Statistical 3D Shape Analysis by Local Generative Descriptors

Umberto Castellani, *Member*, *IEEE*,
Marco Cristani, *Member*, *IEEE*, and
Vittorio Murino, *Senior Member*, *IEEE*

Abstract—In this paper, we propose a new approach for surface representation. Generative models are exploited for encoding the variations of local geometric properties of 3D shapes. Surfaces are locally modeled as a stochastic process which spans a neighborhood area through a set of circular geodesic pathways, captured by a modified version of a Hidden Markov Model (HMM) named multicircular HMM (MC-HMM). The approach proposed consists of two main phases: 1) local geometric feature collection and 2) MC-HMM parameter estimation. The effectiveness of our proposal is demonstrated by several applicative scenarios, all using well-known benchmark data sets, such as multiple view registration, matching of deformable shapes, and object recognition on cluttered scenes. The results achieved are very promising and open up the use of generative models as geometric descriptors in an extensive range of applications.

Index Terms—3D shape analysis, shape representation, Hidden Markov Models, generative modeling.

1 Introduction

SHAPE matching is one of the most critical issues in surface analysis methods, especially in the case of cluttered scenes [13], [11], [1], or when dealing with partial objects [10], or skeletal articulation changes with deformations [21], [5], [20]. Shape descriptors [19] (i.e., *feature*-based techniques) are of paramount importance in such cases, and they are commonly used to tackle these problems. Specifically, description techniques based on statistical reasoning have provided convincing results for both global and local paradigms [11], [20]. The simplest statistical methods are based on the accumulation into a histogram of the geometric properties of shapes, which then becomes the shape descriptor [11]. More advanced techniques are based on the spectral analysis of the shape [5], which captures the *intrinsic* properties of a surface. Such approaches are useful for modeling nonrigid objects, being invariant to different deformations or poses; however, they are likely to fail in the presence of partial views or cluttered scenes. See [20] for an exhaustive survey on shape matching techniques.

In this paper, we propose a statistical description method based on *generative* modeling for point-to-point matching of 3D shapes, represented as meshes. The generative tools employed in our work are Hidden Markov Models (HMM) [17], [7], adapted to deal with 3D data by the combination of geodesic sampling and local geometric feature collection, leading to the *Multicircular Hidden Markov Model* (MC-HMM). The description method proposed is based on two main phases: 1) local geometric feature collection and 2) MC-HMM parameter estimation. In the first phase, the whole 3D object is subsampled by extracting a reduced number of *feature* points [6]. Subsequently, for each feature point, a *neighborhood* set is defined by considering nearby points. Local geodesic pathways are then defined on this neighborhood, and a set of local elementary geometric characteristics [16] is computed for each point of a pathway. In the second phase, for each feature point, a single HMM is trained on the geodesic pathways. This HMM is customized to encode circular sequences defined by the geodesic pathways. The proposed statistical learning approach introduces a novel paradigm for 3D shape analysis. In fact, other methods based on HMMs have been proposed in the past, but only in the 2D domain [9]. In our previous work [6], we introduced some preliminary results on the extension of standard HMMs in the 3D domain. The main technical novelty of this paper is the additional training procedure that enforces the consistency between the pathways and the HMMs. A further contribution is that the descriptor now uses geodesic pathways instead of the spiral pathways used previously, which meant that the method was sensitive to the original mesh sampling. In addition, the experimental evaluation is extended. While [6] only considered the application of point-to-point matching, this paper contains experiments for three shape-matching scenarios and compares the results to the results obtained using previous methods. In the following, the 2-step process for the construction of the MC-HMM descriptor is detailed, and an extensive experimental session is reported.

- *U. Castellani is with the Dipartimento di Informatica, University of Verona, Strada Le Grazie 15, 37134 Verona, Italy. Email: umberto.castellani@univr.it.*
- *M. Cristani and V. Murino are with the Dipartimento di Informatica, University of Verona, Strada Le Grazie 15, 37134 Verona, Italy, and the Istituto Italiano di Tecnologia (IIT), Via Morego 30, 16163 Genova, Italy. Email: {marco.cristani, vittorio.murino}@univr.it.*

Manuscript received 21 Dec. 2009; revised 5 Oct. 2010; accepted 17 Mar. 2011; published online 27 Apr. 2011.
Recommended for acceptance by K. Siddiqi.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2009-12-0841.
Digital Object Identifier no. 10.1109/TPAMI.2011.85.

2 Local Geometric Feature Collection

Local geometric feature collection is a process organized in three consecutive steps: 1) feature point detection, 2) geodesic pathway definition, and 3) collection of local geometric properties.

2.1 Feature Point Detection

During the detection step, a few and significant points on the whole object are selected. The SIFT operator [12] represents the standard method in the 2D domain, allowing the detection of scale invariant feature points. Recently, this approach has been extended to 3D meshes [21], [14], [6] showing its robustness against noise and pose variations.[1] In this work, we employ our previously proposed method [6] although in principle any other detection procedure can be used. In fact, we highlight that in this paper the main focus is on the point description phase.

