EUROGRAPHICS 2008 / G. Drettakis and R. Scopigno
(Guest Editors)

*Volume 27 (2008), Number 2*

# Sparse points matching by combining 3D mesh saliency with statistical descriptors

U. Castellani, M. Cristani, S. Fantoni and V. Murino

Dipartimento di Informatica, University of Verona, Italy

**Abstract**

*This paper proposes new methodology for the detection and matching of salient points over several views of an object. The process is composed by three main phases. In the first step, detection is carried out by adopting a new perceptually-inspired 3D saliency measure. Such measure allows the detection of few sparse salient points that characterize distinctive portions of the surface. In the second step, a statistical* learning *approach is considered to describe salient points across different views. Each salient point is modelled by a Hidden Markov Model (HMM), which is trained in an unsupervised way by using contextual 3D neighborhood information, thus providing a robust and invariant point signature. Finally, in the third step, matching among points of different views is performed by evaluating a pairwise similarity measure among HMMs. An extensive and comparative experimental session has been carried out, considering real objects acquired by a 3D scanner from different points of view, where objects come from standard 3D databases. Results are promising, as the detection of salient points is reliable, and the matching is robust and accurate.*

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling

## 1. Introduction

Recent advancement in research for digitizing and modeling 3D shapes has led to a rapid expansion of the number of available 3D models [MSS*06, SF06, FKMS05]. Such large amount of data poses new challenges for many computer vision and pattern recognition applications such as 3D shape retrieval [FKMS05], objects recognition [JH99, FHK*04], data reduction, and so on. In this context, matching of interesting points is a relevant research topic devoted to the detection of similarities between two or more shapes by considering local information. Two main approaches have been adopted in the literature for shape matching problem, namely, *local* and *global* approaches. Local matching [SF06] is performed between sub-parts or regions of the models. This is in contrast to the global shape matching paradigm, where similarity is measured among entire models [FMK*03].

Typically, the main steps that compose a 3D partial model matching method are: (i) *detection* (ii) *local description*, and (iii) *matching* [SF06]. In the detection step, 3D points with high information content are extracted. In the local description step, interest points are described by including information on their neighborhood area in a compact form, so that a local contextualization of the interest points is provided. Finally, point matching is carried out by defining an appropriate similarity measure among local descriptors. This latter module separates all the interest points in a set of *matching* points, organized as couples, and a set of uncorrelated points.

In this paper, we provide a novel 3D partial matching framework, which deals with a set of partial views of the same object acquired by a 3D scanner. Two main contributions can be highlighted: first, we propose a robust method for the selection of a very small fraction of the whole set of points, extracting those having a strong representativeness with respect to the others. To this aim, we define a 3D saliency measure, able to extract perceptually meaningful interest points from 3D meshes. The proposed approach is theoretically founded and it is inspired by the research on saliency measure on images [IKN98, Lin94, Low04]. In short, the source mesh is decomposed in multiscale repre-


Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

sentations, and salient points are then extracted by distilling opportunely the results gathered on each scale level. The idea is to find robust variations in the mesh which are resolution invariant.

The second contribution consists in introducing a novel local description of interest points based on Hidden Markov Models (HMMs) [Rab89]. For each detected point, multidimensional features are sampled along a 3D geodesic spiral pathway, that lies in a neighborhood zone. Then, a HMM is trained, for each detected point, using the related features in an unsupervised way, providing a reliable model-based point description. Subsequently, points matching among different object views is performed by coupling corresponding interest points using a HMM similarity measure. This provides reliable performances in terms of correct matches and computational time, as compared to state-of-the-art methods.

It is worth noting that the most of partial matching methods select the matched points by introducing some global constraints, which depend by the contextual application. For instance, in a registration framework, 3D points correspondences are combined with global rigid constraint. This strongly improves the matching by safely removing outliers [FKMS05]. Unlike these methods, we focus on *pure* local matching by dealing only with few carefully selected points. This makes our framework versatile for different potential applications such as 3D data categorization, deformable object modelling, shape morphing, and so on.

The rest of the paper is organized as follows. Section 2 describes the state-of-the-art. Section 3 introduces our method for salient points detection, while in Section 4 the HMM framework for interest points description and matching is detailed. Results are shown in Section 5 and, finally, conclusions are drawn in Section 6.

## 2. Related work

We organize this section as follows. First, we introduce the state-of-the-art of interest points' detection by focusing of the notion of saliency. Second, we describe the literature for points description and matching.

**Salient points' detection.** The detection of 3D interest points can be in general faced as extension of the correspondent task performed on 2D images [Low04]. However, such extensions are not straightforward and very few works have shown their effectiveness on both the domains. Actually, while the point description and matching phases are thoroughly addressed in literature, there are few methods oriented to the robust selection of interest points in the 3D domain. The simplest approach is to extract the points by random [FHK*04] or uniform sub-sampling [JH99] the whole set of points, or by adopting a spectral-analysis approach [ZvKD07]. More recently, the interest point selection is focused by exploiting the notion of *saliency* on the 3D domain [LVJ05]. In the literature, well-founded notions of 2D saliency are present which can be divided in two groups. Both perform multi-scale analysis, collecting for each scale filter responses that measure edges and other features [Lin94, Low04, IKN98]. In the first group [Lin94, Low04], named as "independent multi-scale", all the maxima detected over different scales are considered interest points. In the second group [IKN98], named as "joint multi-scale", all the intra-scales features are combined in a single saliency map where maxima are extracted. Both the groups are motivated by perceptual theories: independent multi-scale approaches assume that humans perform automatically and independently a multi-scale analysis, paying more attention to the scales where maxima are present [Lin94]. The joint multi-scale approaches state that humans extract maxima after a natural native smoothing process, aimed at discarding irrelevant maxima.

