Retrieval and classification methods for textured 3D models: a comparative study

S Biasotti, A Cerri, M Aono, A Ben Hamza, V Garro, A. Giachetti, D Giorgi, A Godil, C Li, C Sanada, et al.

HAL Id: hal-01250277

https://hal.archives-ouvertes.fr/hal-01250277

Submitted on 5 Jan 2016

Retrieval and classification methods for textured 3D models: A comparative study

**S. Biasotti · A. Cerri · M. Aono · A. Ben Hamza ·
V. Garro · A. Giachetti · D. Giorgi · A. Godil · C. Li ·
C. Sanada · M. Spagnuolo · A. Tatsuma · S. Velasco-Forero**

Abstract This paper presents a comparative study of six methods for the retrieval and classification of textured 3D models, which have been selected as representative of the state of the art. To better analyse and control how methods deal with specific classes of geometric and texture deformations, we built a collection of 572 synthetic textured mesh models, in which each class includes multiple texture and geometric modifications of a small set of null models. Results show a challenging, yet lively, scenario and also reveal interesting insights in how to deal with texture information according to different approaches, possibly working in the CIELab as well as in modifications of the RGB colour space.

S. Biasotti (✉), A. Cerri (✉) and Michela Spagnuolo
Istituto di Matematica Applicata e Tecnologie Informatiche
"E. Magenes", CNR, Italy
E-mail: {silvia.biasotti,andrea.cerri}@ge.imati.cnr.it

V. Garro
Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", CNR, Italy
Dipartimento di Informatica, Università di Verona, Italy

A. Giachetti
Dipartimento di Informatica, Università di Verona, Italy

A. Ben Hamza
Concordia University, Canada

C. Li and A. Godil
National Institute of Standards and Technology, USA

A. Tatsuma, C. Sanada and M. Aono
Department of Computer Science and Engineering, Toyohashi University of Technology, Japan

D. Giorgi
Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", CNR, Italy

S. Velasco-Forero
Department of Mathematics, National University of Singapore, Singapore

1 Introduction

Thanks to advances in geometric modelling techniques and to the availability of cheaper, yet effective 3D acquisition devices, we are witnessing a dramatic increase in the number of available 3D data [1,36]. How to accurately and efficiently retrieve and classify this data has become an important problem in computer vision, pattern recognition, computer graphics and many other fields. Most methods proposed in the last years analyse geometric and/or topological properties of 3D models [4,22,73], that is, they focus on *shape*. Nevertheless, most sensors are able to acquire not only the 3D shape but also its *texture*; this is the case, for instance, of the Microsoft Kinect device. Also, image-based modelling and multiple-view stereo techniques enable the recovery of geometric and colourimetric information directly from images [66].

Characterizing 3D shapes based on both geometric and colourimetric features can be of great help while defining algorithms for the analysis and the comparison of 3D data. Texture and colourimetric features contain rich information about the visual appearance of real objects: perceptual studies demonstrated that colour plays a significant role in low- and high-level vision [72]. Thus, colourimetric information plays an important role in many shape analysis applications, such as matching and correspondence; it can also provide additional clues for retrieval in case of partial or inaccurate shape scans [28]. An example is given by face recognition, where the combination of geometric and colourimetric properties is a way to achieve better trust-worthiness under un-

controlled environmental conditions (illumination, pose changes, uncooperative subjects) [26].

The attention towards texture properties has grown considerably over the last few years, as demonstrated by the number of techniques for the analysis of geometric shape and texture attributes that have been recently proposed [33,46,54,64,75,80]. Since 2013, a retrieval contest [9] has been launched under the umbrella of the SHREC initiative [76] to evaluate the performances of the existing methods for 3D shape retrieval when dealing with textured models. The contest provided the first opportunity to analyse a number of state-of-the-art algorithms, their strengths as well as their weaknesses, using a common test collection allowing for a direct comparison of algorithms. In 2014 the contest ran over a larger benchmark and was extended to include also a classification task [3]. The two events obtained a positive outcome, indeed they saw the participation of six groups in 2013 and eight groups in 2014.

In this context, we present here a comparative study on the retrieval and classification performance of six state-of-the-art methods in the field of textured 3D shape analysis. The present contribution builds on the dedicated SHREC'14 benchmark [3], and extends the associated track in three main respects:

• Most of the algorithms tested in [3] have been re-implemented with some modifications for performance improvement. Additionally, a new method has been included in the comparative study in order to have a sufficiently detailed picture of the state-of-the-art scenario;

• To help the reader in comparing methods beyond their algorithmic aspects, Section 4.7 presents a taxonomy of methods highlighting the emerging shape structure, the scale at which the shape description is captured, the colour space that is considered to analyse texture shape properties, and how this information is combined with the geometric one;

• The analysis of methods has been strengthened by exploiting the peculiar composition of the dataset, which has been populated by considering multiple modifications of a set of null shapes. This has made possible to evaluate how algorithms cope with specific geometric and colourimetric modifications.

The remainder of the paper is organized as follows: In Section 2 we introduce the related literature. Section 3 describes the collection of textured 3D models and how the comparative study has been organized, while in Section 4 we describe the methods implemented and discuss their main characteristics. Experimental results are presented, analysed and discussed in Section 5, while conclusive remarks and possible future developments are outlined in Section 6.

2 Related literature

While the combination of shape and colour information is quite popular in image retrieval [24] and processing [30,45], most of methods for 3D object retrieval and classification do not take colourimetric information into account [4,73].

The first attempts to devise 3D descriptors for textured objects adopt a 3D feature-vector description and combine it with the colourimetric information, where the colour is treated as a general property without considering its distribution over the shape. For example, Suzuki et al. [71] complemented the geometry description with a colour representation in terms of the Phong's model parameters [57]. Similarly, Ruiz et al. [64] combined geometric similarity based on Shape Distributions [53] with colour similarity computed through the comparison of colour distribution histograms, while in Starck and Hilton [69] the colourimetric and the 3D shape information were concatenated into a histogram.

In the field of image recognition, a popular description strategy is to consider local image patches that describe the behaviour of the texture around a group of pixels. Examples of these descriptions are the Local Binary Patterns (LPB) [52], the Scale Invariant Feature Transform (SIFT) [47], the Histogram of Oriented Gradients (HoG) [13] and the Spin Images [27]. The generalization of these descriptors to 3D textured models has been explored in several works, such as the VIP description [79], the meshHOG [80] and the Textured Spin-Images [12,54]. Further examples are the colour-CHLAC features computed on 3D voxel data proposed by Kanezaki et al. [28]; the sampling method introduced by Liu et al. [46] to select points in regions of either geometry-high variation or colour-high variation, and define a signature based on feature vectors computed at these points; the CSHOT descriptor [75], meant to solve the surface matching problem based on local features, i.e. by point-to-point correspondences obtained by matching shape- and colour-based local invariant descriptors of feature points.

Symmetry is another aspect used to characterize local and global shape properties [51]. For instance Kazhdan et al. [29] introduced the Spherical Harmonic descriptor to code the shape according to its rotational symmetry around axes centred in the centre of mass. In [3], the Spherical Harmonic descriptor has been proposed in combination with colourimetric descriptors to analyse textured 3D models. Giachetti and Lovato [25] introduced the Multiscale Area Projection Transform (MAPT) to couple the local degree of radial symmetry (in a selected scale range) with a saliency notion related to high shape symmetry, following an approach

similar to the Fast Radial Symmetry [48] used in image processing. Colour-weighted variations of MAPT, merging geometric and texture information, have been presented in [3, 9].

In the last years, close attention has been paid to non-rigid 3D shape matching and retrieval. To deal with non-rigid deformations (bendings) it is necessary to adopt shape descriptions that are invariant to isometric shape deformations. A suitable metric for comparing non-rigid shapes is the geodesic one; indeed 3D shape descriptions based on geodesics, such as geodesic distance matrices [68] or geodesic skeleton paths [38], have been successfully adopted for non-rigid shape comparison, see also [44]. In addition to geodesic, more sophisticated choices are possible, such as the diffusion or the commute-time distance [77]. On the basis of the fact that these distances are well-approximated by the Laplace-Beltrami operator, several spectral descriptors were proposed to characterize the geometric features of non-rigid 3D shapes [37], such as the ShapeDNA [62], the Heat Kernel Signature [23, 70], the Wave Kernel Signature [2], the Global Point Signature [65] and the Spectral Graph Wavelet Signature [40]. In the contest of textured 3D meshes, the Photometric Heat Kernel Signatures [31, 32, 33] fuse geometry and colour in a local-global description. The underlying idea is using the diffusion framework to embed the shape into a high-dimensional space where the embedding coordinates represent the photometric information. Following the same intuition, in [5] the authors generalized the geodesic distance as a hybrid shape description able to couple geometry and texture information.

Other invariance classes can be relevant in applications, possibly including non-isometric transformations such as topological deformations or local and global scaling. In this case, topological approaches [16, 21] offer a modular framework in which it is possible to plug in multiple shape properties in the form of different real functions, so as to describe shapes and measure their (dis)similarity up to different notions of invariance. Examples of these descriptions are Reeb graphs [6, 61], size functions [7], persistence diagrams [42, 11] and persistence spaces [10]. Recently, topological descriptors have been shown to be a viable option for comparing shapes endowed with colourimetric information [5].

3 The benchmark

In this Section, we describe the benchmark adopted in the proposed comparative analysis. The dataset and the ground truth are available by following the instructions at http://www.ge.imati.cnr.it/?q=shrec14.

3.1 The dataset

The dataset is made of 572 watertight mesh models, see Figure 1, grouped in 16 geometric classes of 32 or 36 instances. Each geometric class represents a type of geometric shape (e.g., humans, birds, trees, etc). Besides the geometric classification, models are also classified in 12 texture classes. Each texture class is characterized by a precise pattern (e.g., marble, wood, mimetic, etc).

The collection is built on top of a set of *null models*, that is, base meshes endowed with two or three different textures. All the other elements in the dataset come as the result of applying a shape transformation to one of the null shapes, so that a geometric and a texture deformation are randomly combined case by case.

The geometric deformations include the addition of Gaussian noise, mesh re-sampling, shape bending, shape stretching and other non-isometric transformations that do not necessarily preserve the metric properties of shapes (e.g. the Riemannian metric).

As for texture deformations, they include topological changing and scaling of texture patterns, as well as affine transformations in the RGB colour channels, resulting in, e.g., lighting and darkening effects, or in a sort of pattern blending. While the topological texture deformation has been applied manually, affine transformations admit an analytic formulation. A numerical parameter allows to tune each analytic formulation, thus making possible to automatically generate a family of texture deformations. In our dataset, texture transformations are grouped in five families, each family being the result of three different parameter values.

Figure 2 illustrates some geometric and texture deformations in action. The added value in working with a dataset built in this way is that particular weaknesses and strengths of algorithms can be better detected and analysed. Indeed, methods can be evaluated in specific tasks, for example retrieval against (simulated) illumination changing or degradation of texture pattern. This is actually part of the proposed comparative study, see Section 5.1.1 for more details.

Together with the dataset, a training set made of 96 models classified according to both geometry (16 classes) and texture (12 classes) has been made available for methods requiring a parameter tuning phase.

3.2 The retrieval and classification tasks

In our analysis we distinguished two tasks: retrieval and classification. For each task, at most three runs for each method have been considered for evaluation, being the result of either different parameter settings or more sub-

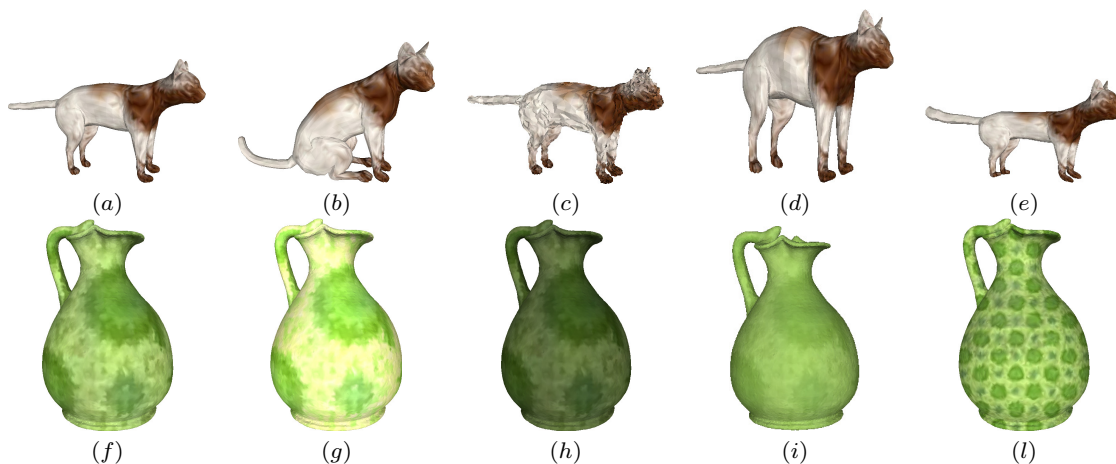**Fig. 1** The collection of textured 3D models used in the comparative study.

