# Computer Graphics and Vision Methods for the Reconstruction, Representation and Retrieval of 3D Objects with Biometric Applications

Georgios Passalis

Computer Graphics Group *,
Department of Informatics and Telecommunications,
University of Athens,
passalis@di.uoa.gr,
WWW home page: http://graphics.di.uoa.gr/

**Abstract.** The increased availability of 3D objects, widens their use in computer graphics, computer vision and biometrics applications. In most applications where 3D object databases are utilized, the goal is retrieval, which is the ability to compare and categorize objects. In this work, the general problem of managing 3D information is tackled, for the reconstruction, representation and retrieval of 3D objects. Initially, a new method for the reconstruction of 3D objects based on Integral Photography is presented. Subsequently, two 3D object representations are utilized by two novel retrieval methods. The first representation, based on depth images, is suitable for inter-class object retrieval while the second, based on geometry images, is suitable for intra-class object retrieval. The intra-class object retrieval method is further specialized and applied in the field of biometrics in order to perform 3D face and 3D ear recognition. The performance of the proposed method is evaluated using the largest publicly available biometric databases. The face recognition results that are presented are considered state-of-the-art worldwide.

## 1 Introduction

The rapid increase in computational power allows the processing of increasingly complex information. This processing was extended from 1D information (text) to 2D information (image) and recently to 3D information (geometry). 3D information is utilized in several domains such as computer graphics, computer vision and biometrics. The use of 3D information creates the need to organize it into databases. Regardless of the dimensionality of the information, certain methods are required for this task. More specifically, there is the need to represent the information in a way that allows efficient storage and comparison in order to perform content-based retrieval.

Content-based retrieval is the task where given a dataset, similar datasets must be retrieved from the database, without any *a priori* knowledge regarding

* Dissertation Advisor: Theoharis Theoharis, Associate Professor

these datasets. In the 3D case, this task is called 3D object retrieval. Retrieval is directly affected by the representation used for the 3D objects. Therefore it is not possible to tackle the 3D object retrieval problem without tackling the 3D object representation problem. The main issue with 3D objects is the lack of a unified representation suitable for all applications. Each representation has its own advantages and shortcomings and their importance depends heavily on the specific application. For example, the most widespread representation, the polygonal mesh, is considered very efficient for rendering and flexible enough for modelling. On the contrary, it is not suitable for retrieval and not very efficient for storage.

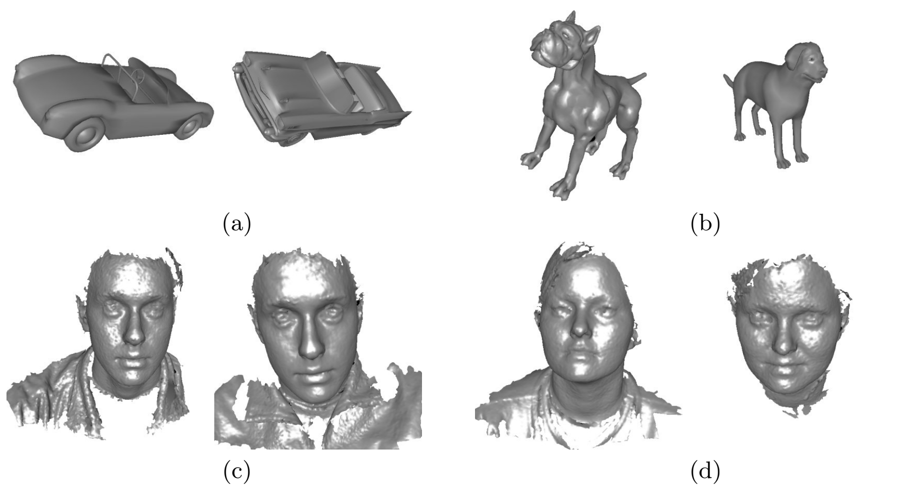



**Fig. 1.** Inter-class database: (a) 2 objects of car class, (b) 2 objects of dog class. Intra-class database: (c) 2 objects of first subjects, (d) 2 objects of second subject.

