imply the non-validity of the former results and the benchmark should be repeated. In the on-line scenario; the database can be updated. However, the previously-mentioned method requires downloading the media, therefore increasing communications between the benchmark server and the MIR client. In addition, since the program is executed on client, it is not possible to compare execution time or CPU/memory usage due to the heterogeneity of computing architectures.

In this paper, we propose an on-line platform called MMW.com (available at http://www.MyMultimediaWorld.com) which addresses the previously-mentioned limitations. In Section 2, we introduce the MMW.com features and describe the original approach for on-line benchmarking. In Section 3, we demonstrate the approach by implementing three different descriptors for 3D graphics description, before presenting the benchmark in Section 4. Concluding remarks and perspectives of this works are finally discussed.

2. MMW.com

The main goal of MMW.com is to be an on-line platform for sharing different media types, including video, image, audio and 3D graphics. This website federates a community around MPEG standards, by using MPEG-4 for media representation and MPEG-7 for media description. Additionally, the platform supports Web2.0 features and an interface allowing to enrich its functionalities. In this paper, we present only the interface for 3D graphics benchmarking.

2.1. Database description

In MMW.com, the database contains all media types including video, image, audio and 3D graphics. In the following, database organization is explained for 3D objects, other media types are organized similarly.

In order to establish an exhaustive classification of 3D objects; we built a hierarchical tree of categories. The basis of this classification was created starting from three existing classifications for 3D objects:

- The Princeton 3D database (http://shape.cs.princeton.edu/benchmark/) classification, containing more than 1800 objects and including 90 hierarchical categories;
- The MPEG 3D database (http://www.gti.ssr.upm.es/~mpeg/3dgc/), containing about 1400 objects classified into 21 flat categories;
- The 3D object database published by Taschen (http://www.taschen.com/), containing 1000 objects and classified into 20 categories.

Based on these categories, we created a hierarchical tree. In addition to the three above-introduced classifications, we considered a motion type feature: static, animated and motion-captured. These types are then further described in semantic categories, allowing for a fine 3D object classification. For example, for a static 3D object representing a hand, MMW.com specifies the following category hierarchy: *Static - Humans - Body_part - Hand.*

2.2. Multimedia description with MPEG-7 extended

In order to ensure a unified management of the media, we define for each MPEG-4 object an extended MPEG-7 description. We have created a specific XML Schema structure which contains a subset of the MPEG-7 elements and introduces other elements specific to the MMW.com platform. In order to specify the structure of a document on MMW.com as an extension of the MPEG-7 Schema, we have defined an extension of the MPEG-7 Schema; this extension allows for describing additional functionalities such as query-by-example description, service information, compression results, and additional parameters inspired from [LPP07]. The high-levelarchitecture of MMW.com is illustrated in Figure 1.

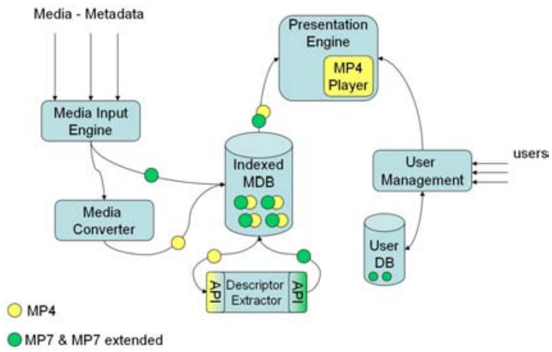

Figure 1: *High-Level architecture of MMW.com.*

2.3. Web2.0 Features

2.3.1. Browsing

The main page introduces the major navigation functionalities, structured in four components: Home, Contents, Upload and Software. By default, the Contents tab is activated. It allows navigation with respect to the medias types (video, image, audio and 3D graphics). For each sub-tab, the categories are presented on the left side and the icons for medias on the right side. The user can change the category, or select a media to be visualized. In anonymous mode, only database browsing is enabled.

2.3.2. Object Visualization

This page calls a plug-in (previously installed and available in the Software page) instantiating an MPEG-4 player rendering 3D object, video, image and sound. The player works on Windows, and is compatible for Internet Explorer and Firefox. On the left side of the page, metadata are presented: title, upload data, owner, category, url pointing to a detailed description, tags and free text description. Let us note the possibility to access all the files of an owner and all the files tagged with similar tags.