By focusing on the 3D domain, the concept of *saliency* is not consolidated, and only few works are presented. In [LVJ05], a definition of *mesh saliency* for mesh simplification and best view point selection are introduced; the focus is here on the joint multi-scale by considering the local curvature as a discriminative feature. In [PKG03] a multi-scale approach is proposed for the extraction of line-type features. For each point, a measure of *surface variation* is introduced by combining the eigenvalues of the local covariance matrix. Then, the line-segments are selected by computing the *persistence* of feature-points over different scales and by performing hysteresis thresholding. A similar approach is proposed by [GMGP05], for selecting the *integral*-volume descriptors. In [GCO06], the authors improve part-in-whole matching by introducing salient geometric features based on curvature properties. In [SF06], an approach for the selection of distinctive 3D shape descriptors is described. After a training phase, the retrieval performance of each descriptors has been evaluated and only the most distinctive are retained. In [WNK06], an interesting parallelism between the 2D and the 3D realms is proposed only regarding the independent multi-scale paradigm [Low04, WNK06].

**Description and matching.** Roughly speaking there are two main approaches for model description and matching: *global* and *local* [SF06, MBO05]. The *global* shape-based approach consists in selecting a set of features that effectively and concisely describe the entire 3D model, and in introducing a distance function between the models descriptions. See [FMK*03] for an exhaustive survey on global shape-based approach. The *local* feature matching paradigm is instead oriented to the detection of part-to-part correspondences by defining a descriptor (or signature) for each 3D interest point. Spin images are introduced [JH99] by creating cylindrical projection of local sets of surface points represented as an image. In [FHK*04] the authors propose the concept of regional point descriptors for the 3D domain. Regional point descriptors lie midway between the classes of

global and local approaches, giving them the advantages of both. In [KPNK03] the shape context is adapted to 3D points distribution, inspired by the work proposed in [BMP02] for the 2D domain. In [MBO06] a novel tensor representation is proposed for robustly describing the points in the context of the automatic pairwise registration of range images. The effectiveness of the method is shown also dealing with low resolution images. In [MPS*03] the paradigm of *blowing bubbles* has been introduced by combining local surface properties at different resolutions. The main idea consists in estimating not only the curvature of a vertex over neighborhoods of variable size, but also in taking into account the topology of the surface in that neighborhood. More examples of *local* feature matching with interesting analysis and discussions, are reported in [MBO05].

In this paper we propose a local description and matching framework based on HMM. Here, the novelty is the introduction of a learning approach to model local geometry variations. Few works propose HMM for shape matching [BM04], and none of them address the 3D domain adopting a local approach.

## 3. Salient points detection

In our framework, we are dealing with 3D partial meshes resulting from a scanning process; therefore, clutter, holes and occlusions due to the acquisition procedure and to the sensor noise poses us in a challenging setting. For this reason, we choose the "joint multi-scale" paradigm, assuming it more robust to noise. Note that the considered meshes are dense and, although the density depends by the point of view, the sampling is locally uniform.

Let $M$ be a given mesh; as preliminary operation, we remesh $M$ at $D$ different levels of decimation (the quadric edge collapse decimation approach has been used [GH97]), obtaining the 3D meshes $M^d, d = 1,\dots,D$ at different resolutions. We call $M^d$ as *octave-d* mesh in order to remark the fact that we consider a variation of resolution as a jump of an octave in the scale space. Our salient points detection is then composed of two main phases, namely *intra-octave* and *inter-octave* phases.

**Intra-octave phase.** The intra-octave phase is based on the processing of the 3D mesh $M^d$ and consists in three main steps: (i) multiscale representation, (ii) 3D saliency measure definition, and (iii) intra-octave points detection.

*(i) Multiscale representation:*

The first step, consists in applying $N$ Gaussian filters on the considered mesh $M^d$, obtaining $N$ multidimensional filtering maps $\{F_i^d\}, i = 1,\dots,N$. Gaussian filtering is applied as follows: let $g(v,\sigma)$ the Gaussian operator with standard deviation $\sigma$, applied on the vertex $v \in M^d$. The *neighborhood region* of $v$, over which the filtering is applied, is built by expanding a $n$-rings search starting from $v$, and collecting all those vertices displaced within a distance equal to $2.5\sigma$. This area can be considered as a good approximation of a geodesic area of radius $2.5\sigma$.

The Difference-of-Gaussians (DoG) operator is defined as:

$$F_i^d(v) = g(v,\sigma_i) - g(v,k\sigma_i) \tag{1}$$

where $\sigma_i$ is the value of the standard deviation associated to scale $i$ and $k$ is a constant equal to 2. We fix six scales of filtering, corresponding to standard deviation values $\sigma_i \in \{1\varepsilon, 2\varepsilon, 3\varepsilon, 4\varepsilon, 5\varepsilon, 6\varepsilon\}$, where $\varepsilon$ amounts to 0.1% of the length of the main diagonal located in the bounding box of the model. Note that, as studied by [Low04], fixing a constant factor $k$ for DoG computation provides a close approximation to the scale-normalized Laplacian of Gaussian, which is required for true scale invariance.

*(ii) 3D saliency measure definition:*

This step aims at obtaining a dense measure of mesh saliency (i.e., associated to each vertex). Note that $F_i^d(v)$ is a 3D vector which denotes how much the vertex $v$ has been moved from its original position after the filtering. In order to reduce such displacement in a scalar quantity, we observe that in general the most significant (in a perceptual sense) motion of the vertices is along the direction perpendicular to their local surface (i.e., along the normals). Therefore, we project the vector $F_i^d(v)$ to the normal $n(v)$ of the vertex $v$. In this fashion we obtain the *scale* map $M_i^d$ as:

$$M_i^d(v) = ||n(v) \cdot (g(v,\sigma_i) - g(v,k\sigma_i))||. \tag{2}$$

Furthermore, this reduces the shrinking effect which rises typically when Gaussian filter is applied to meshes [PKG03]. Moreover, according to the "joint multi-scale" paradigm, each map is normalized by adopting the Itti's approach [IKN98]:

- normalizing the values in the map to a fixed range $[0,\dots,R]$;
- finding the location of global maximum $T$;
- finding all the other local maxima and computing their average $\hat{t}$;
- globally multiplying the map by $(T-\hat{t})^2$ by obtaining the final normalized scale map $\hat{M}_i^d$.

The effect of this normalization is to increase the evidence of the highest peaks.

