3D Image Databases: Acquisition and Retrieval by Content

Carlo Colombo Alberto Del Bimbo Pietro Pala

Dip. di Sistemi e Informatica
Via S. Marta 3
I-50139 Florence, Italy
[colombo,delbimbo,pala]@dsi.unifi.it

ABSTRACT

The availability of a huge amount of cultural heritage material in digital form requires the investigation of new cost-saving, effective and easy-to-use methods for acquisition, annotation and retrieval. In this paper, we will review the state-of-the-art of the field of 3D image databases, and discuss an integrated acquisition-annotation-retrieval approach under development at the Visual Information Processing Laboratory of the University of Florence.

Keywords

3D Databases, Retrieval by Visual Content, Computer Vision, Pattern Recognition

1. INTRODUCTION

Thanks to the latest digital imaging technologies, we can capture, restore, edit, enhance, manipulate, preserve, and retrieve images and other associated information and provide global access to invaluable cultural and historical resources in digital form anywhere anytime. While the benefits and potentials are great, so are the issues and problems. The availability of a huge amount of cultural heritage material in digital form requires the investigation of new cost-saving, effective and easy-to-use methods for acquisition, annotation and retrieval. Image databases and visual information retrieval have been an important research subject in the last five years. Image analysis and processing, pattern recognition and computer vision, multimedia data modelling, multidimensional indexing, psychological modeling of user expectations, man-machine interaction and data visualization are all disciplines that contribute to this area of research. While most of the research activity has been concentrated on retrieval by content of 2D still images and video, very little research has addressed retrieval of 3D visual content. On the other hand, digital 3D information is rapidly being available, particularly in the field of cultural and historical material like vases, sculptures, and man-made objects.

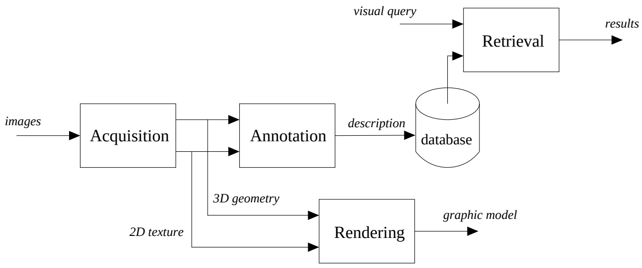

Figure 1: 3D database system.

The creation of digital 3D archives requires that the 3D information is gathered from sensors and stored effectively in order to allow search, retrieval and presentation. Acquisition can be performed using 3D scanners or computer vision, depending on the quality of the digital representation and the goals of the application. Search can be performed in order to find commonalities between 3D objects (e.g. the signature of the artist) or exploring the temporal evolution of a defect (e.g. the amount of bending for wooden tables) as well as to support effective educational programs. Visualization can be complemented with manipulation functions. Fig. 1 summarizes the main functional blocks of a 3D visual database system. 2D images are first acquired and analyzed, so as to characterize the imaged objects in terms of basic physical primitives—3D shape and surface texture. Such primitives are then used both for retrieval and visualization purposes. Visualization is accomplished by recombining (i.e., re-synthesizing) shape and texture information in order to get a 3D graphic model for display. For the purpose of visual retrieval, raw shape and texture information is expressed (annotated) in terms of an internal representation suitable for similarity comparison against a 3D query. In this paper, we will review the state-of-the-art of the field of 3D image databases, and discuss an integrated acquisition-annotation-retrieval approach under development at the Visual Information Processing Laboratory of the University of Florence.

1.1 State-of-the-art

Several different topics are involved in visual 3D database research. In the following, we will discuss background and state-of-the-art of each of them separately.

1.1.1 Acquisition

That of 3D structure reconstruction from images is a key problem in computer vision. A thorough exposition of the solutions proposed in the last ten years can be found in [1], [2]. The different approaches are motivated by different operating conditions. Of relevant interest for many applications are the so called model-based reconstruction approaches, which rely on the a priori knowledge of geometric or motion characteristics in the scene. The most recent model-based approaches work with full perspective, uncalibrated cameras, and yet produce a metric 3D object reconstruction. The most advanced constrained motion approaches are described in [3], where the object is reconstructed based on turntable sequences and corner tracking techniques. The state-of-the-art in static scene reconstruction is [4], offering a methodology for computing the structure of piecewise planar scenes.

