A Feature Vector Approach for Retrieval of 3D Objects in the Context of MPEG-7

Dejan V. Vranić and Dietmar Saupe

University of Leipzig, Department of Computer Science,
P.O. Box 920, D-04009 Leipzig, Germany
Email: vranic@informatik.uni-leipzig.de

ABSTRACT

The topic of this communication is content-based retrieval of 3D objects represented as triangle meshes in the framework of MPEG-7. The idea consists of the use of suitable descriptors, which capture 3D shape. The descriptors are invariant with respect to translation, rotation, scaling, and reflections and robust with respect to level-of-detail. The presented analysis of 3D mesh models possesses advantages in comparison to other approaches.

1. INTRODUCTION

As the world becomes a more and more networked place, efficient data access is increasingly important. Classical database retrieval of structured information, retrieval in collections of textual files and retrieval in collections of binary objects are typical ways of accessing data. The amount of audiovisual data available in digital form is rapidly growing and different forms of representation of multimedia content are also in use. Therefore, the development of systems dedicated to the retrieval of a desired kind of audiovisual information is a current problem which is also addressed by MPEG committees [2].

An object from some binary database can traditionally be accessed using attached information such as textual annotation. Besides the fact that textual annotations cannot encode all the information available in images, video sequences, and 3D models ("an image is worth a thousand words"), the annotations are mostly created manually and subject to the taste of the person who creates them. Therefore, retrieving multimedia material from different collections using its content as a key is an attractive and important problem. Content-based retrieval systems are mainly designed for still image, audio and video libraries, while only a few techniques for content-based 3D model retrieval have recently been presented [1, 4, 6, 7].

2. RELATED WORK

In this section, we briefly describe some previous work on characterizing the global 3D shape of an object. More details about shape descriptors are omitted in order to save space.

The 3D shape descriptor (D) proposed in [6] is invariant with respect to rotations of 90 degrees around the coordinate axes. This restricted rotation invariance is attained by a very coarse shape representation. Furthermore, if an object is rotated around an axis by, e.g. 45 degrees, the feature vector differs significantly.

Cords-based, moments-based, and wavelet transform-based descriptions are presented in [4] as well as a modification of the classical Principal Component Analysis (PCA). The PCA is used to secure translation and rotation invariance, while a modification is needed in order to take into account differing sizes of triangles of the mesh. The modified PCA was applied to the set of centers of mass of all triangles of a mesh model, with additional weighting factors (equal to the surface area of the corresponding triangle) for each center of mass. The definitions of cords-based and moments-based descriptors suggest that these feature vectors are not robust with respect to the level-of-detail of a model. An account of the MPEG-7 description scheme (DS) based on these three descriptors is given in [5].

Independently of [4], we proposed feature vectors based on rays in [7], where the PCA was modified in a different way. It was applied to the set of vertices of a 3D mesh model using weights which are proportional to the areas of the triangles associated with the corresponding vertex. Our other approaches for characterizing the global 3D shape [1] include an enhancement of the ray-based feature vector as well as volume-based, voxel-based, silhouette-based, and depth buffer-based feature vectors. The corresponding descriptors possess invariance with respect to translation, rotation, and scaling of a 3D model.

3. OUR APPROACH

The block-diagram of our 3D model retrieval system is depicted using the processing chain shown in Figure 1. The scope of our research is denoted by the dotted boundaries. The first block inside the scope is *normalization* and is used to ensure invariance with respect to translation, rotation, scaling, and reflections. The necessity of this module can be depicted by the following example: different authors of 3D models have different styles and use different units to create objects. Their models could have substantially similar shape, but different scale, orientation, and position in the 3D space. Our idea [7] is to find a canonical (i.e. normalized) position and orientation of a model.

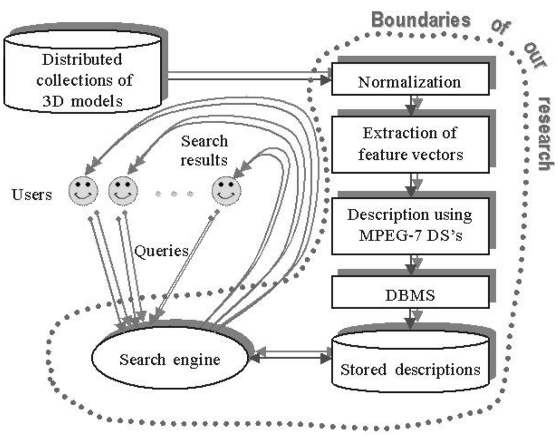

Figure 1: The 3D model retrieval system.

In the analysis, instead of applying the PCA in the classical way [2] (to the set of vertices of a triangle mesh) or using modifications by introducing weights (section 2), we devised a "continuous" version of the PCA, in which each triangle is considered as a set of infinitely many points. As far as we know, this method of PCA has not been applied yet and represents a new result. Numerous tests show superiority of the "continuous" PCA compared to both the classical PCA and the modified PCA. Moreover, some proposals of experiments on 3D shape [3] became irrelevant. More details about the "continuous" PCA will be given in a forthcoming paper.