*(iii) Intra-octave points detection:*

We emphasize the above peaks enhancement, by introducing an adaptive *inhibition-process* on each normalized scale map. From each vertex $v \in M^d$, we consider all the values of the scale map $\hat{M}_i^d$ observed on the neighborhood of $v$. If the $\hat{M}_i^d(v)$ is higher than the 85% of the values in its neighborhood, the value is retained, otherwise $\hat{M}_i^d(v)$ is set to zero. Therefore, the *inhibited* saliency map is obtained by simply adding the contribution of each *inhibited* scale map. Finally, in order to detect salient points a *non-maximum suppression* phase on the *inhibited* saliency map is performed: a point is detected if it is a local maximum and its value is higher than the 30% of the global maximum. Note that, after the inhibition phase, the neighbourhood of a point is adaptively

defined by expanding the local region while new non-zero points are found. Fig. 1, shows a scheme of the *intra-octave* phase. On the left are shown the meshes after the multiscale representation, in the center the output of saliency computation is highlighted and, finally, the intra-octave salient points detection is shown on the right.

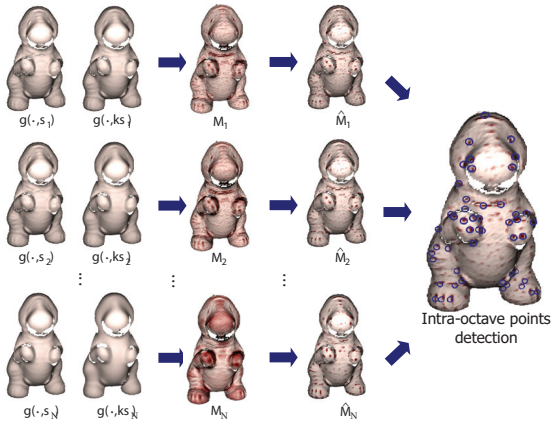


**Figure 1:** *The scheme of the proposed* intra-octave *phase: different gaussian operators $g(\cdot,\sigma_i)$ are applied to the source mesh, then the* scale *maps are computed and normalized. Finally the intra-octave salient points detection is carried out.*

**Inter-octave phase.** In order to improve the robustness of the method to variation of mesh resolution an *inter-resolution* validation process is carried out. We define five levels of decimation $d \in [0, h, 2h, 3h, 4h]$, respectively, where $h = 0.20$. Steps (i),(ii) and (iii) of the intra-octave phase are carried out for each octave $M^d$. Then, a majority criterion is adopted for detecting the *validated* salient points, i.e., only points appearing at least in three octaves are retained.

## 4. Hidden Markov description of interest points

The goal of this step is to build a compact description able to summarize information related to interest points and to their neighborhood area. Let us suppose that $I$ interest points have been extracted, and let us focus on point $v_i$; around it, we build a clockwise spiral pathway $s(v_i)$ connecting vertices which lie at 1-ring distance, then at 2-ring distance and so on, until a fixed geodesic radius $r$ is reached. The radius is fixed to be the 5% of the main diagonal of the bounding box in which the 3D object lies. Connections among vertices which lie on different ring distances are rearranged in order to maintain the area covered by the spiral as regular as possible, obtaining thus a circular geodesic area around $v_i$. If holes are present on the mesh, no data is collected by jumping to the next available point, as visible in Fig.2. Along this pathway, we extract local point information [Pet02] composed by

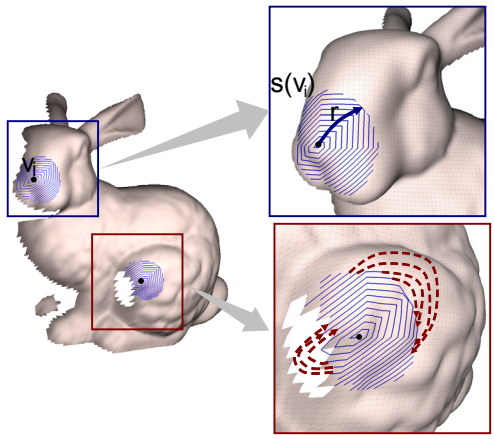


**Figure 2:** *Interest point description: on the left, two interest points are depicted with black dots with their respective spiral pathways. On the top-right, a zoom on point $v_i$, which spiral pathway $s(v_i)$ is limited by the geodesic radius r. On the bottom-right, a spiral built in presence of an hole in the mesh; red dotted arrows give an idea of how the different portions of the spiral are rearranged in a 1D array.*

the saliency level, extracted in the previous step, the maximal and minimal curvature and the normal displacement between the local point and the salient point. Experimentally, we saw that other local features, such as the Gaussian curvature and the shape index, do not improve the description.
Once the data on the spiral $s(v_i)$ is acquired, we observed that all its 5-dimensional entries $\{o\}_i$ form entities which in principle could be quantized in few values, that occur repeatedly themselves along the spiral. For this reason, modelling the spiral as a stochastic process, in which the different entities are thought as discrete states, is a reasonable choice. The model more suited for this idea is the discrete-time Hidden Markov Model (HMM) [Rab89]. A HMM can be viewed as a Markov model whose states are not directly observable: instead, each state is characterized by a probability distribution function, modelling the observations corresponding to that state. More formally, a HMM is defined by the following entities [Rab89]:

- $S = \{S_1, S_2, \cdots, S_N\}$ the finite set of (hidden) states; in our case each state is associated to a particular local geometric configuration that occurs along the spiral.
- the transition matrix $A = \{a_{kj}\}$, $1 \le k, j \le N$ representing the probability of moving from state $S_k$ to state $S_j$,

$$a_{kj} = P[Q_{t+1} = S_j | Q_t = S_k], \quad 1 \le k, j \le N,$$

with $a_{kj} \ge 0$, $\sum_{j=1}^{N} a_{kj} = 1$, and where $Q_t$ denotes the state occupied by the model at time $t$. Here, this matrix encode how the different local configurations succeed along the spiral.