(a) (b) (c) (d) (e)

(f) (g) (h) (i) (l)

Fig. 2 Two base models (a, f) together with some purely geometric $(b - e)$ and purely texture $(g - l)$ modifications.

stantial method variations.

Retrieval task. Each model is used as a query against the rest of the dataset, with the goal of retrieving the most relevant objects. For a given query, a retrieved object is considered *highly relevant* if the two models share both geometry and texture; *marginally relevant* if they share only geometry; *not relevant* otherwise. For this task, a dissimilarity 572×572 matrix was required,

each element (i, j) recording the dissimilarity value between models i and j in the whole dataset.

Classification task. The goal is to assign each query to both its geometric and texture class. To this aim, a nearest neighbour (1-NN) classifier has been derived from the dissimilarity matrices used in the retrieval task. For each run, the output consists of two classification matrices, a 572×16 one for the geometric classifi-

cation and a 572×12 one for the texture classification. In these matrices, the element (i, j) is set to 1 if i is classified in class j (that is, the nearest neighbour of model i belongs to class j), and 0 otherwise.

3.3 The evaluation measures

The following measures have been used to evaluate the retrieval and classification performances of each method.

3.3.1 Retrieval evaluation measures

3D retrieval evaluation has been carried out according to standard measures, namely precision-recall curves, mean average precision, Nearest Neighbour, First Tier, Second Tier, Normalized Discounted Cumulated Gain and Average Dynamic Recall [76].

Precision-recall curves and mean average precision. Precision and recall are common measures to evaluate information retrieval systems. Precision is the fraction of retrieved items that are relevant to the query. Recall is the fraction of the items relevant to the query that are successfully retrieved. Being A the set of relevant objects and B the set of retrieved object,

$$\text{Precision} = \frac{|A \cap B|}{|B|}, \quad \text{Recall} = \frac{|A \cap B|}{|A|}.$$

Note that the two values always range from 0 to 1. For a visual interpretation of these quantities it is useful to plot a curve in the reference frame recall vs. precision. We can interpret the resul as follows: the larger the area below such a curve, the better the performance under examination. In particular, the precision-recall plot of an ideal retrieval system would result in a constant curve equal to 1. As a compact index of precision vs. recall, we consider the mean average precision (mAP), which is the portion of area under a precision recall-curve: from the above considerations, it follows that the maximum mAP value is equal to 1.

Nearest Neighbour, First Tier and Second Tier. These evaluation measures aim at checking the fraction of models in the query's class also appearing within the top k retrievals. Here, k can be 1, the size of the query's class, or the double size of the query's class. Specifically, for a class with $|C|$ members, $k = 1$ for the nearest neighbour (NN), $k = |C| - 1$ for the first tier (FT), and $k = 2(|C| - 1)$ for the second tier (ST). Note that all these values necessarily range from 0 to 1.

Average dynamic recall. The idea is to measure how many of the items that should have appeared before or at a given position in the result list actually have appeared. The average dynamic recall (ADR) at a given position averages this measure up to that position. Precisely, for a given query let A be the set of highly relevant (HR) items, and let B be the set of marginally relevant (MR) items. Obviously $A \subseteq B$. The ADR is computed as:

$$\text{ADR} = \frac{1}{|B|} \sum_{i=1}^{|B|} r_i,$$

where r_i is defined as

$$r_i = \begin{cases} \frac{|\{\text{HR items in the first } i \text{ retrieved items}\}|}{i}, & \text{if } i \leq |A|; \\ \frac{|\{\text{MR items in the first } i \text{ retrieved items}\}|}{i}, & \text{if } i > |A|. \end{cases}$$

Normalized discounted cumulated gain. It is first convenient to introduce the *discounted cumulated gain (DCG)*. Its definition is based on two assumptions. First, highly relevant items are more useful if appearing earlier in a search engine result list (have higher ranks); second, highly relevant items are more useful than marginally relevant items, which are in turn more useful than not relevant items. Precisely, the DCG at a position p is defined as:

$$\text{DCG}_p = \text{rel}_1 + \sum_{i=2}^{p} \frac{\text{rel}_i}{\log_2(i)},$$

with rel_i the graded relevance of the result at position i. Obviously, the DCG is query-dependent. To overcome this problem, we normalize the DCG to get the *normalized discounted cumulated gain (NDCG)*. This is done by sorting elements of a retrieval list by relevance, producing the maximum possible DCG till position p, also called *ideal DCG (IDCG)* till that position. For a query, the NDCG is computed as

$$\text{NDCG}_p = \frac{\text{DCG}_p}{\text{IDCG}_p}.$$

It follows that, for an ideal retrieval system, we would have $\text{NDCG}_p = 1$ for all p.

3.3.2 Classification performance measures.

We also consider a set of performance measures for classification, namely confusion matrix, sensitivity, specificity and the Mattews correlation coefficient [19, 49].

Confusion matrix. Each classification performance can be associated with a confusion matrix CM, that is, a square matrix whose order is equal to the number of classes (according to either the geometric or the texture classification) in the dataset. For a row i in CM,

the element $CM(i,i)$ gives the number of items which have been correctly classified as elements of class i; similarly, elements $CM(i,j)$, with $j \neq i$, count items which have been misclassified, resulting as elements of class j rather then elements of class i. Thus, the classification matrix CM of an ideal classification system should be a diagonal matrix, such that the element $CM(i,i)$ equals the number of items belonging to the class i.

Sensitivity, specificity and Matthews correlation coefficient. These statistical measures are classical tools for the evaluation of classification performances. Sensitivity (also called the true positive rate) measures the proportion of *true positives* which are correctly identified as such (e.g. the percentage of cats correctly classified as cats). Specificity (also known as true negative rate) measures the proportion of *true negatives* which are correctly identified as such (e.g. the percentage of non-cats correctly classified as non-cats). A perfect predictor is 100% sensitive and 100% specific.

The Matthews correlation coefficient takes into account true and false positives and negatives and is generally regarded as a balanced measure which can be used even if the classes are of very different sizes. The MCC is in essence a correlation coefficient between the observed and predicted classifications; it returns a value between -1 and 1. A coefficient of 1 represents a perfect classification, 0 no better than random classification and -1 indicates total disagreement between classification and observation.

4 Description of the methods

Six methods for textured 3D shape retrieval and classification have been implemented. In this section we describe them in detail, focusing also on their possible variations and the choice of the parameters adopted to implement the runs used in our comparative evaluation.

1. Histograms of Area Projection Transform and Colour Data and Joint Histograms of MAPT and RGB data (runs **GG1**,**GG2**,**GG3**), Section 4.1. These runs are based on a multi-scale geometric description able to capture local and global symmetries coupled with histograms of the normalized RGB channels;
2. Spectral geometry based methods for textured 3D shape retrieval (runs **LBG1**, **LBG2**, **LBG3** and **LBGtxt**), Section 4.2. These runs combine an intrinsic, spectral descriptor with the concatenated histogram of the RGB values;
3. Colour + Shape descriptors (runs **Ve1**, **Ve2**, **Ve3**), Section 4.3. These runs adopt combinations (with different weights) of the histogram of the RGB val-

ues with a geometric descriptors represented by the eigenvalues of the geodesic distance matrix;
4. Textured shape distribution, joint histograms and persistence (runs **Gi1**, **Gi2**, **Gi3**), Section 4.4. These runs combine several geometric, colourimetric and hybrid descriptors: namely, the spherical harmonics descriptor, the shape distributions of the geodesic distances weighted with the colourimetric attributes and a persistence-based description based on the CIELab space;
5. Multiresolution Representation Local Binary Pattern Histograms (run **TAS**), Section 4.5. This run captures the geometric information through the combination of a multi-view approach with local binary patterns and combines it with the concatenated histograms of the CIELab colour channels;
6. PHOG: Photometric and geometric functions for textured shape retrieval (runs **BCGS1**, **BCGS2**, **BCGS3**), Section 4.6. These runs combine a shape descriptor based on geometric functions; a persistence-based descriptor built on a generalized notion of geodesic distance that combines geometric and colourimetric information; a purely colourimetric descriptor based on the CIELab colour space.

4.1 Histograms of Area Projection Transform and Colour Data and Joint Histograms of MAPT and RGB data (runs **GG1-3**)

Computing the similarity between textured meshes is achieved according to two different approaches based on histograms of Multiscale Area Projection Transform (MAPT) [25]. MAPT originates from the Area Projection Transform (APT), a spatial map that measures the likelihood of the 3D points inside the shape of being centres of spherical or cylindrical symmetry. For a shape $\mathcal{S}$ represented by a surface mesh S, the APT is computed at a 3D point $\mathbf{x}$ for a radius of interest r:

$$\mathrm{APT}(\mathbf{x}, \mathcal{S}, r, \sigma) = \mathrm{Area}(T_r^{-1}(k_\sigma(\mathbf{x}) \subset T_r(S, \mathbf{n}))),$$

where $T_r(S, \mathbf{n})$ is the surface parallel to S shifted along the inward normal vector $\mathbf{n}$ for a distance r, and $k_\sigma(\mathbf{x})$ is a sphere of radius σ centred in $\mathbf{x}$. APT values at different radii are normalized to have a scale-invariant behaviour, creating the Multiscale APT (MAPT):

$$\mathrm{MAPT}(\mathbf{x}, \mathcal{S}, r) = \frac{\mathrm{APT}(\mathbf{x}, \mathcal{S}, r, \sigma(r))}{4\pi r^2},$$

with $\sigma(r) = c \cdot r$ for $0 < c < 1$.

A discrete version of the MAPT function is implemented following [25]. Roughly, the map is estimated on a grid of voxels with side length s and for a set of

corresponding sampled radius values $r_1, ..., r_t$. This grid partitions the mesh's bounding box, but only the voxels belonging to the inner region of the mesh are considered when creating the histogram.

Histograms of MAPT are very good global shape descriptors, performing state of the art results on the SHREC 2011 non-rigid watertight contest dataset [43]. For that retrieval task, the MAPT function was computed using 8 different scales (radius values) and the map values were quantized in 12 bins; finally the 8 histograms were concatenated creating an unique descriptor of length 96. The voxel size and the radius values were chosen differently for each model, proportionally to the cube root of the object volume, in order to have the same descriptor for scaled versions of the same geometry. The value of c was always set to 0.5.

To deal with textured meshes, the MAPT approach has been modified in two different ways, so to exploit also the colour information.

4.1.1 Histograms of MAPT and Colour Data

MAPT histograms are computed with the same radii and sampling grid values as in [25]: the isotropic sampling grid is proportional to the cube root of the volume V of each model, that is, of side length $s = \sqrt[3]{V}/30$), and the sampled radii are integer multiples of s (10 values from $2s$ to $11s$). The radius σ is taken, as in the original paper, equal to $r_i/2$ for all the sampled r_i. Furthermore, for each mesh the histogram of colour components is computed. With this procedure each mesh is described by two histograms, the first one representing the geometric information and the second one representing the texture information. The total dissimilarity between two shapes $\mathcal{S}_1, \mathcal{S}_2$ is then assessed using a convex combination of the two histogram distances:

$$D(\mathcal{S}_1, \mathcal{S}_2) = \gamma \, d_{geo}(\mathcal{S}_1, \mathcal{S}_2) + (1 - \gamma) \, d_{clr}(\mathcal{S}_1, \mathcal{S}_2), \quad (1)$$

where $0 \le \gamma \le 1$, $d_{geo}(\mathcal{S}_1, \mathcal{S}_2)$ is the normalized Jeffrey divergence between the two MAPT histograms of $\mathcal{S}_1$ and $\mathcal{S}_2$, and $d_{clr}(\mathcal{S}_1, \mathcal{S}_2)$ corresponds to the normalized χ^2-distance of the two colour histograms. The choice of γ in (1) allows the user to decide the relevance of colour information in the retrieval process.

Results shown in Section 5 are obtained by applying two different pre-processing steps to the RGB values, both adopted to have a colour representation that is invariant to illumination changes.

