Pattern Recognition 42 (2009) 2447 – 2459

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

3D object retrieval using the 3D shape impact descriptor ☆

Athanasios Mademlis [a], Petros Daras [b,*], Dimitrios Tzovaras [b], Michael G. Strintzis [a,b]

[a] Information Processing Laboratory, Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Greece
[b] Informatics and Telematics Institute, Centre for Research and Technology Hellas, Greece

ARTICLE INFO

Article history:
Received 13 August 2008
Received in revised form 21 April 2009
Accepted 26 April 2009

Keywords:
3D content-based search
3D content-based retrieval

ABSTRACT

In this paper, the novel 3D shape impact descriptor is introduced, which is based on the resulting gravitational phenomena in the surrounding area of every 3D object. The 3D object is considered as a distributed 3D mass and the descriptor of the 3D object is indirectly computed from the resulting fields. The field is described using both Newton's and general relativity's laws. In the Newtonian approach, histograms of the field values in the surrounding area of the 3D object are computed, while in the relativistic approach the descriptors are histograms of the time–space curvature in the surrounding area of the 3D object. The basic motivation behind the proposed approach is the robustness with respect to object's degeneracies and the native invariance of the resulting descriptors under rotation and translation. Experiments which were performed in various 3D object databases proved that the proposed method can be efficiently used for 3D object retrieval applications.

1. Introduction

3D object retrieval is a relatively new and challenging research field and a major effort of the research community has been devoted to the creation of accurate and efficient 3D object search and retrieval algorithms. Generally, descriptor extraction procedures aim to describe the geometry of the 3D object. The commonly utilized descriptors are global features, local features, histograms, topological features or combinations of the above. Some very good surveys on this topic can be found in [1,2].

The efficiency of the existing algorithms is mainly limited due to two major problems: the degeneracies of the 3D object (e.g. holes, missing polygons, hidden polygons) and the demand for invariance of the description under rotation, since translation and scaling normalization can be easily and robustly resolved. Although the problem of object's degeneracies can be efficiently tackled using a triangulation algorithm (e.g. Delaunnay triangulation) or a hole filling algorithm [3], the resulting mesh may not be a perfect for the 3D object. Concerning the demand for rotation invariance description, two widely acceptable solutions have been presented in the literature: the application of a native rotation invariant 3D object description (e.g. spherical harmonics [4], light field descriptor (LFD) [5], histogram-based descriptors [6]) or the rotation normalization of the object prior to extraction of the geometric descriptors. Both approaches present advantages and drawbacks: the rotation normalization approaches presented so far (e.g. continuous PCA [7], normals PCA [8]), although usually result in descriptors with higher discriminative power, some similar objects are not usually normalized in a similar manner [9]. Natively rotation invariant descriptors, in contrast [4], usually involve an integration-like technique resulting in loss of discriminative accuracy.

The major contribution of this paper is a new perspective in the 3D object description. The descriptors of the 3D object are not derived directly from its geometry. For the first time, a method is proposed which does not work either on the surface or the volume of a 3D object and does not take into account object's topology. Instead, the descriptors are computed on the resulting phenomena derived from the existence of the 3D object in the time–space. More specifically, the 3D object is treated as a distributed mass positioned in the empty space; the gravity laws are then utilized and the resulting field in the surrounding space is computed. The descriptors are extracted from this field. The field is described using two different approaches: expansions of the traditional Newtonian laws and a simplified instance of Einstein's theory for general relativity. This leads to native rotation invariant 3D object description, robust against object degeneracies and with high discriminative power.

The rest of the paper is organized as follows: after an overview of the existing approaches, in Section 2 the complete descriptor

☆This work was supported by the VICTORY project funded by the European Commission and by the PENED project co-funded 75% from European Commission and 25% by the Greek Secretariat for Research and Technology.

* Corresponding author.
E-mail addresses: mademlis@iti.gr (A. Mademlis), daras@iti.gr (P. Daras), Dimitrios.Tzovaras@iti.gr (D. Tzovaras), strintzi@eng.auth.gr (M.G. Strintzis).

0031-3203/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2009.04.024

extraction procedure is presented, while in Sections 3 and 4 the Newtonian and the relativistic descriptor extraction processes are thoroughly described, respectively. The matching approach utilized for the Newtonian and relativistic descriptors is explained in Section 6, while the experimental results are given in Section 8. Finally, the conclusions are drawn in Section 9.

1.1. Related work

In the area of 3D content-based search and retrieval, a variety of methods have been proposed so far. Although there are many ways to organize the existing approaches into broad categories [2,1], the existing methods can be classified into two broad categories: the topology-based approaches, where the object is mainly described using topological features, and the geometry-based approaches, where global or local geometrical descriptors are computed directly from the 3D object. The 3D impact descriptor approach proposed in this paper is clearly a geometrical approach, thus, only methods of this kind are presented in the sequel. These methods are based on global geometrical features of 3D models, which can be extracted either from their surface or from their volumetric representation.

In [10], a fast querying-by-3D-model approach is presented, where the descriptors are chosen so as to mimic the basic criteria that humans use for the same purpose. The descriptors of this approach are simple and their computation is rather straightforward. This method, as well as many early approaches on the field, has not been evaluated using well-known benchmarks.

In [5], an approach that measures the similarity among 3D models by visual similarity is proposed. In [11,12], a method where the descriptor vector is obtained by forming a complex function on the sphere is presented. Then, spherical harmonics are obtained which form the rotation invariant descriptor vector. In [13], the MPEG-7 shape spectrum descriptor is defined as the histogram of the shape index, calculated over the entire surface of a 3D object and is invariant under rotation, translation and scaling. In [14], the spherical harmonic representation is proposed which transforms rotation dependent descriptors into rotation invariant ones. In [15], the theoretical framework for the 3D Zernike moments [16] is extended and applied for 3D content-based search and retrieval. These are computed as a projection of the function which defines the object, onto a set of orthonormal functions within the unit ball. The 3D Zernike descriptors are natively invariant under rotation.

In [17], the extended Gaussian images (EGI) are introduced, where the surface normal orientation is mapped on a sphere, namely the Gaussian sphere. The EGI is obtained by having each triangle vote on the bin corresponding to its normal direction, with a weight equal to the area of the triangle. In [18], EGI has been generalized to the complex extended Gaussian image (CEGI), which stores for each bin also the normal distance of the surface points to the origin.

All the presented approaches are extracting sophisticated descriptors that exploit some significant properties (e.g. rotation invariance transformation [14,15], highly discriminative descriptor sets [9]), however, all of them describe directly the 3D geometry. Thus, the efficiency of the aforementioned approaches can be seriously decreased when they deal with non-perfect polygon meshes. In order to tackle the latter, two solutions have been proposed in the literature: the voxel-based approaches and the view-based approaches. In [19], a voxel-based 3D search and retrieval method based on the generalized radon transform (GRT) is proposed where two forms of the GRT are proposed, while in [20], the spherical trace transform is presented. These approaches are robust against object degeneracies, by implementing a voxelization approach [19].

In [5], the light field descriptor is presented. LFD is a view-based approach, which relies on the assumption that two objects are similar if they look similar from many viewpoints. Thus, every object is described using 2D Fourier and 2D Zernike descriptors extracted from various snapshots, taken from different views. The view-based approaches descriptors ignore the object parts that are hidden from the selected viewpoints (e.g. concave areas of the 3D object), which can decrease their retrieval efficiency.

In [28], a composite descriptor is introduced, based on the combination of the depth buffer, silhouette and radialized extent function descriptors. This descriptor outperforms every single approach that relies on.

More recently, the field of 3D shape search and retrieval has attracted more researchers and the competition on the field has been significantly increased, mainly due to the shape retrieval contest, organized each year by the EU funded project Aim@Shape [21].

In this paper, the 3D shape impact descriptor (SID) is proposed which is based on the physics laws of gravity. The proposed approach offers native invariance with respect to rotation and translation of the object. Additionally, the resulting description, which combines both local and global features, is robust against object degeneracies. Although the proposed 3D shape impact descriptor is computed in the surrounding area of the 3D object, which is common to the view-based approaches, the resulting descriptor also captures the internal structure of the 3D object, which is not taken into account by the view-based approaches. Two instances of the 3D shape impact descriptor are proposed: the Newtonian impact descriptor (NID) and the relativistic impact descriptor (RID). Both approaches require volumetric representation of 3D objects, thus in the preprocessing step all objects are transformed into their volumetric representations.

2. Method overview

Let us assume that the 3D object representation is a binary volumetric function. In cases where the 3D object is described in another form (e.g. polygon mesh), a preprocessing step that transforms the object to a binary volumetric representation is imposed [20]. The key idea of the proposed approach is the description of the resulting phenomena occurred by the insertion of the 3D object in the space. It is expected that similar objects will result in similar physical phenomena. An obvious selection to study this argument is the traditional Newtonian force field. More sophisticated selections involve the generalized Einstein field theory, or the Maxwell electromagnetic field theory. The present paper combines the static fields inspired from the traditional Newton laws with the Einstein's theory of generalized relativity. The surrounding field and, thus, the 3D object description is based both on Newtonian field and time–space curvature. More specifically, the descriptor extraction process can be summarized to the following discrete steps which are depicted in Fig. 1.