Even in the field of 3D object retrieval, there are contradicting needs that the selected representation must cover. Inter-class object retrieval is applied on databases with arbitrary object classes (Fig. 1 (a,b)). In this case, the retrieval method must be general enough to categorize objects of any class. Intra-class object retrieval is applied on databases where all objects belong to the same class (Fig. 1 (c,d)). In this case, the retrieval method must sacrifice generality in order to discern subtle differences among objects of the same class. Intra-class retrieval is widely used in biometric applications, such as 3D face recognition and 3D ear recognition.

In this work, two 3D object retrieval methods are proposed, the first is suitable for inter-class retrieval while the second is suitable for intra-class retrieval. The intra-class retrieval method is subsequently applied on 3D face and 3D ear recognition. Each one is based on a different representation. Both representations utilize the geometric nature of the information and reduce its dimensionality. The first representation is based on depth images which are retrieved by a general method that can also produce a voxel representation. The second representation is based on geometry images which are retrieved by fitting an annotated model using a subdivision-based deformable framework. Additionally a 3D object reconstruction method is presented that utilizes Integral Photography images. For all proposed methods, extensive tests and comparisons are performed that highlight their accuracy and efficiency. The results of the 3D face recognition method are considered state-of-the-art worldwide.

## 2 Related Work

### 2.1 3D Reconstruction using Integral Photography

In the past, two camera setups (stereovision) were mainly used for 3D object reconstruction. Recently, this has been extended to multiple views, either by the use of multiple cameras[**?**] or by the use of IP images. A comparative study between the two approaches is presented by Park [**?**]. A number of modifications to the basic IP camera setup like time multiplexed IP cameras[**?**] has been used in order to increase the resolution of IP cameras[**?**,**?**]. Several attempts[**?**,**?**,**?**,**?**] have been reported in the literature aiming to tackle the problem of 3D object reconstruction using IP images.

The reconstruction algorithms proposed in previous works, are mainly applied to IP images of small objects that do not span many images with their size. This limitation is enforced in order to avoid stitching problems during the reconstruction stage caused by abrupt discontinuities due to depth estimation errors[**?**]. Moreover elemental image modification techniques are proposed in an effort to increase depth accuracy[**?**]. In several works (such as Shin [**?**]) the term *3D object reconstruction* is used to describe the generation of 2D images from multiple views and focus depths from a single IP image. The proposed approach is focused on the estimation of a fully 3D polygonal object and alleviates most of the above problems.

### 2.2 Inter-class Retrieval

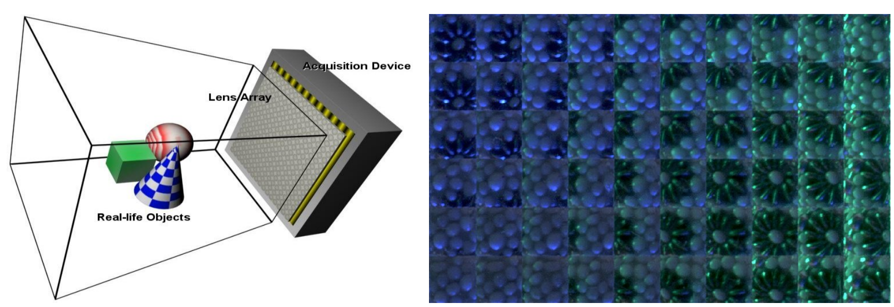
Inter-class retrieval methods are categorized ([**?**]) based on the representation they use into a) histogram-based, b) topology based, c) view-based, and d) shape-based. Histogram-based methods [**?**,**?**], even though they have several advantages (e.g., noise robustness, rotational invariance), in general are considered less descriptive compared to shape-based methods. Topological methods [**?**,**?**], have the drawback that relatively small anomalies on the object's surface can have a significant impact on the topological properties of the object. View-based methods [**?**,**?**] sacrifice part of their descriptiveness when projecting 3D geometry onto a 2D view. Shape-based methods [**?**,**?**,**?**,**?**,**?**] are considered more descriptive and offer state-of-the-art accuracy. Since the proposed method also belongs to this category, it is compared against the top previously proposed methods. To this end, publicly available databases are used under the framework available in [**?**].