The first method, resulting in run **GG1**, is a simple contrast stretching for each RGB channel, mapping the min-max range of each channel to $[0, 1]$. In this case the colour quantization is set to 4 bins for each normalized RGB channels and γ is set to 0.6.

The second model, corresponding to run **GG2**, is the greyworld representation [20] in which each RGB value is divided by its corresponding channel mean value: $(R_g = R/R_{\text{avg}}, G_g = G/G_{\text{avg}}, B_g = B/B_{\text{avg}})$. Here, $\gamma = 0.6$ and the 4 histogram bins are centred in the mean value and are linearly distributed within a range of $[0, 2]$ for each greyworld RGB channel.

4.1.2 Joint Histograms of MAPT and greyworld RGB data

To get run **GG3**, a new descriptor has been designed by concatenating the original APT histogram with those obtained from the 3 components of the selected (normalized) colour space. For each voxel, the APT is evaluated at a certain radius; the procedure is then repeated for all the radius values $(r_1, ..., r_t)$, and the t histograms are finally linearized and concatenated. In the present paper, a sampling grid with side length $s = \sqrt[3]{V}/18$ has been used for each model, together with 9 sampled radii that are integer multiples of s ($t = 9$ values from $2s$ to $10s$). The APT target set has been divided into 8 bins. As for the colour components, the above greyworld RGB representation has been adopted, and each channel has been quantized in 4 bins in the range $[0, 2]$. The dissimilarity between two meshes is obtained with the normalized Jeffrey divergence [15] between the two corresponding linearized and concatenated sets of joint histograms.

4.2 Spectral geometry – based methods for textured 3D shape retrieval (runs **LBG1-3**, **LBGtxt**)

This method is build on the spectral geometry-based framework proposed in [37], suitably adapted for textured 3D shape representation and retrieval.

The spectral geometry approach, which is based on the eigendecomposition of the Laplace-Beltrami operator (LBO), provides a rich set of eigenbases invariant to isometric transformations. Also, these eigenbases serve as ingredients for two further steps: *feature extraction*, detailed in Section 4.2.1, and *spatial sensitive shape comparison* via intrinsic spatial pyramid matching [39], discussed in Section 4.2.2. The cotangent weight scheme [14] was used to discretize LBO. The eigenvalues λ_i and associated eigenfunctions φ_i can be computed by solving the generalized problem

$$C\varphi_i = \lambda_i A \varphi_i, \quad i = 1, 2, \ldots, m,$$

where A is a positive-definite diagonal area matrix and C is a sparse symmetric weight matrix. In the proposed implementation, m is set to 200.

4.2.1 Feature extraction

The first step consists in the computation of an informative descriptor at each vertex of a triangle mesh representing a shape. The spectral graph wavelet signature [40] is used to capture geometry information and colour histogram to encode texture information.

Geometry information. In general, any of the spectral descriptors with the eigenfunction-squared form reviewed in [41] can be considered in this framework for isometric invariant representation. Here, the spectral graph wavelet signature (SGWS) is adopted as local descriptor. SGWS provides a general and flexible interpretation for the analysis and design of spectral descriptors. For a vertex x of a triangle mesh, it is defined as

$$\text{SGWS}_t(x) = \sum_{i=1}^{m} g(t, \lambda_i)\varphi_i^2(x).$$

In a bid to capture both global and local geometry, a multi-resolution shape descriptor is derived by setting $g(t, \lambda_i)$ as a cubic spline wavelet generating kernel and considering the scaling function (see [40, Eq. (20)] for a precise formulation of g). This leads to the multiscale descriptor defined as $\text{SGWS}(x) = \{\text{SGWS}_t(x), t = 1, \ldots, T\}$, with T the chosen resolution and $\text{SGWS}_t(x)$ the shape signature at the resolution level t. In the proposed implementation T is set to 2.

Texture information. Colour histograms (CH) are used to characterize texture information on the surface. Each channel is discretized into 5 bins.

4.2.2 Shape comparison via intrinsic spatial pyramid matching

To incorporate the spatial information, the Intrinsic Spatial Pyramid Matching (ISPM) [39] is considered. ISPM can provably imitate the popular spatial pyramid matching (SPM) [35] to partition a mesh in a consistent and easy way. Then, *Bag-of-Feature* (BoF) and *Locality-constrained Linear Coding* (LLC) [78] can be used to characterize the partitioned regions.

The isocontours of the second eigenfunction (Figure 3) are considered to partition the shape into R regions, with R $= 2^{l-1}$ for the partition at a resolution level l. Indeed, the second eigenfunction is the smoothest mapping from the manifold to the real line, making this intrinsic partition quite stable. Thus, the shape description is given by the concatenation of R sub-histograms of SGWS and CH along eigenfunction values in the real line. To consider the two-sign possibilities in the concatenation, the histogram order is

inverted, and the scheme with the minimum cost is considered as a better matching. Therefore, the descriptive power of SGWS and CH is enhanced by incorporating this spatial information.
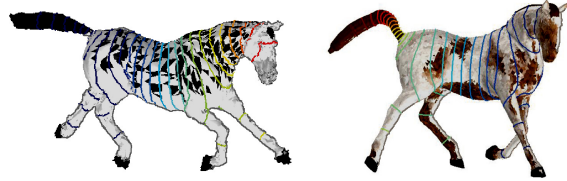

Fig. 3 The isocontours of the second eigenfunction.

Given a SGWS+CH descriptor densely computed on each vertex on a mesh, quantization via the codebook model approach is adopted to obtain a compact histogram shape representation. The classical k-means method is used to learn a dictionary $Q = \{q_1, \ldots, q_K\}$, where words are obtained as the K centroids of the k-means clusters. In the proposed implementation, $K = 100$. In order to assign the descriptor to a word in the vocabulary, approximated LLC is performed for fast encoding, then max-pooling is applied to each region. Finally, ISPM induced histograms for shape representation are derived.

The dissimilarity between two shapes is given by L_2 distance between the associated ISPM induced histograms. Geometry and texture information are handled separately, and the final dissimilarity score is a combination of the geometric and the texture distance.

4.2.3 The runs

The proposed approach has been implemented to derive three different runs for the retrieval task:

- **LBG1** represents LCC strategy with partition level $l = 1$ for geometric information;
- **LBG2** represents LCC strategy with partition level $l = 3$ for geometric information;
- **LBG3** is a weighted combination of geometric and texture information, namely LCC strategy with partition level 3 for SGWS and partition level $l = 5$ for colour histograms, with coefficients 0.8 and 0.2, respectively.

For the classification task, two nearest neighbour classifiers are derived, a geometric one from **LBG2** and a texture one from the texture contribution of **LBG3**. In what follows, the latter is referred to as **LBGtxt**.

4.3 Colour + Shape descriptors (runs **Ve1-3**)

This method is a modification of the "3D Shape + colour" descriptor proposed in [9]. To describe a textured 3D shape $\mathcal{S}$ represented by a surface mesh S, two main steps are considered:

1. Let $\mathbf{G}$ be a $n \times n$ geodesic distance matrix, where n is the number of vertices in S and the element $\mathbf{G}(i,j)$ denotes the geodesic distance from the vertex i to vertex j on S. Building on $\mathbf{G}$, the centralised geodesic matrix [50] is defined as $\mathbf{D} = \mathbf{G} - \mathbf{1}_n\mathbf{G} - \mathbf{G}\mathbf{1}_n + \mathbf{1}_n\mathbf{G}\mathbf{1}_n$, where $\mathbf{1}_n$ denotes a $n \times n$ matrix having each component equal to $1/n$. Following [68], a spectral representation of the geodesic distance is finally adopted as shape descriptor, that is, a vector of eigenvalues $\texttt{Eig}(\mathbf{D}) = (\lambda_1(\mathbf{D}), \ldots, \lambda_n(\mathbf{D}))$, where $\lambda_i(\mathbf{D})$ is the ith largest eigenvalue. As in [9], the first 40 eigenvalues are used as shape descriptor. The vectors of eigenvalues $\texttt{Eig}(\mathbf{D_1})$, $\texttt{Eig}(\mathbf{D_2})$ associated with two shapes $\mathcal{S}_1, \mathcal{S}_2$ are compared through the *mean normalized Manhattan distance*, i.e.,

$$d_{geo}(\mathcal{S}_1, \mathcal{S}_2)) = \sum_{k=1}^{40} \frac{2|\lambda_k(\mathbf{D_1}) - \lambda_k(\mathbf{D_2})|}{\lambda_k(\mathbf{D_1}) + \lambda_k(\mathbf{D_2})}.$$

2. To incorporate texture information in the shape descriptor, the RGB colour histograms are considered as in [9]. Accordingly, the distance $d_{clr}(\mathcal{S}_1, \mathcal{S}_2))$ between the texture shape descriptors associated with $\mathcal{S}_1, \mathcal{S}_2$ is given by the Earth mover's distance (EMD) between the corresponding RGB colour histograms. For two histograms $\mathbf{p}$ and $\mathbf{q}$, the EMD measures the minimum work that is required to move the region lying under $\mathbf{p}$ to that under $\mathbf{q}$. Mathematically, it has been defined as the total flow that minimizes the transport from $\mathbf{p}$ to $\mathbf{q}$. We refer keen readers to [63] for a comprehensive review of EMD formulation, and to [58] for an application to shape retrieval. To concretely evaluate EMD, the fast implementation introduced by [56] has been used with a thresholded ground distance.

Last, the final distance between $\mathcal{S}_1$ and $\mathcal{S}_2$ is defined as follows:

$$D(\mathcal{S}_1, \mathcal{S}_2) = (d_{geo}(\mathcal{S}_1, \mathcal{S}_2))^p + (d_{clr}(\mathcal{S}_1, \mathcal{S}_2))^{1-p},$$

where p is a parameter to control the trade-off between colour and shape information. In the experiments, $p = 0.75$ (run **Ve1**), $p = 0.85$ (run **Ve2**), and $p = 0.95$ (run **Ve3**), following the paradigm that geometric shape properties should be more important than colourimetric ones in the way humans interpret similarity between shapes. An illustration of the proposed description for a textured shape is given in Fig. 4.
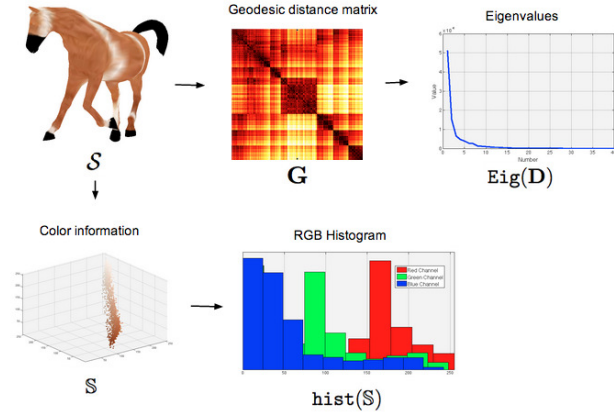
Fig. 4 Proposed method includes a shape descriptor from the geodesic distance matrix and a colour descriptor from the histogram representation of RGB colour information. Details are included in section 4.3

4.4 Textured shape distribution, joint histograms and persistence (runs **Gi1-3**)

The CIELab colour space well represents how human eyes perceive colours. Indeed, uniform changes of coordinates in the CIELab space correspond to uniform changes in the colour perceived by the human eye. This does not happen with some other colour spaces, for example the RGB space. In the CIELab colour space, tones and colours are held separately: the L channel is used to specify the luminosity or the black and white tones, whereas the a channel specifies the colour as either a green or a magenta hue and the b channel specifies the colour as either a blue or a yellow hue.

Run **Gi1**. The Textured Shape Distribution (TSD) descriptor is a colour-aware variant on the classical Shape Distributions (SD) descriptor [53]. Indeed, TSD consists of the distribution of colour-aware geodesic distances, which are computed between a number of sample points scattered over the surface mesh representing the 3D model.

The surface mesh is embedded in the 3-dimensional CIELab colour space, so that each vertex has (L,a,b) coordinates. Then, in order to get colour-aware geodesic distances, a metric has to be defined in the embedding space. To this end, the length of an edge is defined as the distance between its endpoints, namely, the CIE94 distance defined for CIELab coordinates [18]. This distance is used here instead of a classical Euclidean distance as it was specifically defined for the CIELab space, and employs specific weights to respect perceptual uniformity [18]. The colour-aware geodesic distances are computed in the embedding space with

the metric induced by the CIE94 distance. The distances are computed between pairs of points sampled over the surface mesh. A set of 1024 points was sampled in the current implementation, following a farthest-point criterion. The Dijkstra algorithm was used to compute colourimetric geodesic distances between pairs of samples.

