Content-based 3-D Shape Retrieval for Pervasive Computing

Dong Xu\textsuperscript{1,2,3,4}, Hua Li\textsuperscript{1,2,3}, Zongkai Lin\textsuperscript{1}

\textsuperscript{1} Key Laboratory of Intelligent Information Processing, Chinese Academy of Sciences
\textsuperscript{2} Key Laboratory of Computer System and Architecture, Institute of Computing Technology, Chinese Academy of Sciences
\textsuperscript{3} National Research Center for Intelligent Computing Systems, Institute of Computing Technology, Chinese Academy of Sciences
\textsuperscript{4} Graduate University of Chinese Academy of Sciences, P. R. China
\{xudong,lihua,lzk\}@ict.ac.cn

Abstract

3-D models are a new kind of cross-media resource which can be frequently seen in the network. Since the amount of them is very huge now, content-based retrieval can help to recognize a certain object or retrieve similar ones from the giant database. This paper presents a new method for deriving 3-D moment invariants and uses them as shape descriptors for the representation of 3-D models. They are insensitive to surface noise and can be used in pervasive environment conveniently. We also illustrate how to build up experimental system and simulate 3-D shape retrieval in wireless environment.

Keywords: Cross-Media, Pervasive Computing, 3-D Shape Retrieval, Moment Invariants, Wireless Network.

1. Introduction

Cross-media is the general designation of traditional medias (newspaper, magazine, television, broadcast etc) and digital medias spreading over the network (text, image, audio, video etc). Many services such as training and entertainment can be carried out through different digital media channels. Both the formats and quantities of digital medias are increasing quickly accompany with the development of technologies of hardware, software and network. How to manage or retrieval cross-media resources is a problem presented to all of us.

Mark Weiser in [1] described a world in the near future, computers work harmoniously and transparently with human users, serving human users in different aspects of their lives. Since about 83\% of the information obtained by human beings is from vision, and about 11\% is from hearing, it is preferred that more visual and acoustical information is provided to users in pervasive computing environment. The ideal computer system would become invisible to users and the users’ conscious notion of computer hardware would begin to disappear, because both mobile client and infrastructure systems can work seamlessly and unobtrusively [2].

Fortunately, visual and acoustical applications which only exist in PCs within laboratories have gradually gone into the Internet, then the pervasive environment. For instance, face recognition and object tracking, which are two research topics in pattern recognition and computer vision, are now applied for pervasive computing with more friendly interactive functions and more implemental details to concern about. Gao et al [3] described a face recognition system that used significance-based multiview Hausdorff distance (SMVHD) for authentication. Peursum et al [4] presented the concept of identifying and classifying objects floors within a scene by detecting and recognizing activities that humans perform when interacting with these objects. Civilis et al [5] proposed a segment-based policy that predicted an object’s movement according to the road’s shape.

Speech recognition has also been considered into the pervasive computing environment. Starner [6] pointed out there were several challenges to using speech recognition in general contexts, and interface designers must be wary of applying the technology to situations where speech is inappropriate. Alewine et al provided an overview of embedded automatic speech recognition on the pervasive devices and discussed its ability to help develop pervasive applications that met today’s marketplace needs [7].

Now, 3-D models can be found in the Internet frequently, which belong to a new kind of digital media. In this paper, we make a preliminary attempt to extend the 3-D shape retrieval to wireless environment. The goal is that people can retrieval or recognize an object anywhere and anytime. Only some sketches or images of the object are needed as query conditions.
The remainder of the paper is organized as follows. In section 2, we introduce the related work in the area of 3-D shape retrieval. We present a new method for feature extraction in section 3. System constitution and simulation of 3-D shape retrieval in wireless environment are illustrated in section 4. We conclude the paper and open perspectives for future work in the end.

2. Related work

Digital media resources mainly include documents and multimedia materials. The latter now enriches with various of medias, such as images, audios and video, as shown in Fig. 1. With the rapid development of 3-D acquisition techniques and modeling methods, large amounts of 3-D objects appear in front of us. 3-D models have become an inseparable part of multimedia resources besides audio, image and video etc.

Fig. 1. Constitution of digital media resources.

All of these resources can describe the same thing and are mixed together oftentimes. Hence, it is necessary to retrieval in the huge digital media database to get the specific description of the object. In recent decades, people have done much research to semantic-based document retrieval and content-based image retrieval. It is also necessary for us to recognize an unknown object or retrieve similar ones from the giant database quickly.