- the emission matrix $B = \{b(o|S_k)\}$, indicating the probability of emission of symbol $o \in V$ when system state is $S_k$; $V$ can be a discrete alphabet or a continuous set (e.g. $V = \mathbb{R}$), in which case $b(o|S_k)$ is a probability density function. In this paper we used a 5-dimensional Gaussian HMM, *i.e.*

$$b(o|S_k) = \mathcal{N}(o|\mu_k, \Sigma_k).$$

  where $\mathcal{N}(o|\mu,\Sigma)$ denotes a Gaussian density of mean $\mu$ and diagonal covariance matrix $\Sigma$, evaluated at $o$, which represent an entry of the spiral pathway; In our approach, this distribution codifies how probable values on the spiral are connected to a hidden state.
- $\pi = \{\pi_k\}$, the initial state probability distribution,

$$\pi_k = P[Q_1 = S_k], \quad 1 \le k \le N$$

  with $\pi_k \ge 0$ and $\sum_{k=1}^{N} \pi_k = 1$.

For convenience, we represent an HMM by a triplet of parameters $\lambda = (A, B, \pi)$.

Learning the HMM parameters, given an observed sequence $s(v_i)$, is usually performed using the well-known Baum-Welch algorithm [Rab89], which is able to determine the parameters maximizing the likelihood $P(s(v_i)|\lambda)$. An open issue is the correct choice of the number of hidden states. In this paper, the HMM is trained in an unsupervised fashion, using an improved version of the Baum Welch algorithm. This procedure decides the number of states automatically, following an Minimum Description Length (MDL) principle customized for the HMM framework [BMF03]. In practice, the idea is to perform the training section several times, in a serial way, starting by using a large number of states (e.g., 100 states). Each training session starts from a "nearly good" situation, derived from the result of the previous training session by pruning the "least probable" state of the model, if necessary. After the training, using $s(v_i)$ as training sequence, we generate the model $\lambda_i$.

In this way, the HMM gives a statistical encoding of the interest point and its neighborhood, taking into account for the uncertainty in the data. Actually, each HMM state captures a particular geometrical aspect particularly evident near $v_i$. In practice, as shown in the experiments, the expressivity of such a characterization is robust to 1)rotation 2)irregular sampling (for example, due to holes in the mesh) and 3)resolution variation of the mesh over which the interest point lies.

### 4.1. HMM-based matching

Let us suppose to have an object captured by a 3D scanner from two different view-points, obtaining meshes $M$ and $M'$. On the mesh $M$ we find $I$ interest points, each one described by one HMM, thus collecting $I$ models $\lambda_1, \ldots, \lambda_i, \ldots, \lambda_I$. The same applies for mesh $M'$, where we find $J$ interest points, described by HMMs $\lambda_1, \ldots, \lambda_j, \ldots, \lambda_J$.

The goal of the matching step is to find links between points of the two views, such that for each point $v_i$ of $M$ two alternatives are possible: 1) there is only one point $v_j$ in view $M'$ located in the same absolute position of $v_i$ w.r.t. original 3D object *or* 2) such point $v_j$ does not exist. In case 1, we find a *matching* between $v_i$ and $v_j$, otherwise we say that point $v_i$ is *unique* for view $M$ and $M'$. The same applies for points $\{v_j\}$ of view $M'$. After the matching step, we thus have a categorization of all the points $\in M \cup M'$ formed by a set of unique points and a set of matched points.

Such categorization is achieved by using the HMM descriptions: the main ingredient is the symmetric *similarity matrix* $G$, which is a $I \times J$ matrix where each element $g_{ij}$ corresponds to a widely used HMM similarity measure [Smy97]:

$$g_{ij} = \frac{LL(v_i|\lambda_j) + LL(v_j|\lambda_i)}{2} \tag{3}$$

where $LL(v_i|\lambda_j)$ indicates the log-likelihood of the data forming the spiral pathway $s(v_i)$ given the model $\lambda_j$, i.e. $\log P(s(v_i)|\lambda_j)$.

If we consider the similarity measures as weights that characterize links between points, the above categorization can be formally cast as a maximum weighted matching problem (MWMP). Roughly speaking, in this case MWMP translates in selecting for each point of one view only one weighted link to another point on the other view, such that the summation of all the weights is maximal. In order to face this problem, we adopt the classical flow algorithm proposed in [CWC*96].

At the end, we have a set of correspondences which indicates similar local regions displaced on the views of the meshes. Obviously, if $I \neq J$, some points will remain unlabeled; these points concur to form the set of the unique points.

## 5. Experiments

In this section we explore thoroughly the effectiveness of the phases of our local matching strategy. Our "standard" dataset focuses on 8 models; 4 taken from the Stuttgart Range Image Database (SRID) [HLLS01] and 4 selected from the Minolta database [CF98]. Each model includes several range views: we choose 6 distant views for each SRID models and 3 views of the Minolta database models. For each considered view, we produce a triangular mesh (see the first six rows of Fig.9 and the first three rows of Fig.10: below each 3D view, the first number is the view-index, the second number locates the view in the original database, the third one - in bold - is the number of detected salient points).

### 5.1. Points detection

We investigate our interest points detection technique applying it to all the views of the experimental dataset, showing

also results gathered by other state-of-the-art methods. In particular, we consider the umbilical points extraction procedure [Pet02] (UMB), the method proposed in [MSS*06] (MEIG), where points having highest minimum eigenvalue in the scatter matrix are selected, and the method in [LVJ05] (CURV), where the saliency map is defined starting from the curvature map (in order to provide a point detection from the saliency map we apply the intra-octave point detection step to the output of [LVJ05]). In Fig.3, we report detection results on some views of our dataset. To evaluate the results