2.2 Geodesic Pathway Definition

For each feature point v_i of the mesh V, a local neighborhood is defined, and the neighboring surface points are sampled by properly defining several geodesic pathways. In order to control the sampling of the surface, the front propagation procedure [4] is initialized from each of the feature points considered so that different level sets of the geodesic function are collected at different distance steps. The geodesic sampling is then carried out by setting a fixed number of increasing distances $d_1 < \cdots < d_r \cdots < d_K$, each of them defining a level set. In practice, for each level set, an arbitrary finite number N_r of equally spaced surface points is sampled by obtaining the circular sequence $\hat{L}^{d_r} \triangleq \hat{L}^r = [s_1^r, s_2^r, \ldots, s_{N_r}^r]$. Points are sampled in a clockwise direction with respect to the surface normal n_i at the feature point v_i.[2] Finally, the geodesic neighborhood $\mathcal{N}$ of the feature point v_i is defined as $\mathcal{N}(v_i) = \hat{L} = [\hat{L}^1, \ldots \hat{L}^r, \ldots \hat{L}^K]$. In this fashion, a *geodesic pathway* is introduced which allows the neighborhood samples to be observed unambiguously. Moreover, the use of geodesic (i.e., *intrinsic*) pathways makes the local sampling robust to isometric variations. For the sake of clarity, in the following, *ring* will

1. Note that [21] is also isometry-invariant if the detector is used with the Gaussian curvature as scalar field.
2. The starting sample s_1^r is arbitrary.

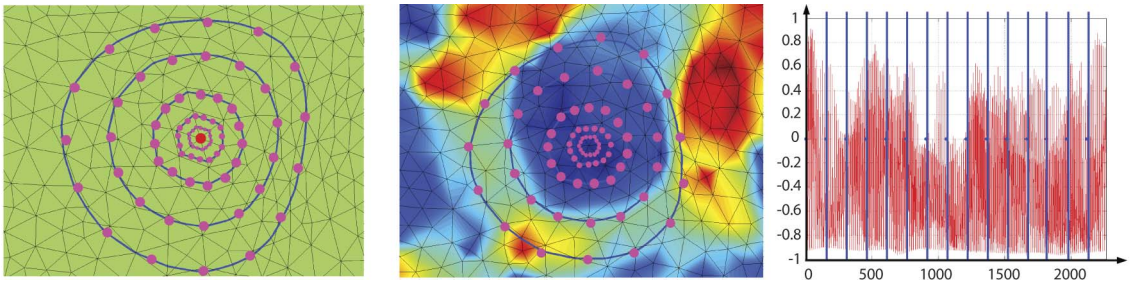

Fig. 1. Feature collection. (Left) Multicircular geodesic pathway: The feature point is highlighted in red, level sets are in blue, and sampled points are pink. (Center) Shape index values observed on the neighborhood of the feature point: Convex regions are colored in red, concave regions are blue. (Right) After the sampling, observations form a multisequence: Features coming from each ring (i.e., each sequence) are concatenated. The jumps between sequences are indicated as vertical blue lines.

indicate the level set. Fig. 1 (left) shows an example of the geodesic pathway around a feature point. Note that new points (in red) are collected on the object surface independently of the original sampling procedure.

2.3 Collection of Local Geometric Properties

To robustly encode the local properties of the sampled area, after the surface sampling step we focus on a set of local surface geometric operators [16]. The following geometric measures are computed: *Shape Index si* [16], *β-Value bv* [11], *Saliency value sv* [6], and the *Normal displacement nd*, i.e., the norm of the difference between the normal n_s of the generic sample and the normal n_i of the feature point. In this way, for each sampled point s_t^r of the geodesic pathway, an n-dimensional ($n = 4$) vector of geometric properties is observed $o_t^r = [si, bv, sv, nd]$.[3] Fig. 1 (center) shows the surface neighborhood of a feature point colored according to the Shape Index. After sampling, the set of feature values sampled according to the geodesic pathway can be interpreted as a multisequence which is then fed to the MC-HMM training procedure (see Fig. 1, right).

3 Multicircular Hidden Markov Models

Each sampled point of the pathway is described by a set of local features. Hence, a geodesic pathway generates an ensemble of feature sets. In order to introduce uncertainty in the interpretation of such observations, local features can be quantized as *hidden states*. Moreover, the spatial ordering introduced by the geodesic pathway can be encoded as a stochastic process which explains how the local area evolves starting from the feature point. These two observations lead us to develop a modified version of the Hidden Markov Model [17], called *MC-HMM. In the MC-HMM, the set of observations* $o_t \in IR^n$ *is organized as*

$$O = \left[o_1^1, \ldots, o_{N_1}^1, \ldots, o_1^r, \ldots, o_{N_r}^r, \ldots, o_1^K, \ldots, o_{N_K}^K\right], \quad (1)$$

where $N_1, \ldots, N_K$ is the number of points sampled at rings $1, \ldots, K$, respectively (as defined by the surface sampling procedure); each ring is closed, therefore $o_t^r = o_{t+N_r}^r$. For the sake of clarity, we use the compact form $O = [O^1, \ldots, O^K]$.