MPEG-7 is multimedia content description interface, which standardizes detailed formatting information and fine-grained descriptions of low-level and high-level features. It is believed that the MPEG-7 standard can unify and manage all the multimedia resources and make the applications of them easy to transplant to any platform. In [8], Paquet and Rioux presented three shape descriptors based on cords, wavelet transform and moments for 3-D data. They also introduced how to integrate them in the framework of the MPEG-7 standard.

Retrieval of document and multimedia resources has already been considered in pervasive environment. Kammanahalli et al [9] proposed a context aware information retrieval system that retrieves and presents information of high utility to a user. Their system could take into account user’s implicit need for information, capabilities of the user’s access devices and other resource constraints of the user.

Content-based retrieval of 3-D objects means we retrieve or recognize the object by its 3-D shapes since there are not adequate semantic tags to describe them. Image-based retrieval is also not enough here because of the different poses of the object. For example, there are maybe two, three or four legs of a horse which can be seen in different images with different viewpoints.

Many methods have been developed for feature extraction of 3-D objects. Some literatures transformed 3-D models to various 2-D graphs, such as Reeb Graph [10], [11], Shock Graph [12], skeleton [13]. LightField [14] or characteristic views [15], [16] were extracted from several directions, and then were regarded as images for comparison. Other literatures were mainly based on the surface or volume features of the 3-D models directly. These features should be independent of the transformation in 3-D space, like translation and rotation. 3-D moments are not invariant under rotation, but the pose of the object can be adjusted by Principal Component Analysis (PCA) [17], [18]. 3-D polar-radius moment invariants [19] and Zernike invariants [20] could be derived from 3-D moments and then used to represent the feature of objects.

Princeton Shape Benchmark [21] provided a publicly available database from World Wide Web and a suite of tools for comparing shape matching and classification algorithms. There have already been several 3-D search engines in the Internet, such as Princeton 3-D model Search Engine [22] (http://shape.cs.princeton.edu/search.html ). The retrieval results are often represented by the typical thumbnail pictures of 3-D models.

3. Feature description by moment invariants

3.1. A method to derive 3-D moment invariants under similarity transformation

By now, the amount of 3-D moment invariants is very small. There were 12 moment invariants derived in paper [23] by group-theoretic technique. In this part, we present a new and intuitive method to derive 3-D moment invariants and use them as shape descriptors for the representation of 3-D solid objects.
3-D moments of order $l+m+n$ of a 3-D density function $\rho(x, y, z)$ are defined by the Riemann integral:

$$M_{lmn} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^l y^m z^n \rho(x, y, z) dx dy dz$$

(1)

Suppose $(x_i, y_i, z_i), (x_j, y_j, z_j), \ldots, (x_k, y_k, z_k)$ are arbitrary points in the object. The below four geometric primitives are distance of two points, area of a triangle between three points, inner product of two vectors between three points and volume of a tetrahedron between four points. They are invariant under rotation and translation.

$$D(i, j) = [(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2]^\frac{1}{2}$$

(2)

$$A(i, j, k) = \frac{1}{2} \| (x_i - x_j, y_i - y_j, z_i - z_j) \times (x_i - x_k, y_i - y_k, z_i - z_k) \|_2$$

(3)

$$An(i, j, k, l) = (x_i - x_j, y_i - y_j, z_i - z_j) \cdot (x_j - x_k, y_j - y_k, z_j - z_k)$$

(4)

$$V(i, j, k, l) = \frac{1}{6} [(x_i - x_j, y_i - y_j, z_i - z_j) \cdot (x_j - x_k, y_j - y_k, z_j - z_k)]$$

(5)

Then, we set $\text{core}(1,2,\ldots,n)$ to be the multiplication of the above four primitives, which involves $n$ participating points. The four primitives in $\text{core}(1,2,\ldots,n)$ can be of different powers. The multiple integral of $\text{core}(1,2,\ldots,n)$ is just the moment invariants if we expand $\text{core}(1,2,\ldots,n)$ by polynomial form

$$\text{core}(1,2,\ldots,n) = \sum_i a_i \prod_{j=1}^{n} x_j^{i_1} y_j^{i_2} z_j^{i_3}.$$  

That is, moment invariants of different orders can be constructed by the following integral formula.

$$Q(\text{core}(1,2,\ldots,n)) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \text{core}(1,2,\ldots,n) \rho(x_1, y_1, z_1) \rho(x_2, y_2, z_2) \ldots \rho(x_n, y_n, z_n)$$

$$dx_1 dy_1 dz_1 dx_2 dy_2 dz_2 \ldots dx_n dy_n dz_n$$

(6)