### 2.3 Intra-class Retrieval and Biometrics

Intra-class retrieval is widely used in the Biometrics domain. However, few works can utilize different object classes (e.g., 3D face and 3D ear). Therefore there has been very little work that combines them. Only Woodward [?] attempted to fuse 3D ear, face and finger data. They achieved 97% rank-one recognition rate on a small database of 85 individuals using all three modalities. To the best of our knowledge, the method proposed in this work outperforms all previous single or multimodal approaches (3D face and 3D ear) that presented results on similarly sized databases.

In the 3D face recognition domain, most recent works utilize the FRGC v2 database, the largest publicly available 3D face database, which is also used in this work. On this database, Chang [?] reported a 92% rank-one recognition rate while Maurer [?] reported 87% verification rate at $10^{-2}$ False Acceptance Rate (FAR). Husken [?] presented a 2D+3D approach that uses hierarchical graph matching (HGM). Their method offers competitive 2D+3D results (96.8% verification rate at $10^{-3}$ FAR), but mediocre 3D resutlts (86.9%). In the 3D ear recognition domain, most recent works utilize the publicly available UND Ear database (used also in this work). Chen [?] using a subset of the UND database, reported 96.4% rank-one recognition rate. Using the full database Yan and Bowyer [?,?] reported 97.6% rank-one recognition rate.

## 3 3D Reconstruction using Integral Photography



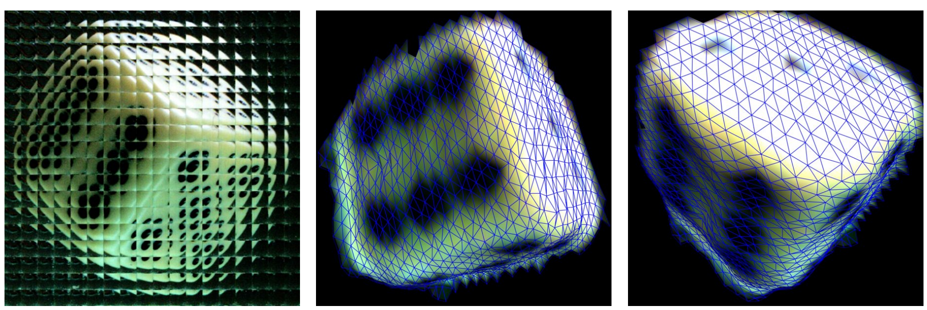

(a) (b)

**Fig. 2.** (a) Hardware setup that acquires IP images and (b) example of an IP image.

A method for 3D object reconstruction using Integral Photography, presented in [?], is summarized in this section. Integral Photography (IP) images are produced by an array of lenses (called lens array) that offers two dimensional parallax. The proposed method uses IP images acquired by real world objects using the hardware setup presented in [?], that employs a scanning CCD and a lens array, as depicted in Fig. 2. The scanning CCD sensor allows the capture of images over a 101 × 127mm area. We utilize the mechanism of a commonly available commercial flatbed scanner as the scanning CCD. This is an inexpensive solution that can offer a scanning area up to 210 × 297mm and an optical resolution of 3200dpi.