Firstly, the static field is computed in the surrounding area of the 3D object. Then, the values of the field's potential and density in the surrounding area of the object form histograms, the NID. The relativistic invariants that describe the curvature of the 4D space (time+3D space) that is caused by the mass of the 3D object (according to the general relativity laws) form the histograms of the RID. NID and RID are combined and form the 3D shape impact descriptor.

2.1. Preprocessing

The aim of the preprocessing step is to provide a common representation for all objects. Thus, all objects are transformed into a volumetric 3D function. In order to transform a given 3D mesh into its volumetric representation, the 3D mesh is firstly scaled in order to fit in the unit sphere. Then, the real subspace $[-1,1]\times[-1,1]\times[-1,1]$ is partitioned into $Q\times Q\times Q$ equally sized volume elements (voxels) and the volumetric representation of the 3D object is a function of

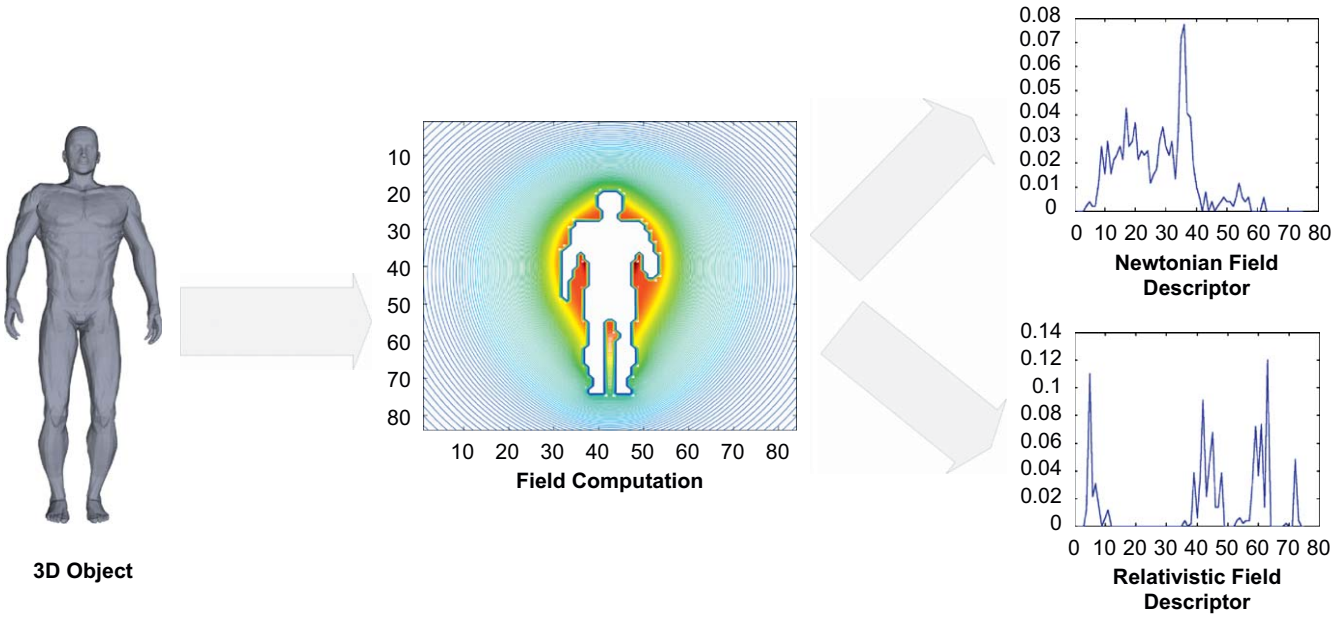

Fig. 1. The SID computation.

the point $\mathbf{x} \in \mathbb{R}^3$ defined as

$$f(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \text{ lies inside the object} \\ 0 & \text{elsewhere} \end{cases} \quad (1)$$

Based on the function $f(\mathbf{x})$, the 3D space is partitioned into two sets, $\mathbf{Y} = \{\mathbf{y}_i = \mathbf{x} | f(\mathbf{x}) = 1\}$, which represents the set of voxels that are occupied by the 3D object, and the complementary set of $\mathbf{Y}$, $\overline{\mathbf{Y}}$, which represents the free space. It is assumed that $\mathbf{Y}$ consists of N voxels, while M is the cardinality of $\overline{\mathbf{Y}}$.

The main benefit of using this kind of volumetric representation of a 3D object is that it is very robust in case of surface degeneracies (e.g. outliers, non-visible polygons, missing polygons), where all basic measurements such as that of the center of mass of a model, cannot be computed correctly using surface-based methods, resulting to various errors.

3. Newtonian impact descriptor computation

Concerning the Newtonian field computation, the object is considered as a discrete volumetric function, thus, for the rest of this section, the discrete formulas of the underlying laws are utilized. In order to compute a static field, a cause for the existence of the field should be present. The field's cause is selected to be the distributed mass of the 3D object, by assuming that every voxel is a point mass $m[\mathbf{y}_i] = 1$. $\mathbf{y}_i$ represents the voxels that are occupied by the object (i.e. $\mathbf{y}_i \in \mathbf{Y}$) and $\bar{\mathbf{y}}_j$ are the free space voxels (i.e. $\bar{\mathbf{y}}_j \in \overline{\mathbf{Y}}$). More specifically, in every point $\bar{\mathbf{y}}_j$ of the 3D space that is not occupied by the object, the density and the potential of the field can be computed according to

$$\mathbf{E}(\bar{\mathbf{y}}_j) = C \sum_{i=0}^{N-1} \frac{1}{|\bar{\mathbf{y}}_j - \mathbf{y}_i|^3} (\bar{\mathbf{y}}_j - \mathbf{y}_i) \quad (2)$$

$$\phi(\bar{\mathbf{y}}_j) = C \sum_{i=0}^{N-1} \frac{1}{|\bar{\mathbf{y}}_j - \mathbf{y}_i|} \quad (3)$$

where $\mathbf{E}(\cdot)$ is the density of the force field, $\phi(\cdot)$ is the potential of the field and C is a constant value.

The field which is described using (2) and (3) is the classical Newtonian field and can be generalized using the following:

$$\mathbf{E}(\bar{\mathbf{y}}_j) = \sum_{i=0}^{N} \frac{1}{|\bar{\mathbf{y}}_j - \mathbf{y}_i|^{K+1}} (\bar{\mathbf{y}}_i - \mathbf{y}_i) \quad (4)$$

$$\phi(\bar{\mathbf{y}}_j) = \sum_{i=0}^{N} \frac{1}{|\bar{\mathbf{y}}_j - \mathbf{y}_i|^{K-1}} \quad (5)$$

where $K \in \mathbb{N}^*$ is a free parameter that defines the field's law. Obviously, when $K = 2$, the generalized field leads to the Newtonian field. For simplicity, and with no loss of the generality of the presented approach, the constant parameter has been selected to be $C = 1$. Although, the parameter C affects the scaling of the resulting field values, the values of the field form histograms that are used as descriptors which discard scaling.

The introduction of the parameter K in the field's equations offers a great flexibility: a variation of the K changes the way that every point of the object contributes to the resulting field. Generally, the static field at a point is mainly the result of the mass that is included in an area centered at this point and its size depends on the value of K, due the quantity $|\bar{\mathbf{y}}_j - \mathbf{y}_i|^{K\pm1}$ in the denominator of (4) and (5). For lower values of K, the area that affects the value of the field in a specific point is bigger, while for greater values of K, the area is smaller. Thus, generally, when the value of K is low, the resulting field captures more global information while greater values of K result in a more local object description.

The field is computed in various points in the exterior of the object. The key point in the presented approach is the selection of the appropriate observation areas in the exterior of the 3D object to create histograms. By examining Eqs. (4) and (5) it is observed that the field vanishes and tends to be homogenous as the point under suspicion in the exterior of the 3D object is moved away from the object. This effect is clearly depicted in the equipotential areas around the object (Fig. 2). Thus, the field at points that are closer to the surface of the object presents more variations and, thus, the resulting descriptor corresponding to these points is intuitively more discriminative.

