A Comparative Survey on 3D Models Retrieval Methods

Sarah Ali, Trung-Thien Tran, Denis Laurendeau

Laval University, Quebec (QC), Canada

Correspondence: Denis Laurendeau, denis.laurendeau@gel.ulaval.ca

Abstract– In computer vision many studies have been conducted in order to perform the matching and comparison of 3D models of objects. The main goal of matching is to group the models into different categories according to their similarity in order to allow their retrieval for recognition purposes and for further usage. So, in most of the cases, the comparison is run on a large dataset containing various models whether they belong to the same type of object or not and generally having similar or different shapes and poses. The objects' nature and characteristics are important factors to be taken into consideration before performing the comparison step. We distinguish between two main categories of objects: rigid objects and deformable objects whose treatment and handling differ in the modeling as well as in the comparison phases. In this paper, we will be focusing on the comparison of deformable objects, and thus dealing with objects whose shapes might vary in different instances. For this purpose two main approaches used in the retrieval of 3D deformable models will be reviewed and implemented: the spectral approach and the bag-of-features approach. The deformation or variation in shape involves different aspects depending on the type of object. It could be a change in the posture of an articulated or bendable model, or it could result from a variation (loss or gain) in the total mass leading to a change in the surface and thus in the shape of the object. Even more complex situations occur when both cases are combined together.

Keywords– shape representation, shape analysis, global shape retrieval, partial matching, features extraction, bag of features, spectral embedding.

1 Introduction

In general, a comparison of 3D models can be based on their topology, their geometry or both. A comparison based on topology, which refers to "the anatomical structure of a specific area or body part", gives information about the skeletal structure of the model. Topological features can be used in the retrieval of objects that represent the same models but in different positions. Methods based on topology are generally computationally costly [1]. However, methods based on geometry are used more frequently since they usually rely on descriptors and features that allow a high discrimination between shapes.

Comparison methods which use geometric features are classified into three major categories according to the type of shape feature [2]: (1) global features, (2) histograms of local features, and (3) spatial maps. Global features describe the shape with moments, aspect ratio, or volume-to-surface ratio. Histograms of local shape features consist of bins, each bin storing the probability of occurrence of a feature. Generally, histograms are invariant to rotation, reflection, and uniform scaling of objects; they are used to represent various features such as angles, distances, areas, volumes and curvatures. Examples of histograms such as spin images [3] and shape contexts [4], which represent the relative positions of the data points, have also been used for matching and recognition tasks. Finally, spatial maps represent the spatial information of an object's features meaning that they usually store information about the location of

the features in an object. For example, distance maps and surface penetration approaches introduced by [2] are spatial maps able to describe the geometry of the object's shape, the topology and concavity of the entire object.

Most of the comparison methods try to meet several invariance criteria such as invariance to: (1) similarity transformations, (2) shape representations, (3) geometrical and topological noise, and (4) articulation or global deformation. Such methods might also use shape features that are invariant to geometrical transformations, pose normalization or a combination of both to achieve their goals. In this work, a survey on the main approaches used for retrieving 3D models based on geometric and topological features is presented in addition to experimental comparison for some of these methods to compare global 3D models and partially match them as well.

After giving a general definition of the main types of comparison methods and their basic requirements, this paper is organized as follows: retrieval methods using global descriptors are reviewed, and the spectral approach is discussed as one relevant methodology for retrieving deformable 3D objects in Section 2. Then a review on important local features used in the retrieval of 3D models is provided, and the bag-of-features approaches that use these local features in global and partial matching is presented in Section 3. Retrieval based on spatial maps is introduced in Section 4. Examples of topological features combined with geometric features are discussed in Section 5. Some experimental results
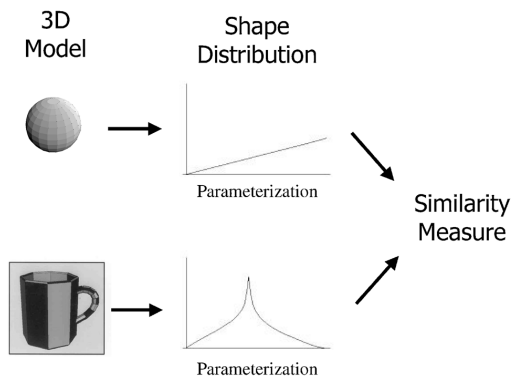
Figure 1. Shape distributions facilitate shape matching because they represent 3D models as functions with a common parameterization (taken from [6]).

using the spectral and the bag-of-features approaches in comparing deformable objects of the same time are given in Section 6. Finally a conclusion is provided in Section 7.

2 Global Features

Over the last few years, the number of methods that involve comparing and retrieving 3D models has increased significantly. These methods mainly use geometric descriptors extracted from 3D shape representations (e.g., 3D polygonal meshes and point clouds), represent and compress these features in a certain manner to enable their comparison, and then to compare the different models in large databases.

Methods that use global descriptors for retrieving 3D models have gained considerable interest over the few last years. The most popular methods in the retrieval of 3D models involve the use of the following global descriptors: the D1 descriptor [5], the D2 descriptor [6], the spherical harmonic descriptor SHD [7], the 3D wavelet descriptor [8], the skeleton descriptor [9], the Reeb graph descriptor [10], the depth buffer images, silhouettes, and ray-extents DESIRE [11], etc. For example, the D2 shape distribution of [6] represents the distribution of Euclidian distances between pairs of randomly sampled points. Therefore, 3D models are transformed into parameterized functions which facilitate their comparison using an appropriate similarity measure (see Figure 1). The main advantage of this approach is its simplicity because the shape matching problem is reduced to the following tasks: random sampling, normalization of models, and comparison of their probability distributions. This is to be compared to methods that require the reconstruction of manifold surfaces from degenerate 3D data, the registration of pose transformations, the matching of features and the fitting of high-level models. Despite the fact that the D2 shape distribution enabled good results in retrieving rigid objects, it fails sometimes in comparing complex shapes because some features might be missing in the random sampling.

Another interesting example of a global shape de-

scriptor is spectral embedding. The spectral approach derived from the graph theory of matching can be applied in 3D geometry for purposes such as finding point correspondence, segmenting and compressing 3D meshes and retrieving articulated 3D models [12]. The global geometric descriptors used in the spectral approach are invariant to shape articulation and bending where a spectral embedding representation of the 3D shape is proposed by only considering the first k normalised eigenvectors which are scaled by the square root of their appropriate eigenvalues. These eigenvalues can be used as effective descriptors which enable the retrieval of articulated 3D models of the McGill University benchmark database (MSB) [13].