Formally, an MC-HMM is made up of the following entities [17]:

- $S = \{S_1, S_2, \ldots, S_N\}$, the finite set of (hidden) states; in our case each state is associated with a particular local geometric configuration that occurs along the surface. Therefore, a possible state sequence is organized as

  $$Q = \left[q_1^1, \ldots, q_{N_1}^1, \ldots, q_1^r, \ldots, q_{N_r}^r, \ldots, q_1^K, \ldots, q_{N_K}^K\right], \quad (2)$$

  where $q_t^r \in S$ denotes the state occupied by the model at ring r and position t, respectively.

- The transition matrix $A = \{a_{ij}\}, 1 \le i, j \le N$ representing the probability of moving from states S_i to S_j:

  $$a_{ij} = P[q_{t+1}^r = S_j | q_t^r = S_i], \quad 1 \le i, j \le N,$$

  with $a_{ij} \ge 0$, $\sum_{j=1}^N a_{ij} = 1$. This matrix encodes how the different local configurations progress along the geodesic pathway.
- The emission matrix $B = \{b(o|S_i)\}$, indicating the probability of emission of symbol o when the system state is S_i:

  $$b(o|S_i) = \mathcal{N}(o|\mu_i, \Sigma_i),$$

  where $\mathcal{N}(o|\mu, \Sigma)$ denotes a Gaussian probability density function of mean μ and diagonal covariance matrix Σ, evaluated at o. In our approach, this distribution codifies how likely it is that the features on the surface have been generated by a hidden state.

For convenience, we represent an MC-HMM by a couple of parameters $\lambda = (A, B)$. Note that differently from standard HMM [6], [17], start and endpoints in rings coincide, and there is no notion of probability of starting from a well-defined state.

For the training of the MC-HMM, the sequences O^r are considered as i.i.d., therefore, there is no connection between the ring r and the following ring $r + 1$. This fact is modeled without considering a Markovian relation between states $q_{N_r}^r$ and q_1^{r+1}. Indeed, we define a modified Maximum Likelihood training procedure that is slightly different from the standard Baum-Welch algorithm [17], [6]. We need to maximize

$$P(O|\lambda) = \prod_{r=1}^{K} P(O^r|\lambda),$$

where $P(O^r|\lambda)$ is defined as (omitting the index r in the right-hand side of the equation)

$$P(O^r|\lambda) = \sum_{Q} a_{(q_N, q_1)} b_{(q_1)}(O_1) a_{(q_1, q_2)} b_{(q_2)}(O_2), \ldots, a_{(q_{N-1}, q_N)} b_{(q_N)}(O_N), \quad (3)$$

where the ring nature of the sequence O^r is adequately captured by the term $a_{(q_N, q_1)}$.

In the following, the likelihood of observing the multicircular sequence of feature point v_i, given the MC-HMM parameters λ_i, is indicated by $P(v_i|\lambda_i)$. This training procedure is carried out separately for each feature point by obtaining a robust probabilistic point description which is used for matching purposes. Similarly, to obtain a visual feedback of the proposed generative procedure, the Viterbi path [17] $Vt(v_i, \lambda_i)$ is evaluated. Fig. 2 shows the Viterbi paths $Vt(v_i, \lambda_i)$ and $Vt(v_j, \lambda_j)$ of two geodesic pathways associated with two feature points v_i and v_j after training. As expected, the sequence of states is coherent with the variation of local surface configurations. To get a better insight into the nature of hidden

3. Note that other features with different properties can be included in the geometric vector.

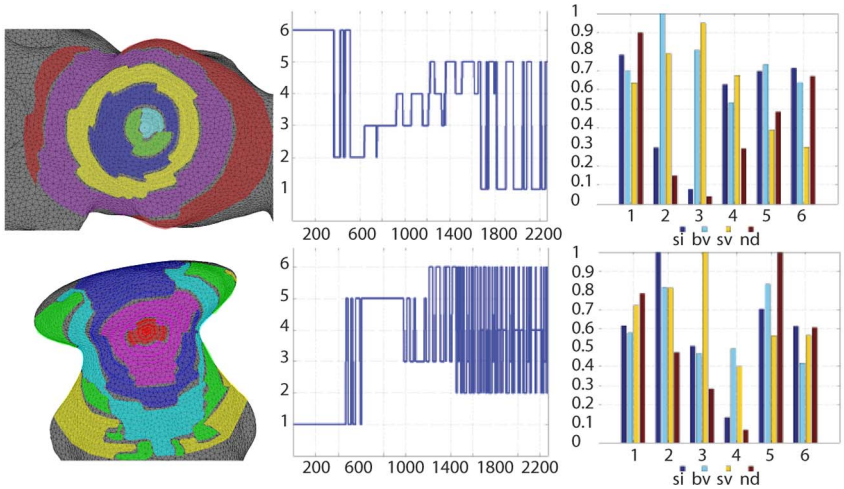

Fig. 2. Training phase: Estimation of MC-HMM parameters for two different points (top and bottom rows). Left: Each surface point is colored according to the Viterbi estimation. Center: The Viterbi path profile. Right: The most likely feature values for each state. The number of states is $N = 6$.