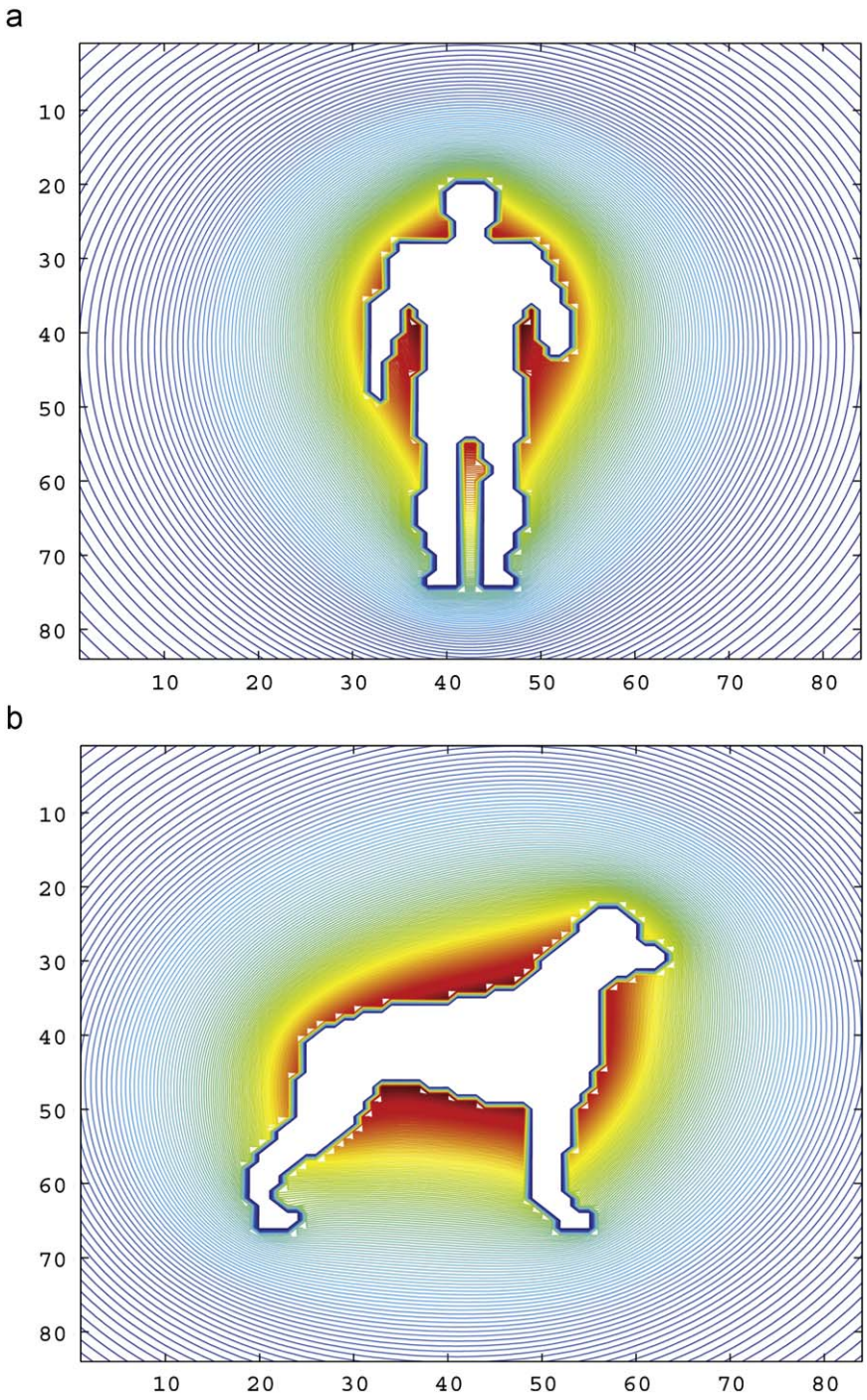

Fig. 2. The field's potential $\phi(\mathbf{x})$ produced from (a) human and (b) animal. Red areas represent higher potential values while blue areas correspond to lower potential value. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.1. Newtonian histogram computation

The Newtonian impact descriptor, in the proposed approach, is composed of three major histograms created by:

- The field potential values, computed in points that are equidistant from the object surface, i.e.

$$\{\phi(\bar{\mathbf{y}}_j) : \bar{\mathbf{y}}_j \in \mathbb{R}^3, L(\bar{\mathbf{y}}_j, \mathbf{Y}) = d\} \tag{6}$$

- The field density Euclidean norms, computed in points that are equidistant from the object surface:

$$\{|\mathbf{E}(\bar{\mathbf{y}}_j)| : \bar{\mathbf{y}}_j \in \mathbb{R}^3, L(\bar{\mathbf{y}}_j, \mathbf{Y}) = d\} \tag{7}$$

- The radial component of the field density, computed in points that are equidistant from the object surface:

$$\{\mathbf{E}(\bar{\mathbf{y}}_j)^T \cdot \mathbf{n}_r(\bar{\mathbf{y}}_j) : \bar{\mathbf{y}}_j \in \mathbb{R}^3, L(\bar{\mathbf{y}}_j, \mathbf{Y}) = d\} \tag{8}$$

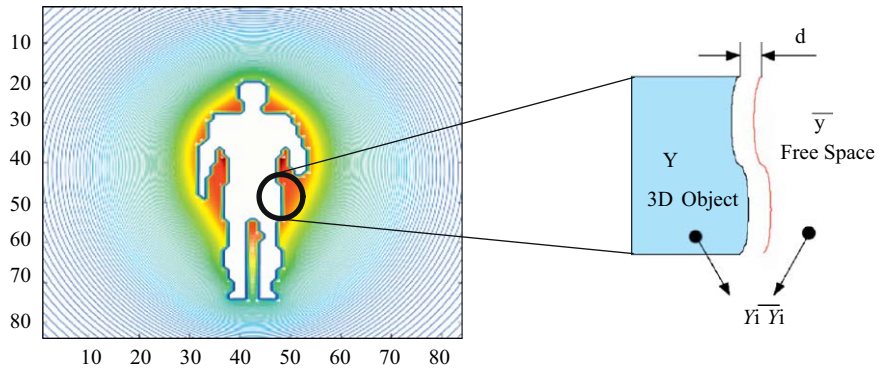

Fig. 3. The red surface is composed by points which are equidistant from the object surface. The three descriptor histograms are constructed based on the values of the field in the equidistant surface. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where $L(\bar{\mathbf{y}}_j, \mathbf{Y})$ is the distance of point $\bar{\mathbf{y}}_j$ from the surface of the 3D object Y and $\mathbf{n}_r(\bar{\mathbf{y}}_j) = (\bar{\mathbf{y}}_j - \mathbf{y}_c)/|\bar{\mathbf{y}}_j - \mathbf{y}_c|$ and $\mathbf{y}_c$ is the mass center of the 3D object. The equidistant points are depicted in Fig. 3.

Lemma. *The Newtonian impact descriptor is invariant under scaling, rotation and translation.*

Let us assume that O is a 3D object represented by the volumetric function $f(\mathbf{x})$, $\mathbf{Y} = \{\mathbf{y} = \mathbf{x} | f(\mathbf{x}) = 1\}$ and $\overline{\mathbf{Y}} = \mathbb{R}^3 - \mathbf{Y}$ are the sets that represent the points of the object and the surrounding space, respectively. Moreover, let O_R be the 3D object arbitrarily rotated by a rotation matrix $\mathbf{R}$ and translated by a translation vector $\mathbf{t}_t$ ($\mathbf{Y}_R$ and $\overline{\mathbf{Y}}_R$ are the corresponding point sets). Thus, if $\bar{\mathbf{y}}_j \in \overline{\mathbf{Y}}$, $\mathbf{y}_i \in \mathbf{Y}$, $\bar{\mathbf{y}}_j^R \in \overline{\mathbf{Y}}_R$ and $\mathbf{y}_i^R \in \mathbf{Y}_R$, then

$$\bar{\mathbf{y}}_j^R = \mathbf{R}\bar{\mathbf{y}}_j + \mathbf{t}_t \tag{9}$$

$$\mathbf{y}_i^R = \mathbf{R}\mathbf{y}_i + \mathbf{t}_t \tag{10}$$

$$\begin{aligned}\mathbf{E}(\bar{\mathbf{y}}_j^R) &= \sum_{i=0}^{N} \frac{1}{|\bar{\mathbf{y}}_j^R - \mathbf{y}_i^{R}|^{K+1}}(\bar{\mathbf{y}}_j^R - \mathbf{y}_i^R) \\ &= \sum_{i=0}^{N} \frac{1}{|\mathbf{R}\bar{\mathbf{y}}_j + \mathbf{t}_t - \mathbf{R}\mathbf{y}_i - \mathbf{t}_t|^{K+1}}(\mathbf{R}\bar{\mathbf{y}}_j + \mathbf{t}_t - \mathbf{R}\mathbf{y}_i - \mathbf{t}_t) \\ &= \sum_{i=0}^{N} \frac{1}{|\mathbf{R}|^K|\bar{\mathbf{y}}_j - \mathbf{y}_i|^{K+1}}\mathbf{R}(\bar{\mathbf{y}}_j - \mathbf{y}_i) \\ &= \mathbf{R}\cdot\mathbf{E}(\bar{\mathbf{y}}_j)\end{aligned} \tag{11, 12}$$