The final descriptor encodes the distribution of these distances. In the current implementation, the distribution was discretized using a histogram of 64 bins. Histograms were compared using the L_2 norm. Therefore, the distance between two models is the distance between their descriptors, namely the L_2 norm between the corresponding histograms. TSD encodes the distribution of colour distances, yet it also takes into account the connectivity of the underlying model, as distances are computed by walking on the surface model. In this sense, TSD can be considered as a hybrid descriptor, taking into account both colourimetric and geometric information.

Run **Gi2**. Though TSD retains some information about the shape of 3D models, in terms of the connectivity of the mesh representing the object, it still loses most of the geometric information about the object, as it does not take into account the length of the edges in the Euclidean space. This geometric information can be recovered by using a joint distribution, which takes into account both colourimetric geodesic distances and classical geodesic distances computed on the surface embedded in the Euclidean space. In this run, the joint distribution has been discretized by computing a 16×16 bi-dimensional joint histogram (JH) for each 3D model. The L_2-norm is used for comparison. The distance matrix is the sum of the distance matrix obtained using the TSD descriptor and the distance matrix obtained using the JH descriptor.

Run **Gi3**. In [5] the authors proposed a signature which combines geometric, colourimetric, and hybrid descriptors. In line with this idea, Run Gi3 combines TSD with a geometric descriptor, namely the popular Spherical Harmonic (SD) descriptor [29], and a colourimetric descriptor, namely the persistence-based descriptor of the PHOG signature in [5], using the CIELab colour space coordinates. The distance matrix corresponding to this run is the sum of the three distance matrices obtained using the TSD descriptor, the SH descriptor, and the persistence-based descriptor of PHOG, respectively.

4.5 Multi-resolution Representation Local Binary Pattern Histograms (run **TAS**)

The Multi-resolution Representation Local Binary Pattern Histograms (MRLBPH) is proposed here as a novel 3D model feature that captures textured features of rendered images from 3D models by analysing multi-resolution representations using Local Binary Pattern (LBP) [52].

Figure 5 illustrates the generation of MRLBPH. A 3D model is normalized via Point SVD [74] to be contained in a unit geodesic sphere. From each vertex of the sphere, depth and colour buffer images with 256×256 resolution are rendered; a total of 38 viewpoints are defined. A depth channel and each CIELab colour channel are then processed as detailed in what follows.

To obtain multi-resolution representations, a Gaussian filter is applied to an image with varying standard deviation parameters. The standard deviation parameter σ_l at level l is evaluated by using the following equation involving the left factorial of l:

$$\sigma_l = \sigma_0 + \alpha \cdot !l, \ l = 0, \dots, \Lambda$$

where σ_0 is the initial value of the standard deviation parameter, and α is the incremental parameter. This equation has been derived from the optimal standard deviation parameters obtained through preliminary experiments. In the proposed implementation, $\sigma_0 = 0.8$ and $\alpha = 0.6$, while the number of levels Λ is set to 4.

For each scale image, a LBP histogram is evaluated. To incorporate spatial location information, the image is partitioned into 2×2 blocks and the LBP histogram at each block is computed. The LBP histogram of each scale image is obtained by concatenating the histograms of these blocks. Let g_c denote the image value at arbitrary pixel (u, v), and let $g_1, \dots, g_8$ be the image values of each of the eight neighbourhood pixels. The LBP value is then calculated as

$$\text{LBP}(u, v) = \sum_{i=1}^{8} s(t, g_i - g_c) \cdot 2^{i-1},$$

where $s(t, g)$ is a threshold function defined as 0 if $g < t$ and 1 otherwise. In the proposed implementation, the threshold value t is set to 0, and the LBP values are quantized into 64 bins.

An MRLBP histogram is generated by merging the histograms of scale images through the selection of the maximum value of each histogram bin. Let $h_i^{(l)}$ be the ith LBP histogram element of a scale image at level l. The ith MRLBP histogram element h_i is defined as

$$h_i = \max_l h_i^{(l)}, \quad l = 0, \dots, \Lambda.$$
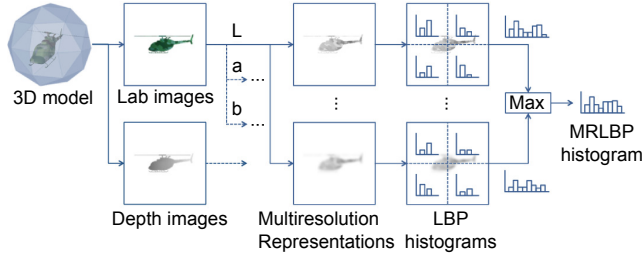
Fig. 5 Overview of our Multiresolution Representation Local Binary Pattern Histograms (MRLBPH)

The MRLBP histogram is finally normalized using the L_1 norm.

The feature vector associated with a 3D model is obtained by calculating the MRLBP histogram of the depth and CIELab channel for each viewpoint.

To compare two shapes $\mathcal{S}_1$ and $\mathcal{S}_2$, the Hungarian method [34] is applied to all dissimilarities between the associated MRLBP histograms. In evaluating the final dissimilarity score, the histograms of the depth and CIELab channels are combined by calculating the weighted sum of each dissimilarity. Let d_d, d_L, d_a, and d_b denote the dissimilarity of each channel, and let w_d, w_L, w_a, and w_b be the weight of each channel. The dissimilarity $D(\mathcal{S}_1, \mathcal{S}_2)$ is defined as the dissimilarity between the corresponding MRLBP histograms:

$$D(\mathcal{S}_1, \mathcal{S}_2) = w_d d_d(\mathcal{S}_1, \mathcal{S}_2) + w_L d_L(\mathcal{S}_1, \mathcal{S}_2)$$
$$+ w_a d_a(\mathcal{S}_1, \mathcal{S}_2) + w_b d_b(\mathcal{S}_1, \mathcal{S}_2).$$

In this implementation, w_d is set to 0.61, w_L to 0.13, w_a to 0.13, and w_b to 0.13. For the dissimilarity between two histograms, the Jeffrey divergence is used [15].

4.6 PHOG: Photometric and geometric functions for textured shape retrieval (runs **BCGS1-3**)

The combination of colourimetric properties and geometric properties represented in terms of scalar and multi-variate functions has been explored in PHOG [5], a shape signature consisting of three parts:

- A colourimetric descriptor. CIELab colour coordinates (normalized L, a, b channels) are seen as either scalar or multi-variate functions defined over the shape. The CIELab colour space is considered due the perceptual uniformity of this colour representation;
- A hybrid descriptor. Shape and texture are jointly analysed by opportunely weighting the colourimetric information (L, a, b channels) with respect to the underlying geometry and topology;
- A geometric description relying on a set of functions representing as many geometric shape properties.

Functions are first clustered; then, a representative function is chosen for each cluster. The goal here is to select functions that are mutually independent, thus complementing each other via the geometric information they carry with them.

Figure 6 shows a pictorial representation for the generation of a PHOG signature.

Run **BCGS1**. Following the PHOG original setting, the colourimetric description is included in the persistence framework. Indeed, the a, b coordinates are used to jointly define a bivariate function over a given shape, whereas L is used as a scalar function. In this way, colour and intensity are treated separately. Precisely, for a shape $\mathcal{S}$ represented by a triangle mesh S, the two functions $f_L : S \to \mathbb{R}$ and $f_{a,b} : S \to \mathbb{R}^2$ are considered, the former taking each point $x \in S$ to the L-channel value at x, the latter to the pair given by the a- and the b-channel values at x, respectively. The values of f_L and $f_{a,b}$ are then normalized to range in the interval $[0,1]$. Last, $\mathcal{S}$ is associated with the $0th$ persistence diagram $\mathrm{Dgm}(f_L)$ and the $0th$ persistence space $\mathrm{Spc}(f_{a,b})$: these descriptors encode the evolution of the connectivity in the sublevel sets of f_L and $f_{a,b}$ in terms of birth and death (i.e. merging) of connected components, see [5] for more details.

The hybrid description comes from a geodesic distance $f_{geod} : S \to \mathbb{R}$ defined in a higher dimensional embedding space, similarly to the approach proposed in [31,33], and used as a real-valued function in the persistence framework to associate $\mathcal{S}$ with the persistence diagram $\mathrm{Dgm}(f_{geod})$. The definition of the joint geometric and colourimetric integral geodesic distance is straightforward and implemented through the Dijkstra's algorithm, which is based on edge length.

The geometric description is based on the *DBSCAN* clustering technique [17]. Once a set of functions $\{f_i : S \to \mathbb{R}\}$ (from an original set of 70 geometric functions, see [5] for the complete list) is selected, a matrix $MDM(S)$ with entries

$$MDM(i,j) := 1 - \frac{1}{\mathrm{Area}(S)} \sum_{t \in S} \left| < \frac{\nabla^t f_i}{\|\nabla^t f_i\|}, \frac{\nabla^t f_j}{\|\nabla^t f_j\|} > \right|$$

is used to store the distances between all the possible couple of functions, with $\nabla^t f_i$, $\nabla^t f_j$ representing the gradient of f_i and f_j over the triangle t of the mesh S.

To assess the similarity between two shapes $\mathcal{S}_1$ and $\mathcal{S}_2$, the corresponding colourimetric, hybrid and geometric descriptions are compared. In particular, the colourimetric distance $d_{clr}(\mathcal{S}_1, \mathcal{S}_2)$ is the normalized sum of the Hausdorff distance between the 0th persistence diagrams of f_L and that between the 0th persistence
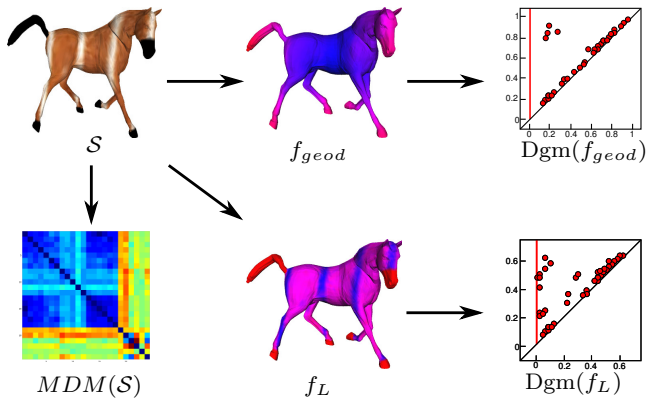
Fig. 6 Generating a PHOG signature. First row: A shape $\mathcal{S}$ (left), the function f_{geod} (center) and the corresponding persistence diagram Dgm(f_{geod}). Second row: the mutual distance matrix $MDM(\mathcal{S})$, the function f_L and the corresponding persistence diagram Dgm(f_L).

spaces of $f_{a,b}$; the hybrid distance $d_{hbd}(\mathcal{S}_1, \mathcal{S}_2)$ is the Hausdorff distance between the corresponding persistence diagrams of f_{geod}; the geometric distance $d_{geo}(\mathcal{S}_1, \mathcal{S}_2)$ is computed as the Manhattan distance between the matrices $MDM(S_1)$ and $MDM(S_2)$. The final dissimilarity score between $\mathcal{S}_1$ and $\mathcal{S}_2$ is the normalized sum $d_{clr}(\mathcal{S}_1, \mathcal{S}_2) + d_{hbd}(\mathcal{S}_1, \mathcal{S}_2) + d_{geo}(\mathcal{S}_1, \mathcal{S}_2)$.

Variations. Several variations of the PHOG framework are possible, for instance exploring the use of different distances between feature vectors or dealing with variations of the three (colourimetric, hybrid, geometric) shape descriptions. For the current implementation the following changes have been proposed:

- run **BCGS2**. The original hybrid description is replaced by a histogram-based representation of the geodesic distance. While getting rid of the additional geometric contribution provided by persistence, the hybrid perspective is maintained as the considered geodesic distance takes into account both geometric and texture information.
- run **BCGS3**. The stability properties of persistence diagrams and spaces imply robustness against small variations in the L, a, b values. This also holds when colour perturbations are widely spread over the surface model, as in the case of slight illumination changes. On the other hand, colour histograms behave well against localized colourimetric noise, even if characterized by large variations in the L, a, b values. Indeed, in this case colour distribution is not altered greatly. In this view, the idea is to replace the hybrid contribution with CIELab colour histograms, so to improve the robustness properties of the the persistence-based description. Histograms are ob-

tained as the concatenation of the L, a, b colour channels.