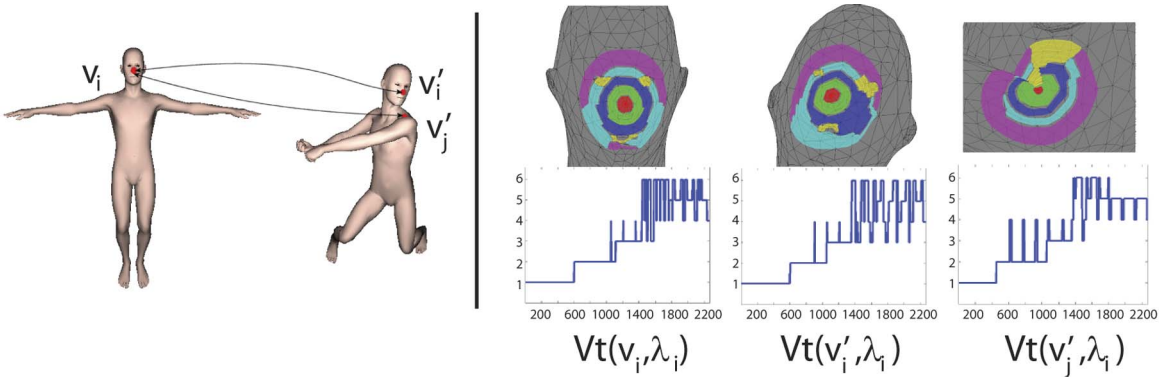

Fig. 3. A matching example. On the left, feature point v_i is compared with the corresponding point v'_i and with another feature point v'_j. On the right, MC-HMM parameters of v_i are evaluated on both v'_i and v'_j: Points are colored according to the Viterbi estimation (top) and the Viterbi paths (bottom).

states, we also plot the most likely feature values for each state for both points (i.e., in practice, the μ value components of each Gaussian for each state). It is interesting to observe that such a set of representative local geometric feature patterns is estimated in an adaptive fashion by automatically defining which local features or which combination of them are most suitable to characterize the local shape at that local area.

After the training phase, the matching of two feature points v_i and v_j is computed by comparing the two MC-HMM signatures λ_i, λ_j in a probabilistic sense [18]. Fig. 3 shows one example of feature points match, v_i and v'_i, and a wrong match, v_i and v'_j. The Viterbi paths $Vt(v_i, \lambda_i)$, $Vt(v'_i, \lambda_i)$, and $Vt(v'_j, \lambda_i)$ are also shown. Also, in this case, the sequence of states for corresponding points is coherent, whereas, in the noncorresponding case, it is not. Note that false positives can occur with our MC-HMM since, in principle, noncorresponding feature points can also generate similar MC-HMM descriptors. For instance, since our generative model is stationary [17], any permutation among the level sets of a geodesic pathway generates the same MC-HMM signature. This could be avoided by nonstationary HMMs, but a more complicated transition procedure between the HMM states should be estimated.

4 Experiments

The approach proposed was applied to different tasks: 1) point-to-point matching on deformable objects, 2) multiple view registration for full 3D model reconstruction, and 3) 3D object detection and recognition on cluttered scenes with occlusions. The free parameters were manually fixed and equal for all the experiments. The sampling parameters were set to $K = 15$ and $N_1 = N_2 = \cdots = N_K = 150$. Each ring r was sampled at distance $(d_r = r \cdot INC)$, where $INC = (d_K/K)$ and d_K was at 5 percent of the main diagonal of the bounding box in which the 3D object was contained.[4] The only free MC-HMM parameter was defined by fixing the number of states $N = 6$. Note that the size of the neighborhood d_K should be estimated adaptively for each feature point. To this aim, local scale properties should be evaluated during the feature point detection phase. However, the employed detector does not provide us this information.

We also compared our approach by exploiting other local shape descriptors, namely, *Spin Images* [11] and the *Exponential Map Descriptor* (EMD) [15]. Specifically, we inserted these descriptors into the proposed applicative scenarios and designed for each of them a proper matching strategy. To make a fair comparison for the Spin Image parameter computation we replaced the euclidean distance, which is likely to fail, especially on deformable scenarios, with the geodesic distance [4], and we called this descriptor the *Geodesic* Spin Image (GSI). The Exponential Map Descriptor was implemented as described in [15]. A local reference system is estimated on the feature point from which the so-called *geodesic polar coordinates* are computed. As in [15], the rotated surface normals were encoded in each sampled point. Again, for purposes of comparison, since the detector is unable to deal with variable neighborhood size selection, we fixed the same scale for all the points. Therefore,

4. The choice of d_K is a trade-off between the discriminativeness of the descriptors and their robustness against occlusions. We choose a small support in order to deal with strong occlusions, i.e., we follow a conservative approach.

