3D Database Population from Single Views of Surfaces of Revolution

C. Colombo, D. Comanducci, A. Del Bimbo, and F. Pernici

Dipartimento di Sistemi e Informatica,
Via Santa Marta 3, I-50139 Firenze, Italy
{colombo, comandu, delbimbo, pernici}@dsi.unifi.it

Abstract. Solids of revolution (vases, bottles, bells, ...), shortly SORs, are very common objects in man-made environments. We present a complete framework for 3D database population from a single image of a scene including a SOR. The system supports human intervention with automated procedures to obtain good estimates of the interest entities needed for 3D reconstruction. The system exploits the special geometry of the SOR to localize it within the image and to recover its shape. Applications for this system range from the preservation and classification of ancient vases to advanced graphics and multimedia.

1 Introduction

3D object models are growing in importance in several domains such as medicine, architecture, cultural heritage and so on. This is due to their increasing availability at affordable costs, and to the establishment of open standards for 3D data interchange (e.g. VRML, X3D). A direct consequence of this is the creation of 3D databases for storing and retrieval; for this purpose many different descriptions of 3D shape have been devised.

Based on the 3D shape representation used, various 3D object retrieval methods have been developed. These can be grouped into methods based on *spatial decomposition*, methods working on the *surface* of the object, *graph-based* methods, and *2D visual similarity* methods. Spatial decomposition methods [1][2] subdivide the volume object into elementary parts and compute some statistics of occurrence of features extracted from them (also in histogram form [3]). Surface-based methods [4] rely on the extraction of salient geometric features of the object: they can range from basic features as bounding box and axes of inertia to more sophisticated measurements such as moments and position of curvature salient points. In [5], the 3D problem is converted into an image indexing problem, representing the object shape by its curvature map. Graph-based approaches employ the object topology represented by a graph; thus, the similarity problem is reduced to a graph comparison problem. Finally, 2D visual similarity methods compare the appearance of the objects using image-based descriptions, typically derived from several views.

3D shape acquisition of a real object can be done through CAD (fully manual authoring) or through automatic techniques such as 3D laser scanner and

imaged-based modeling. Laser scanners use interference frequencies to obtain a depth map of the object. Image-based modeling methods rely instead on camera calibration and can be subdivided into active and passive methods. Active methods employ structured light projected onto the scene. Though conceptually straightforward, structured-light scanners are cumbersome to build, and require expensive components. Computationally more challenging, yet less expensive ways to obtain 3D models by images are passive methods, e.g. classic triangulation [8], visual hulls [6], geometry scene constraints [7]. A central role is assumed here by self-calibration, using prior knowledge about scene structure [9] or camera motion [10].

To obtain a realistic 3D model, it is important to extract the texture on the object surface too. Basically, one can exploit the correspondence of selected points in 3D space with their images, or minimize the mismatch between the original object silhouette and the synthetic silhouette obtained by projecting the 3D object onto the image. Texture can also be used for retrieval purposes, either in combination with 3D shape or separately.

In this paper we propose a 3D acquisition system for SORs from single uncalibrated images, aimed at 3D object database population. The paper is structured as follows: in section 2, the system architecture is described; section 3 provides an insight into the way each module works. Finally, in section 4, results of experiments on real and synthetic images are reported.

2 System Architecture

As shown in Fig. 1, the system is composed by three main modules:

1. Segmentation of SOR-related image features;
2. 3D SOR shape recovery;
3. Texture extraction.

Module 1 looks for the dominant SOR within the input image and extracts in an automatic way its *interest curves*, namely, the “apparent contour” and at least two elliptical “imaged cross-sections.” The module also estimates the parameters of the projective transformation characterizing the imaged SOR symmetry.

The interest curves and the imaged SOR symmetry are then exploited in Module 2 in order to perform camera self-calibration and SOR 3D shape reconstruction according to the theory developed in [11]. The particular nature of a SOR reduces the problem of 3D shape extraction to the recovery of the SOR *profile*—i.e., the planar curve generating the whole object by a rotation of 360 degrees around the symmetry axis. The profile can be effectively employed as a descriptor to perform SOR retrieval by shape similarity.

Module 3 exploits the output of Module 1 together with raw image data to extract the visible portion of the SOR texture. In such a way, a description of the object in terms of its photometric properties is obtained, which complements the geometric description obtained by Module 2.

Image retrieval from the database can then take place in terms of shape-related and/or texture-related queries.