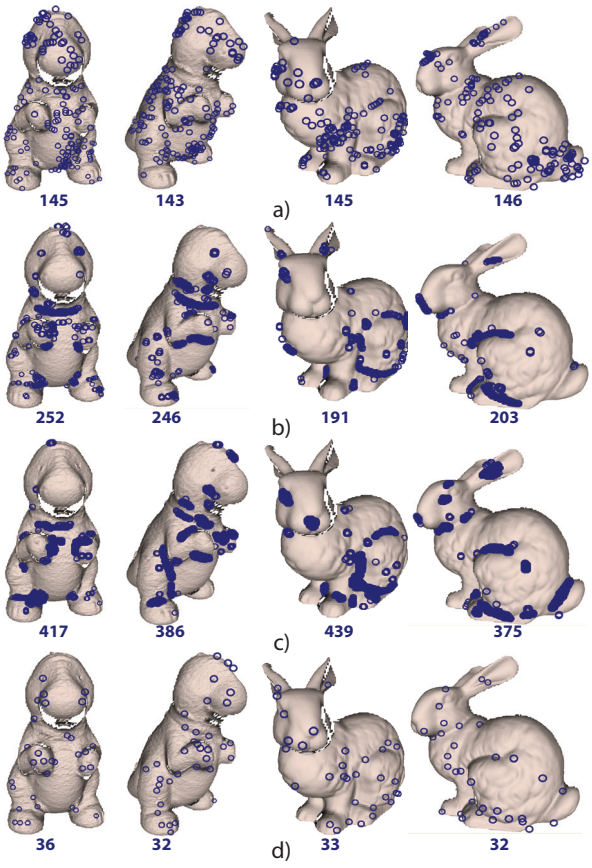


**Figure 3:** *Detection results: each row shows the detection results obtained on two Minolta 3D views (1–072, 2–108) and on two SRID 3D views (4–99, 5–98), by using a) UMB [Pet02], b) MEIG [MSS*06], c) CURV [LVJ05], and d) our method, respectively. Below on each view the number of extracted points is shown.*

several meaningful considerations can be done. In general, each point extracted by our method represents a particular and distinctive portion of the 3D view. Fewer and more meaningful points are extracted than those produced by the other techniques, which are tightly collapsed on the edges, or diffused uniformly on the surface of the captured object. Our method is able to detect most significant parts of the subjects such as the eyes, the nastrils of the nose, the fingers of the paws, and so on.

Note that for all the comparative methods, an accurate parameters' tuning phase has been applied, in order to propose the fairest comparison. Note also that our method does not need of any tuning of parameters (few general parameters are set globally for all the models of both the databases). Moreover, using our method several detected points are in correspondence among different views, facilitating the point matching phase (see for example the paws of the animals in Fig. 3).

### 5.2. Point Description

This section provides some insight on the robustness of the HMM salient point description against changes in the 3D mesh resolution. Here, we consider one view of the "Bunny" model (view 4–99) at two different resolutions, i.e., $M^0$ and $M^{0.6}$, respectively. In Fig.4, we show two close-ups on the nose of the Bunny. The detection phase finds two salient points in the same absolute position. Spiral pathways are automatically built, and two HMMs $\lambda_H$ and $\lambda_L$ are trained, for the high and the low resolution, respectively. Four considerations can be done: 1) the unsupervised learning process produces in both cases 6 states; 2) the observation densities associated to each state permit to delineate a clear correspondence (in a Malahanobis sense) between the states of the different HMMs; 3) by means of this correspondence, transition matrices are equivalent, even if in $\lambda_L$ the auto-transition probabilities are higher, due to the minor length of the low resolution spiral; 4) we plot the Viterbi path of the associated HMM along the two spirals: equal colors stand for corresponding states, in the sense explained before; indeed, it is possible to see that corresponding zones are described by equivalent states.

### 5.3. Point Matching

The validation accomplished in this section is the most dense: actually, it exhibits results that witness the robustness and accuracy of all the phases of the proposed framework. Let us initially focus on the Bunny model, concentrating on the results obtained on a particular view pair, $M$ and $M'$ (Fig.5).
Among all the points detected we can identify those "positive" points $P$ which have a real correspondence in the other view (true correspondences have been evaluated by hand), and the remaining "negative" points $N$ which represent locations which have no correspondent in the other view. After the matching, we can define $TP$ (green circles with an outgoing solid link) and $FP$ (red circles with an outgoing dashed link) as the numbers of points for which an exact or a wrong

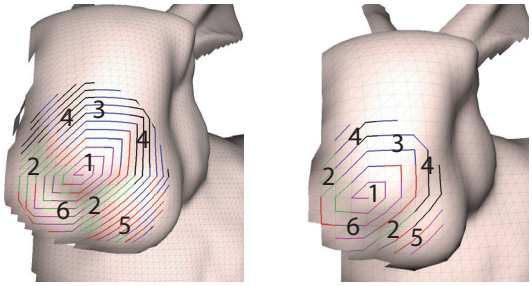


**Figure 4:** *Qualitative robustness of the HMM descriptions: the Viterbi path of the HMMs built on the two different-length spirals are plotted; equal color in different HMMs corresponds to similar states (see text). For visual clarity, a state-identifying number is positioned on the area which exhibits mainly the presence of that state. Note that similar states lie in corresponding areas.*

correspondence has been found, respectively. In the same fashion, we can define $TN$ (green circles) as the number of unique (see Sec.4.1 for a definition of *unique*) points found and $FN$ (red solid circles) the number of points detected as unique for which a (correct) correspondence does exist. Analyzing Fig.5, it is possible to see that: 1) a big amount of

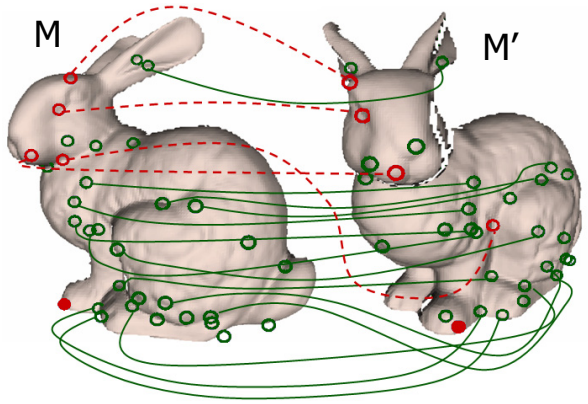


**Figure 5:** *Matching results: green circles with an exiting solid link mean right correspondences found, red circles with an exiting dashed link mean wrong correspondences found; green circles and red solid circles mean correct and uncorrect unique points found, respectively.*

interest points have been matched correctly, showing that the salient point detection is robust to view-changes; 2) points not in correspondence are mainly due to the different geometric displacement of the views which occluded the parts; 3) on the left ear of the bunny, on $M'$, the mesh is incomplete, but nonetheless, a salient point has been detected, and the HMM description has been able to capture and model the geometric area near that point, associating it exactly to the correspondent one on $M$.
In order to compactly summarize the performances of our algorithm, we build the *global matching index* $GB = (TP + TN)/(P+N)$. This quantity expresses accurately the ability of the system to capture the existent correspondences between views; the maximum value $GB = 1$ means that all the right correspondences and unique points have been discovered. The GB values are given as percentages, in order to ease the understanding.