In Runs **BCGS2-3**, histograms are compared through the Earth Mover's distance (EMD). The DBSCAN clustering technique for selecting representative geometric functions is replaced by the one used in [8], which is based on the *replicator dynamics technique* [55]. The modified geometric descriptors are compared via the EMD as well, after converting the MDM matrices into feature vectors.

4.7 Taxonomy of the methods

The methods detailed above can be considered as representatives of the variety of 3D shape retrieval and classification techniques overviewed in Section 2. Indeed, they range from local feature vector descriptions coded as histograms of geometric and/or colour properties, to spectral and topological based descriptions, including also a spatial pyramid matching framework. In what follows, we group the properties of these methods on the basis of the key characteristics they exhibit, e.g., the geometric and colourimetric structure they capture, at which scale level the shape description is formalized, or which colour space has been chosen for texture analysis. These characteristics are briefly described in the following and summarized in Table 1.

Intrinsic vs. extrinsic. Studying the geometric shape of a 3D model relies on the definition of a suitable metric between its points. Among the possible options, two particular choices appear quite natural.

The first one is to consider the Euclidean distance, which in turn reflects the *extrinsic* geometry of a shape. Extrinsic shape properties are related to how the shape is laid out in an Euclidean space, and are therefore invariant to rigid transformations, namely rotations, translations an reflections.

A second choice is to measure the geodesic distance between points, that is, to consider the *intrinsic* geometry of a shape. Intrinsic shape properties are invariant to those transformations preserving the intrinsic metric, including rigid shape deformations but also non-rigid ones such as shape bendings.

Methods associated with runs **GG(1-3)**, **TAS** and **Gi3** are examples of extrinsic approaches; spectral methods (runs **LBG(1-3)**) and those based on geodesic distances (runs **Ve(1-3)** and **Gi(1-2)**) well represent intrinsic approaches. The PHOG method (runs **BCGS(1-3)**) can be seen as a "mostly" intrinsic approach, being based on a collection of geometric functions which

prevalently describe intrinsic shape properties.

Global vs. multi-scale. Shape descriptors generally encode information about *local* or *global* shape properties. Local properties reflect the structure of a shape in the vicinity of a point of interest, and are usually unaffected by the geometry or the topology outside that neighbourhood. Global properties, on the other hand, capture information about the whole structure of a shape.

Another option is to deal with shape properties at different scales, thus providing a unifying interpretation of local and global shape description. Such an approach is usually referred to as *multi-scale.*

The methods presented in this contribution can be classified as global and multi-scale ones, both for geometric and texture information, see Table 1 for details.

RGB vs. non-RGB. The more natural colour space to be used in the analysis of colourimetric shape properties appears to be the RGB one. This was actually the choice carried on by methods associated with runs **LBGtxt** and **Ve(1-3)**. However, other options are possible. For instance, runs **GG(1-3)** are based on normalized and averaged RGB channels, while methods related to runs **TAS**, **BCGS(1-3)**, **Gi(1-3)** study colourimetric shape properties in the CIELab colour space.

Feature vectors vs. topology. Once geometric and texture shape properties have been analysed and captured, they have to be properly represented through suitable shape descriptors. The most popular approach is to use *feature vectors* [73], and most of the methods implemented in this paper actually adopt this description framework. Feature vectors generally encode shape properties expressed by functions defined on the shape, and are usually represented as *histograms*. While being very efficient to compute, histograms might forget part of the structural information about the considered shape property. To overcome this limitation, it is possible to consider the shape connectivity directly at the function level, see for instance shape distributions (runs **Gi1** and **Gi2**). Alternatively, one can move to more informative histogram variations: *bi-dimensional histograms*, such as the geodesic distance matrix used in runs **Ve(1-3)** and the mutual distance matrix adopted in runs **BCGS(1-3)**, or *concatenated histograms* obtained at different resolution levels, as in the case of runs **GG(1-3)**, **LBG(1-3,txt)** and **TAS**.

A different way to preserve the structure of geometric information is provided by descriptors rooted in *topology* (runs **BCGS(1-3)** and **Gi3**). Indeed, they keep track of the spatial distribution of a considered shape property, and possibly encode the mutual relation among shape parts of interest, that is, regions that are highly characterized by the considered property. The reader is referred to Table 1 for details about the approaches adopted by methods under evaluation.

Hybrid vs. combined. Finally, methods can be distinguished by the way the geometric and texture information are merged together. Typically, this can be done either *a priori* or *a posteriori.* The first case results in a *hybrid* shape descriptor, as for runs **GG3** and partially for runs **BCGS(1-3)** and **Gi(1-3)**. In the second case, a pure geometric and a pure texture descriptor are obtained and compared separately, while the final dissimilar score is a weighted *combination* of the two distances.

5 Comparative analysis

The methods detailed in Section 4 have been evaluated through a comparative study presented in what follows. Each run has been processed in terms of the output specified in Section 3.2 and according to the evaluation measures described in Sections 3.3.1 and 3.3.2.

5.1 Retrieval performances.

Following [76], the retrieval performance of each run has been evaluated according to the following relevance scale: if a retrieved object shares both shape and texture with the query, then it is highly relevant; if it shares only shape, it is considered marginally relevant; otherwise, it is not relevant. Note that, because of the multi-level relevance assessment of each query, most of the evaluation measures have been split up as well. *Highly relevant* evaluation measures relate to the highly relevant items only, while *relevant* evaluation measures are based on all the relevant items (highly relevant items + marginally relevant items).

5.1.1 Highly relevant evaluation.

In the highly relevant scenario, the main goal is to evaluate the performance of algorithms when models vary by both geometric shape and texture.

Figure 7 shows the performances of the six methods in terms of the *average* precision-recall curve, which is obtained as the average of the precision-recall curves computed over all the queries. To ease results visualization, the plot in Figure 7 includes only the best run for each method, that is, runs with the highest *mean average precision* (mAP) score. We remind that, for ideal retrieval systems, the mAP score equals to 1.
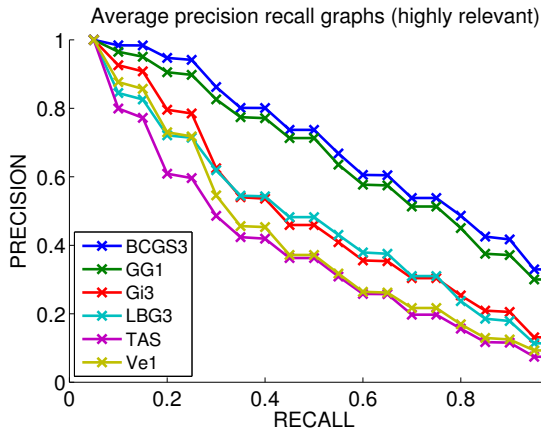
Fig. 7 Highly relevant precision-recall curves for the best run of each method.

The runs of Figure 7 have been analysed also in terms of the *weighted average* mAP score computed over the 48 classes that represent the highly relevant scenario, where weights are determined by the size of the classes, see the second column of Table 2. Moreover, we considered also the percentage of classes whose mAP score is larger that some threshold values, namely 0.40, 0.55, 0.70, 0.85, see columns (3-6) in Table 2.

Table 2 Highly relevant analysis for the runs in Figure 7: weighted average mAP score (first column), and how many of the 48 highly relevant classes have a mAP score exceeding values 0.40, 0.55, 0.70, 0.85 ((third - last column, respectively; results are reported in percentage points). The best two results are in gold and silver text, respectively.

Runs	mAP	> 0.40 (%)	> 0.55 (%)	> 0.70 (%)	> 0.85 (%)
BCGS3	0.7225	100.00	79.17	52.08	27.08
GG1	0.6980	100.00	81.25	50.00	18.75
Gi3	0.5365	79.17	39.58	18.75	4.17
LBG3	0.5256	77.08	33.33	10.42	2.08
TAS	0.4380	64.58	12.50	2.08	0.00
Ve1	0.4671	68.75	29.17	4.17	0.00

To further analyse retrieval performances against texture deformations, we restrict the mAP analysis to each specific class of colourimetric transformations described in Section 3. More precisely, we first let the algorithms run exclusively on set of null models. Then, we add only those elements that come as a result of one of the five texture transformations used to generate the entire dataset. Note that, according to the procedure used to create the benchmark, each texture deformation is always applied together with a geometric one, hence it still makes sense to apply the highly relevant

paradigm along the evaluation process. Table 3 summarizes the results.

Looking at how the performances degrade across the different families of transformations, it can be noted that all methods appear to be not too sensitive against transformations of type 2, while the worst results are distributed among transformations of type 1 (runs **Gi3**, **LBG3**, **TAS** and **Ve2**), and 3 (runs **BCGS3** and **GG1**).

Table 4 reports the best highly relevant performances in terms of Nearest Neighbour, First Tier and Second Tier evaluation measures. Additionally, the last column of Table 4 records the ADR measures. All the scores, which range from 0 (worst case) to 1 (ideal performance), are averaged over all the models in the dataset.

Table 4 Best NN, FT, ST and ADR values for each method. Numbers in parenthesis indicate the run achieving the corresponding value. For each evaluation measure, the best two results are in gold and silver text, respectively.

Runs	NN	FT	ST	ADR
BCGS	0.967(3)	0.620(3)	0.760(3)	0.496(3)
GG	0.930(3)	0.600(1)	0.740(1)	0.478(1)
Gi	0.894(2)	0.455(3)	0.590(3)	0.383(3)
LBG	0.686(3)	0.440(3)	0.592(3)	0.369(3)
TAS	0.558	0.375	0.527	0.318
Ve	0.735(1)	0.396(1)	0.539(1)	0.342(1)

Finally, Figure 8 shows the best run of each method according to the NDCG measure as a function of the rank p. In the present evaluation, the NDCG values for all queries are averaged to obtain a measure of the average performance for each submitted run. Remind that, for an ideal run, it would be NDCG $\equiv 1$.
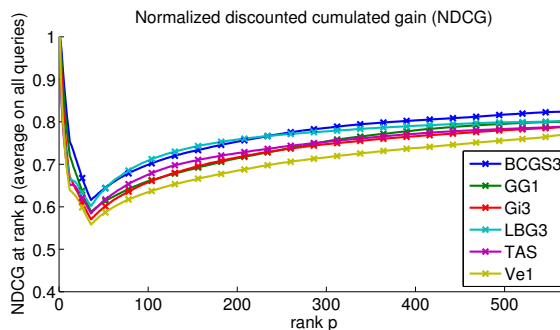
Fig. 8 Performances of the best runs w.r.t. the NDCG measure (run **GG1** is almost totally covered by runs **Gi3**, **LBG3** and **TAS**).

The NDCG measure takes geometric retrieval performances into larger account than texture ones. In-

deed, geometric shape similarity is involved in the definition of both relevant and highly relevant items. This means that runs characterized by moderate geometric retrieval performances will be more penalized than others (see also Section 5.1.2). This is the case of run **GG1**, which is indeed definitely tuned for the highly relevant rather then the relevant scenario.

Discussion. Trying to interpret the outcome of the highly relevant evaluation, we are led to the following considerations:

• The algorithm design associated with runs **BCGS3** and **Gi3** proposes a similar combination of geometric and texture information. Indeed, both methods rely on a hybrid shape description, in which texture contribution is in part based on a geometric–topological analysis of colourimetric properties and carried out in the CIELab colour space. In other words, a "structured" analysis of the colour channels is paired with the choice of a colour space that better reflects, with respect to the RGB one, human colour perception. Moreover, the geometric–topological approach allows for keeping track of the underlying connectivity of 3D models, thus providing additional information about the spatial distribution of colourimetric shape properties;

• Run **GG1** represents a combined shape descriptor, whose texture contribution is based on considering a normalized version of the RGB colour space. Such a choice seems to imply a good robustness against texture affine transformations. Incidentally, it should be noted that, in spite of presenting only the best runs to ease readability and visualization of results, runs **GG2** and **GG3**, which are based on the greyworld RGB channels normalization, exhibit results which are comparable with those of run **GG1**;

• As for runs **LBG** and **Ve**, texture description is accomplished through standard histograms of RGB colour channels, although runs **LBG** incorporate some additional information as the result of considering a multi-resolution approach applied to shape sub-parts. However, it seems that dealing with colourimetric information in other colour spaces, such as the CIELab one or variations of the RGB colour space, allows for a representation of colour that is more robust to the texture deformations proposed in this benchmark.