Spectral embedding

In the spectral approach, the input models are given by 3D meshes containing hundreds of thousands of faces [12]. In order to facilitate feature extraction, the resolution of the meshes is reduced using mesh simplification techniques. Next the affinity matrix A of the graph $GM = (V, EM)$ of a specific mesh M having V as the set of vertices and EM as their connected edges is constructed. The affinity matrix is of size $n \times n$, where n is the number of vertices of the mesh M. In the affinity matrix A the entry d_{ij} corresponds to the geodesic distance between the ith and jth vertices of the mesh. The geodesic distances give the smallest curve distances between any 2 points of the 3D mesh model. Unlike the Euclidian distances, the geodesic distances are invariant to certain shape bending. Therefore, their use is shown to be relevant especially when retrieving deformable 3D models that undergo shape articulation and bending. After computing the affinity matrix of the geodesic distances, it is normalized using a Gaussian distribution, and it is given by the Gaussian equation

$$A_{ij} = \exp(-d_{ij}^2/2\sigma^2), \qquad (1)$$

where σ is defined by the Gaussian width and it is set $\sigma = \max(i, j)\{d_{ij}\}$. In the next step, the spectral embedding (i.e. the eigenvalues and eigenvectors) are computed. It is sufficient to consider only the first k largest eigenvalues and their corresponding eigenvectors (generally the eigenvectors are scaled by the square root of their eigenvectors). Generally the first eigenvector is constant, so it can be excluded and only the $(k-1)$ embedding derived from $v_2, \ldots, v_k$ are considered. Finally two corresponding dissimilarity measures that combine the extracted spectral embedding can be used to compare the 3D models and achieve the retrieval of articulated shapes. The first dissimilarity measure is the eigenvalue descriptor EVD. In fact, the eigenvalues are the indicators of the shape variation along the axis which are given by the corresponding eigenvalues. Let P and Q be two meshes, with their respective eigenvalues λ_i^P and $\lambda_i^Q, i = 1, \ldots, 20$, the EDV is given by the χ^2-distance as a measure of dissimilarity according to

$$Dist_{EVD} = 1/2 \sum_{i=1}^{20} \frac{[|\lambda_i^P|^{1/2} - |\lambda_i^Q|^{1/2}]^2}{|\lambda_i^P|^{1/2} + |\lambda_i^Q|^{1/2}}. \qquad (2)$$

Figure 2. Articulated shapes from the McGill database (top row) with their respective spectral embedding (bottom row).

The second dissimilarity measure is the cost correspondence descriptor measure and is given by

$$Dist_{CCD}(P,Q) = \sum_{p \in P} \| V_P(p) - V_Q(\texttt{match}(p)) \|, \quad (3)$$

where $V_P(p)$ and $V_Q(q)$ are the pth and qth rows of V_P and V_Q, respectively, p represents a vertex of P, and $\texttt{match}(p)$ is the vertex in Q corresponding to p based on the Euclidian distance in the embedding space [12]. Figure 2 shows the 3D spectral embedding of some articulated shapes taken from the McGill database.

Although the spectral method enabled high performances in retrieving deformable objects in general and articulated shapes in particular, this method fails in computing large graphs of models since the computation of geodesic distances is time consuming, and software programs that perform such computations often run out of memory. For this reason, the resolution of 3D meshes should be reduced by compressing the number of faces and point clouds (mesh simplification). In addition to the problem of large graphs, computing eigenvalues, reordering them, and scaling the eigenvectors is expensive in time especially if we have to deal with large databases including a variety of shapes and deformations.

3 LOCAL DESCRIPTORS

It was shown in most cases, that global descriptors when used alone, can limit the retrieval performance. However, when local descriptors (e.g., 3D spin image [3], harmonic shape context [14], 2.5D SIFT [15], etc.) are used, this limitation is overcome since they are proven to be very successful in many applications. In addition, the use of local descriptors in 3D shape retrieval is promising because of the fact that local features have intrinsic properties in solving problems of partial shape retrieval and articulated shape retrieval as well [16]. Section 3.1 discusses curvature maps as local descriptors and an advanced version of surface curvature used in the comparison of 3D surfaces. Section 3.2 introduces important examples of local shape descriptors that are stored and represented in histograms; this is followed by a review on the bag-of-features approaches using these local descriptors in Section 3.3.

3.1 Curvature Maps

Many studies using surface curvatures have been conducted in order to compare 3D shapes and to compute local surface similarity as well (local surface similarity is used to decide if a region of a given surface has a similar shape to another region). Curvature is an intrinsic property of discrete surfaces which might be segmented into regions based on their curvature features. However, using such concepts is in general not easy since there are many difficulties and challenges to overcome, when the goal is to compare 3D models because surface curvatures are usually sensitive to noise and mesh resolution, and they cannot describe information about a local region at a given vertex. However, curvature maps which are combinations of functions containing surface curvatures are able to describe the local region at a given vertex. Thus they can be used in the matching and comparison of similar surfaces. In Gatzke et al., [17] the authors incorporate curvature maps to obtain information around a given point by performing a sampling of the vertices on the surface.The sampling is done by either defining rings on a surface mesh or by using geodesic fans. A curvature map around a given vertex is then generated and used. The curvature map at a given vertex describes the shape information of the local region around that vertex. It is a set of piecewise linear functions applied to either the mean or the Gaussian curvature. The curvature map can be a one dimensional (1D) map which only describes the distance between vertices, or a two-dimensional (2D) map that consists of both the distance and the orientation of the normal to the surface. The (0D) curvature map is simply the surface curvature at a given vertex.

The authors in [17] propose different functions of curvature maps exploiting both mean and Gaussian curvatures. They compare between two curvature maps by using a local shape similarity operator. For instance, they applied the square root and logarithmic functions to the average Gaussian curvature, and the logarithmic function to the average of the mean curvature. Their experiments showed that the average mean curvature combined with the square root of the average Gaussian curvature led to the best discriminatory results between local shapes. The authors concluded that ring-based methods are more suitable for large regions while the fan-based 1D methods are convenient for comparing small local regions. In addition, the comparison method relying on the 0D curvature is noisy while the 1D ring-based and fan-based methods are much more capable of identifying regions of surfaces that are similar. However, the computationally complex but more exact 2D method can be applied for the same purposes. So it is preferable to use the ring-based 1D method in comparing large regions and then apply the slower 2D method for the final stage if the exact matching is needed.