2.3.3. User management /content upload

The Home tab allows to log in or to create a user account. Once logged-in, the user can manage his/her account, visualize his/her own content classified in categories, and upload new content. The media converter allows to obtain MP4 representation and supports a large family of formats.

The conversion/upload interface enables specifying metadata, categories and public/private flag.

2.3.4. Additional functionalities

Once logged-in, when visualizing the content, additional functionalities are exposed to the user:

- download the MP4 file;
- simplify the 3D content for visualization on mobile phones or less powerful terminals;
- annotate the file;
- insert the content into a web page or a blog.

2.4. API for 3D Object Description and Retrieval

While MPEG-7 [MSS02] defines a complex and almost complete scheme for audio/video content description, it does not specify methods for descriptor extraction. Moreover, descriptor diversity and complexity make the implementation of a multi-descriptor system by a single contributor almost impossible. On MMW.com, based on an open Application Programming Interface (API), developers can integrate descriptor extraction algorithms as plug-in libraries.

The first use of the MMW.com API is the online benchmarking of content descriptions, with the goal of retrieving similar medias. The API is organised into three components with the following functionalities (Figure 2):

- access the media in a raw (decoded) format and plug-in description extractors that fills the sub-parts of the description scheme (*ComputeDescriptor*);
- submit description (*DumpXML*);
- submit description scheme extensions if the MPEG-7 XSD does not cover the needed descriptor (*ExtendSchema*);
- plug-in similarity measures (*ComputeDistance*).

For supporting media description and retrieval, we provide an abstract class as a Descriptor Library (Visual C and gcc projects are available). The third-party contributor needs to derive the class and implements the four interfaces.

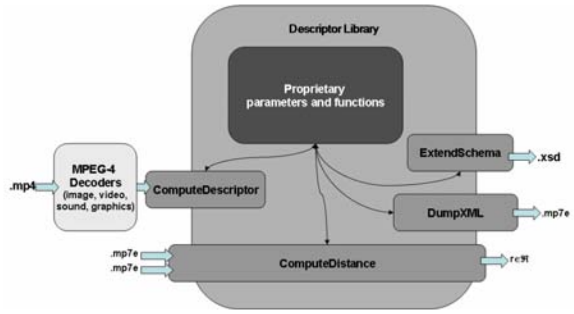

Figure 2: *QBE Descriptor library.*

The MMW.com system decompresses the 3D object contained in the MPEG-4 file before inputing it as a vertex buffer through the API. This buffer is then inserted in a simple structure containing a list of vertices and triangles.

For the output, developers should provide an XML file of the descriptor. By calling the function *ExtendSchema*, the proposed descriptor can be inserted in the extended MPEG-7 Schema. An example is illustrated on-line (`http://vega.int-evry.fr/3dod_php_dev/MMW_API/3DObjectRetrieval/MMW_3DObjectRetrievalAPI.zip`).

To compare two objects, MMW.com inputs the two XML descriptions in the *ComputeDistance* function that should provide a real value indicating the distance between these descriptions.

When the MPEG-4 file contains several 3D objects, the description is performed at two levels: one descriptor is computed for each connected component and a global descriptor is computed for the entire object. This allows to address more complicated scenarios as illustrated in the following example:

- a virtual human body is created as a collection of objects representing different anatomical parts (*e.g.* one for each arm). When searching for similar avatars, the system should use the global -descriptor associated with the whole avatar.
- a virtual human body is enriched with artefacts such as a blade. When searching for similar blades, the system should use only the descriptor of the blade.

3. 3D descriptors implemented in MMW.com

In the following, we explain three different descriptors using the above described API.