Our algorithm estimates the 3D shape and texture of an object from a single IP image. To this end a two step process is applied: first, 3D points (vertices) on the surface of the object are computed and second, these points are connected in a polygonal (e.g. triangular) mesh:

– *Vertex Grid Computation*: Vertices are computed using the central pixel of each lens, forming a rough regularly sampled vertex grid. Our algorithm extends the classical correspondence problem of stereo vision [**?**]: instead of having only a single image pair, multiple pairs are used.
– *Grid Refinement and Triangulation*: The grid is subdivided, new vertices are computed and the refined grid is triangulated. The idea is that the new vertices are inserted in-between the central vertices, without violating their order. This allows seamless triangulation.
– *Post-Processing*: The final grid is filtered in order to improve reconstruction quality, and the following filters are sequentially applied: median-cut, smoothing and subsampling.



(a) (b) (c)

**Fig. 3.** Reconstruction of a dice: (a) Input IP image (52×52 pixels/lens, f=3.3mm) and (b,c) Reconstructed 3D object rendered with triangulation superimposed.

The proposed algorithm reconstructs both the shape and texture of an object as depicted in Fig. 3. Apart from the parameters specified by the acquisition hardware (pixels/lens and focal length) two additional parameters affect the reconstruction: the number of neighboring lenses and the size of the search window that are used to solve the extended correspondence problem.

## 4 Inter-class 3D Object Retrieval

A method for inter-class 3D object retrieval presented in [**?**], is summarized in this section. The proposed method is shape-based as it uses a depth image (also called zbuffer) representation. The conversion from the polygonal representation to the depth image representation (as well as to a voxel representation) is described in detail in [**?**,**?**]. The idea behind this representation is to describe the object with multiple depth image pairs from several directions. For the retrieval application, this is simplified to a single depth image pair from three orthocanonical directions, as depicted in Fig. 4.

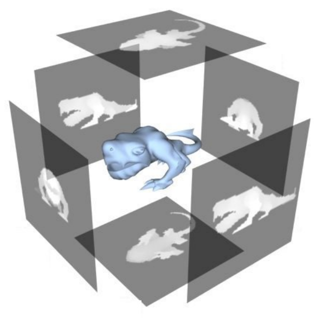


**Fig. 4.** 3D object represented by three depth image pairs.

The proposed method consists of two steps, a preprocessing step that needs to be applied once on each object in the database to produce a feature vector (or *shape descriptor*), and the retrieval step, where feature vectors from different objects are compared to determine similarity.

- *Preprocessing.* For each 3D object:
  - Normalize the scale of the object and translate its center of mass to $(0, 0, 0)$. Vertex weights based on triangle areas are used to ensure invariance to different tessellation.
  - Align the object using symmetry information and Principal Component Analysis [**?**], thus achieving rotational invariance. The novelty in the alignment is the symmetry computation which alleviates many of the problems found in PCA-only alignments.
  - Acquire the depth images (three pairs from orthocanonical directions). For each pair, the difference and the sum of the images is stored. Additionally per direction weights are applied based on the eigenvalues obtained by PCA.
  - Transform the buffers from the spatial to the spectral domain. Weighting is applied to suppress the contribution of high frequencies.

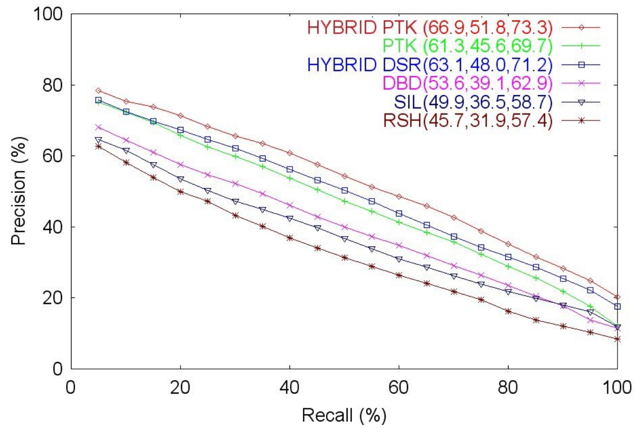



**Fig. 5.** Precision-recall curves for proposed method ($PTK$, $HYBRID\ PTK$) and methods presented in [**?**] ($DBD$, $SIL$, $RSH$, $HYBRID\ DSR$).