3.2 Histogram of Local Features

In his work [3], Johnson created a new surface representation method which was used in surface matching and 3D object recognition. This surface representation, which is stored into a histogram, is called a spin image. Such a spin image describes and encodes all the proper-

ties of the 3D object's surface in an 'object-centered' system and not in a 'viewer-centered' system. An 'object-oriented' system is a stable system that belongs to the surface or the object, while a 'viewer-centered' system represents coordinates in a system that depends on the observer's viewpoint. Spin images describe the relative position of points on a rigid object with respect to a set of points that belong to the same object, and they are independent of rigid transformations applied to the object. As a result, they can describe the shape of an object independently of the different poses taken by that object. Thus spin images are invariant to rigid transformations applied to an object, and they are truly object centered shape descriptions [3].

Scale Invariant Feature Transform (SIFT), introduced by Lowe [15], is another example of a local image descriptor represented by a histogram. The SIFT descriptor is used for many applications in computer vision, such as in point matching between different views of a 2D or 3D scene and also in object recognition. It is invariant to translations, rotations and scaling transformations, and it is robust to certain kinds of transformations and illumination variations. Initially the SIFT descriptor includes a method for matching interest points of the images in gray levels where information on local gradient orientations of image intensities are stored into a histogram in order to describe the local region around each interest point.

The spin image local descriptor introduced by Johnson [3] is robust and simple to use in matching applications. However, it is not scale invariant since a spin image is computed at each vertex using a constant predefined support. The support is the vertical or the horizontal support range of the spin image. When the support is constant, it means that all of the oriented points (which are surface point associated to a given direction of a normal and a position) have the same vertical and horizontal support range, and the computation expands over most of the mesh, which imposes to uniform or global scale for the 3D model. However, the scale invariant spin images (SISI) mesh descriptor proposed in [18] is an improved version of the spin image descriptor that can be computed over a local scale and can be directly extracted from a 3D mesh. Figure 3 shows a comparison between the SISI and spin image descriptors computed using a constant support based on the object's size or scale. Thus the SISI descriptors can be used as local features in the retrieval of 3D models with different scales.

In addition to SISI descriptors, the authors in [18] proposed an extension of Lowe's SIFT descriptor that can be extracted directly on 3D meshes, called the LD-SIFT descriptor which is also scale invariant. The LD-SIFT uses the difference of Gaussian (DOG) operators in order to detect the interest points and estimate the local scale. The DOG operator is defined as a Gaussian filter on the mesh geometry which enables the computation of a set of filtered meshes, represented by the mesh octaves. The consecutive octaves are subtracted to form the DOG operator, where the local maxima (in location and scale) represent the feature points. Similar to the
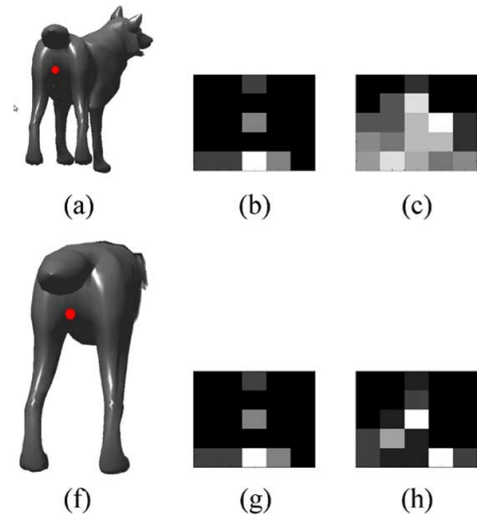

Figure 3. SISI descriptors compared with the spin image descriptors with a constant support on the model's size. The SISI descriptors in (b) and (g) computed with respect to the marked point in (a) and (b) are similar. However, the spin images with constant support in (c) and (h) computed at the same marked points are not similar (taken from [18]).

SISI descriptor, the LD-SIFT descriptor can also be used to achieve the retrieval of 3D models.

3.3 Bag-Of-Features

Among the methods that use local geometric features in retrieving 3D shapes and represent them into histograms, the bag-of-features (bof) approaches demonstrate excellent retrieval performance for both articulated and rigid objects. Most of these methods usually extract local descriptors (SIFT descriptors, spin images, SISI, LD-SIFT and others) from 3D models and represent them in a probabilistic and statistical approach using unsupervised learning and clustering techniques (for instance the simplest clustering technique of the features that we can mention is the k-means), in order to compress and group the features into a dictionary of visual words. The 'visual word' is a small patch on the local feature (array of pixels resulting from the feature extraction), which can carry any kind of interesting information in any feature space (color changes, texture changes, etc). In the next steps, these 'visual words' are counted according to the frequency of their occurrence using the histogram of words. This histogram, which is a discrete probability distribution vector, enables a unique representation of each of the models and becomes the feature vector of the 3D shape. Finally, in order to compare and retrieve 3D models, their histograms are compared using a convenient distance measure (Kullback-Leibler divergence, $L2$ norm, cosine distance norm etc.) which gives the percentage of dissimilarity between the models under comparison.

The Bag-Of-Features proposed in [20] is used for both global comparison and partial matching. It relies on the extraction of spin image signatures which are later grouped in clusters (using k-means). Each cluster is a "word distribution" and has its own label or
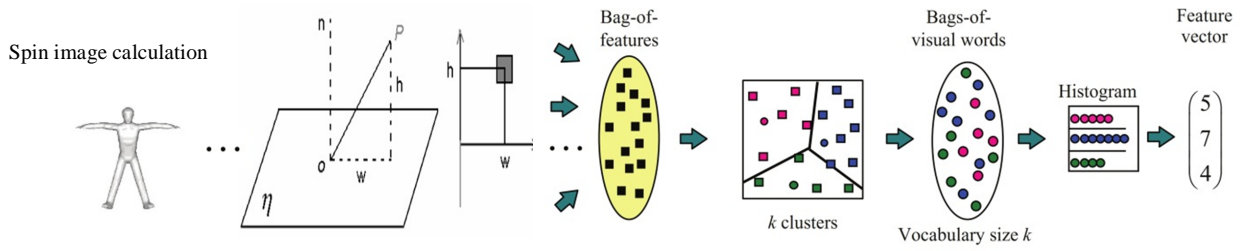
Figure 4.   Bag-Of-Features for spin image descriptors (part of this figure was taken from [19]).

code. By counting the label frequencies of the words, a histogram or feature vector representation of the model is built. This step is called vector quantization, and it is a part of the Bag-of-Features process. The histograms of the features are then compared using the Kullback-Leibler divergence. The partial matching by Liu et al. [20] (see Figure 4) can be performed without aligning the models, which is an important contribution of this approach. In addition, the authors accelerated the partial matching retrieval by introducing a small set of probability distributions known as "shape topics". In this process, each local feature has a probability to belong to a certain class or "shape topic". It is then classified according to that class.