1.1.2 Annotation

The object recognition technique using the so-called "spin-images" was recently used for the purpose of 3D structure representation [5]. According to this technique, 3D objects are represented invariantly with respect to rigid motion by means of a suitable collection of images gathering information about the morphology of the 3D object's surface. Spin-images can be considered an evolution of geometric hashing and structural indexing in that they overcome the typical problems of these techniques. Feature-based representation of shapes have been proposed as well [8], [9]. To overcome limitations related to the use of global shape descriptors, techniques based on part decomposition of the shape have been proposed. Local surface descriptors are based on the extraction of primitives such as contours, planar surfaces and curves [6] and their visual features such as the curvature parameters, the orientation changes, and so on. Multiscale shape curvature analysis is also used for shape description in [10], [11].

1.1.3 Retrieval

Retrieval of 3D VRML object models based on textual metadata was early proposed in [12]. In [13] geometric and visual features like color and texture, of the object surface are exploited to support retrieval of 3D VRML object models. In [16] retrieval of CAD object models is supported by describing the surface of 3D objects through a 2D representation of their contours, taken from different viewpoints. Lamdan and Kriegel have addressed the definition of both structure and surface descriptors: in [14] different orientations of the surface of the object are stored in a hash table mapping triplets of surface points into their orientations; geometric hashing techniques are used for model based recognition of 3D objects. In [15] retrieval of 3D objects based on similarity of surface segment is addressed. Description of 3D volumetric data for the purpose of retrieval by shape similarity has also been addressed [17]. In [7], 3D object recognition is obtained by means of the comparison of surface graphs extracted by single view of range images. In [18] the retrieval of 3D objects is provided by searching of 2D views.

2. OUR APPROACH

Our current research goal is to investigate relevant aspects related to the creation and organization of digital libraries of three-dimensional objects supporting retrieval based on similarity. In the following, we will review our latest achievements in the following areas: (1) acquisition and visualization of 3D models from one or more images; (2) object annotation through the modeling of surface visual features; (3) definition of similarity measures and 3D retrieval by visual similarity.

2.1 Acquisition

We have recently focused on the development and implementation of computer vision algorithms for automatic 3D structure reconstruction of objects modeled as surfaces of revolution (SOR). The reconstruction task is carried out using one or more uncalibrated perspective images acquired under non structured light conditions. The final goal is to obtain a point-wise description of the curvature characteristics of the SOR, as reflected by the 2D shape of any of its meridians. The decorative, superficial content of polychromous objects such as black- or red-figured archaeological vases is also extracted by image resampling, and associated pixel by pixel to the reconstructed surface so as to obtain a full visual annotation of the object, taking into account both its 3D geometric characteristics and its 2D photometric appearance. Having reduced the 3D structure to a planar curve (the profile), and the superficial texture to a single image, the SOR annotation is totally bidimensional, thus allowing to perform search by content of 3D textured objects using even standard image retrieval engines. In order to obtain an uncalibrated metric reconstruction of the objects, the existing theory of self-calibration for planar scenes was extended to the case of SOR, thus allowing the SOR reconstruction methodology to be applied to any kind of images, ranging from new photographic material to photographic archives, art catalogues, and so on.

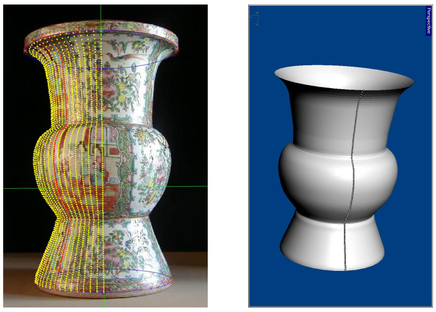

Figure 2: *Left*: a chinese vase and its meridians. *Right*: the reconstructed 3D model.