3.1. MPEG-7 3D Shape Spectrum Descriptor

The MPEG-7 3D Shape Spectrum Descriptor (3DSSD) [CKO*01] provides an intrinsic shape description related to local geometric properties of 3D object surface known as the shape index. The 3DSSD was previously used for indexing 2D images [Nas97] and 3D range data [DJ97]. The shape index was introduced by Koenderink [Koe90] as a function of local principal curvatures. 3DSSD was adopted in the MPEG-7 standard with the following syntax:

```
<complexType name="Shape3DType" final="#all">
 <complexContent>
  <extension base="mpeg7:VisualDType">
   <sequence>
    <element name="Spectrum">
     <simpleType>
      <restriction>
       <simpleType>
        <list itemType="mpeg7:unsigned12"/>
       </simpleType>
       <maxLength value="255"/>
      </restriction>
     </simpleType>
    </element>
    <element name="PlanarSurfaces"
     type="mpeg7:unsigned12"/>
    <element name="SingularSurfaces"
     type="mpeg7:unsigned12"/>
   </sequence>
   <attribute name="bitsPerBin"
    type="mpeg7:unsigned4" use="optional"
    default="12"/>
  </extension>
 </complexContent>
</complexType>
```

3.2. 3D Hough Transform Descriptor

The 3D Hough Transform Descriptor (3DHTD) [ZP01] is a shape descriptor intrinsically invariant to topological representation issues. Since 3DHTD is not standardized in the MPEG-7 Schema, we extended the latter by adding a new ComplexType, called *Hough3DType*, implemented as an extension of the MPEG-7 *VisualDType* type. The resulting syntax is the following one:

```
<complexType name="Hough3DType">
 <complexContent>
  <extension base="mpeg7:VisualDType">
   <sequence>
    <element name="Spectrum"
     type="mpeg7:doubleVector"/>
   </sequence>
   <attribute name="Ns" type="mpeg7:unsigned8" *
    use="optional" default="20"/>
   <attribute name="Nniv" type="mpeg7:unsigned8"
    use="optional" default="2"/>
  </extension>
 </complexContent>
</complexType>
```

3.3. BM Descriptor

The BM descriptor (BMD) [BDTK06] has been proposed by the Computer Science and Distance Education Laboratory (Faculty of Electrical Engineering - Skopje University), to address the issues of non-uniform sampling of vertices and difference in polygon topology between separate mesh-objects. BMD introduces a pre-processing parameterization scheme, which outputs a mesh with uniform (parameterized) topology using spherical mapping. The description scheme for BMD is the following:

```
<complexType name="BMDType">
 <complexContent>
  <extension base="mpeg7:VisualDType">
   <sequence>
    <element name="Spectrum"
     type="mpeg7:doubleVector"/>
   </sequence>
  </extension>
 </complexContent>
</complexType>
```

4. Benchmarking results

MMW.com presents benchmark results in two flavours: quantitatively and qualitatively.

4.1. Qualitative benchmark by on-line visualization

Based on the previously-introduced 3D shape descriptors, MMW.com allows to visualize on-line the result of the similarity search. Starting from the visualization page of the 3D object, a user can trigger the search procedure. Let us note that the descriptors and the distance between any two descriptors are computed when the content is uploaded. The result is presented in a separate web-page where, for each descriptors, the retrieved 3D objects are displayed according to increasing distances.

To refine visualization results, users can use filters. Filters allow to modify the search with different parameters:

- Modify the numbers of retrieved objects;
- Restrict search to sub-databases like Princeton, Taschen or MPEG;
- Retrieve MPEG-4 files containing only one 3D object;
- Retrieve only whole 3D objects (and not sub-parts).

Figure 3 illustrates benchmark results with qualitative visualization.

4.2. Quantitative measurements

Together with retrieval results, MMW.com provides quantitative measurements. We have implemented the most recognized measure from the SHREC benchmark [Rem06]. The latter is based on a set of highly-relevant items and a set of marginally-relevant items. A highly-relevant item represents

Figure 3: *Qualitative visualization of benchmark results.*

the number of 3D objects classed as similar and belonging to the same category, and a marginally-relevant item is the number of 3D objects classed as similar and belonging to the parent category.

For each query, various evaluation metrics are computed:

- The data set size.
- The number of highly-relevant classified items.
- The number of marginally-relevant classified items.
- The total number of relevant classified items.