Thus,

$$|\mathbf{E}(\bar{\mathbf{y}}_j^R)| = |\mathbf{E}(\bar{\mathbf{y}}_j)| \tag{13}$$

$$\begin{aligned}\mathbf{n}_r(\bar{\mathbf{y}}_j^R) &= \frac{\bar{\mathbf{y}}_j^R - \mathbf{y}_c^R}{|\bar{\mathbf{y}}_j^R - \mathbf{y}_c^R|} \\ &= \frac{\mathbf{R}\bar{\mathbf{y}}_j + \mathbf{t}_t - \mathbf{R}\bar{\mathbf{y}}_c - \mathbf{t}_t}{|\mathbf{R}\bar{\mathbf{y}}_j + \mathbf{t}_t - \mathbf{R}\bar{\mathbf{y}}_c - \mathbf{t}_t|} \\ &= \frac{\mathbf{R}\bar{\mathbf{y}}_i - \mathbf{R}\bar{\mathbf{y}}_c}{|\mathbf{R}\bar{\mathbf{y}}_j - \mathbf{R}\bar{\mathbf{y}}_c|} \\ &= \frac{1}{|\mathbf{R}|}\mathbf{R}\frac{\bar{\mathbf{y}}_j - \mathbf{y}_c}{|\bar{\mathbf{y}}_j - \mathbf{y}_c|} \\ &= \mathbf{R}\mathbf{n}_r(\bar{\mathbf{y}}_j)\end{aligned} \tag{14}$$

$$\mathbf{E}(\bar{\mathbf{y}}_j^R)^T \cdot \mathbf{n}_r(\bar{\mathbf{y}}_j^R) = \mathbf{E}(\bar{\mathbf{y}}_i)\mathbf{R}^T\mathbf{R}\mathbf{n}_r(\bar{\mathbf{y}}_j) \tag{15}$$

$$= \mathbf{E}(\bar{\mathbf{y}}_j)\mathbf{n}_r(\bar{\mathbf{y}}_j) \tag{16}$$

Thus,

$$\mathbf{E}^T(\bar{\mathbf{y}}_i^R)\mathbf{n}_r(\bar{\mathbf{y}}_i^R) = \mathbf{E}(\bar{\mathbf{y}}_i)\mathbf{n}_r(\bar{\mathbf{y}}_j) \tag{17}$$

$$\begin{aligned}\phi(\bar{\mathbf{y}}_j^R) &= \sum_{i=0}^{N} \frac{1}{|\bar{\mathbf{y}}_j^R - \mathbf{y}_j^{R}|^{K-1}} \\ &= \sum_{i=0}^{N} \frac{1}{|\mathbf{R}\bar{\mathbf{y}}_j + \mathbf{t}_t - \mathbf{R}\mathbf{y}_i - \mathbf{t}_t|^{K-1}} \\ &= \sum_{i=0}^{N} \frac{1}{|\mathbf{R}|^{K-1}|\bar{\mathbf{y}}_j - \mathbf{y}_j|^{K-1}} \\ &= \phi(\bar{\mathbf{y}}_j)\end{aligned} \tag{18}$$

Thus,

$$\phi(\bar{\mathbf{y}}_j^R) = \phi(\bar{\mathbf{y}}_j) \tag{19}$$

The rotation and translation invariance have been proved for the norm of the field density of (7), the fields potential and the radial component of the field density.

Thus, the resulting histograms are invariant under translation and rotation. As a consequence, every object requires normalization only with respect to scaling. The latter can be easily achieved by uniform scaling of the 3D object so as its bounding sphere has a predefined radius.

4. Relativistic impact descriptor computation

The relativistic impact descriptor computation is inspired by the work presented in [22]. RID captures the way that a 3D object curves the surrounding time space. In order to be invariant for the representation of 3D objects, the method is not applied directly to the 3D object, but to its surrounding field.

Initially, it is assumed that the surrounding time–space of the object gets curved due to the 3D object's mass. Then, the surrounding space is sampled and at every sample the Einstein's gravity equation (assuming that the cosmological constant $\Lambda = 0$) is solved:

$$R_{\mu\nu} - \tfrac{1}{2}Rg_{\mu\nu} = \kappa\breve{T}_{\mu\nu} \tag{20}$$

where $g_{\mu\nu}$ is the metric of the curved space, $\breve{T}_{\mu\nu}$ is the input stress-energy tensor computed directly from the field primitives (Eqs. (4), (5)) and $R_{\mu\nu}R$ are the Ricci tensor and the Ricci scalar, respectively, that are functions of the $g_{\mu\nu}$ and its derivatives. Also, $\mu, \nu = [1, 2, 3, 4]$.

Due to the symmetry of the $g_{\mu\nu}$, i.e. $g_{\mu\nu}=g_{\nu\mu}$, (20) represents a system of partial differential equations. For the needs of this paper, it is assumed that the time–space is weakly curved, in order to have an analytic solution for (20) [23]:

$$g_{\mu\nu}=\int\int\int_D \frac{\breve{T}_{\mu\nu}(t-|\mathbf{x}-\mathbf{x}'|,\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}ds \tag{21}$$

Using the values of $g_{\mu\nu}$, which are computed according to (21), two invariants [22] that characterize the time–space curvature are computed:

$$V_1=R=-\sum_{\mu=1}^{4}\sum_{\nu=1}^{4}\kappa\breve{T}_{\mu\nu}g^{\mu\nu} \tag{22}$$

$$V_2=\sum_{\mu=1}^{4}\sum_{\nu=1}^{4}R_{\mu\nu}R^{\mu\nu} \tag{23}$$

where $R_{\mu\nu}$ is computed directly from (20). The values V_1 and V_2 are calculated for every point in the surrounding area of the 3D object. These invariants represent measures of the Riemannian curvature.

The values V_1 and V_2 can be considered as the Euclidean analogous of κ_1 and κ_2 in the Riemannian geometry and capture the local curvature of the time–space near a point $\mathbf{x}$. The V_1 and V_2 are proved that are invariant under rotation scaling and translation in the Riemannian theory [24]. The intuitive proof can be easily derived by the fact that $V_1(\mathbf{x})$ and $V_2(\mathbf{x})$ are computed locally for every point using as input the Newtonian field which is proved that is invariant under rotation and translation (Section 3) and involves distances and coordinates that are relative to the point $\mathbf{x}$.

Finally, two histograms are constructed using the values V_1 and V_2. These histograms capture the curvature of the surrounding time–space, due to the insertion of the object in the time–space. The mathematical relations between the metric $g_{\mu\nu}$ and the Ricci scalar and Ricci tensor are provided in the Appendix. For more sophisticated analysis, the reader is referred to any book concerning general relativity (e.g. [24]).

5. The 3D shape impact descriptor

The 3D shape impact descriptor is the combined novel descriptor introduced in this paper. It is composed of 36 independent histograms. Given a 3D object A, the 3D SID is computed as follows: firstly, the NID is computed for various values of d and K. For every pair (d,K) three separate histograms are formed. The values used for the computation of NID histograms are also used for the estimation of the values required for the RID histogram computation. Then, the RID histograms are stored along with the NID histograms. The 3D SID is composed by a set of independent histograms which are separately stored and combined during the matching process.

6. Matching method

The matching method of the presented approach is based on histogram metrics. The widely adopted L_n Minkowski distance is not a sufficient dissimilarity metric for histogram-based matching because histograms present differences and require special metrics. The need for more sophisticated metrics can be clearly depicted in Fig. 4, where two histograms of similar objects are presented along with a histogram corresponding to a third, different object. Intuitively, the distance $d(H_1,H_2)$ between the histograms H_1 and H_2 should be smaller than the distance $d(H_1,H_3)$ between H_1 and H_3. However, the distances computed using traditional metrics (e.g. L_1) lead to completely different results. A more comprehensive study

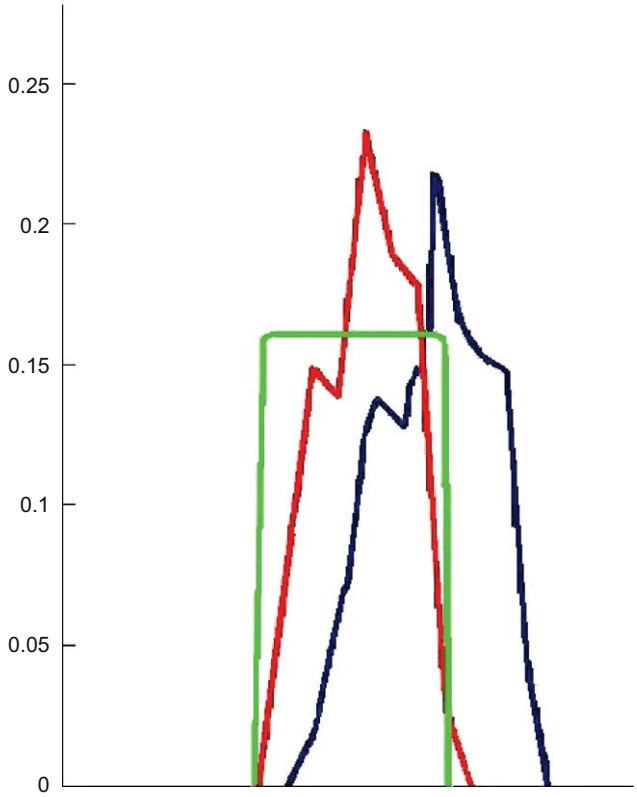

Fig. 4. Three histograms. H_1: blue, H_2: red, H_3: green. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

concerning histogram comparison can be found in [25]. In this paper, the following metrics have been utilized:

(a) The normalized distance, presented in [20], which is computed as

$$d(H_1,H_2)=\sum_{i=1}^{k}\frac{2|H_1(i)-H_2(i)|}{H_1(i)+H_2(i)} \tag{24}$$

where k is the number of histogram bins.

(b) The modified normalized distance, which is computed as

$$d(H_1,H_2)=\sum_{i=1}^{k}\frac{2(H_1(i)-H_2(i))^2}{H_1(i)+H_2(i)} \tag{25}$$

where k is the number of histogram bins.