Ohbuchi et al. [19] propose a method for retrieving rigid models using the PSB benchmark [21] and articulated models as well using the MSB benchmark [13]. The method first renders the range images by generating an orthographic projection of the 3D model at multiple view directions after achieving pose normalization. Usually pose normalization occurs after scaling by finding the smallest polyhedron that encloses the model, such that its centroid coincides with the origin of the global coordinate system. The viewpoints are captured at vertices placed at equal distance on the polyhedron circumscribing the model. Next local features from each range image of the different views are extracted using the Scale Invariant Feature Transform (SIFT) algorithm proposed by Lowe [15]. Thus the 3D model becomes associated with thousands of local features, which are later vector quantized into visual words using a visual codebook which is learned by using k-means clustering. Next the frequencies of visual words are counted and stored into a histogram with N_v bin (where $N_v = k = 1500$). The histograms of the models are also compared using the Kullback-Leibler divergence dissimilarity. The results of the experiments showed 75% R-precision for the articulated shapes of the MSB and 45% R-Precision for the rigid shapes of the PSB. Figure 5 illustrates the bag-of-features approach using the local SIFT features.

This work was later extended and improved [22] by extracting significantly more local visual features via a dense random sampling of each depth image and using a decision tree to encode these features into visual words. In the random and dense sampling of SIFT features in range images, samples are concentrated on or near the 3D object, and not on the background. Another improvement was on the encoding of local

features using the Extremely Randomized Clustering Trees (ERC-trees), which iteratively divides the space of features into two parts using the tree nodes. Each subdivision is done, first by choosing a dimension (or axis) and a point (a scalar value) on the axis at which a separating hyperplane is placed. The subdivision of the feature space continues until the number of data points per subspace is below a given parameter that does not change the number of words in the vocabulary of the codebook N_v. These two improvements accelerated the SIFT-BOF algorithm and increased the retrieval performance from 45.1% to 55.8% for the PSB benchmark, while the retrieval performance on the MSB was unchanged.

In addition, Ohbuchi et al. [23] proposed a more advanced algorithm that uses unsupervised distance metric learning with a combination of appearance-based features. They employed a set of local visual features combined with a set of global features. The local visual features are SIFT features computed using salient dense points while the global visual features are also SIFT features sampled only at the center of each (2D) range image. Then an unsupervised distance metric learning based on the Data-Adaptive Distance via Manifold Ranking is used in order to compute the distances between these features. With the Data-Adaptive Distance via Manifold Ranking, the performance of the dense sampling 3D model retrieval using a combination of local dense SIFT features and global SIFT features, has significantly increased from 75% to 91% for the MSB. Furthermore, this algorithm achieved the highest scores in SHREC 2012 – Shape Retrieval Contest based on Generic 3D Dataset.

Daromand Keller, [18] tested their proposed local features (SISI and LD-SIFT) for the retrieval of 3D models using the TOSCA database [24] by adopting the Bag-of-Feature approach. The authors extracted feature points from the models, and obtained a large set of features which was used to compute a dictionary of 2500 words by using the k-means algorithm. They conducted two experiments on the retrieval of 3D models. In the first experiment, they used the models of the TOSCA dataset in their original form (in which all models are approximately of the same scale). In the second experiment, they applied a scaling transform from 0.25 to 4 for each model. In both experiments, they tested the SISI, LD-SIFT, and standard spin image descriptors of Johnson [3]. The SISI and LD-SIFT descriptors achieved similar results in the two experiments, validating their
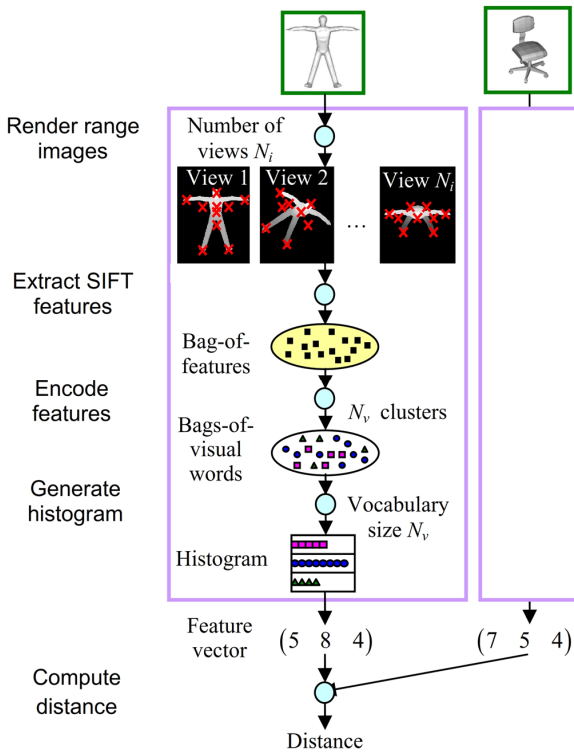
Figure 5.   The bag-of-features using local SIFT descriptors (taken from [22]).

scale invariance. The SISI and the LD-SIFT descriptors outperformed the standard spin images in the first experiment, and the SISI were also better than the LD-SIFT in the retrieval of 3D models. In the second test, the performance of the standard spin image descriptors by Johnson et al. degrades because the models do not have the same scale. However, the experiment of Darom et al. showed that the proposed features (SISI) in the second test were robust to most scale transforms, alsoshowing good results with relatively small feature support. Increasing the feature support ($C = 3$) rendered both (SISI and LD-SIFT) features more robust to noise (the support for the SIFT features is the gradient of scaling factor and for spin images, it is the vertical support αmax which is equal to the horizontal support βmax).

One main advantage of the Bag-of-Features approach is that it reduces the cost of storage of the features by which a 3D model is usually associated and represented, since rather than computing the distance or the dissimilarity between two sets of thousands of features representing two given models under comparison, which is costly in time (the complexity is of order $O(n^2)$, where n is the number of features per model), all the features of the model are rather integrated in a feature vector by using the bag-of-features. On the other hand, similar to the spectral approach, the bag-of-features is convenient in retrieving articulated models rather than rigid ones, especially when such models reveal shape complexities such as occlusions. The bag-of-features cannot provide information on features when such internal structures are hidden due to occlusions.