Another experiment suitable to show the robustness of our model against changes in the mesh resolution is performed simply by considering one view $M^0$, and its decimated versions $M^{0.2}, M^{0.4}, M^{0.5}, M^{0.6}, M^{0.7}, M^{0.8}$, respectively. Thus, our technique is applied. In Fig.6 matching results between the original view $M^0$ and the decimated one $M^{0.6}$ are shown. Please note that some salient points in the right view are missing, due to the effective geometric change in the mesh. Anyway, note that the remaining salient points are effectively discovered by our method, and the points on the high-res view not present in the low-res view are mostly detected as unique points. On Fig.7, the GB value are presented for

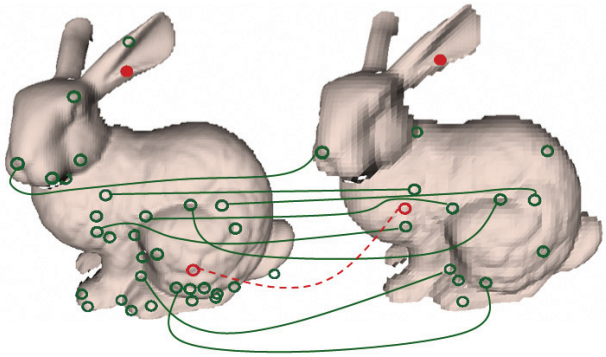

**Figure 6:** *Resistance against down-sampling of our technique: on the left, the original view $M^0$; on the right, the same decimated view $M^{0.6}$. For the meaning of the points' correspondences see caption of Fig.5.*

each resolution. After the decimation level 0.7 the performances fall drastically, mainly due to the fact that the geometric aspect of the low-res view is greatly different from the original one. After that, we consider all the models, and all the views of our datasets (Fig.9-10 ), and, for each model, we perform detection and description of the interest points on all the views (6 for the SRID models and 3 for the Minolta models). Subsequently, we apply the matching algorithm, evaluating correspondences among all the possible couples of views, given that model. For each couple of views, we calculate the GB value, and we put all the values in a matrix form, shown in Fig. 9, $7^{th}$ row and Fig.10, $4^{rd}$ rows. Note that,

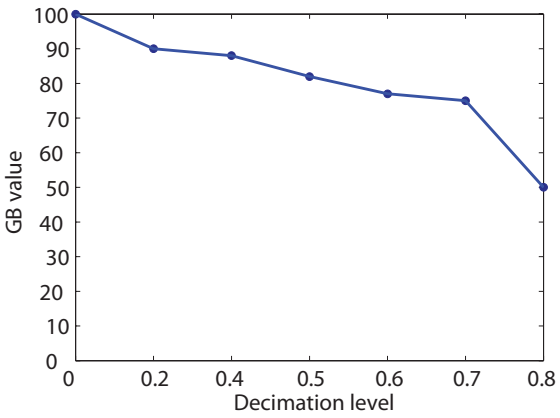


**Figure 7:** *Resistance against down-sampling evaluation: the GB values (on the y-axis) calculated by considering one view and its decimated version by different levels are shown (on the x-axis).*

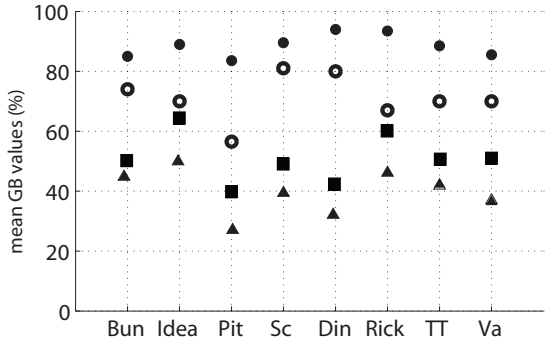


**Figure 8:** *Global experiments. The circles represent the mean GB values obtained by our technique, for each the 3D models. The triangles and the squared represent the results obtained using Spin Images and 3D Shape Contexts respectively. The rings represent the mean GB values using the view dataset with decimated views by a level 0.6. Note that our technique applied on the "mixed" dataset overtakes both the Spin Images and 3D Shape Contexts techniques applied on the "standard" dataset.*