The next two steps in Figure 1, *extraction of feature vectors* and *description using MPEG-7 description schemes*, are explained in subsections 3.1 and 3.2. The system includes a module for managing the database (*DBMS*) of *stored descriptions*. Finally, the *search engine* (application) is based on the description scheme.

3.1. Feature vectors

In this subsection, we stress some general properties (relevant to MPEG-7 DS) of our feature vectors which are aimed at describing spatial characteristics of a model. The most important problems and requirements in defining 3D shape descriptors are the following:

- Invariance with respect to translation, rotation, scaling, and reflections;
- Robustness with respect to level-of-detail;
- Changeable dimensions of feature vectors;
- Provision of fundamentally different types of descriptors.

The first requirement is secured by the normalization, while the remaining three should be fulfilled by the definitions of the descriptors. The last two requirements are needed because there is no theory about the optimal way of describing a 3D object.

Besides those presented in [7, 1], we created and implemented some new feature vectors and incorporated spherical harmonic representations of those that can be interpreted as functions on a sphere. Most of the feature vectors are actually real vectors of dimension n. The parameter n, which provides the changeability of the dimension of each feature vector, cannot have an arbitrary value. Currently, we are using the l_1 or l_2 norm for calculating the distance between feature vectors. Nevertheless, the choice of the most effective way to calculate this distance is an open question.

We use a different data structure to store feature vectors derived from a voxelization of a 3D model. The voxelization is performed in the following steps: the determination of the dimensions of a voxel, the subdivision of the whole 3D space into voxel cells, the calculation of the proportion of the total surface area of the 3D model inside each voxel, and the consideration of the calculated proportion as a voxel value. The dimensions of a voxel are determined in three different ways: using the bounding box (BB), using the "average" box (AB) (the average distances of triangle points along the PCA axes fix the dimensions of the AB), and using the "canonical" cube (CC) (the length of the diagonal of the average box is taken as the length of the edge of the CC). Finally, the dimensions of voxel cells are fixed by dividing each dimension of BB, AB, or CC by 2^r, where r is the so-called resolution of voxelization. Obviously, of all voxels inside BB, AB or CC, the fraction which have values greater than zero decreases with increasing r. Therefore, the most suitable way of storing a voxel-based feature vector is an octree structure.

3.2. Description of 3D-models in the context of MPEG-7

The only step in Figure 1 that is within the scope of MPEG-7 is the description. We created a set of descriptors and a description scheme, both of which are compatible with MPEG-7. Descriptors for the two fundamentally different classes of feature vectors described in the previous subsection are embedded in a simple description scheme depicted in Figure 2 using a pseudo-code.

```
3D-Model {
  Vector(dim) {
    int dimension=dim;
    float component[dimension];
  }
  OctreeNode {
    Byte data;
    Byte links;
  }
  Shape {
    VectorBased {
      String featureVectorName;
      int dimension;
      Vector feature(dimension);
    }
    int nOfVectorBased;
    VectorBased fvV[nOfVectorBased];

    OctreeBased {
      String featureVectorName;
      String treeTraversal;
      int numberOfNodes;
      int numberOfLevels;
      Vector quant(numberOfLevels);
      OctreeNode node[numberOfNodes];
    }
    int nOfOctreeBased;
    OctreeBased fvO[nOfOctreeBased]:
  }
}
```

Figure 2: OO representation of D and DS.

Two instances in which the DS is applied may have a different number of vector-based and octree-based Ds, depending on the chosen set of the corresponding feature vectors. Consequently, new descriptors from these classes can easily be added. Also, if some descriptors become surpassed, they can be excluded. Each descriptor contains the name of the corresponding feature vector; this name encodes the information about the descriptor's semantics.

An octree node is stored using two bytes. One byte contains information about child nodes, while the other byte encompasses information about the value of the corresponding voxel. Voxel values are quantized and the values of quanta for each octree level are stored. The octree-based D possesses also a field storing information concerning the tree traversal method (postorder or preorder). The given fields are sufficient to create the octree representation of a feature vector.

The matching (search) procedure is not standardized and represents a challenging problem. As mentioned in subsection 3.1, we can engage the l_p norm in the case of vector-based Ds. If we consider an octree-based feature vector as a real vector of infinite dimension (but with finitely many non-zero components), the same norm can also be applied. However, our results show that the l_p norm possesses drawbacks, particularly in the case of octree-based Ds. We expect that search effectiveness will be improved by a modification of the Hausdorff distance for voxel-based Ds. In the future, the Ds shown in Figure 2 could have a separate field for recommending the matching criterion.

4. EVALUATION OF D AND DS

The set of criteria for evaluating descriptors and descriptor schemes has been established by MPEG-7. The most significant criteria used by MPEG-7 for D and DS are the application domain, the effectiveness, the scalability, the multilevel representation, etc.