2.1.1 Surface shape reconstruction

In the case of a SOR, the problem of 3D reconstruction is equivalent to that of reconstructing the scaling function, i.e. the shape of any meridian. This can be achieved in three main steps: (1) evaluation of image projective symmetries; (2) camera calibration by exploitation of SOR scene constraints; (3) metric rectification of the imaged meridian. Details on the 3D metric reconstruction algorithm can be found in [19].

Fig. 2 (left) shows a textured SOR, with several imaged meridians superimposed to the original photograph; the figure also shows the horizontal and vertical visual symmetry axes. The reconstructed model obtained by imaged meridian rectification is shown in Fig. 2 (right).

2.1.2 Flattened texture acquisition

The knowledge of the scaling function can be exploited so as to perform texture acquisition in the form of a flat (planar) image. Flat images are required to extend the object representation and allow using traditional image database technology for the retrieval of textured 3D objects. Another objective is to build texture maps which are compliant with commercial 3D modeling-rendering software standards. We have elected to use one of the most common image mapping techniques, the *normal cylindrical projection*, as it fits well with the kind of objects we are dealing with. Cylindrical maps are constructed by aligning a sampling cylinder so that it is coaxial with the SOR. In such a way, parallels and meridians transform as a rectangular grid, meridians being equally spaced. The 3D reconstruction information is used, of course, in order to move metrically over the imaged meridians and parallels and map them to the texture image space. In the case of SOR, the normal cylindrical projection map is readily obtained from the natural surface parametrization. Fig. 3 (*left*) shows the image of textured

Figure 3: *Left*: a chinese vase. *Right*: the extracted texture.

SOR (a chinese vase); the flattened texture is given in Fig. 3 (*right*). The flattened texture preserves the local geometry of the original 3D object, thus allowing to reproduce correctly also highly deformed and nearly invisible particulars in the original photo.

2.2 Annotation

Representation of the geometric structure of a 3D object is accomplished through the following steps: *Curvature estimation*, *Warping*, *Curvature mapping*. In the first step, the 3D object model is represented as a 3D mesh of points. For each vertex of the mesh the curvature is estimated. Information about curvature is used to annotate each vertex. In the second step, the 3D object mesh is subject to a deformation process. By acting on the (x,y,z) position of each mesh vertex, this process aims at collapsing the 3D object mesh onto a 3D sphere. At the end of the deformation process the 3D object mesh roughly approximate a sphere and each vertex of the mesh keeps track of its curvature on the original 3D object. Once warping is complete, the annotated spherical mesh is projected into an image that encodes position and curvature of points on the original object surface. This image is used as descriptor of the geometric structure of the original 3D object. These three steps are described in detail in the following.

2.2.0.1 Curvature estimation.

An approach to evaluation of curvature of polygonal meshes is described by Taubin [20]. A different approach is proposed in [21], where the estimation is not limited to adjacent triangles, but rather extends to a geodesic neighborhood. While this latter approach reduces sensitivity to noise, it requires an algorithm for evaluation of the neighborhood. In our approach, estimation of surface curvature at a generic vertex v_i of the mesh is accomplished by considering variations of surface normal over the *platelet* V^{v_i} of vertex v_i. This guarantees less sensitivity to noise and acquisition errors.

2.2.0.2 Deformation.

Mesh warping is accomplished through its convolution with a Gaussian kernel. Actually, our analysis is restricted to a particular class of surface meshes, that is, surfaces that can be described as functions on the sphere.

The iterative application of the Gaussian smoothing kernel to every vertex of the mesh is equivalent to a deformation process that changes the position of each vertex in order to reduce mesh curvature at that vertex.