(c) The diffusion distance [25], where the difference between two histograms H_1 and H_2 is treated as an isolated temperature field and a metric for its diffusion is computed. The latter can be mathematically be expressed as

$$d(H_1,H_2)=\int_0^{+\infty}L_1(T(i,t))\,dt \tag{26}$$

where

$$T(i,t)=\frac{H_1(i)-H_2(i)}{t\sqrt{2\pi}}*e^{-i^2/2t^2}$$

For every proposed histogram descriptor a different comparison metric has been utilized. The histograms that compose NID and RID are based on different underlying laws (NID is based on the classical

Newtonian theory and RID on the laws of general relativity). Also, NID is composed of different histograms that represent the fields density, potential and radial density component. Thus, every histogram captures information concerning the 3D shape in a unique way and for this reason requires different similarity metric for comparison. More specifically, for the potential related histograms the normalized distance of (25) has been utilized and the diffusion distance of (26), for the other two types of histograms. The basic motivation behind the selection of different metric for the histograms created by the values of the potential is that it is a scalar field value while the values compose the other NID histograms are based on non-scalar descriptors.

6.1. Combined matching

Let A, B be two 3D objects which are described by the 3D shape impact descriptors histograms, i.e. $\{H_i^A,\ i=1,\ldots,X\}$ and $\{H_i^B,\ i=1,\ldots,X\}$. The final score for RID, NID and SID is computed using

$$D(A,B)=\sum_{i=1}^{X} w_i d(H_i^A, H_i^B) \tag{27}$$

where X is the number of the total number of independent histograms, $d(\cdot)$ is the appropriate metric and w_i are experimentally selected weights.

7. Computational aspects and practical considerations

A major aspect in every search and retrieval method is the computational time. The approach presented in this paper involves a serious number of computations and, thus, a performance analysis should be considered.

Firstly, in case of NID, let N be the number of voxels occupied by the object and M the number of points where the field values require computation. Thus, according to (4) and (5) the complexity of the approach is $O(NM)$. These values can also be exploited by RID, for a direct estimation of the metric $g_{\mu\nu}$ using (21). The field is considered as close-to-static, so the field tensor is practically constant during time (i.e. $\partial T_{\mu\nu}/\partial t \simeq 0$). Also, the metric is a local feature, thus, the integration area D is much more smaller when compared to the size of the object, resulting in a fast computation of the metric. Concluding, RID, assuming that the field values have been pre-computed during NID, has a complexity of $O(N)$. Thus, the whole descriptor extraction approach has complexity of $O(MN)$.

The $O(MN)$ complexity can be considerably time consuming, when the values of N and M are relatively high. Both N and M values depend on the original 3D object and the selected level-of-detail in the voxelization approach. However, this leads in a trade-off between execution time and accuracy of the resulting representation. Higher values for Q result in a more discriminative discrete volumetric function $f(\mathbf{x})$, however, the values of N and M are high and thus, the execution time is considerably higher. In contrast, selecting lower values for Q leads to faster extraction of NID and RID with, however, lower discrimination. Towards selecting the best trade-off between accuracy and execution time, the value of Q has been experimentally selected to be $Q=48$, while the mean execution time is kept below 60 s using a dual core machine (2 GHz per core) running Windows XP.

A very interesting characteristic of the proposed approach is that it offers a native way to parallelize the computations. The computation of the field values required both by NID and RID and the metric $g_{\mu\nu}$ computation can easily be computed in parallel for every point.

8. Experimental results

Firstly, the robustness of the whole approach was evaluated with respect to object degeneracies, using an example 3D object which consists of 14 602 vertices and 29 202 triangles. Nine different versions of the same object were generated, using mesh simplification, mesh subdivision and noise addition. For every version of the 3D mesh, its voxelized representation was computed and then its SID descriptor. In Table 1 the maximum, the mean and the cummulated difference are presented. It is obvious that the proposed method is robust against mesh subdivision, while mesh simplification and noise addition do not seriously affect the results. It should be mentioned that both 50% mesh simplification and 2% noise addition seriously affect the quality of the 3D mesh.

Then, various NID descriptors for different values of K and d were compared using the ITI database. Fig. 5 depicts the comparative retrieval performance of the NID descriptors for $K=1,4,5,6$, $d=1,2$ and combinations of them.

By examining Fig. 5, it is obvious that the best results are achieved using the combination of all histograms. A very interesting result of the comparison is the significant increase of the retrieval performance for versions that combine NID histograms computed using a lower value of K to histograms computed using a high value of K. The latter can be explained by the fact that the shape information captured by NID histograms with lower value of K is "more global", when compared to the shape information included in NID histograms with greater value of K and, thus, the combination of local and global shape information results in better retrieval performance. Also, lower values of d result in higher retrieval performance, which is expected, as the field primitives tend to be homogenous as d increases.

The method was also evaluated towards applying the best value of Q with respect to both time efficiency and retrieval accuracy. In Fig. 6 the retrieval performance of SID for various sizes of the voxel model $Q=32,40,48,64$ is presented. It is obvious that the best performance is achieved when $Q=64$, with a slight advantage over $Q=48$. However, the time required for the SID extraction using $Q=64$ is significantly greater while the retrieval accuracy is not increased proportionally (Table 2).

The proposed approach was evaluated for its retrieval performance using the Princeton shape benchmark (PSB) [26], the engineering shape benchmark (ESB) [27] and the ITI 3D model's database [19] and was compared to well-known approaches of Gaussian Euclidean distance transform (GEDT), which is based on the comparison of a 3D function, whose value at each point is given by composition of a Gaussian with the Euclidean distance transform of the surface [4], the light field descriptor, where a representation of a model as a collection of images rendered from uniformly sampled positions on a view sphere is utilized [5], the 3D Zernike descriptor, the Radialized spherical extent function (REXT), where a collection of spherical functions giving the maximal distance from center of mass as a function of spherical angle and radius is utilized [7], the DSR descriptor proposed by Vranic [28]. The results for the aforementioned

Table 1
SID difference between various versions of the same object.

Object	Voxels different	SID mean difference
1. Mesh simplification 90%	0	0
2. Mesh simplification 75%	0	0
3. Mesh simplification 60%	0	0
4. Mesh simplification 50%	0.9%	0.3%
5. Catmul–Clark subdivision 1 loop	0	0
6. Catmul–Clark subdivision 2 loops	0	0
7. Catmul–Clark subdivision 3 loops	0	0
8. Uniform noise 0.5%	0	0
9. Uniform noise 1%	0.2%	0.1%

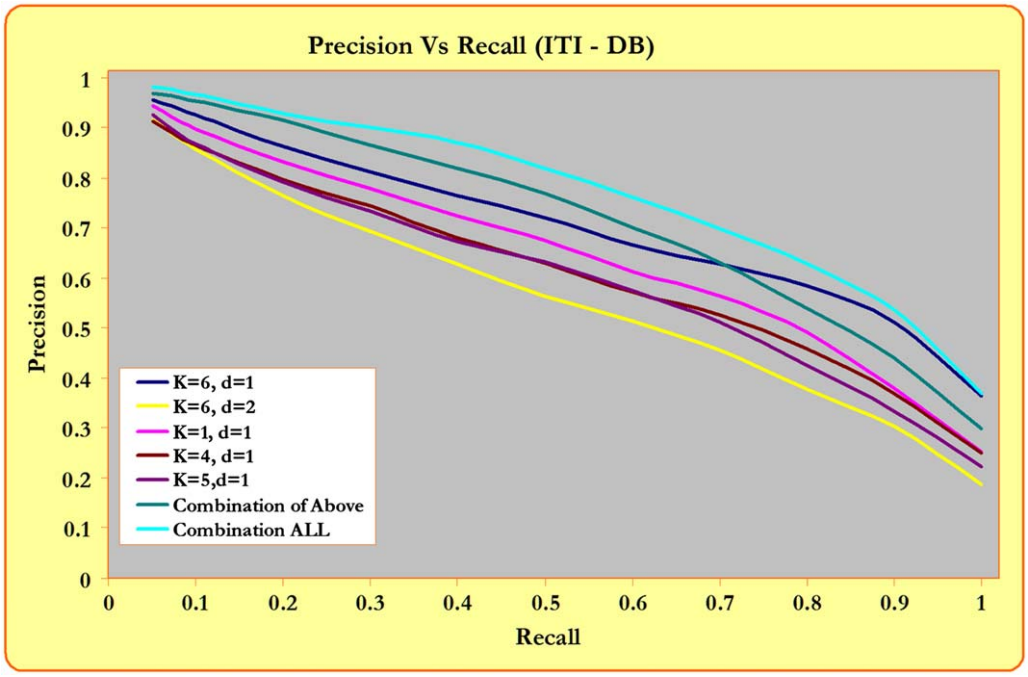

Fig. 5. Comparative precision–recall diagrams for the ITI database for various versions of the NID descriptor.