With the classified items, we can compute significant metrics about the descriptor result. For all descriptors, we compute some performance measures for highly-relevant items and marginally-relevant items:

- **True positive** is the number of 3D objects retrieved using the descriptor.
- **First tier** is a percentage representing the number of correct results in the first tier of all results.
- **Second tier** is a percentage representing the number of correct results in the second tier of all results.
- **Precision** is the number of true positive items divided by the length of ranked list (*i.e.* the number of 3D objects returned for the descriptor).
- **Recall** is the number of true positive items divided by the number of relevant classified items.

All measures are automatically computed for all descriptors and updated according to the filter parameters.

Figure 4 illustrates the benchmark with quantitative measurements.

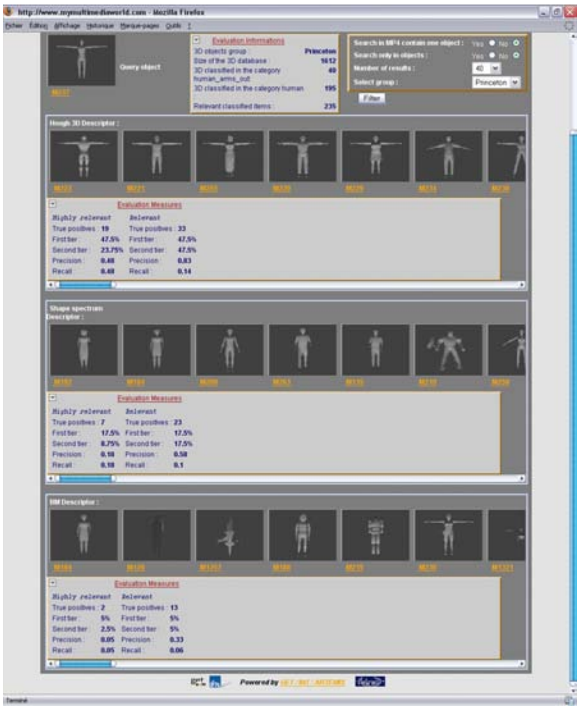

Figure 4: *Quantitative measurements of benchmark results.*

5. Conclusion and perspectives

In this paper, we have described a new strategy for 3D object retrieval benchmarking based on an on-line platform. Additionally to web and web2.0 facilities, the originality of the platform consists in providing an open API allowing to include media descriptor extractors by third-parties. We have demonstrated the relevance of this API by implementing three descriptors, and have implemented qualitative and quantitative measurements for benchmarking the descriptors.

Currently, when new content is uploaded, the benchmark is automatically executed and the results are presented in real-time. Similarly, when a new algorithm is integrated, it is executed for all the files in the database and the benchmark is automatically updated.

As a perspective, we commit in assisting interested researchers in integrating their algorithms in MMW.com with the goal of establishing a reference system, easy to use and update, for the fast development of meaningful 3D shape descriptors.

References

[BDTK06] BLERIM M., DAVCEV D., TRAJKOVIK V., KALAJDZISKI S.: 3d object matching using spherical mapping. In *Proceedings of the 32nd IEEE Annual Conference on Industrial Electronics, IECON 2006* (Paris, France, 2006), pp. 3450–3454.

[CCS98] CHANG S.-F., CHEN W., SUNDARAM H.: Videoq: A fully automated video retrieval system using motion sketches. In *WACV '98: Proceedings of the 4th IEEE Workshop on Applications of Computer Vision (WACV'98)* (Washington, DC, USA, 1998), IEEE Computer Society, p. 270.

[CKO*01] CIEPLINSKI L., KIM M., OHM J., PICKERING M., YAMADA A.: Multimedia content description interface - part 3 visual, 2001. ISO/IEC JTC1/SC29/WG1; MPEG2001/N4062.

[DJ97] DORAI C., JAIN A. K.: Shape spectrum based view grouping and matching of 3d free-form objects. *IEEE Trans. Pattern Anal. Mach. Intell. 19*, 10 (1997), 1139–1146.

[GG07] GONZÁLEZ J. L. V., GÓMEZ J.: Trec: Experiment and evaluation in information retrieval. *JASIST 58*, 6 (2007), 910–911.