Fig. 1. System architecture

3 Modules and Algorithms

Module 1 employs the automatic SOR segmentation approach described in [12]. The approach is based on the fact that all interest curves are transformed into themselves by a single four degrees of freedom projective transformation, called *harmonic homology*, parameterized by an axis (2 dof) and a vertex (2 dof). The approach then consists in searching simultaneously for this harmonic homology and for the interest curves as the solution of an optimization problem involving edge points extracted from the image; this is done according to a multiresolution scheme. A first estimate of the homology is obtained by running the RANSAC algorithm at the lowest resolution level of a Gaussian pyramid, where the homology is well approximated by a simple axial symmetry (2 dof). New and better estimates of the full harmonic homology are then obtained by propagating the parameters through all the levels of the Gaussian pyramid, up to the original image (Fig. 2(a)). In particular, the homology and the interest curves consistent with it are computed from the edges of each level, by coupling an *Iterative Closest Point* algorithm (ICP) [13] with a graph-theoretic curve grouping strategy based on a Euclidean Minimum Spanning Tree [14]. Due to the presence of distractors, not all of the image curves thus obtained are actually interest curves; this calls for a curve pruning step exploiting a simple heuristics based on curve length and point density. Each of the putative interest curves obtained at the end of the multiresolution search is then classified into one of three classes: “apparent contour,” “cross-section,” and “clutter.” The classification approach exploits the tangency condition between each imaged cross-section and the apparent contour.

As mentioned above, SOR 3D reconstruction (Module 2) is tantamount to finding the shape of its profile. This planar curve is obtained as the result of

Fig. 2. From images to descriptions. (a): The input image. (b): The computed interest curves (one of the two halves of the apparent contour, two imaged cross-sections), the harmonic homology parameters (imaged SOR axis, vanishing line), and the recovered imaged meridian (dotted curve). (c): The reconstructed SOR profile. (d): The flattened texture acquired from the input image.

a two-step process: (a) transformation of the apparent contour into an imaged profile (see Fig. 2(b)); (b) elimination of the projective distortion by the rectification of the plane in space through the profile and the SOR symmetry axis (see Fig. 2(c)). Step *a* requires the knowledge of the homology axis and of the vanishing line of the planes orthogonal to the SOR symmetry axis; the latter is computed in Module 1 from the visible portions of at least two imaged

Table 1. System failures (automatic procedure) for a dataset of 77 images

type of system failure	%
bad edge extraction	24.68
harmonic homology not found	5.19
bad curve-grouping	6.49
bad ellipse extraction	12.99
total failures	49.35

cross-sections. Step *b* requires calibration information, also extracted from the homology and from the imaged cross-sections.

Texture acquisition is performed in Module 3. It exploits the calibration information obtained in Module 2, together with the vanishing line of the SOR cross-sections and the homology obtained in Module 1. The coordinates of the texture image are the usual cylindrical coordinates. Texture acquisition is performed by mapping each point in the texture image onto the corresponding point in the original image. This is equivalent to projecting all the visible SOR points onto a right cylinder coaxial with the SOR, and then unrolling the cylindrical surface onto the texture image plane. Fig. 2(d) shows the texture acquired from the original image of Fig. 2(a). Notice that, since the points that generate each half of the apparent contour generally do not lie on the same meridian, the flattened region with texture is not delimited with vertical lines.

4 Experimental Results

In order to assess the performance of the system, experiments have been conducted both on synthetic and real images. The former allowed us to use a ground truth and simulate in a controlled way the effect of noise and viewpoint on system percentage of success and accuracy. Experiments with real images helped us to gain an insight into system dependency on operating conditions and object typologies.

We indicate as “system failures” all the cases in which the system is unable to complete in a fully automatic way the process of model (shape, texture) extraction. System failures are mainly due to the impossibility for Module 1 to extract either the harmonic homology or the interest curves from raw image data: this can happen for several reasons, including large noise values, scarce edge visibility and poor foreground/background separation, specular reflections of glass and metallic materials, distractors as shadows and not-SOR dominant harmonic homologies, and the absence of at least two visible cross-sections. Tab. 1 shows the relative importance of the principal causes of system failures for a set of 77 images taken from the internet. The table indicates that most of the failures are due to the impossibility, in almost a quarter of the images, to obtain a sufficient number of edge points to correctly estimate the interest curves. The percentage of success for fully automatic model extraction is then slightly greater than 50%.

Table 2. Profile reconstruction error ε_r : mean value, standard deviation, min value, max value and number of failures; due to RANSAC initialization, the system has a random behavior also at zero noise level

σ	mean	std	min	max	failures
0	0.0010	0.0002	0.0008	0.0016	0
0.1	0.0011	0.0002	0.0008	0.0015	1
0.2	0.0013	0.0003	0.0009	0.0021	1
0.4	0.0014	0.0003	0.0010	0.0032	1
0.8	0.0021	0.0015	0.0011	0.0110	0
1.6	0.0039	0.0061	0.0013	0.0334	21

To cope with system failures, the system also includes a semi-automatic and a fully manual editing procedures. The former procedure involves only a slight intervention on the user part, and is aimed at recovering only from the edge extraction failures by manually filling the gaps in the portions of interest curves found automatically. Thanks to the semi-automatic procedure, system success increase to about 75% of the total, residual failures being due to bad homology estimates and misclassifications. The fully manual procedure involves the complete specification of the interest curves, and leads to a percentage of success of 100%. The three kinds of model extraction procedures give the user the opportunity to rank the input images into classes of relevance; in particular, models of the most relevant objects can in any case be extracted with the fully manual procedure.