2.2.0.3 Mapping.

Projection of a curved surface is a well known problem in cartography [22]. There are many different projections used to map (a part of) the globe onto a plane, but their description is far beyond the scope of this paper. In our approach, we have elected an area preserving projection, the Archimedes projection (also known as the Lambert equal-area projection). Similarly to the Mercator projection, the Archimedes projection is a cylindrical projection. This mapping is equivalent to wrapping a cylinder around the equator and then projecting along lines of constant latitude. When the cylinder is unrolled, a flat coordinate system is produced. A cylinder can be unrolled without creating distortion in the east-west directions, the distortion being limited to north/south only. Areas close to the equator exhibit little distortion either way.

Once a 3D model is represented through a 2D map, any approach supporting image retrieval by visual similarity can be used to evaluate the similarity between two 3D models. In fact, this can be achieved by computing the similarity of the corresponding maps. To take full advantage of properties of the Archimedes projection, the similarity between two image maps should be computed based on region area and their spatial arrangement. To accomplish these goals representation of map content is carried out using histograms. To support local as well as global descriptions of the models we have chosen to work with three levels of detail: at the highest level of detail the image is partitioned according to a regular grid comprising 32 tiles, at the intermediate level the grid comprises 8 tiles, and at the lowest level a single tile covers the whole image. Following the above discussion, histograms at the lower levels can be computed from histograms at the highest level.

2.3 Retrieval

Since the content of a map is represented at several resolution levels, the computation of the similarity between two maps relies on matching descriptors at equal resolution levels. At a generic resolution level, each map is partitioned into n tiles and each tile is represented through a 100-bins curvature histogram. Computing the distance between two maps requires to find the best tiles correspondence function. Actually, a suboptimal solution is computed through a *heuristic search* approach that requires to scan all tiles in the first map in a predefined order and associate to each tile

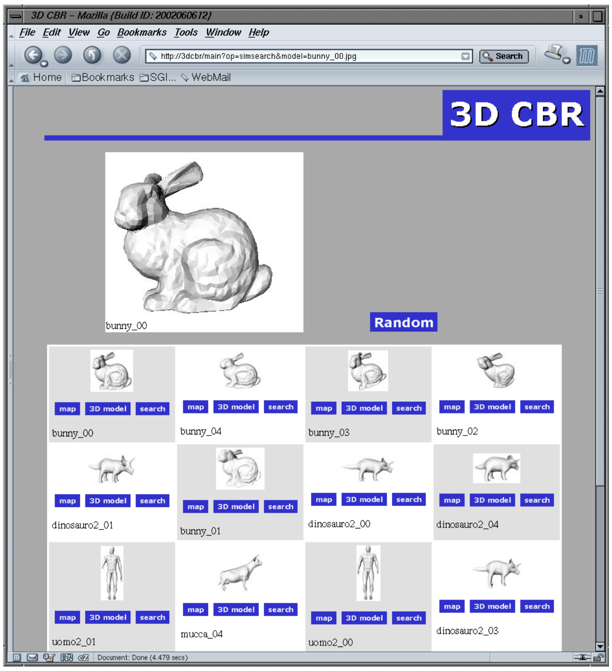

Figure 4: The system interface: a query and the retrieval results.

the most similar tile not yet associated in the second map.

Fig.4 shows a retrieval example where the model of a bunny was selected as a query template (upper left). The database includes 120 models either taken from the web, manually authored, or acquired with the SOR reconstruction technique referred to in section 2.1. Retrieved models are also shown in Fig.4. The 5 bunny models included in the database were all retrieved.

3. ACKNOWLEDGMENTS

The authors wish to thank Gianni Antini, Jürgen Assfalg and Federico Pernici for their help in the development of this research.

4. REFERENCES

[1] O. Faugeras and Q.-T. Luong, The Geometry of Multiple Images. MIT Press, Cambridge, Massachusetts, 2001

[2] R.I. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge, England, 2000

[3] G. Jiang, H. Tsui, L. Quan, A. Zisserman, Single axis geometry by fitting conics. In *Proc. 7th European Conf. on Comp. Vis. ECCV 2002*, Copenhagen, Denmark, May 2002, pages I:537–550, LNCS 2350, Springer-Verlag, 2002

[4] D. Liebowitz, *Camera Calibration and Reconstruction of Geometry from Images*, Ph.D. thesis, Univ. of Oxford, 2001