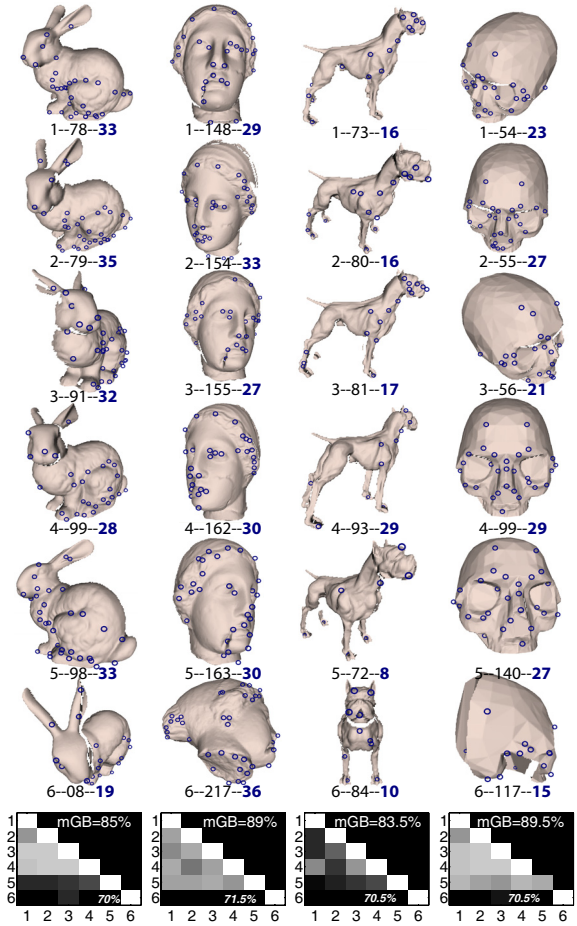


**Figure 9:** *Matching results for SRID models: the first six rows, the views of the 3D objects considered, for each view we report the view index and the number of extracted salient point (in bold); in the seventh row, (symmetric) matching matrices, in which only the off-inferior diagonal elements are present; the minimum GB value is shown in the corresponding entry ij; brighter entries mean higher similarity values; in the top/right the mean GB value achieved for that 3D model is reported; all the mean GB values are summarized in Fig.8.*

intuitively, higher values are collected when two consecutive views are taken into account. Additionally, for each 3D model we calculate the mean GB value (*mGB*), reported in the top right the GB matrices, and the minimum *GB* value (in italic). Note that all the GB values are bigger than 70%. As comparison, we perform the same extended matching experiment by changing salient point descriptors, calculating for all the salient points of all the views considered the related spin images and 3D *shape context*, according to the method described in [JH99] and in [KPNK03] respectively. We estimate the free parameters (such as the support-size of the spin images) in order to obtain the best results. Then, we adopt the same policy adopted for the HMM descriptions, substituting to the HMM similarity measure the correlation value among spin images or shape contexts. In Fig.8 the mean GB values of the involved techniques adopted for each 3D model are compared. Note that the proposed method clearly outperforms both the matching based on the spin image, and the matching based on the shape context.

Finally, in order to summarize all the issues faced in this

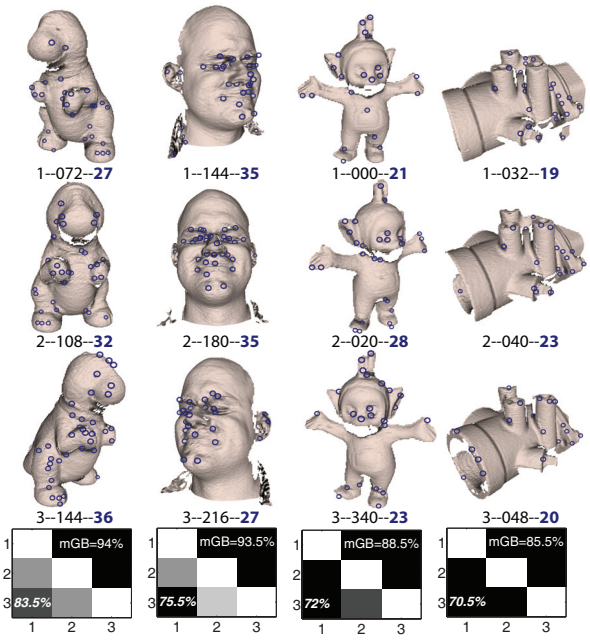


**Figure 10:** *Matching results for Minolta models: On the first three rows, the views of the 3D objects considered, for each view we report the view index and the number of extracted salient point (in bold); in the fourth row, (symmetric) matching matrices, for which the same considerations made in Fig.9 do hold.*

section, we create a novel "mixed" dataset, in which for each 3D model we decimate by a level 0.6 the $1/3$ of the present views, maintaining the remaining views unchanged. Then we apply our framework to all the models and all the views, using our HMMs descriptors. The results, in terms of mean GB, are reported on Fig.8.

### 5.4. Performance evaluation

In order to evaluate the feasibility of our technique on real applications, we report the computational efforts spent on the main involved steps. Table 1 summarises both the computational complexity and the running times. Tipicaly, a mesh is composed by $20K$ triangles. Points detection and feature extraction are carried out for a single mesh where $V$ is the number of vertices and $\eta$ is the *mean* size of the neighborhood (tipically around 50 points but it changes with $\sigma$). The HMM training and testing phases are performed for each salient point. More in details, HMM training is related to the computation of $\lambda = (A,B,\pi)$, while HMM testing addresses the computation of $LL(v_i|\lambda_j)$. The table refers to a single point where $N$ is the number of hidden states and $\tau$ is the length of the sequence associated to that point. Note that the training phase is repeated for several value of the hidden states $N$. Then, the *best* value $N$ is chosen according to the MDL principle [BMF03]. Note further that the training is an off-line phase. The matching step refers to a pair of meshes having $I$ and $J$ salient points respectively. It is related to the solution of the MWMP problem applied to the similarity matrix $G$ of Eq. 3. According to the MWMP algorithm the matrix $G$ is considered as a graph of $X = I + J$ nodes and $E = I \cdot J$ edges. Experiments have been carried out on a Intel Core 2 Duo, $E6300$, 1.86Ghz. Points detection and feature extraction have been implemented in $C++$, while the HMM training, testing and the matching have been developed in Matlab.

| Step | Complexity | Run. times (sec.) |
|---|---|---|
| Points detection | $O(V \cdot \eta)$ | 9.3 |
| Feature extraction | $O(V \cdot \eta)$ | 0.8 |
| HMM training | $O(N \cdot \tau)$ | 4.65 |
| HMM testing | $O(N \cdot \tau)$ | 0.08 |
| Matching | $O(X \cdot (E + X log X))$ | 30.3 |

**Table 1:** *Performances of the main steps of the proposed framework. Running times are related to mean values.*

## 6. Conclusions

In this paper, a new approach for 3D points matching is proposed. Few and sparse interest points are selected robustly by exploiting visual saliency principles on 3D meshes. Then, we propose a Hidden Markov model-framework that combines at the same time points description, organized as a spiral pathway around the interest point, and matching. Therefore, a thoroughly experimental section is reported by analyzing real partial views of 3D objects acquired by a 3D scanner. Although such kind of data are particularly challenging because of noise, holes and occlusions, the proposed results are very promising. The proposed detection method is able to evidence the most significant parts of each view (i.e., eyes, nose, knee, and so on) in a more stable fashion with respect to other techniques in the literature. Moreover, the matching performance are always higher than 70%, which means that our method safely detects the large majority of the correspondences without any global constraint, outperforming similar methods, based for example on spin-images or 3D shape context. Future work will address the extension of the proposed methodology to 3D object retrieval by exploiting effective *local-to-global* representation of the involved objects. Moreover, we will investigate the extension of the proposed approach to general models, non necessarly coming from a scanning process (i.e., CAD-like models).