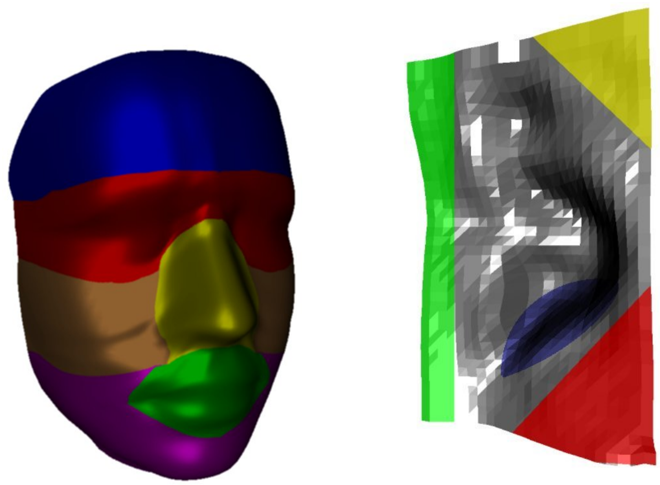
– *Retrieval.* Compare two feature vectors by computing their difference using an $L^1$ or $L^2$ metric.

In order to compare the proposed method with the state-of-the-art methods presented in [**?**], the framework available in [**?**] was used. This framework also provides a publicly available database with 1841 3D objects. The results are given in precision-recall curves (see Fig. 5) and for each shape descriptor three numbers are provided, in addition to its curve. The first is the average precision for recall 5% to 50%, the second is the average precision for recall 5% to 100% and the third is the percentage of correctly classified nearest neighbors. The proposed method, referred to as $PTK$, outperforms the previously proposed non-hybrid methods (including the previously proposed depth image based method referred as $DBD$). The hybrid shape descriptor is a combination of shape descriptors. By replacing the coefficients of the previously proposed depth image-based method ($DBD$) with our novel method's coefficients ($PTK$), the performance of the hybrid shape descriptor was increased.

## 5 Intra-class Retrieval and 3D Biometrics

The application to 3D biometrics of the intra-class retrieval method described in [**?**] is summarized in this section. The proposed method is applied to 3D face recognition [**?**,**?**,**?**,**?**,**?**], 3D ear recognition [**?**] and multimodal face and ear recognition [**?**,**?**].

The basic idea behind the proposed method is the creation of a 3D annotated model representative of the class. Two examples from the face and ear class are given in Fig. 6. The annotated model specializes the intra-class retrieval method to a specific 3D object class. The model has special properties that allows it to be converted to a geometry image. Additionally, the area annotation allows the



(a) (b)

**Fig. 6.** (a) Annotated Face Model; (b) Annotated Ear Model. Annotation areas are marked with different color.

incorporation of *a priori* knowledge into the model. Using the subdivision-based deformable model described in [**?**], the model is fitted to each 3D object, thus acquiring its shape. The fitted model is represented as a geometry image from which a normal image is derived. This process is depicted in Fig. 7. Finally, these images are analyzed using a Walsh Packet Wavelet transform [**?**]. The wavelet coefficients are weighted based on the area annotation. If more than one modality is available (face and ear) the coefficients are concatenated, and the final feature vector can be directly compared with feature vectors from other subjects.

In order to evaluate the performance of the proposed method in 3D face recognition, FRGC v2 was used, which is the largest available 3D face database. It consists of 4007 laser scans of 466 subjects and includes facial expressions. Along with the database a set of verification experiments is provided by the FRGC [**?**] framework, referred to as ROC I,II and III. In these experiments, for the 3D data, the proposed method has the highest reported performance: 97.3%, 97.2% and 97.0% verification rates at $10^{-3}$ FAR for ROC I,II and III respectively.