[5] A. Johnson, M.Hebert, Using spin images for efficient object recognition in 3D cluttered scenes, IEEE Trans. on PAMI, vol. 21, no.25, pp. 433-449, May 1999

[6] A. Hoover et al., An Experimental Comparison of Range Image Segmentation Algorithms, IEEE Trans. on PAMI, vol. 18, n. 7, Jul. 1996, pp. 673-689

[7] J.H. Yi, D.M. Chelberg, Model based 3D object recognition using bayesian indexing, Comp. Vis. and Image Underst., vol. 69, n. 1, Jan. 1998, pp. 87-105

[8] Chin-Hsiung Wu, Shi-Jinn Horng, Pei-Zong Lee, A new computation of shape moments via quadtree decomposition, Patt. Rec., vol. 34, no. 7, 2001, 1319-1330

[9] J. Zunic, N. Sladoje, Efficiency of Characterizing Ellipses and Ellipsoids by Discrete Moments, IEEE Trans. on PAMI, 22(4):407, 2000

[10] F. Mokhtarian, S. Abbasi, Shape similarity retrieval under affine transforms, Pattern Recognition, vol. 35, no. 1, 2002, 31-41

[11] S. Berretti, A. Del Bimbo, P. Pala. Retrieval by shape similarity with perceptual distance and effective indexing IEEE Trans. on Multimedia, V.2 N.4 , Dec. 2000, pp: 225 -239

[12] T.T. Elvins, R. Jain. Web-based Volumetric Data Retrieval. In Proc. ACM, 1995

[13] E. Paquet, M. Rioux, "Nefertiti: a query by content system for three-dimensional model and image database management", *Image Vision and Computing*, 17:157-166, 1999.

[14] Y. Lamdan, H.J. Wolfson. Geometric Hashing: A General and Efficient Model-Based Recognition Scheme. In Proc. IEEE Int. Conf. on Comp. Vis. ICCV'88, Tampa, Florida, 1988, pp.238-249

[15] H.P. Kriegel, T. Seidl. Approximation-Based Similarity Search for 3D Surface Segments. GeoInformatica Journal, 2(2):113-147, Kluwer Academic Publisher, 1998

[16] S. Berchtold, H.P. Kriegel, S3: Similarity Search in CAD Database Systems. In Proc. of the ACM SIGMOD Int. Conf. on Management of Data, May 11-15, 1997, Tucson, AZ, USA, pp. 564-567

[17] M. Ankerst, G. Kastenmuller, H.P. Kriegel, T. Seidl, 3D Space Histograms for Similarity Search and Classification in Spatial Databases. Proc. 6th Int. Symposium on Spatial Databases (SSD'99) Hong Kong, China, July 1999

[18] J.S. Beis, D.G. Lowe, Indexing Without Invariants in 3D Object Recognition, IEEE Trans. on PAMI, 21(10):1000-1015, Oct. 1999

[19] C. Colombo, A. Del Bimbo and F. Pernici, "Uncalibrated 3D metric reconstruction and flattened texture acquisition from a single view of a surface of revolution." In *Proc. 1st Int. Symp. on 3D Data Proc. Vis. and Trans. 3DPVT 2002*, Padua, Italy, June 2002, pages 277–284, IEEE 2002.

[20] G. Taubin, "Estimating the Tensor of Curvature of a Surface from a Polyhedral Approximation", *5th Int. Conf. on Comp. Vis. ICCV'95*, 1995.

[21] D. L. Page, A. Koschan, Y. Sun, J. Paik, and M. A. Abidi, "Robust Crease Detection and Curvature Estimation of Piecewise Smooth Surfaces from Triangle Mesh Approximations Using Normal Voting", in *Proc. of the Comp. Vis. and Patt. Rec. Conf. CVPR'01*, Kauai, Hawaii, 11-13 December 2001.

[22] J.P. Snyder, L.M. Bugayevski, *Map projections—a reference manual*, Taylor & Francis, 1995.