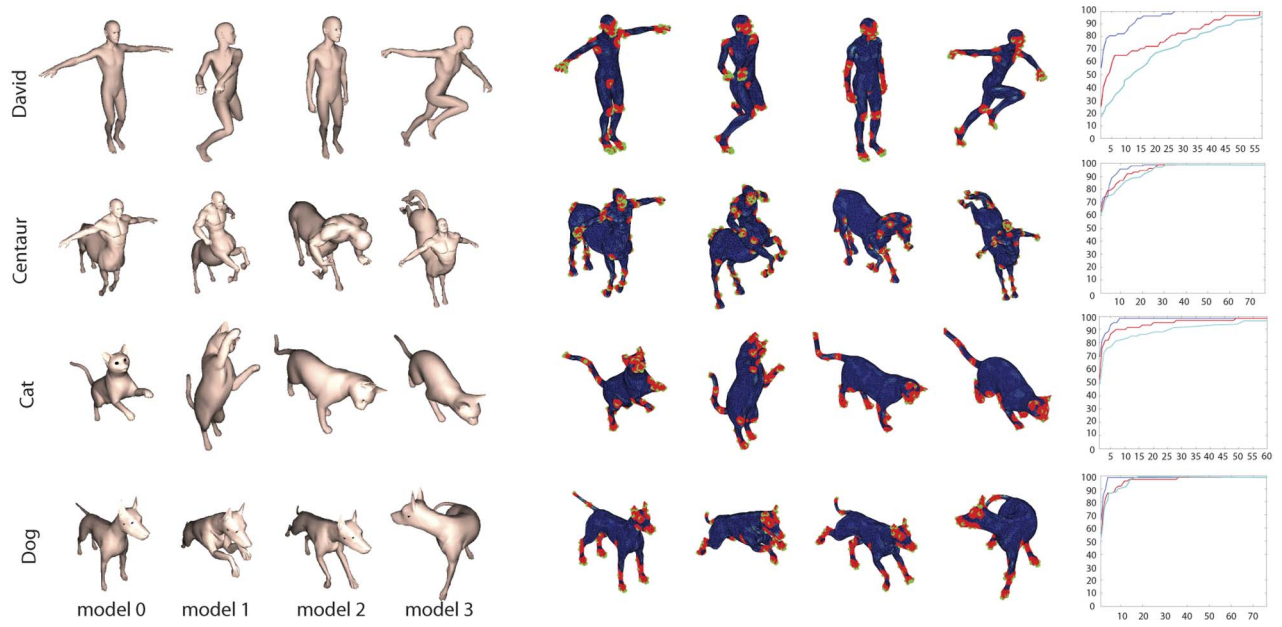

Fig. 4. Experiments on the TOSCA data set. On the left, four models: *David*, *Centaur*, *Cat*, and *Dog*. In the middle, the detection results. For each feature point (in green), the neighborhood area is highlighted in red. On the right, matching accuracies for the four analyzed objects. Blue lines correspond to our method, the red and cyan lines correspond to results obtained with GSI and EMD, respectively.

the local support of both GSI and EMD is the same as the MC-HMM descriptor which is determined by d_K.

4.1 Point-to-Point Matching on Deformable Objects

The first experiment was carried out on nonrigid objects available from the TOSCA data set [4], [3]. Objects are organized as regular triangular meshes. Each object contains approximately 3,000 vertices. The data set is composed of articulated objects in different poses and deformations. Four objects were selected (Fig. 4 (left)), namely, *David*, *Centaur*, *Cat*, and *Dog*. Each object appears in four different instances for which the ground truth point-to-point correspondences are available.

Shape matching was carried out by considering pairs of models of the same object. For each object, *Model 0* was considered as *reference*, whereas *Models 1, 2, and 3* were *test* models, respectively. According to the proposed pipeline, feature points were selected from both reference and test models and the MC-HMM training procedure was computed to estimate the MC-HMM descriptors. Note that in this experiment, the extraction of feature points was performed only on the reference model. Then, corresponding feature points on the test model were recovered from the ground truth. In this fashion, since we focused on the evaluation of the descriptors, we reduced the conditioning of the detection phase that can cause missdetections, false positives, or wrongly located points. Note also that, for the detector used, the isometry-invariant property is not theoretically demonstrated. A few feature points were extracted from each object: 57 for *David*, 75 for *Centaur*, 60 for *Cat*, and 75 for *Dog*. Therefore, each feature point of the reference model was compared with all feature points of the test model. Hence, *precision*[5] [19] was computed from 1 to M (where M is the number of feature points). Fig. 4 (center) shows the feature points in green, with the associated neighborhood areas shown in red. The main anatomical parts of the subjects, such as eyes, nose, fingers, knees and so on, were extracted. Note that reference and test models are very different: these modifications are mainly caused by the pose variations of the articulated parts of the objects. Moreover, some models appear with strong deformations on local surfaces (for example, see the tails and the leg articulations of the dog, which are stretched and shrunk), changing the local geometric properties. In spite of these variations, the point-to-point matching is quite satisfactory, thus showing the capability of the MC-HMM in managing such kinds of deformation *noise*. Note that since we employ Gaussian-HMM, the robustness to deformation is derived from the assumption that small deformations can behave as Gaussian noise. Fig. 4 (right) shows the matching accuracy for the four analyzed objects. In each graph, the x-axis represents the number of feature points, which is the same for the four models. The y-axis represents the precision by taking into account the four models (i.e., one reference and three test models). The blue lines refer to the results of our method, the red lines correspond to those obtained with Geodesic Spin Images, and the cyan lines are the results of the Exponential Maps. In general, more than 50 percent of correct matchings were observed at the first trial and more than 80 percent success was achieved within the first 15 trials. Note that error in early trials is justified by the presence of several symmetries and similar subparts. It is also worth noting that in all of the experiments, our method outperformed both GSI and EMD descriptors. Our method was quite stable with all the objects. Moreover, our method reached the full correct matching much faster than the GSI and EMD techniques.