In order to construct invariants independent of scaling, the above formula can be then divided by a certain power of $M_{000}$, which is the volume of the 3-D object. Here, we give an explicit expression of a fourth-order moment invariant below.

$$I = Q(D^4(1,2))$$

$$= \frac{2}{10} \frac{l!}{l_0! l_1! l_2! l_3!} (M_{000} + M_{001} + M_{010} + M_{011} + M_{020} + M_{021} + M_{030} + M_{031} + M_{100} + M_{101} + M_{110} + M_{111} + M_{120} + M_{121} + M_{130} + M_{131} + M_{200} + M_{201} + M_{210} + M_{211} + M_{220} + M_{221} + M_{230} + M_{231} + M_{300} + M_{301} + M_{310} + M_{311} + M_{320} + M_{321} + M_{330} + M_{331})$$

(7)

Since high-order moment invariants often have a huge number of moment terms, now they are generated automatically by symbolic computation softwares, like Maple.

### 3.2. Feature extraction of 3-D models

Density distribution function $\rho(x, y, z)$ determines the shape of an object uniquely, and it can be recovered by inverse Fourier transform if all orders of moments have been obtained. Hence, each object can be uniquely determined by an infinite sequence of moment invariants theoretically.

To each 3-D solid model, we use a voxelization method in computer graphics to get $N$ discrete inner sampling points. Moments of all orders are approximately computed by the following discrete form

$$M_{lmn} = \sum_{i=1}^{N} x_i^l y_i^m z_i^n \Delta x \Delta y \Delta z$$

where $\Delta x$, $\Delta y$ and $\Delta z$ are sampling intervals in three directions.

In practice, low-order moment invariants are more useful because the computation time will be less and the results will be stable to noise and distortion. Based on the above analysis, we choose moment invariants below fifth-order as shape descriptors. In other words, each object is represented by a 54-dimensional vector approximately.

### 3.3. Predominance of moment invariants for pervasive computing

It was pointed out in [22] that the principal axes might quite different to some similar models, so 3-D moments after PCA are not stable and invariant shape descriptors. Moreover, the moments in [17] and [18] are surface moments, which are more sensitive to noise. Polar-radius moment invariants in [19] can be derived by the method presented here. They also require that the moments should be central moments.

In contrast, moment invariants are very suitable for shape representation. They are insensitive to surface noise for the moments here are the integrals of the whole 3-D solid model. The numerical values of moment invariants keep invariant under arbitrary similarity transformation.

The second advantage of moment invariants is their efficient discrimination. Similar objects often have close feature vectors. The most similar retrieval results can be firstly presented to users with top ranks.

Each object is represented by a 54-dimensional vector in this paper. The dimension can be controlled dynamically because more high-order moment invariants can be added in if necessary. This kind of shape descriptor is a compact expression of a complicated 3-D model. Hence, the third advantage is...
that feature vectors of all the models in the database can be pre-computed offline and saved with little storage space in pervasive environment.

4. Shape retrieval in pervasive environment

One of the goals for pervasive computing is that users can accomplish their tasks anywhere, anytime and ignore the existence of computer system. Moreover, it should have friendly user interface. In this section, we propose a proper solution for shape retrieval in wireless environment.

4.1. Experimental requirement

4.1.1 Equipments. In order to conduct the shape retrieval experiment in wireless environment, we need the following basic equipments and environments—Wireless Local Area Network (WLAN), SmartPhones or Personal Digital Assistants (PDAs) (with digital pc camera or handheld pen) and server (including processor and memorizer). Both the mobile devices and server should have transmitter and receiver. If the data for transmission are huge or users require less transfer time, we also need encoder and decoder components for data compress and uncompress.

The basic functional components of the system are illustrated in Fig. 2.

4.1.2 Database. Most of the 3-D models in Princeton Shape Benchmark [21] are “polygon soups” [22], which contain missing, overlapping, and/or wrongly oriented polygons. These models can’t form closed volumes and are hard to repair [22]. Many efforts have been done for adjusting the polygons and for the voxelization of these models approximately. Moreover, we build our database containing 119 common objects for this experiment. Most of the objects are cell phones, cars, pens and cups.

Feature vectors of all the models in the database are pre-computed offline by the method proposed in section 3.2 and we build a table to store them. If the table is very huge, the nearest neighbor search can be accelerated by building k-d tree [24] or auxiliary index tables for the feature vectors.