5.1.2 Relevant evaluation.

In this Section we analyse the performances of methods with respect to their capability of retrieving relevant items; in this case shape (dis)similarity depends only on geometric shape properties. In analogy to Section 5.1.1, Figure 9 shows the best runs for all methods in terms of
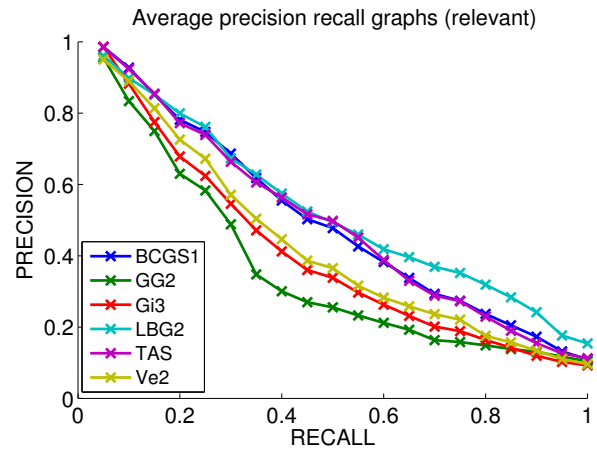
Fig. 9 Relevant precison-recall curves for the best run of each method.

the *average* precision-recall curve. Table 5 reports the mAP scores of those runs, which are now averaged over the 16 geometric classes composing the dataset. Also in this case, we report for how many classes the mAP score exceeds values 0.40, 0.55, 0.70, 0.85.

Table 5 Relevant analysis for the runs in Figure 9: weighted average mAP score (first column), and how many of the 16 relevant classes have a mAP score exceeding values 0.40, 0.55, 0.70, 0.85 (third - last column, respectively; results are reported in percentage points). The best two results are in gold and silver text, respectively.

Run	mAP	> .40 (%)	> .55 (%)	> .70 (%)	> .85 (%)
BCGS1	0.5181	81.25	37.50	6.25	0.00
GG2	0.3811	37.50	6.25	0.00	0.00
Gi3	0.4266	56.25	12.50	0.00	0.00
LBG2	0.5480	81.25	43.75	18.75	0.00
TAS	0.5158	81.25	37.50	12.50	0.00
Ve2	0.4479	56.25	25.00	12.50	0.00

Table 6 reports the best relevant retrieval performances according to the Nearest Neighbour, First Tier and Second Tier evaluation measures. All scores are averaged over all the models in the dataset.

Discussion. Apart from the nearest neighbour scores (Table 6, first column), it seems that the overall relevant performance is still not ideal here. This is actually not surprising, since most of methods considered for evaluation have been specifically tuned for dealing with *both* texture and geometric shape modifications. Nevertheless, the relevant evaluation can be used as a lever for further comments about the benchmark and the considered methods.

Table 6 Best NN, FT and ST values of each method. Numbers in parenthesis indicate the run achieving the corresponding value. For each evaluation measure, the best two results are in gold and silver text, respectively.

Runs	NN	FT	ST
BCGS	0.972(2)	0.463(1)	0.601(1)
GG	0.938(3)	0.374(1)	0.374(2)
Gi	0.968(3)	0.383(3)	0.504(3)
LBG	0.939(2)	0.471(2)	0.626(3)
TAS	0.9768	0.4578	0.5847
Ve	0.918(2)	0.398(1)	0.499(1)

As an overall comment, it is worth mentioning that a large part of geometric shape modifications are *not* metric-preserving. Indeed, the geometric deformations used to create the benchmark alter both the extrinsic and the intrinsic properties of shapes. This suggests the need for the development of more general techniques for 3D shape analysis and comparison. This is actually one of the most recent trends in the field, see e.g. [60,59].

More in detail, we observe that:

• The good performance of run **LBG2** relies on two main motivations. First, this run is completely determined by a geometric contribution, and therefore it is not affected by any other, possibly misleading, information about texture shape properties. Second, the method represented by run **LBG2** is spectral-based, and therefore is able to capture intrinsic shape properties. As a consequence, it is invariant to rigid shape transformations, as well as some non-rigid deformations such as pose variations and bendings, which are all present in the dataset. Finally, differently from runs **Gi3** and **Ve2**, whose geometric contribution is intrinsic as well, the descriptive power of the spectral-based approach is improved by additional spatial information provided by the intrinsic spatial pyramid matching (see Section 4.2 for details);

• A similar reasoning about the invariance under rigid and non-rigid deformations also holds for run **BCGS1**, whose associated method can be considered as "mostly" intrinsic. Indeed, the geometric contribution relies in this case on a collection of descriptors which are mainly intrinsic, considering either spectral-based functions or geodesic distances or Gaussian curvature;

• The relatively good performance of run **TAS** can be explained by the fact that it mixes an extrinsic approach with a "view-based" strategy that is widely acknowledged as the most powerful and practical approach for rigid 3D shape retrieval [67]. Even when a 3D model is articulated, non-uniformly deformed or partially occluded, the number of views (38, actually) used in this

implementation should limit the noisy effect possibly generated in the images captured around an object;

• Run **GG2** mainly focuses on radial and spherical local symmetries. While the approach appears to be robust when the analysis is restricted to a single class of geometric transformations, in the most general case (i.e. the whole dataset) results are affected by a non optimal trade-off between geometric and colourimetric contribution. In this way, the relevant retrieval performance is partially conditioned by the inter-class texture variations.

5.2 Classification performances.

In the classification task, each run results in two classification matrices, one for geometry and one for texture, which are derived from the 1-NN classifier associated with the dissimilarity matrices used for the retrieval task. Hence, for a classification matrix C, the element $C(i,j)$ is set to 1 if model i is assigned to class j, meaning that j is the nearest neighbour of i, and 0 otherwise.

Figures 10 and 11 represent the confusion matrices for the best runs of each method. For visual purposes, we have normalized the matrices with respect to the number of elements in each class, so that possible values range from 0 to 1.

Tables 7 and 8 provide a quantitative interpretation of the visual information contained in the confusion matrices. Indeed, the true positive rate (TPR), the true negative rate (TNR) and the Matthews correlation coefficient (MCC) can be directly computed from the elements of a confusion matrix.

More precisely, given a confusion matrix CM and a class $\bar{\imath}$, it is possible to derive the associated TPR, TNR and MCC as follows. It is first convenient to introduce the *number of true positive* TP, the *number of false negative* FN, the *number of false positive* FP and the *number of true negative* TN that are defined as TP $=CM(\bar{\imath}.\bar{\imath})$, FN $=\sum_{j\neq\bar{\imath}}CM(\bar{\imath},j)$, FP $=\sum_{j\neq\bar{\imath}}CM(j,\bar{\imath})$, TN $=\sum_{i,j}CM(i,j)-(\text{TP}+\text{TN}+\text{FP})$. Then, we get TPR, TNR and MCC by the relations

$$\text{TPR} = \frac{\text{TP}}{\text{TP}+\text{FN}}, \ \text{TNR} = \frac{\text{TN}}{\text{FP}+\text{TN}},$$

$$\text{MCC} = \frac{\text{TP}\times\text{TN}-\text{FP}\times\text{FN}}{\sqrt{(\text{TP}+\text{FP})(\text{TP}+\text{FN})(\text{TN}+\text{FP})(\text{TN}+\text{FN})}}.$$

In Tables 7 and 8, the reported values are averaged over all the considered classes (16 geometric and 12 textured ones).

Discussions. Dealing with 1-NN classifiers, the classification results resemble somehow the nearest neighbour performances registered in Tables 4 and 6. Note however, that the dataset classifications considered in this
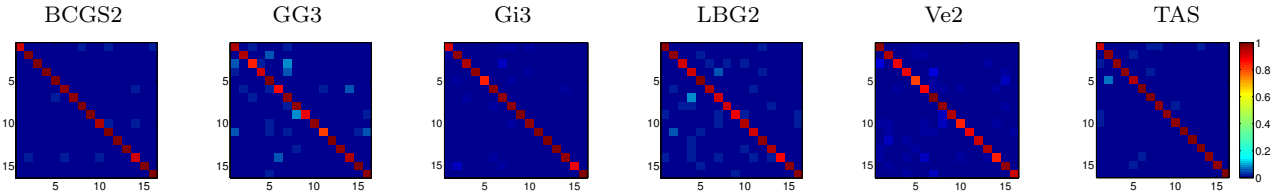
Fig. 10 Geometric confusion matrices for the best runs of the considered methods.
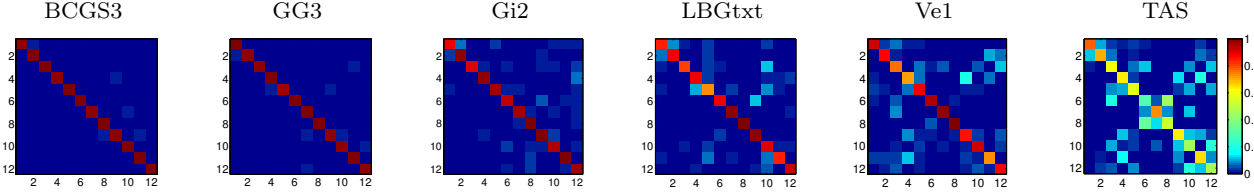

Fig. 11 Texture confusion matrices for the best runs of the considered methods.

Table 7 Averaged geometric TPR, TNR and MCC for the best run of all considered methods. The best two results are in gold and silver text, respectively.

Run	TPR	TNR	MCC
BCGS2	0.9725	0.9988	0.9814
GG3	0.9371	0.9959	0.9341
Gi3	0.9685	0.9979	0.9668
LBG2	0.9388	0.9959	0.9355
TAS	0.9773	0.9985	0.9759
Ve2	0.9178	0.9945	0.9138

Table 8 Averaged texture TPR, TNR and MCC for the best run of all considered methods. The best two results are in gold and silver text, respectively.

Run	TPR	TNR	NCC
BCGS3	0.9913	0.9992	0.9904
GG3	0.9860	0.9987	0.9847
Gi2	0.9161	0.9922	0.9093
LBGtxt	0.8811	0.9885	0.8714
TAS	0.5874	0.9621	0.5524
Ve1	0.8129	0.9820	0.7952

task are a purely geometric and a purely texture ones. In particular, the latter does not coincide with the classification adopted in the highly relevant retrieval task, since geometric similarity is not involved. Also, results in Tables 7 and 8 are averaged on the dataset classes: this explains the slight discrepancy between the geometric TPR and the relevant NN measure that is averaged over all the elements in the dataset.

As shown by Figure 10, geometric confusion matrices reveal a good classification performance of the methods. Indeed, all matrices appear almost diagonal, meaning that almost all elements in the dataset should be correctly classified. This qualitative intuition is con-

firmed by the TPR scores reported in Table 7. Furthermore, TNR values are even higher, thus revealing that all methods are close to the optimal performance in detecting true negatives (e.g., "non-tables" correctly identified as such). As much in the same way, the MCC measure assigns scores very close to 1 for all methods.

Nevertheless, Figure 11 shows that, while the geometric classification of the considered runs are roughly comparable, in the texture scenario a sort of transition occurs, in such a way that three methods (**BCGS3**, **GG3**, **Gi2**) perform substantially better than the others. The numerical details in Table 8 highlight that the main differences are at the TPR and MCC level, revealing that the confusion highlighted in Figure 11 is essentially in the localization of true positives (e.g., "tables" correctly identified as tables).

Finally, it is worth noting how, in the texture classification task, best performances still come from those methods dealing with texture information in colour spaces which differ from the standard RGB one, that is, CIELab and the greyworld normalized RGB colour space.

6 Discussion and conclusions

In this paper, we have provided a detailed analysis and evaluation of state-of-the-art retrieval and classification algorithms dealing with an emerging type of content, namely textured 3D objects, which we believe deserve attention from the research community. The increasing availability of textured models in Computer Graphics, the advances in 3D shape acquisition technology which are able to acquire textured 3D shapes, the importance of colour features in 3D Shape Analysis applications together call for shape descriptors which take into consideration also colourimetric information.