To evaluate the performance in 3D ear recognition and in multimodal recognition, the FRGC v2 was appended by the UND 3D ear database [**?**]. This publicly available database was acquired with the same sensor as FRGC v2 and they share 324 common subjects. For each of these subjects two multimodal dataset pairs (face and ear) are created. In order to perform identification experiments, one dataset pair is considered as gallery and the other as probe for each subject. The rank-one recognition rate for the multimodal fusion is 99.7%, for face only it is 97.5% while for ear only is 95.0% (as depicted in Fig. 8). Note that even though the proposed method's performance in ear recognition is not

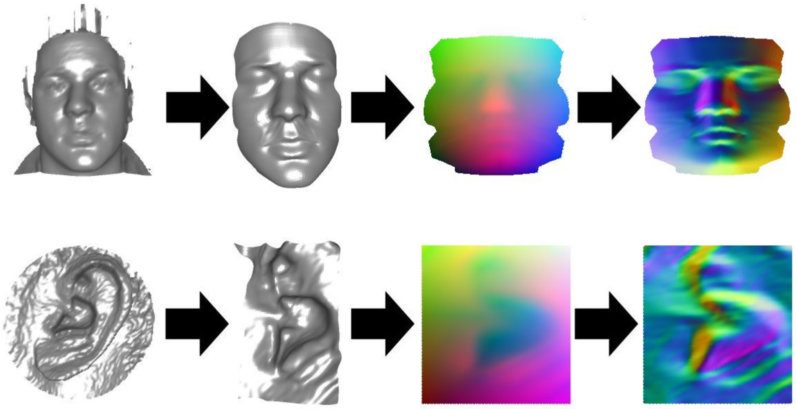


**Fig. 7.** From left to right, for face (top) and ear (bottom) datasets of the same subject: Raw data → Fitted Model → Extracted geometry image → Computed normal image.

state-of-the-art, the performance of the face and ear multimodal is extremely challenging. This is attributed to the low correlation of the failure sets of the two modalities.

## 6 Conclusions

In this work, two 3D object retrieval methods were summarized that utilize two different representations. In both cases the dimensionality of 3D information was reduced without however violating its geometric nature. Additonally, a 3D object reconstruction method was presented, that allows the extraction of 3D information from Integral Photography images. In all cases, the results are challenging, but especially in the biometric application, the results are considered state-of-the-art worldwide.

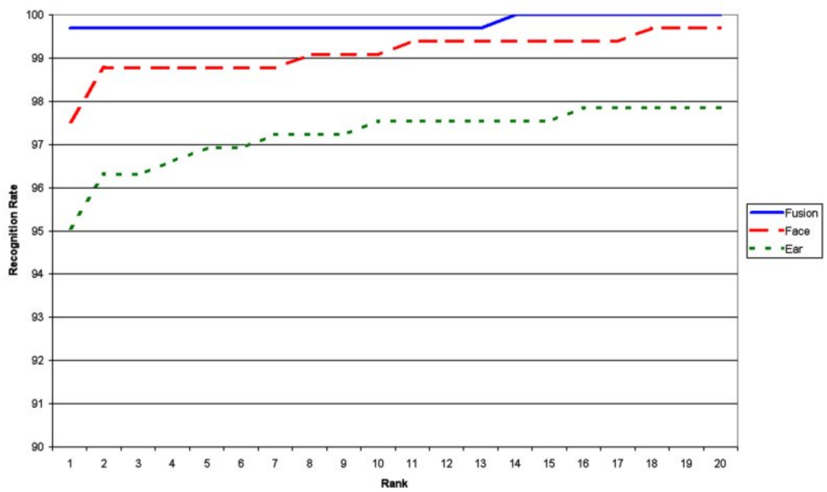



**Fig. 8.** Rank-one recognition rates of the proposed method on the combination of FRGC v2 and UND Ear databases.