[KCG*03] KIRANYAZ S., CAGLAR K., GULDOGAN E., GULDOGAN O., GABBOUJ M.: Muvis: a content-based multimedia indexing and retrieval framework. In *Seventh International Symposium on Signal Processing and Its Application* (Finland, 2003), vol. 1, pp. 1–8.

[Koe90] KOENDERINK J.: *Solid shape*. The MIT Press, Cambridge, Massachusetts, 1990.

[LBK03] LUX M., BECKER J., KROTTMAIER H.: Caliph & emir: Semantic annotation and retrieval in personal digital photo libraries. In *Forum at 15th Conference on Advanced Information Systems Engineering* (Velden, 2003), pp. 85–89.

[LE05] LOY G., EKLUNDH J.: A review of benchmarking content based image retrieval. In *Workshop on Image and Video Retrieval Evaluation* (2005).

[LPP07] LE BONHOMME B., PREDA M., PRÊTEUX F.: Extended mpeg-7 schema for 3d on demand. In *Workshop on Multimedia Semantics - The Role of Metadata* (Aachen, Germany, March 2007), pp. 218–229.

[LSDJ06] LEW M. S., SEBE N., DJERABA C., JAIN R.: Content-based multimedia information retrieval: State of the art and challenges. *ACM Trans. Multimedia Comput. Commun. Appl. 2*, 1 (2006), 1–19.

[MDH*05] MEZARIS V., DOULAVERAKIS H., HERRMANN S., LEHANE B., OŠCONNOR N., KOMPATSIARIS I., STRINTZIS M. G.: The schema reference system: An extensible modular system for content-based information retrieval. In *Proc. Workshop on Image Analysis For Multimedia Interactive Services (WIAMIS), Montreux, Switzerland* (April 2005).

[MMMM*00] MÜLLER W., MÜLLER H., MARCHAND-MAILLET S., PUN T., SQUIRE D., PECENOVIC Z., GIESS C., DE VRIES A. P.: Mrml: A communication protocol for content-based image retrieval. In *VISUAL '00: Proceedings of the 4th International Conference on Advances in Visual Information Systems* (London, UK, 2000), Springer-Verlag, pp. 300–311.

[MMMMP00] MÜLLER W., MARCHAND-MAILLET S., MÜLLER H., PUN T.: Towards a fair benchmark for image browsers. In *SPIE Photonics East, Voice, Video, and Data Communications* (Boston, MA, USA, nov 5–8 2000).

[MSS02] MANJUNATH B., SALEMBIER P., SIKORA T.: *Introduction to MPEG-7: Multimedia Content Description Interface*. John Wiley & Sons, Inc., New York, NY, USA, 2002.

[Nas97] NASTAR C.: *The Image Shape Spectrum for Image Retrieval*. Tech. Rep. RR-3206, INRIA, 1997.

[PPS96] PENTLAND A., PICARD R. W., SCLAROFF S.: photobook: content-based manipulation of image databases. *int. j. comput. vision 18*, 3 (1996), 233–254.

[SC96] SMITH J. R., CHANG S.-F.: Visualseek: a fully automated content-based image query system. In *MULTIMEDIA '96: Proceedings of the fourth ACM international conference on Multimedia* (New York, NY, USA, 1996), ACM, pp. 87–98.

[SKMF04] SHILANE P., KAZHDAN M., MIN P., FUNKHOUSER T.: The princeton shape benchmark. In *SMI '04: Proceedings of the Shape Modeling International 2004 (SMI'04)* (Washington, DC, USA, 2004), IEEE Computer Society, pp. 167–178.

[SOK06] SMEATON A. F., OVER P., KRAAIJ W.: Evaluation campaigns and trecvid. In *MIR '06: Proceedings of the 8th ACM international workshop on Multimedia information retrieval* (New York, NY, USA, 2006), ACM, pp. 321–330.

[VH07] VELTKAMP R. C., HAAR F. B. T.: *SHREC2007: 3D Shape Retrieval Contest*. Tech. Rep. UU-CS-2007-015, Department of Information and Computing Sciences, Utrecht University, 2007.