Beyond the extensive analysis that has been carried out throughout the paper, we hope that the experimental results presented here may offer interesting hints for further investigation. We list a few as follows:

• On the one hand, the retrieval performances are positive for all methods, in either the relevant or the highly relevant scenario, or both. On the other hand, the NDCG and ADR measures are specifically designed for interpreting a multi-level dataset classification as in this case, and thus offer a complementary evaluation to the above ones. As can be seen by Figure 8 and Table 4, results in this sense are quite far from being optimal: indeed, the best possible value for the ADR score is 1, while the highest registered scores fluctuate around 0.5; similarly, the highest possible area under a NDCG curve equals 1, while the best scores in this contribution are around 0.75. In other words, the benchmark was challenging and call for further improvements and new strategies able to deal with non-isometric geometric deformations, as well as affine texture deformations;

• Results achieved by some of the proposed runs suggest that a structured colourimetric analysis could be more informative that a purely histogram-bases one. However, there is probably still a long road ahead in this sense. For example, an interesting question could be how to generalize, in a reasonable colourimetric sense, well-know extrinsic and intrinsic geometric properties, for example as in the case of the colour-aware geodesic distances, see Section 4.4 for details;

• An issue deserving further investigation is to understand which approach is preferable for textured shape analysis, either a combined or a hybrid one. For instance, the aforementioned colour-aware geodesic distances and the topological approach for the analysis of colour information appear to be promising hybrid solutions to extract useful information from texture shape attributes. Nevertheless, it should be noted that such approaches have been complemented with purely geometric and colourimetric contributions, in order to achieve satisfactory retrieval and classification performances. Another meaningful example is the one provided by MAPT-based algorithms, which obtained top-rank results in the highly relevant scenario: however, a combined approach generally performed better than a hybrid one in a retrieval context, being the opposite in the classification task. Based on the above remarks, it seems that the overall picture is still quite unclear, calling for a deeper understanding of how geometric and colourimetric shape properties can be jointly analysed;

• Some of the best retrieval and classification results have been accomplished through the use of the CIELab colour space, as well as variations of the more classical RGB colour space. Indeed, the CIELab space well represents human perception of colour, and hence appears as a more natural choice. As for the considered variations of the RGB colour space, it seems that they allow to better cope with the affine colourimetric transformations that have been included in the proposed benchmark. Obviously, these are not the only possible alternative to the RGB colour space. Also, the affine colourimetric transformations considered here are just a subset of all the possible texture modifications. Therefore, it could be interesting to investigate how changing the choice of colour representation might affect performance results in retrieving and classifying textured 3D models, as well as which particular choice in the colour space is better suited to face with certain classes of texture deformations.

References

1. TurboSquid. http://www.turbosquid.com/
2. Aubry, M., Schlickewei, U., Cremers, D.: The wave kernel signature: A quantum mechanical approach to shape analysis. In: Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on, pp. 1626–1633 (2011)
3. Biasotti, S., Cerri, A., Abdelrahman, M., Aono, M., Ben Hamza, A., El-Melegy, M., Farag, A., Garro, V., Giachetti, A., Giorgi, D., Godil, A., Li, C., Liu, Y.J., Martono, H.Y., Sanada, C., Tatsuma, A., Velasco-Forero, S., Xu, C.X.: Retrieval and Classification on Textured 3D Models. In: Eurographics Workshop on 3D Object Retrieval (3DOR), pp. 111–120 (2014)
4. Biasotti, S., Cerri, A., Bronstein, A., Bronstein, M.: Quantifying 3D Shape Similarity Using Maps: Recent Trends, Applications and Perspectives. In: S. Lefebvre, M. Spagnuolo (eds.) EG 2014 - STARs, pp. 135–159 (2014)
5. Biasotti, S., Cerri, A., Giorgi, D., Spagnuolo, M.: PHOG: Photometric and geometric functions for textured shape retrieval. Comput. Graph. Forum **32**(5), 13–22 (2013)
6. Biasotti, S., Giorgi, D., Spagnuolo, M., Falcidieno, B.: Reeb graphs for shape analysis and applications. Theor. Comput. Sci. **392**(1-3), 5–22 (2008)
7. Biasotti, S., Giorgi, D., Spagnuolo, M., Falcidieno, B.: Size functions for comparing 3D models. Pattern Recogn. **41**(9), 2855–2873 (2008)
8. Biasotti, S., Spagnuolo, M., Falcidieno, B.: Grouping real functions defined on 3D surfaces. Comput. Graph. **37**(6), 608 – 619 (2013)
9. Cerri, A., Biasotti, S., Abdelrahman, M., Angulo, J., Berger, K., Chevallier, L., El-Melegy, M., Farag, A., Lefebvre, F., Giachetti, A., Guermoud, H., Liu, Y.J., Velasco-Forero, S., Vigouroux, J., Xu, C.X., Zhang, J.B.: SHREC'13 Track: Retrieval on Textured 3D Models. In:

Eurographics Workshop on 3D Object Retrieval (3DOR), pp. 73–80 (2013)

10. Cerri, A., Landi, C.: The persistence space in multidimensional persistent homology. In: R. Gonzalez-Diaz, M.J. Jimenez, B. Medrano (eds.) Discrete Geometry for Computer Imagery, *Lecture Notes in Computer Science*, vol. 7749, pp. 180–191. Springer Berlin Heidelberg (2013)

11. Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Stability of persistence diagrams. Discr. Comput. Geom. **37**(1), 103–120 (2007)

12. Cortelazzo, G., Orio, N.: Retrieval of Colored 3D Models. In: 3D Data Processing, Visualization, and Transmission, Third International Symposium on, pp. 986–993 (2006)

13. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition (CVPR), 2005 IEEE Conference on, vol. 1, pp. 886–893 (2005)

14. Desbrun, M., Meyer, M., Schröder, P., Barr, A.H.: Implicit fairing of irregular meshes using diffusion and curvature flow. In: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '99, pp. 317–324. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA (1999)

15. Deza, M.M., Deza, E.: Encyclopedia of Distances. Springer Berlin Heidelberg (2009)

16. Edelsbrunner, H., Harer, J.: Persistent homology—a survey. In: Surveys on discrete and computational geometry, *Contemp. Math.*, vol. 453, pp. 257–282. Amer. Math. Soc., Providence, RI (2008)

17. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD, pp. 226–231 (1996)

18. Fairchild, M.D.: Color Appearance Models. John Wiley & Sons Ltd (2005)

19. Fawcett, T.: An introduction to ROC analysis. Pattern Recogn. Lett. **27**(8), 861 – 874 (2006)

20. Finlayson, G.: Coefficient Color Constancy. Simon Fraser University (1995)

21. Frosini, P., Landi, C.: Size theory as a topological tool for computer vision. Pattern Recogn. and Image Anal. **9**(4), 596–603 (1999)

22. Funkhouser, T., Kazhdan, M., Min, P., Shilane, P.: Shape-based retrieval and analysis of 3d models. Commun. ACM **48**(6), 58–64 (2005)

23. Gębal, K., Bærentzen, J.A., Aanæs, H., Larsen, R.: Shape analysis using the auto diffusion function. Comput. Graph. Forum **28**(5), 1405–1413 (2009)

24. Gevers, T., Smeulders, A.W.: Pictoseek: combining color and shape invariant features for image retrieval. IEEE T. Image Process. **9**(1), 102–119 (2000)

25. Giachetti, A., Lovato, C.: Radial symmetry detection and shape characterization with the multiscale area projection transform. Comput. Graph. Forum **31**(5), 1669–1678 (2012)

26. Giorgi, D., Attene, M., Patane, G., Marini, S., Pizzi, C., Biasotti, S., Spagnuolo, M., Falcidieno, B., Corvi, M., Usai, L., Roncarolo, L., Garibotto, G.: A critical assessment of 2d and 3D face recognition algorithms. In: Advanced Video and Signal Based Surveillance (AVSS), 2009 IEEE International Conference on, pp. 79–84 (2009)

27. Johnson, A.E., Hebert, M.: Using spin images for efficient object recognition in cluttered 3D scenes. IEEE T. Pattern Anal. **21**(5), 433–449 (1999)

28. Kanezaki, A., Harada, T., Kuniyoshi, Y.: Partial matching of real textured 3D objects using color cubic higher-order local auto-correlation features. Visual Comput. **26**(10), 1269–1281 (2010)

29. Kazhdan, M., Funkhouser, T., Rusinkiewicz, S.: Rotation invariant spherical harmonic representation of 3D shape descriptors. In: Proceedings of the First Eurographics/ACM SIGGRAPH Symposium on Geometry Processing, SGP '03, pp. 156–164. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland (2003)

30. Kimmel, R., Malladi, R., Sochen, N.: Images as embedded maps and minimal surfaces: Movies, color, texture, and volumetric medical images. Int. J. Comput. Vision **39**(2), 111–129 (2000)

31. Kovnatsky, A., Bronstein, M.M., Bronstein, A.M., Kimmel, R.: Photometric heat kernel signatures. In: Proceedings of the Third International Conference on Scale Space and Variational Methods in Computer Vision, SSVM'11, pp. 616–627. Springer-Verlag, Berlin, Heidelberg (2012)

32. Kovnatsky, A., Bronstein, M.M., Bronstein, A.M., Raviv, D., Kimmel, R.: Affine-Invariant Photometric Heat Kernel Signatures. In: Eurographics Workshop on 3D Object Retrieval (3DOR), pp. 39–46 (2012)

33. Kovnatsky, A., Raviv, D., Bronstein, M., Bronstein, A.M., Kimmel, R.: Geometric and Photometric Data Fusion in Non-Rigid Shape Analysis. Numer. Math. Theor. Meth. Appl. **6**(1), 199–222 (2013)

34. Kuhn, H.W.: The hungarian method for the assignment problem. Nav. Res. Logist. Q. **2**, 83–97 (1955)

35. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Computer Vision and Pattern Recognition (CVPR), 2006 IEEE Conference on, vol. 2, pp. 2169–2178 (2006)

36. Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., Koller, D., Pereira, L., Ginzton, M., Anderson, S., Davis, J., Ginsberg, J., Shade, J., Fulk, D.: The digital michelangelo project: 3D scanning of large statues. In: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '00, pp. 131–144. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA (2000)

37. Li, C.: Spectral geometric methods for deformable 3D shape retrieval. Master's thesis, Concordia University (2013)

38. Li, C., Ben Hamza, A.: Symmetry discovery and retrieval of nonrigid 3D shapes using geodesic skeleton paths. Multimed. Tools Appl. **72**(2), 1027–1047 (2014)

39. Li, C., Ben Hamza, A.: Intrinsic spatial pyramid matching for deformable 3D shape retrieval. Int. J. Multimed. Inf. Retr. **2**(4), 261–271 (2013)

40. Li, C., Ben Hamza, A.: A multiresolution descriptor for deformable 3D shape retrieval. Visual Comput. pp. 1–12 (2013)

41. Li, C., Ben Hamza, A.: Spatially aggregating spectral descriptors for nonrigid 3D shape retrieval: a comparative survey. Multimedia Syst. pp. 253–281 (2014)

42. Li, C., Ovsjanikov, M., Chazal, F.: Persistence-based structural recognition. In: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, pp. 1995–2002 (2014)

43. Lian, Z., Godil, A., Bustos, B., Daoudi, M., Hermans, J., Kawamura, S., Kurita, Y., Lavoué, G., Nguyen, H.V., Ohbuchi, R., Ohkita, Y., Ohishi, Y., Porikli, F., Reuter, M., Sipiran, I., Smeets, D., Suetens, P., Tabia, H., Vandermeulen, D.: SHREC '11 track: Shape retrieval on nonrigid 3D watertight meshes. In: Eurographics Workshop on 3D Object Retrieval (3DOR), pp. 79–88 (2011)

44. Lian, Z., Godil, A., Bustos, B., Daoudi, M., Hermans, J., Kawamura, S., Kurita, Y., Lavoué, G., Van Nguyen, H., Ohbuchi, R., Ohkita, Y., Ohishi, Y., Porikli, F., Reuter, M., Sipiran, I., Smeets, D., Suetens, P., Tabia, H., Vandermeulen, D.: A comparison of methods for non-rigid 3D shape retrieval. Pattern Recogn. **46**(1), 449–461 (2013)

45. Ling, H., Jacobs, D.: Deformation invariant image matching. In: Computer Vision (ICCV), 2005 IEEE International Conference on, vol. 2, pp. 1466–1473 (2005)

46. Liu, Y.J., Zheng, Y.F., Lv, L., Xuan, Y.M., Fu, X.L.: 3D model retrieval based on color + geometry signatures. Visual Comput. **28**(1), 75–86 (2012)

47. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision **60**(2), 91–110 (2004)

48. Loy, G., Zelinsky, A.: Fast radial symmetry for detecting points of interest. IEEE T. Pattern Anal. **25**(8), 959–973 (2003)