4.1.3 Algorithms. Traditional 3-D object formats contain information of 3-D coordinates and polygonal patches, which can not obtained easily in mobile devices as query condition. In this experimental system, the input can be multiple views of the object taken by pc camera, or sketches drawn by users with handheld pens. These are two common input components in current mobile devices and can be manipulated easily.

By now, the computational ability and storage space are limited in mobile devices. Hence, it is necessary that the database and the computation are carried out on the server end. Images or graphs collected by the input devices in mobile phones, are then compressed and sent to the server. We reconstruct a coarse 3-D model from them by a three-dimensional sculpture method. Meanwhile, the voxels of the object are obtained and moment invariants of low orders can be computed quickly. The feature vector of the query object is calculated from these moment invariants on the server.

Then, we compute the Euclidean distances of feature vectors between the query object and all the models in the database and sort the distances in descending orders. Models with high similarities are then compressed and sent to the mobile devices, while each model is represented by one of its multiple views.

Algorithms for implementing the online shape retrieval in wireless environment are summarized as the following five steps:

(1) Image or graph acquisition
(2) 3-D reconstruction of the query object
(3) Feature extraction
(4) Shape retrieval from the database
(5) Display of the retrieval results

4.2. Wireless programming

There are several ways to program in mobile devices to make them with more functions such as games. In
In this part, we discuss two possible schemes to implement the retrieval system in mobile devices.

Wireless application protocol (WAP) is an application environment and set of communication protocols for wireless devices designed to enable manufacturer-vendor-and-technology-independent access to the Internet and advanced telephony services. The wireless device contains a microbrowser, while content and applications are hosted on Web servers. The reuse of existing Web technologies eases the development of WAP applications, and makes it similar to developing HTML-based Web applications since it is browser-based. Cgogo Inc has provided mobile search services such as “WAP Search”, “Internet Search”, “Product Search” and “Mobile WAP Yellow Pages” to mobile end users (wap.cgogo.com). These searches belong to the category of information search or semantic search. It is not difficult that users can browse the shape retrieval results in mobile devices by editing the retrieval results in Web servers with Wireless Markup language (WML). The implementation also needs the support from special service providers.

Another approach to developing wireless applications is to use the Java 2 Platform Micro Edition (J2ME) Mobile Information Device Profile (MIDP). MIDP is also an open specification that adapts existing technologies such as Java and the Web. Developing MIDP-based applications (also known as MIDlets) is similar, but not identical, to developing Java Applets in the sense they share a similar programming model.

An introduction of J2ME for pervasive computing is in [25]. Pervasive Java is a surge of wireless and mobile platform standardization activities led by Sun Microsystems. It attempts to bridge the gap between disparate devices and platforms, and profitable business and application development. It is transforming mobile devices into essential players and integrated elements in the pervasive computing world.

4.3. Simulation

In this part, we simulate the 3-D shape retrieval in mobile phones. The basic components and algorithms for implementing the system have been discussed above in section 4.1. Here, we perform the simulation of this process with J2ME in a personal computer and give a clear illustration of the retrieval interface.

The retrieval result of a car can be seen in Fig. 3. Most of the similar cars in the database are retrieved with top ranks, which proves the discrimination ability of our shape descriptor. The images are of small size to reduce transfer delay and for convenient display in the screen of the wireless devices. Images of about 15 objects with high similarities are sent back to the mobile phone in descending orders. If users are interested in one of the retrieval results, they can select it as a new object for a new round of retrieval. In this situation, it takes less computational time because we do not need the 3-D reconstruction process and the sorted orders of this object have already been pre-computed offline.

5. Conclusion and future work

The main contribution of the paper is that we propose a new method to derive 3-D moment invariants under similarity transformation, and use them as shape descriptors for shape retrieval in pervasive environment. The system has a friendly user interface and is very helpful if users want to recognize a special object or find similar ones.

Face recognition technique for identity authentication can be carried out in pervasive environment in a similar way. Taking a picture of a suspicious person, we can get the retrieval results from the database and then give the judgment. Face recognition systems in pervasive environment can applied into a broader area conveniently, compared with the immovable ones.

In the future, users can conduct mixed cross-media retrieval in portable devices, which means both the query condition and the retrieval results can be the mixture of multimedia and document resources, such
as sounds, images, videos and texts. For example, detailed document description can be returned to users and help them to know the characteristics of the object more completely. We can also embed the experimental system with feedback mechanism to enhance more interactive functions.

Acknowledgements

This work is supported by National Natural Science Foundation of China (60533090).

References