### Acknowledgments

This work was supported by the Italian Ministry of Research and Education under projects *Three-Dimensional Shape Indexing and Retrieval Techniques*, and *Similarity-based methods for Computer Vision and Pattern Recognition: theory, algorithms, applications*.

**References**


[BM04] BICEGO M., MURINO V.: Investigating hidden markov models' capabilities in 2d shape classification. *IEEE Trans. Pattern Anal. Mach. Intell. 26*, 3 (2004), 281–286.

[BMF03] BICEGO M., MURINO V., FIGUEIREDO M.: A sequential pruning strategy for the selection of the number of states in Hidden Markov Models. *Pattern Recognition Letters 24*, 9–10 (2003).

[BMP02] BELONGIE S., MALIK J., PUZICHA J.: Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell. 24*, 4 (2002).

[CF98] CAMPBELL R., FLYNN P.: A WWW-accessible 3D image and model database for computer vision research. In *Empirical Evaluation Methods in Computer Vision* (1998).

[CWC*96] CHENG Y., WU V., COLLINS R., HANSON A., RISEMAN E.: Maximum-Weight Bipartite matching technique and its application in image feature matching. In *Proc. SPIE Visual Comm. and Image Processing* (1996), vol. 27.

[FHK*04] FROME A., HUBER D., KOLLURI R., BULOW T., MALIK J.: Recognizing objects in range data using regional point descriptors. In *ECCV* (2004).

[FKMS05] FUNKHOUSER T., KAZHDAN M., MIN P., SHILANE P.: Shape-based retrieval and analysis of 3D models. *Communications of the ACM 48*, 6 (2005).

[FMK*03] FUNKHOUSER T., MIN P., KAZHDAN M., CHEN J., HALDERMAN A., DOBKIN D.: A search engine for 3D models. *ACM Transactions on Graphics 22* (2003).

[GCO06] GAL R., COHEN-OR D.: Salient geometric features for partial shape matching and similarity. *ACM Transaction on Graphics 25*, 1 (2006).

[GH97] GARLAND M., HECKBERT P. S.: Surface simplification using quadric error metrics. In *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques* (1997).

[GMGP05] GELFAND N., MITRA N. J., GUIBAS L. J., POTTMANN H.: Robust global registration. In *Proceedings of Eurographics symposium on Geometry processing* (2005).

[HLLS01] HETZEL G., LEIBE B., LEVI P., SCHIELE B.: 3D object recognition from range images using local feature histograms. In *CVPR* (2001).

[IKN98] ITTI L., KOCH C., NIEBUR E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell. 20*, 11 (1998).

[JH99] JOHNSON A. E., HEBERT M.: Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Trans. Pattern Anal. Mach. Intell. 21*, 5 (1999).

[KPNK03] KORTGEN M., PARK G.-J., NOVOTNI M., KLEIN R.: 3d shape matching with 3d shape contexts. In *The 7th Central European Seminar on Computer Graphics* (2003).

[Lin94] LINDEBERG T.: Scale-space theory: A basic tool for analysing structures at different scales. *Journal of Applied Statistics 21*, 2 (1994).

[Low04] LOWE D. G.: Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision 60*, 2 (2004).

[LVJ05] LEE C. H., VARSHNEY A., JACOBS D.: Mesh saliency. In *ACM SIGGRAPH* (2005).

[MBO05] MIAN A. S., BENNAMOUN M., OWENS R.: Automatic correspondence for 3D modeling: An extensive review. *Int. Journal of Shape Modeling (IJSM) 11*, 2 (2005).

[MBO06] MIAN A. S., BENNAMOUN M., OWENS R. A.: A novel representation and feature matching algorithm for automatic pairwise registration of range images. *Int. Journal of Computer Vision 66*, 1 (2006).

[MPS*03] MORTARA M., PATANÉ G., SPAGNUOLO M., FALCIDIENO B., ROSSIGNAC J.: Blowing Bubbles for Multi-Scale Analysis and Decomposition of Triangle Meshes. *Algorithmica 38*, 1 (2003), 227–248.

[MSS*06] MATEI B., SHAN Y., SWHNEY H., TAN Y., KUMAR R., HUBER D., HEBERT M.: Rapid object indexing using locality sensitive hashing and joint 3D-signature space estimation. *IEEE Trans. Pattern Anal. Mach. Intell. 28*, 7 (2006).

[Pet02] PETITJEAN S.: A survey of methods for recovering quadrics in triangle meshes. *ACM Comput. Surv. 34*, 2 (2002).

[PKG03] PAULY M., KEISER R., GROSS M.: Multi-scale feature extraction on point-sampled surfaces. *Computer Graphics Forum 22*, 3 (2003).

[Rab89] RABINER L.: A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proc. of IEEE 77*, 2 (1989).

[SF06] SHILANE P., FUNKHOUSER T.: Selecting distinctive 3D shape descriptors for similarity retrieval. In *SMI* (2006), IEEE Computer Society.

[Smy97] SMYTH P.: Clustering sequences with Hidden Markov Models. In *NIPS* (1997), Mozer M., Jordan M., Petsche T., (Eds.), vol. 9, MIT Press.

[WNK06] WESSEL R., NOVOTNI M., KLEIN R.: Correspondences between salient points on 3D shapes. In *VMV* (2006), Akademische Verlagsgesellschaft.

[ZvKD07] ZHANG H., VAN KAICK O., DYER R.: Spectral methods for mesh processing and analysis. In *Proc. Eurographics State-of-the-art Report* (2007).