Furthermore, another drawback of the bag-of-features is the loss of spatial information, thus when features are accumulated and aggregated in a bag of features leading to a feature vector that represents the entire model, information about the particular sections and locations of the models might be dismissed. Therefore it is impossible to weight the features according to the deformation impact on the subparts of the model. This problem might be overcome using some feature representations, called spatial maps, which will be discussed in the next section.

4 Spatial Maps

Spatial maps are used as representations that store information about the location of the features on an object. The spatial map entries give information about the physical locations of features or sections of a particular object. They are structured such that the relative positions of the features on the object are always conserved. Since spatial maps vary according to different linear transformations, the Fourier transform is often needed to transform spatial maps into invariant descriptors [2]. Some researchers used spatial maps to describe their features. For instance, Kriegeland Seidl [25] and Suzuki et al. [26] partitioned the object into cells or surface segments, and counted the number of points within each cell to become the features used for their surface representation. Vranic et al. [27] used 2D maps of spherical harmonics and Novotni and Klein [28] used 3D maps of distances to compute and represent the features of the objects. In addition, Yu et al. presented a comparison method of 3D shapes based on the use of spatial maps and on morphing 3D shapes [2].

We now explain the method of Yu et al. [2] since it was used in the retrieval of 3D models. In this method, authors morph an object into a canonical one, such as a sphere or a cylinder, by computing spatial feature maps. They use this amount of morphing as a way to measure the similarity and compare the 3D object to the canonical object. First the scaled 3D object is embedded into a unit sphere such that the center of the sphere coincides with the centroid of the model. Next, Principal Component Analysis is applied on the 3D points of the models in order to align the objects (by aligning the major and minor axes of the object to the first and second eigen-vectors of PCA). Afterward, a ray is cast from the center of the sphere through each point of the object to the sphere's surface (Figure 6) storing the Distance Map (DM) which is the distance traveled by the ray from this point of the object to the surface of the sphere. The Distance Map is a spatial map since it records the spatial locations of the points. It measures the energy (displacement) required to morph an object into a sphere. In addition to the distance maps, a second spatial map is used, which is known as the Surface Penetration Map (SPM), in which the number of surfaces penetrated by the traveled ray is recorded. Surface Penetration also describes the topology and concavity of the object by counting the number
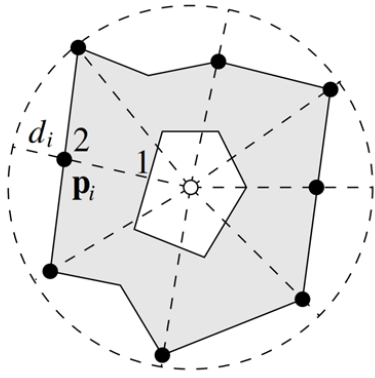
Figure 6. Computing feature maps. Rays (dashed lines) are cast from the center (white dot) of a bounding sphere (dashed circle) through the object points (black dots) to the sphere's surface. The distance di traveled by the ray from a point pi to the sphere's surface and the number of object surfaces penetrated by the ray since it leaves the sphere's center, are recorded in the feature maps (taken from [2]).

of surfaces containing holes, disconnected components, and concave parts. After computing the (SM) and (SPM) feature maps, Fast Fourier Transform (FFT) is applied on these maps in order to insure their invariance to rotation and reflection. Finally the dissimilarity between two models is computed using weighted normalized Euclidean distance of the Fourier transforms of the feature maps, given by the following equation:

$$dist(D, D') = \sum_{u,v} \max\{D(u,v), D'(u,v)\} \quad (4)$$
$$\times \ [D^*(u,v) - D'^*(u,v)]^2,$$

where $D(u,v)$ and $D'(u,v)$ are the amplitudes of the FFTs of the Distance Map features. D^* and D'^* are the normalized values of D and D'. The same dissimilarity equation can be applied on the Surface Penetration Maps by replacing $D(u,v), D'(u,v)$ with $S(u,v)$ and $S'(u,v)$, which are the amplitudes of the FFTS of the Surface Penetration Maps. In their experiments the authors only used 52 sampled objects from 34 categories. The results of their experiments showed that the retrieval precision is above 0.86 even at recall rate of 1.0.

5 COMBINING TOPOLOGICAL AND GEOMETRIC FEATURES

Topological features can also be used in the retrieval of 3D models. They give information about the skeletal structure of the model such as the connected and disconnected parts of the object and the existence of holes and their number in the model. Topological features were used by Bespalov et al. in Reeb graph representations, in which vertices of the model are resampled, short-cut edges are generated and geodesic distances are computed [10]. Later Sundar el al. [9] compared 3D models using a method based on the "skeleton" of a 3D shape which is a graph-based representation of the object. The skeletal graph describes the topological

information of the 3D object since it includes all the local shape descriptors, which are localized at each node in the graph. The computation of the skeletal graph is simple and direct. It is obtained after skinning the volumetric model (at each voxel) according to a thinness parameter. The local shape descriptors obtained after the skinning process, are stored in the form of point clouds at each node, which are needed in the matching process as they contain necessary information about the mean, radius, degrees of freedom about the joint (for topological matching), or the degree of importance of a particular joint and/or node. Moreover, each node contains topological signature vectors which are used in indexing. The joint of a skeleton corresponds to the edge in the skeletal graph. The skeleton is an effective shape descriptor because it provides local representations for parts and components which make it a flexible tool for the user since the skeleton can be automatically computed, and its resolution can be modified using the slider bar of the matcher interface and visualizer. This helps in refining a particular search query to perform the partial matching with other objects.

Combining both topological and geometric features can lead to good results in the retrieval of 3D models. The method of Tam et al. [1] that consists of two main steps: the extraction of topological points which are points that depend on the topology and the structure of a specific area or body part, and the use of geometric information to enable the computation of the similarity measure for these topological points and for models in general. First the topological points are extracted from the 3D model based on the "Level Set Diagram" (LSD) which describes the critical points (maxima, minima and saddles of a region) based on the topology change of the waterfronts [29], [30]. In the second step, the geometric information is calculated by computing the normalized sum of Geodesic Distance between the extracted topological points. Next the global surface curvature change with respect to a topological point t is computed by using a feature vector $V(t)$ of dimension n. Afterwards, the vertices of the two surfaces of models to be compared are partitioned into Pi bands, and the curvature at each partition are approximated and then normalized in order to compute the feature vector $Vi(t)$ which is equal to the sum of these normalized curvature over each partition $Pi(t)$. Finally, in order to estimate the similarity measure for two models of point sets A and B respectively, the authors construct a bipartite graph $G = (V, E)$ where $V = A \cup B$, and every edge in E has two endpoints, one point in the set A and the other point in the set B. Let M be a matching in E, the number of edges of M is denoted by $|M|$ and the cost function defined by