As a further experiment, we evaluated the robustness of matching against noise. Four versions of *David0* with Gaussian noise were defined as test models, with $\sigma = \{0.25, 0.5, 0.75, 1\}$ (Fig. 5, top). Fig. 5 (bottom) shows matching results. Our matching was quite stable for the first two levels. Note that, from $\sigma = 0.75$, many feature points lost meaning, for instance, see the fingers that merge with each other. Also in this case, our method outperformed both GSI and EMD. In particular, EMD was also very sensitive with low noise levels. These experiments show that although, in general, the local geometric features of the MC-HMM are quite sensitive to deformation and Gaussian noise, the intermediation of the MC-HMM model makes our matching robust.

5. Precision P means that the correct matching is estimated within the Pth tentative according to the likelihood ordering. Note that wrong correspondences will provide the same error independently by the fact that the computed corresponding point is close or far away from the correct one.

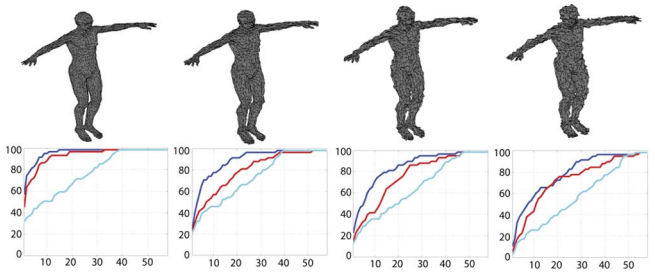

Fig. 5. Matching with noisy models. Top: Four versions with Gaussian noise of *David0*. Bottom: Performance evaluation. Blue lines correspond to the proposed method, whereas the red and cyan lines correspond to results obtained with GSI and EMD, respectively.

4.2 Partial View Registration

In the second experiment, we tested the effectiveness of the proposed MC-HMM descriptor in addressing the problem of the partial view registration [10]. Given a set of partially overlapping 3D views of a single object, the goal of the partial view registration is the estimation of the set of rigid transformations which brings each partial view into a common coordinate system. In this context, our MC-HMM matching method was used to estimate a robust pre-alignment which is independent of the starting pose and reliable in the case of small overlap. More specifically, we designed a simple pairwise registration pipeline as follows: 1) extraction of feature points from each partial view and estimation of MC-HMM descriptors, 2) bipartite graph matching [7] to estimate a starting set of candidate correspondences, and 3) outlier rejection by RANSAC algorithm [8], which introduces the rigid constraint in the alignment (note that we have tuned the RANSAC parameters to allow the three methods to perform well as possible). Finally, each view is moved to the global reference system by simple concatenation of the estimated pairwise motion matrices. We tested eight objects obtained from the Stuttgart Range Image Database.[6] Each object is observed from 16 contiguous views. Note that ICP [2] registration failed on these view pairs due to the large distance between the views at the starting pose and the small overlap. Fig. 6a shows a range image for each object. Objects are characterized by well-structured areas, the presence of wide planar surfaces and symmetries. Moreover, the scanning acquisition phase caused several holes,[7] noise and occlusions. Fig. 6b shows the results obtained with our pre-alignment procedure, where each view appears with a different color. In only a few cases, namely, *Porsche* and *Mole*, a small misalignment can be observed which is justified by the lack of reliable features on these view pairs. To evaluate the accuracy of the proposed approach, we counted the number of *inliers*[8] after the correct rigid alignment between each view pair. In fact, this gave an estimation of the correct matches that were computed with our descriptor. Note that for each view, only a few feature points were extracted (the value of M ranged from 13 to 75). Fig. 6c shows the graphs of inliers for each object. The total number of candidate matchings is also plotted (in green), i.e., the minimum between the number of feature points extracted from the first view, and the number of feature points extracted from the second view. Note that since the views partially overlap, the number of candidate matchings is, in general, much larger than the number of true correspondences (which are unknown). In general, our method shows a higher number of inliers[9] and never causes view-divergence. On the contrary, both GSI and EMD fail in the registration on several pairs, for example *Porsche*, *Mole*, and *Female*.

6. University of Stuttgart, http://range.informatik.uni-stuttgart.de/htdocs/html/.

7. Note that we adapted the standard Fast Marching algorithm [4] to work on triangular meshes with holes.

8. According to the RANSAC algorithm, an *inlier* is a correspondence which has voted for the chosen rigid transformation.