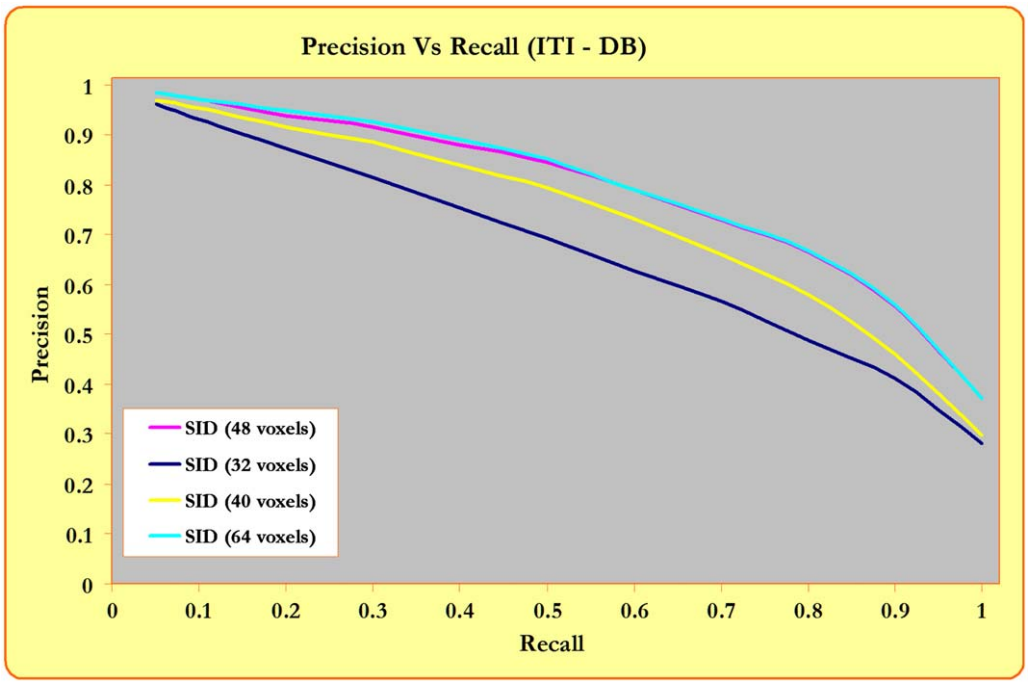

Fig. 6. Comparative precision–recall diagrams for the ITI database.

Table 2
SID mean execution times for various values of Q.

Q	Mean execution time (s)
32	10
40	17
48	25
64	50

approaches were computed using the executables provided by the authors. For the 3D Zernike approach, which was applied to voxel-based representation, the voxelization approach followed is the one described in this paper. The retrieval accuracy of all methods was evaluated using the precision–recall diagrams, where precision is defined as the ratio of the relevant retrieved objects against the total number of the retrieved objects, and recall is the ratio of

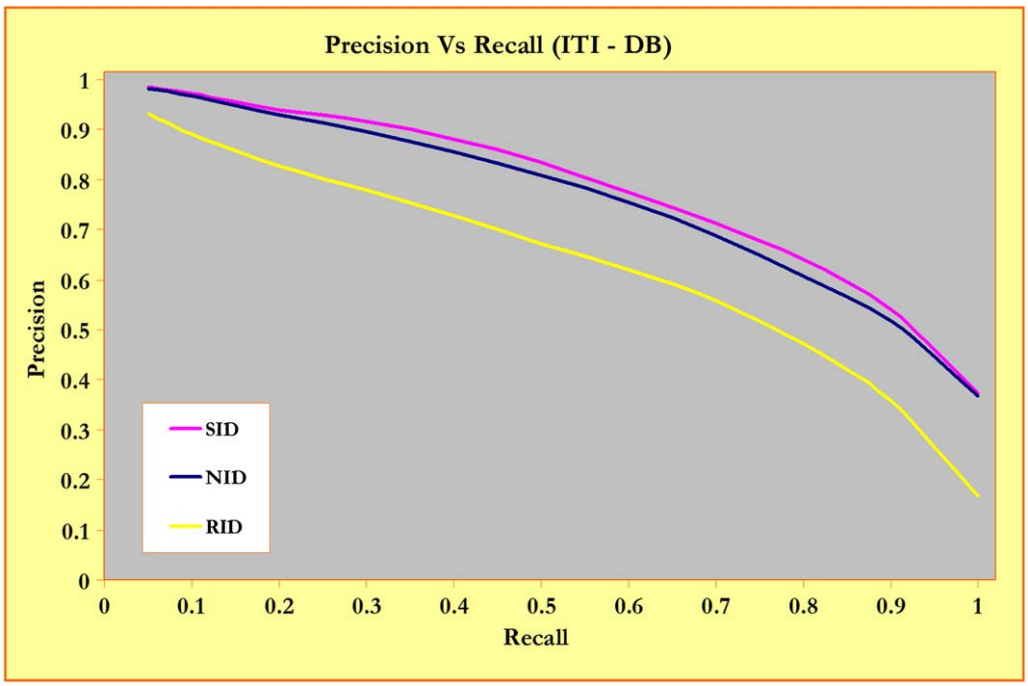

Fig. 7. Comparative precision–recall diagrams for the ITI database.

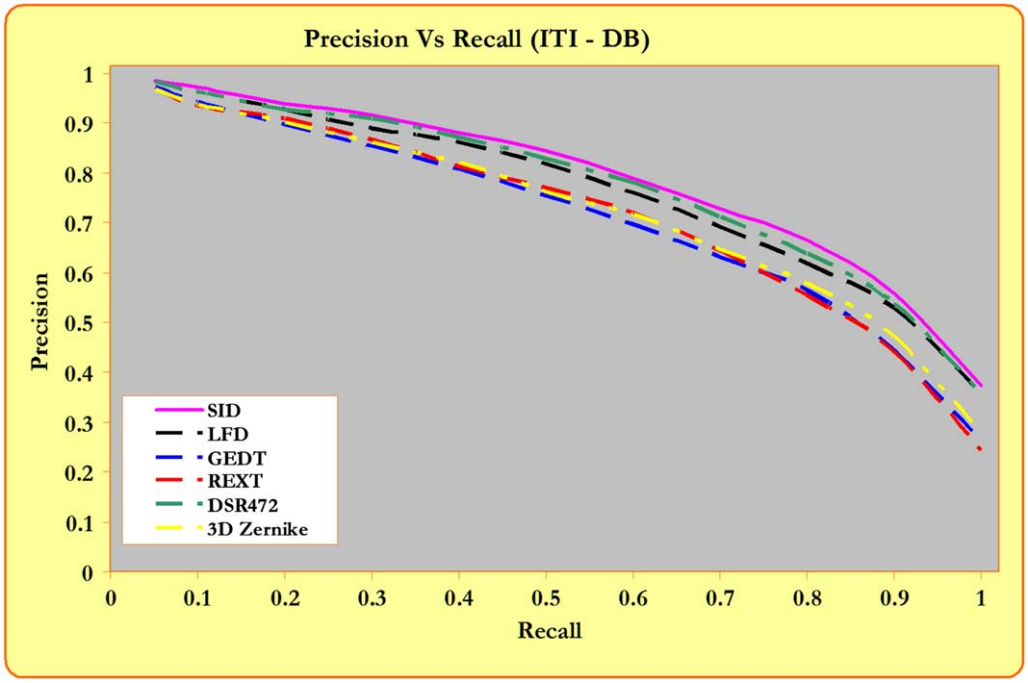

Fig. 8. Comparative precision–recall diagrams for the ITI database.

the relevant retrieved objects against the total relevant objects in the database.

The evaluation of the proposed scheme was performed separately for the NID, RID and the combined shape impact descriptor composed by RID and NID. Fig. 7 depicts the comparison of the retrieval performance of NID, RID and SID, in order to select the most appropriate descriptor vector. Although, the results of the single RID are not competitive to NID, due to the different nature of the computed descriptors, the SID descriptor that combines RID and NID is proved to be the best descriptor. The worse performance of the RID when compared to NID is mainly explained by the assumption of the weakly curved time–space.

The performance of the SID was compared to the performance of REXT, LDF and GEDT in Figs. 8, 9 and 10 for the ITI database, PSB and ESB, respectively. The presented precision recall diagrams can lead to very useful results. By comparing the precision recall diagrams for

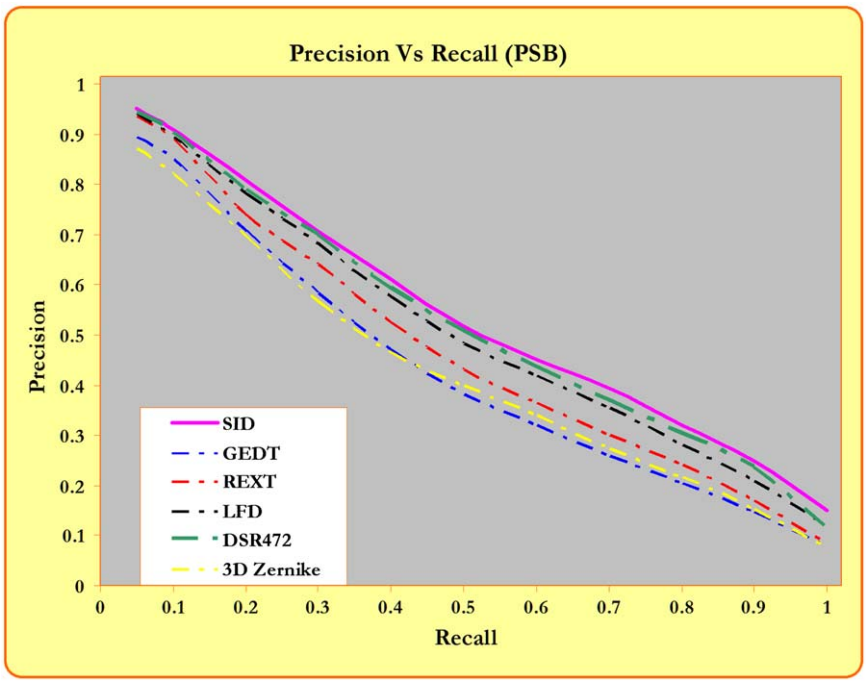

Fig. 9. Comparative precision–recall diagrams for the Princeton shape benchmark.