[ZP01] ZAHARIA T., PRÊTEUX F.: Hough transform-based 3d mesh retrieval. In *Proceedings SPIE Conference 4476 on Vision Geometry X* (San Diego, CA, August 2001), pp. 175–185.

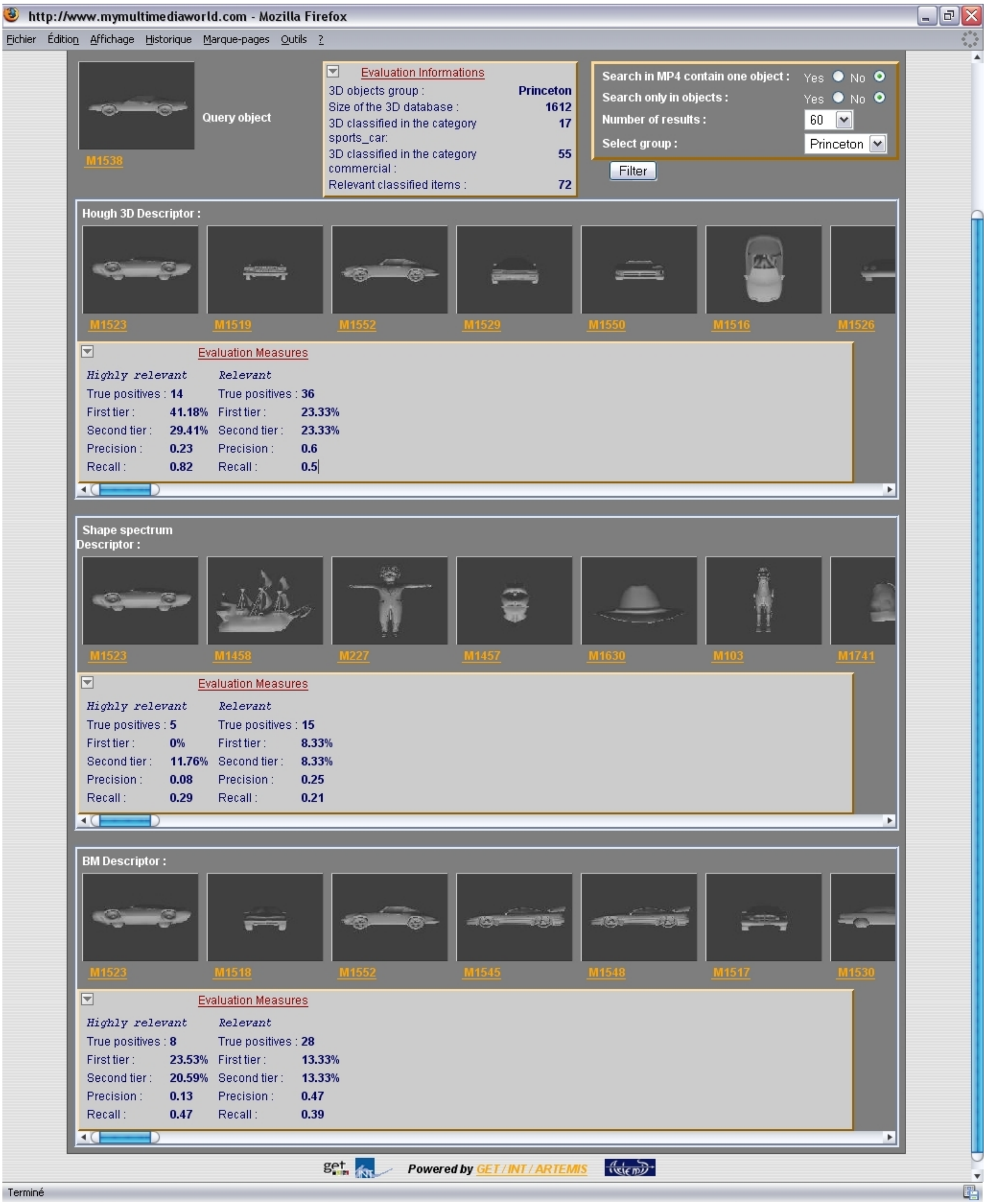

Figure 5: *Example of 3D object retrieval benchmarking with MyMultimediaWorld.com.*