9. This allows RANSAC to reach the convergence with less iterations.

4.3 Object Recognition in Cluttered Scenes

In the third experiment, we focused on 3D object recognition in complex scenes characterized by the presence of multiple objects. In general, in an object recognition problem, the complete 3D models of objects appearing on a scene are available in a database. From this database, several pieces of information can be extracted, for example, the feature points and the related descriptors (representing the *training* phase). Here, a *recognition-by-fitting* [7] approach was applied by simultaneously matching feature points of trained objects with feature points extracted from a test scene (the *testing* phase). The rigid constraint was introduced, as was the case for the partial view registration, by combining bipartite graph matching with robust outlier rejection. The object with the highest number of inliers (i.e., the best fit) was declared recognized and was removed from the scene. The process was repeated until the scene was completely segmented. Experiments were carried out on the UWA Data Set[10] [13]. The database is composed of five objects, namely, *Chef*, *Parasaurolophus*, *T-rex*, *Chicken*, and *Rhino*. These objects are randomly combined in different poses, generating 50 cluttered scenes. Each scene is composed of four or five objects, as shown in Fig. 7. Objects are only partially visible in a scene with strong occlusions and noise. Moreover, the objects in the scene are in different poses with respect to the scanner and with locally different sampling rates. Therefore, recognition was challenging since the number of matches of correct points was very few with respect to the total number of possible point correspondences. Moreover, the clutter caused the local neighborhood of one point of an object to overlap with local portions of another object. Fig. 7 shows some results of object recognition and pose estimation by superimposing each recognized object onto the scene for each test case. We evaluated the performance of our method according to [13], by computing the recognition rate as a function of the occlusion for all 50 scenes. In addition to comparison with GSI and EMD, we compared our method with the *Scale-Hierarchical* (SH) method proposed by Bariya and Nishino [1] and with the *Tensor Matching* (TM) method proposed by Mian et al. [13] which represent the state of the art on this data set (Fig. 7, right).[11] Note that both SH and TM were purposely designed to work with cluttered scenes by exploiting the scale estimation of feature points [1] and by using the interrelations between points [13]. Moreover, the matching procedure is more effective. Although these properties were not exploited in the MC-HMM descriptor, our method achieved performance comparable to SH and TM. The overall recognition rate of MC-HMM with up to 84 percent occlusion was 92.44 percent, while SH and TM performed 97.57 and 96.6 percent, respectively. Furthermore, our method outperformed both GSI and EMD, which provided 86.63 and 84.30 percent accuracy, respectively. Note that the comparison with GSI and EMD was carried out starting from the same feature points and adopting the same matching procedure. Note, in particular, that the EMD descriptor was the same adopted in [1] with a fixed scale for all points. This highlights the fact that, as expected, the higher performances of the SH and TM methods are mainly due to the use of scale information and to better matching strategies rather than to the nature of the chosen descriptors. Furthermore, note that, in our method, more failures were observed on less-structured objects, like Parasaurolophus and Chicken, where local information was not sufficient to recognize such objects, especially at the largest occlusion levels. Conversely, on more structured objects, our method performed better. In particular, our method can correctly detect the Chef model in all of the scenes (with more than 90 percent occlusion).

10. University of Western Australia, http://www.csse.uwa.edu.au/ajmal/recognition.html.

11. As suggested in [1], we excluded Rhino in the performance evaluation for comparison purposes.

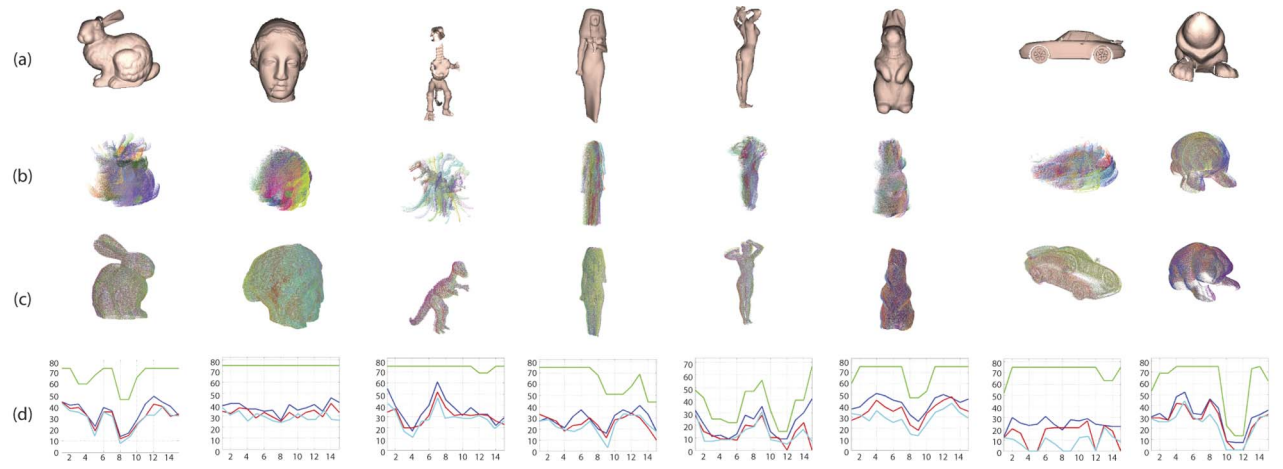

Fig. 6. Partial view registration on the Stuttgart Data Set. (a) From left to right, the original *Bunny*, *Igea*, *Dino*, *Isis*, *Female*, *Hasi*, *Porsche*, and *Mole* models. (b) All views at the starting pose. (c) Registered views: Each view appears as a cloud of points with a different color. (d) Accuracy of partial view registration: number of inliers obtained using MC-HMM descriptors (blue), GSI (red), and EMD (cyan). Green lines show the total candidate matching number.