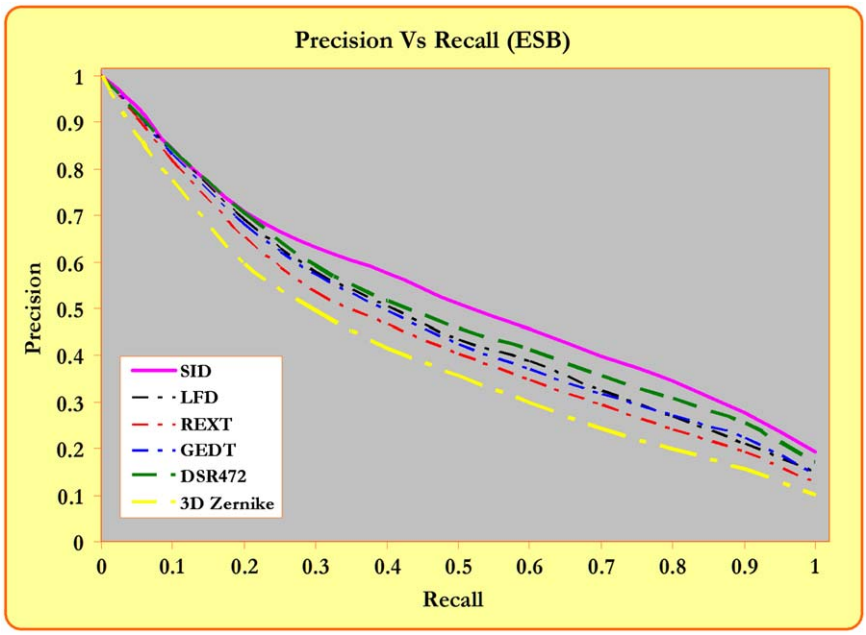

Fig. 10. Comparative precision–recall diagrams for the engineering shape benchmark.

the three different databases, a first and very obvious result is that all the approaches present completely differently behavior in a different database. The latter is basically the effect of the different 3D content included in each database, as well as the number of the 3D objects and the way that the 3D objects have been classified. Also, the relative performance of the competitive methods varies significantly when a different database is used to evaluate the retrieval performance. This can be easily explained by the dataset classification with respect to performance of every method for every class.

It is obvious that SID outperforms all other approaches in all datasets. Another important fact in the presented approach is identified on the nature of the descriptors. Histogram-based descriptors have been generally considered as descriptors with lower discriminative power, mainly due to their statistical nature. SID is a histogram-based descriptor, which proves that the appropriate selection of the values that construct the histograms is crucial for the discriminative power of the resulting histogram-based descriptor and that histogram-based descriptors can potentially provide highly discrimination.

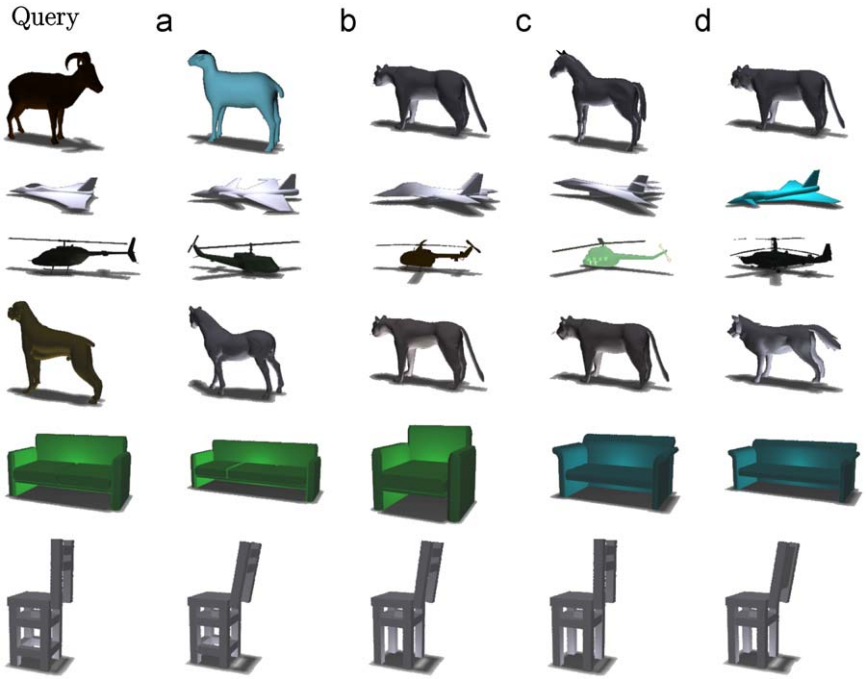

Fig. 11. Indicative retrieved results. The first column depicts the query 3D object while the rest are the first four retrieved objects.

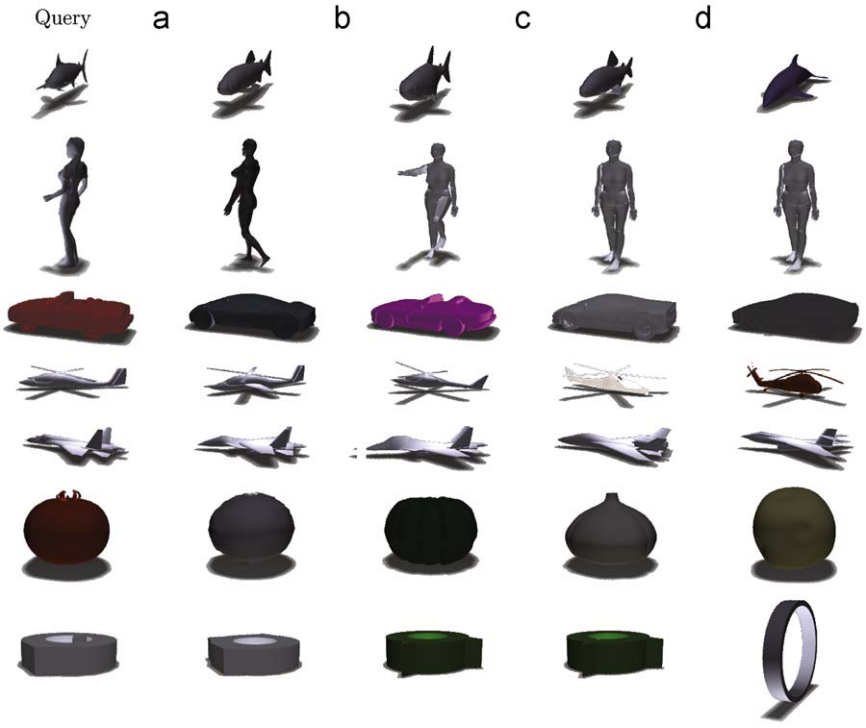

Fig. 12. Indicative retrieved results. The first column depicts the query 3D object while the rest are the first four retrieved objects.

In Figs. 11 and 12 some representative retrieved results for specific queries are shown. In the last line an interesting result is presented: for a query that belongs in the class of "Tubes", a "belt" is ranked 4th. Although this result is semantically irrelevant to the query, the geometric similarity between them is obvious. Results of this kind are generally expected in all approaches that do not take into account any high-level information, that could semantically discriminate the 3D objects.

9. Conclusions

In this paper, the shape impact descriptor has been introduced. To the best knowledge of the authors, this is one of the first approaches where the descriptor indirectly captures geometrical features of the 3D object, by describing its impact in the surrounding space. The experimental results proved the efficiency of the proposed descriptors in performing geometry-based 3D object search and retrieval.

Although geometry-based retrieval approaches present very good results, 3D objects that are semantically similar may not be retrieved. In these cases, the geometry-based results should be combined in a semantic-based framework where the system is enhanced with external knowledge in order to improve the retrieval performance.

Appendix A. Relativistic relations

The Ricci scalar (R) is related to Ricci tensor ($R_{\alpha\beta}$) and the covariant expression of the Riemannian tensor ($R_{\mu\alpha\nu\beta}$):

$$R=\sum_{\alpha=1}^{4}\sum_{\beta=1}^{4}g^{\alpha\beta}R_{\alpha\beta}=\sum_{\alpha=1}^{4}\sum_{\beta=1}^{4}\sum_{\mu=1}^{4}\sum_{\nu=1}^{4}g^{\alpha\beta}g^{\mu\nu}R_{\mu\alpha\nu\beta} \tag{28}$$

The covariant expression of the Riemannian tensor is related to the original expression as

$$R_{\kappa\beta\nu\mu}=\sum_{\lambda=1}^{4}g_{\kappa\lambda}R^{\lambda}_{\beta\nu\mu} \tag{29}$$

where the Riemannian tensor is

$$R^{\nu}_{\mu\sigma\tau}=\frac{\partial\Gamma^{\nu}_{\mu\tau}}{\partial x^{\sigma}}-\frac{\Gamma^{\nu}_{\mu\sigma}}{\partial x^{\tau}}+\sum_{\rho=1}^{4}\Gamma^{\nu}_{\rho\sigma}\Gamma^{\rho}_{\mu\tau}-\Gamma^{\nu}_{\rho\tau}\Gamma^{\rho}_{\mu\sigma} \tag{30}$$

$$\Gamma^{\nu}_{\mu\tau}=\sum_{\sigma=1}^{4}\frac{1}{2}g^{\nu\sigma}\left(\frac{\partial g_{\mu\sigma}}{\partial x^{\tau}}+\frac{\partial g_{\tau\sigma}}{\partial x^{\mu}}-\frac{\partial g_{\mu\tau}}{\partial x^{\sigma}}\right) \tag{31}$$

where x^1, x^2, x^3, x^4 are the four coordinates of the time–space.