$$c(M) = \sum_{e \in M} c(e), \quad (5)$$

which is the "Maximum Weight Maximum Cardinality Bipartite Matching" MWMCB. Let n_1 and n_2 be the number of vertices in A and B respectively. If $n_1 \leq n_2$ the point set similarity is normalized as follows

$$\text{Sim}_{pointset}(A, B) = c(M)/n^2 \quad (6)$$

Finally the similarity measure for two models is

$$\mathrm{Sim}_{model}(A,B) = W_s \times \mathrm{Sim}_{pointset}(A_{saddle}, B_{saddle})$$
$$+(1 - W_s) \times \mathrm{Sim}_{pointset}(A_{max}, B_{max}), \quad (7)$$

where W_s is a given ratio and (A_{max}, B_{max}), (A_{saddle}, B_{saddle}) represent the sets of local maximum points and saddle points respectively of the previously extracted topological points.

In general topological features are computationally expensive. When used alone, they might not lead to the best performance in retrieving 3D models. However, by combining geometric features with topological features, not only do the resulting extracted features become invariant to rotation, non-uniform and multi-resolution scale and pose deformation, but they also depend on the model topology rather than on the tessellation which limits the number of matching points and accelerates the comparison process.

6 Experimental Results

We have experimentally tested some of the approaches mentioned in this survey on several deformable objects that we believe are relevant for our research. The retrieval of models was not performed in a large database though since our goal is to compare a model M1 of a deformable object at time $t1$ and a model M2 of the same object after it undergoes a certain deformation. The following methods were tested: the D2 shape distribution method of Osada et al. [6], the bag-of-features (bof) of Daromand Keller [18] using the SIFT and spin images descriptors, and the spectral approach of Jain and Zhang [12].The results obtained, enabled a comparison of the performance of these methods based on the type of deformable objects tested.

First simple objects such as a sphere were selected and deformed into four different ellipsoids elongated toward the x-axis (the length axis). The sphere and the four ellipsoids all have the same volume. In order to compare them, the bof using SIFT descriptors was first implemented (using the parameter $k = 4$ for the k-means and the $L2$ norm for measuring the dissimilarity), followed by the bof using spin images (using the parameter $k = 8$ for the k-means and the $L2$ norm for measuring the dissimilarity) and finally the D2 shape distribution (using the Chi-square distance to measure the dissimilarity between probability distributions of the models). Figures 7 and 8 show respectively the sphere and the different ellipsoids superposed on each other, and the curves of dissimilarity ratios using the different approaches. Table I gives the measurements of the different axes and Table II gives the results of comparison using the three approaches. We notice from the results that the more the ellipsoids are elongated toward the x-axis, the higher their dissimilarity to the sphere increases. Furthermore, we claim that the D2 shape distribution gives the best dissimilarity ratio compared to the bof-sift and bof-spin images since the curve dissimilarity ratios were continuously increasing (see Figure 8). The reason the D2 shape distribution

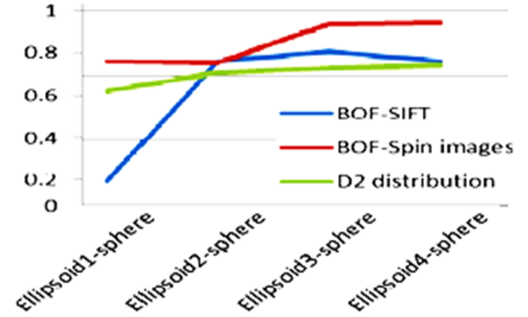Figure 7. The sphere and the elongated ellipsoids.

Figure 8. The curves of dissimilarity ratios using the three tested approaches.

performs better than the bof is that the first approach is more suitable in comparing simple 3D models, while the second approach works better with more complex 3D shapes.

In the second set of experiments we compared two models of tiger toys, which are more complex shapes than the previous ones. The two models are very similar having slight deformations in the pose. These models of tigers were downloaded from the Princeton Shape Database (http://shape.cs.princeton.edu/benchmark.). The bof-sift, the bof-spin images, the spectral approach and the D2 distribution were implemented to compare the two similar models. The results of dissimilarity ratios are 0.11, 0.15, 0.12 and 0.01 for the bof-sift ($k = 4$), the bof-spin images ($k = 8$), the spectral method (where the number of first k spectral eigenvalues is 20) and the D2 distribution respectively. This shows that on one hand the results using the first three methods are

Table I
Measurements of the Sphere and Ellipsoids

	Axis Dimensions in mm		
	Length (x-axis)	Width (y-axis)	Height (z-axis)
Sphere	4	4	4
Ellipsoid1	8	4	2
Ellipsoid2	10	3.2	2
Ellipsoid3	12	2.67	2
Ellipsoid4	16	2	2

Table II
Dissimilarity Ratios between the Sphere and the Different Ellipsoids

Dissimilarity between sphere and	Bof-SIFT	Bof-spin images	D2 shape distribution
Ellipsoid1	0.4747	0.8430	0.7539
Ellipsoid2	0.8422	0.8394	0.8098
Ellipsoid3	0.8744	0.9615	0.8222
Ellipsoid4	0.8449	0.9625	0.8813

Table III
DISSIMILARITY RATIOS OF THE DEFORMABLE MODELS OF FLOWER WITH
AND WITHOUT HOLES

	Models with holes	Models without holes
Bof-sift ($k = 64$)	0.3147	0.3612
Bof-spin images ($k = 16$)	0.3497	0.0407
CCD spectral descriptor ($k = 30$)	0.3048	0.3696
EVD spectral descriptor ($k = 3$)	0.4394	0.3387

close and coherent, and on the other hand the bof and the spectral approach were more suitable than the D2 distribution method in comparing complex models.

In the third set of experiments, a 3D model of an artificial flower was captured, and the flower was deformed. Another 3D model was captured using two different sensors, the ARTEC and the Creaform Handyscan 3D. The ARTEC is a white light scanner while the Handy scan 3D is a laser scanner. The first two models of flowers using the ARTEC were free of holes, while with the other scanner the models contained some holes. The bof and the spectral approaches were implemented in both cases (with and without holes). The dissimilarity ratios varied between 0.31 and 0.43 (see Table III). Furthermore, the spectral method that relied on the CCD descriptor [12] is robust to holes and gave good results in matching parts of the flowers (petals to petals, stem to stem etc.) since the colors of the parts of the deformed flowers matched in both cases when the models were with holes and free of holes (see Figures 9 and 10).