49. Matthews, B.: Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim. Biophys. Acta **405**(2), 442 – 451 (1975)

50. Mika, S., Schölkopf, B., Smola, A., Müller, K.R., Scholz, M., Rätsch, G.: Kernel pca and de-noising in feature spaces. In: Advances in Neural Information Processing Systems II, pp. 536–542. MIT Press (1999)

51. Mitra, N.J., Pauly, M., Wand, M., Ceylan, D.: Symmetry in 3D geometry: Extraction and applications. Comput. Graph. Forum **32**(6), 1–23 (2013)

52. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern Recogn. **29**(1), 51–59 (1996)

53. Osada, R., Funkhouser, T., Chazelle, B., Dobkin, D.: Shape distributions. ACM T. Graphics **21**(4), 807–832 (2002)

54. Pasqualotto, G., Zanuttigh, P., Cortelazzo, G.M.: Combining color and shape descriptors for 3D model retrieval. Signal Process-Image **28**(6), 608 – 623 (2013)

55. Pavan, M., Pelillo, M.: Dominant sets and pairwise clustering. IEEE T. Pattern Anal. **29**(1), 167 –172 (2007)

56. Pele, O., Werman, M.: Fast and robust Earth Mover's Distances. In: Computer Vision (ICCV), 2009 IEEE International Conference on, pp. 460–467 (2009)

57. Phong, B.T.: Illumination for computer generated pictures. Commun. ACM **18**(6), 311–317 (1975)

58. Rabin, J., Peyré, G., Cohen, L.D.: Geodesic shape retrieval via optimal mass transport. In: Proceedings of the 11th European Conference on Computer Vision: Part V, ECCV'10, pp. 771–784. Springer-Verlag, Berlin, Heidelberg (2010)

59. Raviv, D., Bronstein, A., Bronstein, M., Waisman, D., Sochen, N., Kimmel, R.: Equi-affine invariant geometry for shape analysis. J. Math. Imaging Vis. **50**(1-2), 144–163 (2014)

60. Raviv, D., Bronstein, A.M., Bronstein, M.M., Kimmel, R., Sochen, N.: Affine-invariant geodesic geometry of deformable 3D shapes. Comput. Graph. **35**(3), 692 – 697 (2011)

61. Reeb, G.: Sur les points singuliers d'une forme de Pfaff complètement intégrable ou d'une fonction numérique. Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences **222**, 847–849 (1946)

62. Reuter, M., Wolter, F.E., Peinecke, N.: Laplace-Beltrami spectra as "Shape-DNA" of surfaces and solids. Comput. Aided Design **38**(4), 342–366 (2006)

63. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover's distance as a metric for image retrieval. Int. J. Comput. Vision **40**(2), 99–121 (2000)

64. Ruiz, C., Cabredo, R., Monteverde, L., Huang, Z.: Combining Shape and Color for Retrieval of 3D Models. In: INC, IMS and IDC (NCM'09). Fifth International Joint Conference on, pp. 1295–1300 (2009)

65. Rustamov, R.M.: Laplace-beltrami eigenfunctions for deformation invariant shape representation. In: Proceedings of the Fifth Eurographics Symposium on Geometry Processing, SGP '07, pp. 225–233. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland (2007)

66. Seitz, S., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: Computer Vision and Pattern Recognition (CVPR), 2006 IEEE Conference on, vol. 1, pp. 519–528 (2006)

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

68. Smeets, D., Fabry, T., Hermans, J., Vandermeulen, D., Suetens, P.: Isometric deformation modelling for object recognition. In: Computer Analysis of Images and Patterns, *Lecture Notes in Computer Science*, vol. 5702, pp. 757–765. Springer Berlin Heidelberg (2009)

69. Starck, J., Hilton, A.: Correspondence labelling for wide-timeframe free-form surface matching. In: Computer Vision (ICCV), 2007 IEEE International Conference on, pp. 1–8 (2007)

70. Sun, J., Ovsjanikov, M., Guibas, L.: A concise and provably informative multi-scale signature based on heat diffusion. Comput. Graph. Forum **28**(5), 1383–1392 (2009)

71. Suzuki, M.: A Web-based retrieval system for 3D polygonal models. In: IFSA World Congress and 20th NAFIPS International Conference. Joint 9th, vol. 4, pp. 2271–2276 (2001)

72. Tanaka, J., Weiskopf, D., Williams, P.: The role of color in high-level vision. Trends Cogn. Sci. **5**, 211–215 (2001)

73. Tangelder, J., Veltkamp, R.: A survey of content-based 3D shape retrieval methods. Multimed. Tools Appl. **39**(3), 441–471 (2008)

74. Tatsuma, A., Aono, M.: Multi-fourier spectra descriptor and augmentation with spectral clustering for 3D shape retrieval. Visual Comput. **25**(8), 785–804 (2008)

75. Tombari, F., Salti, S., Di Stefano, L.: A combined texture-shape descriptor for enhanced 3D feature matching. In: Image Processing (ICIP), 2011 IEEE International Conference on, pp. 809–812 (2011)

76. Veltkamp, R., Ruijsenaars, R., Spagnuolo, M., van Zwol, R., ter Haar, F.: Shrec2006: 3D Shape Retrieval Contest . Tech. Rep. CS-2006-030, UU (2006)

77. Wang, C., Bronstein, M.M., Bronstein, A.M., Paragios, N.: Discrete minimum distortion correspondence problems for non-rigid shape matching. In: Proceedings of the Third International Conference on Scale Space and Variational Methods in Computer Vision, SSVM'11, pp. 580–591. Springer-Verlag, Berlin, Heidelberg (2012)

78. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pp. 3360–3367 (2010)

79. Wu, C., Clipp, B., Li, X., Frahm, J.M., Pollefeys, M.: 3D model matching with Viewpoint-Invariant Patches (VIP). In: Computer Vision and Pattern Recognition (CVPR), 2008 IEEE Conference on, pp. 1–8 (2008)

80. Zaharescu, A., Boyer, E., Horaud, R.: Keypoints and local descriptors of scalar functions on 2D manifolds. Int. J. Comput. Vision **100**(1), 78–98 (2012)

Table 1 Methods classification according to the taxonomy provided in Section 4.7. *Geometric description:* each run is characterized by type of invariance (type of information); scale; type of descriptor. *Texture description:* each run is characterized by scale; type of descriptor. *Colour space:* the space used for colour representation. *Final descriptor:* either hybrid or combined; "hybrid/combined" means that a hybrid contribution is combined *a posteriori* with other shape descriptions.

Run	geometric description	texture description	colour space	final descriptor
BCGS1	mostly intrinsic (mixed); global; 2D histogram;	multi-scale; topological;	CIELab;	hybrid/combined;
BCGS2	mostly intrinsic (mixed); global; 2D histogram;	multi-scale, global; topological, histogram;	CIELab;	hybrid/combined;
BCGS3	mostly intrinsic (mixed); global; 2D histogram;	multi-scale, global; topological, histogram	CIELab;	hybrid/combined;
GG1	extrinsic (symmetry); multi-scale; concatenated histogram;	global; histogram;	normalized RGB;	combined;
GG2	extrinsic (symmetry); multi-scale; concatenated histogram;	global; histogram;	averaged RGB;	combined;
GG3	extrinsic (symmetry); multi-scale; concatenated histogram;	multi-scale; histogram;	averaged RGB;	hybrid;
Gi1	intrinsic (geodesic); global; histogram;	global; histogram;	CIELab;	hybrid;
Gi2	intrinsic (geodesic); global; 2D histograms;	global; histogram;	CIELab;	hybrid/combined;
Gi3	extrinsic (spherical harmonics); global; histogram;	global; topological;	CIELab;	hybrid/combined;
LBG1	intrinsic (spectral); multi-scale; concatenated histogram;	− −	− −	− −
LBG2	intrinsic (spectral); multi-scale; concatenated histogram;	− −	− −	− −
LBG3	intrinsic (spectral); multi-scale; concatenated histogram;	multi-scale; histogram;	RGB;	combined;
TAS	extrinsic (image-based); multi-scale; concatenated histogram;	multi-scale; histogram;	CIELab;	combined;
Ve1	intrinsic (geodesic); global; 2D histogram;	global; histogram;	RGB;	combined;
Ve2	intrinsic (geodesic); global; 2D histogram;	global; histogram;	RGB;	combined;
Ve3	intrinsic (geodesic); global; 2D histogram;	global; histogram;	RGB;	combined;

Table 3 Highly relevant analysis of the five families of affine texture deformations used to build the benchmark.

No transformations (only null models)

Run	mAP	> 0.40 (%)	> 0.55 (%)	> 0.70 (%)	> 0.85 (%)
BCGS3	72.41	100.00	97.92	64.58	6.25
GG1	71.91	100.00	91.67	66.67	6.25
Gi3	62.45	100.00	75.00	27.08	0.00
LBG3	61.84	100.00	66.67	25.00	6.25
TAS	60.30	100.00	66.67	18.75	2.08
Ve1	64.46	100.00	72.92	39.58	0.00

Table 3.0 Highly relevant analysis of the mAP score associated with null models. The best two results are in gold and silver text, respectively.

Tran. ♯1

Run	mAP	> 0.40 (%)	> 0.55 (%)	> 0.70 (%)	> 0.85 (%)
BCGS3	0.6340	97.92	72.92	31.25	4.17
GG1	0.6057	95.83	72.92	20.83	2.08
Gi3	0.5162	93.75	33.33	10.42	0.00
LBG3	0.5030	85.42	22.92	10.42	0.00
TAS	0.4713	77.08	22.92	4.17	0.00
Ve1	0.5104	87.50	31.25	8.33	0.00

Table 3.1 Highly relevant analysis of the mAP score associated with affine texture deformation ♯1. A visual interpretation of this kind of transformation, and of the considered strength levels, is given by the pictures above.

Tran. ♯2

Run	mAP	> 0.40 (%)	> 0.55 (%)	> 0.70 (%)	> 0.85 (%)
BCGS3	0.6506	81.25	77.08	62.50	14.58
GG1	0.6261	81.25	70.83	52.08	18.75
Gi3	0.5609	79.17	58.33	37.50	10.42
LBG3	0.5611	81.25	56.25	35.42	12.50
TAS	0.5564	77.08	58.33	39.58	8.33
Ve1	0.5766	81.25	62.50	47.92	6.25

Table 3.2 Highly relevant analysis of the mAP score associated with affine texture deformation ♯2. A visual interpretation of this kind of transformation, and of the considered strength levels, is given by the pictures above.

Tran. ♯3

Run	mAP	> 0.40 (%)	> 0.55 (%)	> 0.70 (%)	> 0.85 (%)
BCGS3	0.6234	95.83	68.75	37.50	4.17
GG1	0.6072	95.83	64.58	25.00	4.17
Gi3	0.5227	93.75	41.67	6.25	2.08
LBG3	0.5172	87.50	39.58	8.33	0.00
TAS	0.4758	83.33	25.00	4.17	0.00
Ve1	0.5312	85.42	45.83	14.58	0.00

Table 3.3 Highly relevant analysis of the mAP score associated with affine texture deformations ♯3. A visual interpretation of this kind of transformation, and of the considered strength levels, is given by the pictures above.

Tran. ♯4

Run	mAP	> 0.40 (%)	> 0.55 (%)	> 0.70 (%)	> 0.85 (%)
BCGS3	0.6401	97.92	72.92	41.67	4.17
GG1	0.6247	97.92	66.67	31.25	8.33
Gi3	0.5498	93.75	45.83	10.42	0.00
LBG3	0.5197	87.50	33.33	10.42	0.00
TAS	0.4974	79.17	35.42	6.25	0.00
Ve1	0.5324	87.50	43.75	12.50	2.08

Table 3.4 Highly relevant analysis of the mAP score associated with affine texture deformations ♯4. A visual interpretation of this kind of transformation, and of the considered strength levels, is given by the pictures above.

Tran. ♯5

Run	mAP	> 0.40 (%)	> 0.55 (%)	> 0.70 (%)	> 0.85 (%)
BCGS3	0.6262	95.83	64.58	33.33	12.50
GG1	0.6021	97.92	62.50	35.42	4.17
Gi3	0.5442	87.50	50.00	16.67	2.08
LBG3	0.4964	81.25	25.00	10.42	4.17
TAS	0.5003	81.25	31.25	14.58	0.00
Ve1	0.5264	87.50	45.83	12.50	2.08

Table 3.5 Highly relevant analysis of the mAP score associated with affine texture deformations ♯5. A visual interpretation of this kind of transformation, and of the considered strength levels, is given by the pictures above.