If the system has completed its task on a given image (either with the automatic, semi-automatic, or manual procedure), it is reasonable then to measure the accuracy with which the shape of the profile is acquired. The profile reconstruction error is defined as

$$\varepsilon_r = \int_0^1 |\rho(t) - \hat{\rho}(t)| dt, \quad (1)$$

where $\rho(t)$ and $\hat{\rho}(t)$ are respectively the ground truth profile and the estimated one. Tab. 2 shows results using the ground truth profile

$$\rho(t) = \frac{1}{10}(\cos(\frac{\pi}{2}(\frac{19}{3}t + 1)) + 2) \quad (2)$$

of Fig. 3, for increasing values of Gaussian noise (50 Monte Carlo trials for each noise level $\sigma = 0, 0.1, 0.2, 0.4, 0.8, 1.6$).

If the error ε_r is regarded as the area of the region between $\rho(t)$ and $\hat{\rho}(t)$, we can compare it with the area given by the integral of $\rho(t)$ on the same domain: 0.1812. The error appears to be smaller by several orders of magnitude than this value, even at the highest noise levels. We can also notice how the number of failures increase abruptly for $\sigma = 1.6$; the failures are mainly due to a bad RANSAC output at the starting level.

Fig. 3. Synthetic SOR views for $\sigma = 0$ (left) and $\sigma = 1, 6$ (right)

5 Discussion and Conclusions

A 3D database population system has been presented, in which both shape and texture characteristics of SOR objects are acquired. Applications range from advanced graphics to content-based retrieval. Suitable representations of SOR models can be devised for retrieval purposes. Specifically, the SOR profile can be represented through its spline coefficients, and a measure of profile similarity such as that used in Eq. 1 can be used for shape-based retrieval. Texture-based retrieval can be carried out by standard image database techniques [15].

The experiments show that not all photos can be dealt with automatically, and that in half of the cases either a semi-automatic or a fully manual procedure have to be employed to carry out model extraction successfully. In order to obtain a fully automatic model extraction in any case, further research directions should include better edge extraction algorithms (color edges, automatic edge thresholding), a better grouping algorithm than EMST at the end of ICP, and some heuristics to discriminate whether a SOR is present in an image or not. In fact, EMST is good for outlier rejection inside ICP for its speed, but a better curve organization before curve classification can improve ellipse extraction; on the other hand, a SOR-discriminant heuristics can make the system able to work on a generic data-set of images.

References

1. Paquet, E., Rioux, M., A.Murching, Naveen, T., Tabatabai, A.: Description of shape information for 2D and 3D objects. *Signal Processing: Image Communication* **16** (2000) 103–122
2. Kazhdan, M., Funkhouser, T.: Harmonic 3D shape matching. In: *SIGGRAPH '02: Technical Sketch*. (2002)
3. Ankerst, M., Kastenmüller, G., Kriegel, H.P., Seidl, T.: 3D shape histograms for similarity search and classification in spatial databases. In: *SSD '99: Proceedings of the 6th International Symposium on Advances in Spatial Databases* (1999) 207–226

4. Zaharia, T., Prêteux, F.: Shape-based retrieval of 3d mesh-models. In: ICME '02: International Conference on Multimedia and Expo. (2002)
5. Assfalg, J., Del Bimbo, A., Pala, P.: Curvature maps for 3D Content-Based Retrieval. In: ICME '03: Proceedings of the International Conference on Multimedia and Expo. (2003)
6. Szeliski, R.: Rapid octree construction from image sequences. *Computer Vision, Graphics, and Image Processing* **58** (1993) 23–32
7. Mundy, J., Zisserman, A.: Repeated structures: Image correspondence constraints and ambiguity of 3D reconstruction. In Mundy, J., Zisserman, A., Forsyth, D., eds.: *Applications of invariance in computer vision*. Springer-Verlag (1994) 89–106
8. Hartley, R.I., Zisserman, A.: *Multiple View Geometry in Computer Vision*. Cambridge University Press (2003)
9. Pollefeys, M.: Self-calibration and metric 3D reconstruction from uncalibrated image sequences. PhD thesis, K.U. Leuven (1999)
10. Jiang, G., Tsui, H., Quan, L., Zisserman, A.: Geometry of single axis motions using conic fitting. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (**25**) (2003) 1343–1348
11. Colombo, C., Del Bimbo, A., Pernici, F.: Metric 3D reconstruction and texture acquisition of surfaces of revolution from a single uncalibrated view. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (**27**) (2005) 99–114
12. Colombo, C., Comanducci, D., Del Bimbo, A., Pernici, F.: Accurate automatic localization of surfaces of revolution for self-calibration and metric reconstruction. In: *IEEE CVPR Workshop on Perceptual Organization in Computer Vision (POCV)* (2004)
13. Zhang, Z.Y.: Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision* **13** (1994) 119–152
14. de Figueiredo, L.H., Gomes, J.: Computational morphology of curves. *The Visual Computer* **11** (1995) 105–112
15. Del Bimbo, A.: *Visual Information Retrieval*. Morgan Kaufmann (1999)