Finally a new technique of partial matching is proposed and performed using the bag-of-features. A complete model of an artificial foot was partially compared to partial 3D scans of the same object taken at different views using the ARTEC sensor. The same experiments performed on the model of flower were now used in the third set of experiments. First, the number of vertices was reduced without altering the shape of the object using a down-sampling of the vertices called the Fast Farthest Sampling [31]. The Fast Farthest Sampling is a new version of sampling that speeds up the farthest sampling which is adopted by [32]. The farthest sampling enables a uniform and adaptive distribution of sample points and a reconstruction of the mesh (remeshing) so that the correlation between pairs of vertices and the overall error of reconstruction are minimised. In our case the fast farthest sampling also enables a compression of the vertices of the meshes. Thus feature extraction is facilitated without altering the initial shape of the model. After performing the sampling of vertices and the reconstruction of the meshes, the next step is feature extraction and representation. SISI spin images of Daromand Keller [18] were extracted from the global and the partial scans. These features are vector quantized and encoded into visual words using the k-means clustering ($k = 16$ for the foot model and 8 for the flower model). Then the visual words are accumulated into a histogram which represents the feature vector of the model. The same tasks were performed to obtain the feature vectors of
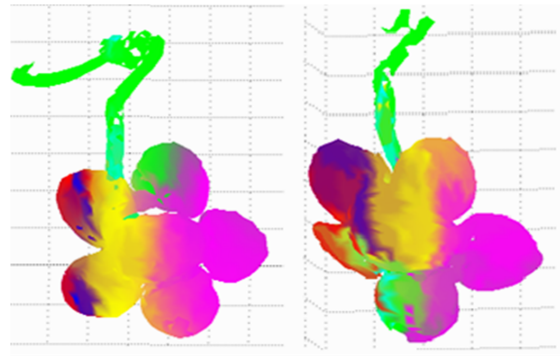

Figure 9.   Matching parts of the deformed models of flowers with holes.
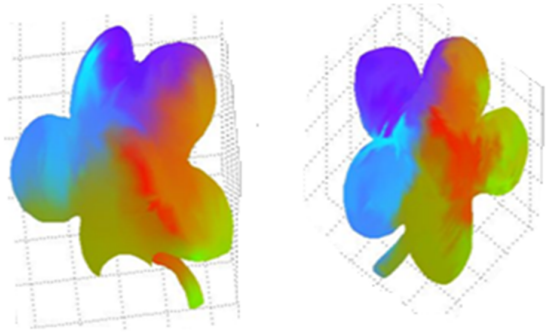

Figure 10.   Matching parts of the deformed models of flowers free of holes.

Table IV
PARTIAL DISSIMILARITY RATIOS OF COMPLETE MODELS AND THEIR
PARTIAL VIEW SCANS

Partial Dissimilarity ratios	Complete model of foot	Complete model of flower
Partial view 1	0.19	0.12
Partial view 2	0.22	0.21
Partial view 3	0.35	0.32

the partial scans of the artificial foot and the flower as well. Finally, using the Kullback-Leibler divergence as a dissimilarity measure, the partial views could be matched to their global models. We also noticed that the more the overlap between the partial view and its complete model is significant, the smaller the dissimilarity ratios. Table IV shows dissimilarity ratios of the different partial views 1, 2 and 3 having respectively about 60%, 45% and 30% of overlap with their complete models: the foot and the flower. Figure 11 shows the partial scans and the full model of the foot.

7 CONCLUSION

This paper provides a comprehensive and comparative survey of the different techniques and approaches used in the retrieval of 3D shapes. The different types of features used in the comparison of models were defined and some of algorithms were implemented on examples of models that are relevant for our research.

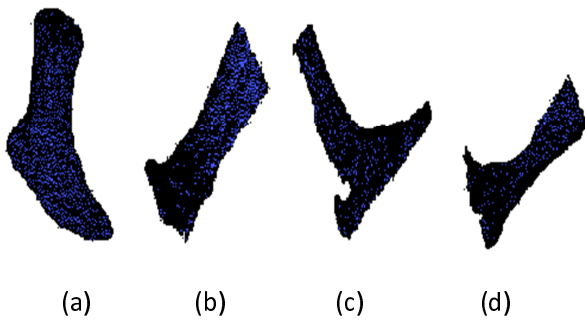After investigating the methods in retrieving 3D

Figure 11. The 3D models of a foot. (a) Complete model, (b) partial view 1, (c) partial view 2, (d) partial view 3.

models and according to the results obtained, we conclude that the bof approaches are more suitable in comparing complex shapes than simple shapes while the D2 shape distribution was more accurate in comparing simple shapes such as the spheres and ellipsoids. Furthermore, the CCD descriptor of the spectral approach was shown to be robust to holes and it enabled good matching of the parts of deformable flower. Surprisingly, similar dissimilarity ratios were also obtained using both the bof and the spectral approach in comparing complex shapes (flower and tiger).

In the future we plan to investigate different types of deformable models, and our aim will be to identify the local regions of deformation in comparing 3D models.

Acknowledgement

References

[1] K.-L. Tam, L. R. W.H, and C.-W. Ngo, "Deformable object model matching by topological and geometric similarity," in *Proc. of the Computer Graphics International*, Jun. 2004, pp. 335–342.

[2] M. Yu, I. Atmosukarto, W. K. Leow, Z. Huang, and R. Xu, "3D model retrieval with morphing-based geometric and topological feature maps," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 2003, pp. 656–661.

[3] A. Johnson, "Spin-images: a representation for 3-D surface matching," Ph.D. dissertation, The Robotics Institute, Carnegie Mellon University, USA, Aug. 2007.

[4] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, 2002.

[5] M. Ankerst, G. Kastenmüller, H.-P. Kriegel, T. Seidl *et al.*, "Nearest neighbor classification in 3D protein databases," in *Proc. of the Seventh International Conference on Intelligent Systems for Molecular Biology*, Jun. 1999, pp. 34–43.

[6] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, "Shape distributions," *ACM Transactions on Graphics*, vol. 21, pp. 807–832, 2002.

[7] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz, "Rotation invariant spherical harmonic representation of 3D shape descriptors," in *Proceedings of the 2003 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, 2003, pp. 156–164.

[8] H. Laga, H. Takahashi, and M. Nakajima, "Spherical wavelet descriptors for content-based 3D model retrieval," in *Proc. of Shape Modelling and Applications*, 2006, pp. 15–25.