The *application domain* of the proposed D and DS covers a broad range of real-world applications (CAD, visualization, VR, art, education, etc.). The *effectiveness* means that the descriptor characterizes an important property of the content. Since our descriptors are aimed at characterizing the 3D shape of mesh models, the effectiveness criterion is also fulfilled. Our descriptors are *scalable* because the size of the descriptor is determined only by its definition. The complexity of a 3D object (i.e. number of vertices and triangles) does not affect the dimension of the feature vector. Finally, the *multilevel representation* is related to the representation of features at multiple levels of abstraction. In the case of vector-based Ds, the abstraction level is determined by the dimension of the corresponding feature vector. Reducing the dimension decreases the precision of the description but increases the level of abstraction, and vice versa. In the case of voxel-based Ds, the resolution defines the level of abstraction. Note that, because of the employed data structure, all higher levels of abstraction are automatically embedded for a given resolution.

5. EXAMPLE

The 3D model database used for experiments contains around 1500 models in different 3D file formats (VRML, DXF, 3DS, OFF, etc.). The models were gathered from the Internet (3dcafe, viewpoint, etc.). In order to evaluate the performance of the feature vectors, an evaluation system has been implemented. The most important options that are included in the evaluation system are: verification of parameters used in the normalization step, visualization of 3D mesh models, search using a selected D, and evaluation using a type of precision/recall test.

We have manually classified models by shape (e.g. cars, planes, bottles, chairs, etc.) and used this classification in our precision/recall test. Briefly, precision is the proportion of retrieved models that are relevant and recall is the proportion of the relevant models actually retrieved. By examining the precision/recall diagrams for different queries (and classes) we obtain a measure of the retrieval performance for a selected descriptor and matching criterion.

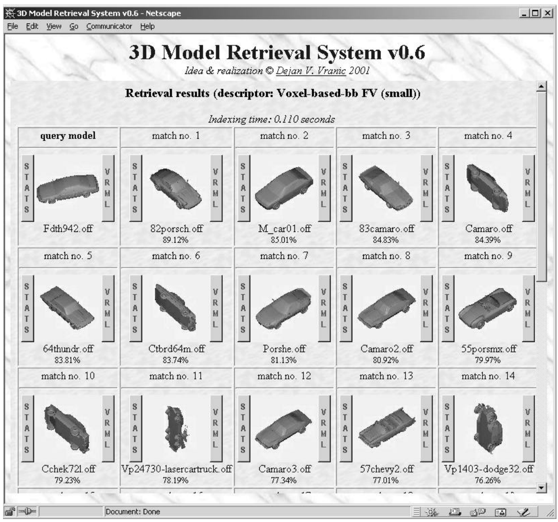

Figure 3: Query for a car.

Besides the off-line evaluation system, we are implementing an on-line retrieval system. The retrieval system provides a user friendly interface with options for selecting the type of descriptor and matching criterion, browsing the database, and displaying a model in a VRML browser. A retrieval example generated by this on-line prototype is shown in Figure 3. A model of a car is used as the query, while the l_1 norm was applied to the bounding box voxel-based feature vector with the resolution $r = 2$. The models depicted in Fig. 3 are visualized from the same direction in the original frame.

6. CONCLUDING REMARKS

In this paper, the system for content-based 3D model retrieval is presented. The "continuous" PCA is included in the analysis, original feature vectors are used for capturing 3D shape, and a new 3D shape description scheme is proposed. Evaluation results are compared using a type of precision/recall test and are used to conclude which type of feature vector (and which dimension) is the most suitable for a given class of 3D objects.

We are studying the trade-off between complexity in time and space vs. retrieval performance for all descriptors (type and dimension).

7. REFERENCES

[1] M. Heczko, D. Keim, D. Saupe, and D.V. Vranić, "A method for similarity search of 3D objects" (in German), *Proc. of BTW 2001*, Oldenburg, Germany, March 2001.

[2] MPEG Video Group, "MPEG-7 Visual part of eXperi-metation Model (version 8.0)," Doc. ISO/MPEG N3673, MPEG La Baule Meeting, October, 2000.

[3] T. Murao and E. Paquet, "Core Experiment Proposal on 3D shape for MPEG-7," DOC. ISO/MPEG M4679, MPEG Vancouver Meeting, July, 1999.

[4] E. Paquet, A. Murching, T. Naveen, A. Tabatabai, and M. Rioux, "Description of Shape Information for 2-D and 3-D Objects," *Signal Processing: Image Communication*, 16:103–122, 2000.

[5] E. Paquet and M. Rioux, "The MPEG-7 Standard and the Content-based Management of Three-dimensional Data: A Case Study," *IEEE Int. Conf. on Multimedia Computing and Systems*, Florence, Italy, pp. 375–380, June 1999.

[6] M.T. Suzuki, T. Kato, and N. Otsu, "A Similarity Retrieval of 3D Polygonal Models Using Rotation Invariant Shape Descriptors," *IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC2000)*, Nashville, Tennessee, pp. 2946–2952, October 2000.

[7] D.V. Vranić and D. Saupe, "3D Model Retrieval," *Proc. of Spring Conference on Computer Graphics and its Applications (SCCG2000)*, Budmerice Manor, Slovakia, pp. 89–93, May 2000.