References

[1] J.W.H. Tangelder, R.C. Veltkamp, A survey of content based 3d shape retrieval methods, in: SMI '04: Proceedings of the Shape Modeling International 2004, IEEE Computer Society, Washington, DC, USA, 2004, pp. 145–156.

[2] B. Bustos, D. Keim, D. Saupe, T. Schreck, Content-based 3d object retrieval, IEEE Computer Graphics and Applications 27 (4) (2007) 22–27.

[3] T. Weyrich, M. Pauly, R. Keiser, S. Heinzle, S. Scandella, M. Gross, Post-processing of scanned 3d surface data, in: Eurographics, Granada, Spain, 2004.

[4] M. Kazhdan, T. Funkhouser, S. Rusinkiewicz, Rotation invariant spherical harmonic representation of 3D shape descriptors, in: Symposium on Geometry Processing, 2003.

[5] D.-Y. Chen, M. Ouhyoung, X.-P. Tian, Y.-T. Shen, M. Ouhyoung, On visual similarity based 3d model retrieval, in: Eurographics, Granada, Spain, 2003, pp. 223–232.

[6] M. Ankerst, G. Kastenmüller, H. Kriegel, T. Seidl, 3D shape histograms for similarity search and classification in spatial databases. In: Proceedings of the sixth International Symposium on Advances in Spatial Databases, July 20–23, 1999.
R.H. Güting, D. Papadias, F.H. Lochovsky, (Eds). Lecture Notes in Computer Science, vol. 1651. Springer, London, pp. 207–226.

[7] D.V. Vranic, An improvement of rotation invariant 3d-shape based on functions on concentric spheres, in: ICIP (3), 2003, pp. 757–760.

[8] P. Papadakis, I. Pratikakis, S. Perantonis, T. Theoharis, Efficient 3D shape matching and retrieval using a concrete radialized spherical projection representation, Pattern Recognition 40 (9) (2007) 2437–2452.

[9] D. Vranic, 3d model retrieval, Ph.D. Thesis, University of Leipzig, 2004.

[10] L. Kolonias, D. Tzovaras, S. Malassiotis, M. Strintzis, Fast content-based search of vrml models based on shape descriptors, in: Proceedings of the 2001 International Conference on Image Processing, vol. 2, 7–10 October 2001, 2001, pp. 133–136.

[11] D. Vranic, D. Saupe, Description of 3d-shape using a complex function on the sphere, in: ICME '02: Proceedings of the 2002 IEEE International Conference on Multimedia and Expo, vol. 1, 2002, pp. 177–180.

[12] D. Vranic, D. Saupe, J. Richter, Tools for 3d-object retrieval: Karhunen–Loeve transform and spherical harmonics, in: 2001 IEEE Fourth Workshop on Multimedia Signal Processing, 2001, pp. 293–298.

[13] MPEG Video Group, MPEG-7 Visual part of eXperimentation Model (version 9.0), ISO/MPEG N3914, 2001.

[14] M. Kazhdan, T. Funkhouser, S. Rusinkiewicz, Rotation invariant spherical harmonic representation of 3D shape descriptors, in: Symposium on Geometry Processing, 2003.

[15] M. Novotni, R. Klein, 3d Zernike descriptors for content based shape retrieval, in: SM '03: Proceedings of the Eighth ACM Symposium on Solid Modeling and Applications, ACM, New York, NY, USA, 2003, pp. 216–225.

[16] N. Canterakis, 3D Zernike moments and Zernike affine invariants for 3D image analysis and recognition, in: Scandinavian Conference on Image Analysis, 1999.

[17] B. Horn, Extended Gaussian images, Proceedings of the IEEE 72 (12) (1984) 1671–1686.

[18] S. Kang, K. Ikeuchi, Determining 3-D object pose using the complex extended Gaussian image, in: CVPR'91: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1991, 1991, pp. 580–585.

[19] P. Daras, D. Zarpalas, D. Tzovaras, M. Strintzis, Efficient 3-d model search and retrieval using generalized 3-d radon transforms, IEEE Transactions on Multimedia 8 (1) (2006) 101–114.

[20] P. Daras, D. Zarpalas, D. Tzovaras, M.G. Strintzis, Efficient 3-d model search and retrieval using generalized 3-d radon transforms, IEEE Transactions on Multimedia 8 (1) (2006) 101–114.

[21] Aim@Shape, Shape retrieval contest ⟨http://www.aimatshape.net/event/SHREC/⟩.

[22] Eric Paquet, Representation of 3D and 4D objects based on an associated curved space and a general coordinate transformation invariant description, EURASIP Journal on Advances in Signal Processing (2007), Article ID 42505, 10, doi: 10.1155/2007/42505.

[23] E. Paquet, Content-based description of multi-dimensional objects using an invariant representation of an associated Riemannian space, MMIR MultiMedia Information Retrieval: metodologie ed esperienze internazionali di content-based retrieval per l'informazione e la documentazione, ISBN 88-901144-9-5, NRC 46532, 2004 (Publisher: AIDA), doi: 10.1400/24193.

[24] C. Bona, C. Palenzuela-Luque, Elements of Numerical Relativity: From Einstein's Equations to Black Hole Simulations, in: Lecture Notes in Physics, Springer, Berlin, 2005.

[25] H. Ling, K. Okada, Diffusion distance for histogram comparison, in: CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, Washington, DC, USA, 2006, pp. 246–253.

[26] P. Shilane, P. Min, M. Kazhdan, T. Funkhouser, The Princeton shape benchmark, in: SMI '04: Proceedings of the Shape Modeling International 2004, IEEE Computer Society, Washington, DC, USA, 2004, pp. 167–178.

[27] S. Jayanti, Y. Kalyanaraman, N. Iyer, K. Ramani, Developing an engineering shape benchmark for cad models, Computer-Aided Design 38 (2006) 939–953.

[28] D. Vranic, DESIRE: a composite 3D-shape descriptor, in: Proceedings of the 2005 IEEE International Conference on Multimedia and Expo, ICME 2005, Amsterdam, The Netherlands, July 6–9, 2005, 2005, pp. 962–965.

About the Author—ATHANASIOS MADEMLIS was born in Thessaloniki, Greece, in 1980. He received the Diploma degree in electrical and computer engineering in 2004 from Aristotle University of Thessaloniki, where he is currently pursuing the Ph.D. degree in electrical and computer engineering. Mr. Mademlis is a member of the Technical Chamber of Greece.

About the Author—PETROS DARAS was born in Athens, Greece, in 1974. He received the Diploma degree in electrical and computer engineering, the M.Sc. degree in medical informatics and the Ph.D. degree in electrical and computer engineering, all from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 1999, 2002 and 2005, respectively. He is a Senior Researcher at the Informatics and Telematics Institute, Thessaloniki. His main research interests include computer vision, search and retrieval of 3D objects, and medical informatics. He has been involved in more than 15 European and national research projects. Dr. Daras is a member of the Technical Chamber of Greece.

About the Author—DIMITRIOS TZOVARAS received the Diploma degree in electrical engineering and the Ph.D. degree in 2D and 3D image compression from Aristotle University of Thessaloniki, Thessaloniki, Greece, in 1992 and 1997, respectively. He is a Senior Researcher in the Informatics and Telematics Institute, Thessaloniki. Previously, he was a Senior Researcher, working on 3D imaging, at Aristotle University of Thessaloniki. His main research interests include virtual reality, assistive technologies, 3D data processing, medical image communication, 3D motion estimation, and stereo and multiview image sequence coding. His involvement with those research areas has led to the co-authoring of more than 35 papers in refereed journals and more than 80 papers in international conferences. He has served as a regular reviewer for a number of international journals and conferences. Since 1992, he has been involved in more than 40 projects in Greece, funded by the European Commission, and the Greek Secretariat of Research and Technology. Dr. Tzovaras is an Associate Editor of the EURASIP Journal of Applied Signal Processing and a member of the Technical Chamber of Greece.

About the Author—MICHAEL G. STRINTZIS received the Diploma degree in electrical engineering from the National Technical University of Athens, Athens, Greece, in 1967, and the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, in 1969 and 1970, respectively. He then joined the Electrical Engineering Department, University of Pittsburgh, Pittsburgh, PA, where he served as an Assistant Professor (1970–1976) and Associate Professor (1976–1980). Since 1980, he has been a Professor of electrical and computer engineering at the University of Thessaloniki, Thessaloniki, Greece, and, since 1999, Director of the Informatics and Telematics Research Institute, Thessaloniki. His current research interests include 2D and 3D image coding, image processing, biomedical signal and image processing, and DVD and Internet data authentication and copy protection. Dr. Strintzis has served as an Associate Editor for the IEEE Transactions on Circuits and Systems for Video Technology since 1999. In 1984, he was awarded one of the Centennial Medals of the IEEE.