[9] H. Sundar, D. Silver, N. Gagvani, and S. Dickinson, "Skeleton based shape matching and retrieval," in *Proc. of the Conference on Shape Modeling International*, 2003, pp. 130–139.

[10] D. Bespalov, W. C. Regli, and A. Shokoufandeh, "Reeb graph based shape retrieval for CAD," in *Proc. of ASME design engineering technical conferences, computers and information in engineering conference*, 2003.

[11] D. Vranic, "Desire: a composite 3D-shape descriptor," in *Proc. of IEEE International Conference on Multimedia and Expo*, 2005.

[12] V. Jain and H. Zhang, "A spectral approach to shape-based retrieval of articulated 3D models," *CAD*, vol. 39, pp. 398–407, 2007.

[13] J. Zhang, R. Kaplow, R. Chen, and K. Siddiqi, "The McGill shape benchmark," 2005, www.cim.mcgill.ca/ shape/benchMark/.

[14] A. Frome, D. Huber, R. Kolluri, T. Bülow, and J. Malik, "Recognizing objects in range data using regional point descriptors," in *Proc. on European Conference on Computer Vision*, 2004, pp. 224–237.

[15] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int.J.Comput.Vis*, vol. 60, no. 2, pp. 91–110, 2003.

[16] L. Zhouhui, G. Afzal, and S. Xianfang, "Visual similarity based 3D shape retrieval using bag-of-features," in *Proc. of the 2010 Shape Modeling International Conference*, 2010, pp. 25–36.

[17] T. Gatzke, C. Grimm, M. Garland, and S. Zelinka, "Curvature maps for local shape comparison," in *Proc. of SMI'05*, 2005, pp. 244–256.

[18] T. Darom and Y. Keller, "Scale-invariant features for 3-D mesh models," *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2758–2769, 2012.

[19] R. Ohbuchi, K. Osada, T. Furuya, and T. Banno, "Salient local visual features for shape-based 3D model retrieval," in *Proc. of SMI'08*, 2008, pp. 93–102.

[20] Y. Liu, H. Zha, and H. Qin, "Shape topics: A compact representation and new algorithms for 3D partial shape retrieval," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006, pp. 2025–2032.

[21] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser, "The Princeton shape benchmark," in *Proc. of the Shape Modeling International*, 2004, pp. 167–178.

[22] T. Furuya and R. Ohbuchi, "Dense sampling and fast encoding for 3D model retrieval using bag-of-visual features," in *Proc. of the ACM International Conference on Image and Video Retrieval*, July 2009, pp. 26:1–26:8.

[23] R. Ohbuchi and T. Furuya, "Distance metric learning and feature combination for shape-based 3D model retrieval," in *Proc. of the ACM workshop on 3D object retrieval*, 2010, pp. 63–68.

[24] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching," in *Proc. of the National Academy of Sciences*, 2006, pp. 1168–1172.

[25] H.-P. Kriegel and T. Seidl, "Approximation-based similarity search for 3-D surface segments," *Geoinformatica Journal*, vol. 2, no. 2, pp. 113–147, Jun 1998.

[26] M. T. Suzuki, T. Kato, and H. Tsukune, "3D object retrieval based on subjective measures," in *Proc. of Ninth International Workshop on Database and Expert Systems Applications*, 1998, pp. 850–856.

[27] D. V. Vranic, D. Saupe, and J. Richter, "Tools for 3D object retrieval: Karhunen-Loeve transform and spherical harmonics," in *Proc. of IEEE Workshop on Multimedia*

Signal Processings, 2001, pp. 293–298.

[28] M. Novotni and R. Klein, "A geometric approach to 3D object comparison," in *Proc. of the International Conference on Shape Modeling & Applications*, 2001, pp. 167–175.

[29] F. Lazarus and A. Verroust, "Level set diagrams of polyhedral objects," in *Proc. of the fifth ACM symposium on Solid modeling and applications*, 1999, pp. 130–140.

[30] D. Steiner and A. Fischer, "Cutting 3D freeform objects with genus-n into single boundary surfaces using topological graphs," in *Proc. of the seventh ACM symposium on Solid modeling and applications*, 2002, pp. 336–343.

[31] C. Moenning and N. A. Dodgson, "Fast farthest point sampling for point clouds and implicit surfaces," 2003, university of Cambridge, Computer Laboratory Technical Report No. 565, Cambridge,UK.

[32] A. Elad and R. Kimmel, "On bending invariant signatures for surfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 10, pp. 1285–1295, Oct. 2003.

Sarah Ali received the BS in Pure Mathematics at the Lebanese University, Tripoli, Lebanon and the MA in Applied Mathematics in Beirut, Lebanon. She is currently working toward her PhD in Electrical Engineering at Laval University, Quebec, Canada. Her interests are in computer vision, 3D modeling, and comparison of 3D models.

From 2009 to 2010 she was a high school teacher at the International School of Akkar in Lebanon and she is currently a teaching assistant at Laval University. She can be joint at sarah.ali.1@ulaval.ca.

Trung-Thien Tran received B.Sc. degree in electrical engineering from Hanoi University of Science and Technology, Vietnam in 2009. He also received the M.S. degree in electrical engineering from University of Ulsan, South Korea in 2011. He is currently working towards his PhD at Laval University in the Computer Vision and System Laboratory. His research interests include computational geometry, 3D modelling and analysis.

Denis Laurendeau holds a Bachelor's degree in Engineering Physics (1981) and a M.Sc. degree (1983) and Ph.D. degree (1986) in Electrical Engineering from Laval University, Québec, Canada. In 1987 he was a visiting scientist at Hydro-Quebec Research Institute (IREQ) where he developed computer vision applications for live-line maintenance at IREQ's Robotics Division. In 2001, he was a visiting researcher at ABB-Bomem where he worked as project leader in the field of Fourier Transform Spectrometry. In 2005 he was a visiting scientist at RDDC-Valcartier (Defence Research Lab) where he worked in Virtual Reality and simulation of complex situations.

Dr. Laurendeau's research interests include 3D image processing, 3D modeling for Virtual Reality, simulation in VR, object tracking, and biomedical applications of computer vision.

Dr. Laurendeau is chairholder of an NSERC-Creaform Industrial Research Chair on 3D Scanning. He is Director of the REPARTI research center and Head of the Computer Vision and Systems Laboratory at Laval University. He was president of the International Association for Pattern Recognition (IAPR) from 2010 to 2012. He is a member of the IEEE.