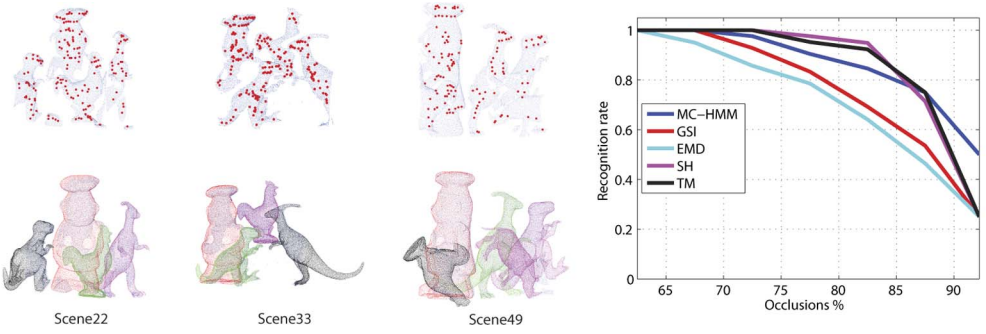

Fig. 7. Object recognition on cluttered scenes. On the left, a selection of scenes is portrayed, enumerated according to [13]. Each test scene is shown before (top) and after (bottom) the recognition of the objects. Test scenes are shown in blue with the extracted feature points highlighted in red. Recognized objects are superimposed on the scene with different colors. On the right, recognition rates with respect to occlusion percentage.

5 Conclusions

In this paper, we propose a generative model, called MC-HMM, as a descriptor for encoding the statistical properties of local 3D surfaces. We show how the MC-HMM model can adaptively merge different local geometric properties into a unified descriptor. Only a few *feature* points are considered for each object, which is standard for 2D images but less known and used in the domain of 3D shape analysis. The approach presented is versatile and effective in the treatment of deformable shapes, strong occluded objects, and cluttered scenes. We compared our descriptor with Geodesic Spin Image and Exponential Map Descriptor in a number of tasks, showing that MC-HMM outperforms them in all the experiments. Moreover, it performs comparably with both *Scale-Hierarchical* and *Tensor Matching* methods on cluttered scenes despite the fact that we adopted a simpler matching strategy in comparison to the more effective procedures adopted by them. We expect an improvement of our descriptor if scale properties of the feature point will be exploited to obtain an adaptive estimation of the local support.

References

[1] P. Bariya and K. Nishino, "Scale-Hierarchical 3d Object Recognition in Cluttered Scenes," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition*, pp. 1657-1664, 2010.

[2] P. Besl and H. McKay, "A Method for Registration of 3-D Shapes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239-256, Feb. 1992.

[3] A.M. Bronstein, M.M. Bronstein, and R. Kimmel, "Efficient Computation of Isometry-Invariant Distances Between Surfaces," *SIAM J. Scientific Computing*, vol. 28, no. 5, pp. 1812-1836, 2006.

[4] A.M. Bronstein, M.M. Bronstein, and R. Kimmel, *Numerical Geometry of Non-Rigid Shapes*. Springer Verlag, 2007.

[5] M.M. Bronstein and A.M. Bronstein, "Shape Recognition with Spectral Distances," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 1065-1071, May 2011.

[6] U. Castellani, M. Cristani, S. Fantoni, and V. Murino, "Sparse Points Matching by Combining 3D Mesh Saliency with Statistical Descriptors," *Computer Graphics Forum*, vol. 27, pp. 643-652, 2008.

[7] R. Duda, P. Hart, and D. Stork, *Pattern Classification*. Wiley, 2001.

[8] M. Fischler, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Comm. ACM*, vol. 24, no. 6, pp. 381-395, 1981.

[9] R. Huang, V. Pavlovic, and D.N. Metax, "Embedded Profile Hidden Markov Models for Shape Analysis," *Proc. IEEE 11th Int'l Conf. Computer Vision*, pp. 1-8, 2007.

[10] D. Huber and M. Hebert, "Fully Automatic Registration of Multiple 3D Data Sets," *Image and Vision Computing*, vol. 21, no. 7, pp. 637-650, 2003.

[11] A. Johnson and M. Hebert, "Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 433-449, May 1999.

[12] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int'l J. Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.

[13] A. Mian, M. Bennamoun, and R. Owens, "A Novel Representation and Feature Matching Algorithm for Automatic Pairwise Registration of Range Images," *Int'l J. Computer Vision*, vol. 66, no. 1, pp. 19-40, 2006.

[14] J. Novatnack and K. Nishino, "Scale-Dependent 3D Geometric Features," *Int'l Conf. Computer Vision*, 2007.

[15] J. Novatnack and K. Nishino, "Scale-Dependent/Invariant Local 3D Shape Descriptors for Fully Automatic Registration of Multiple Sets of Range Images," *Proc. 10th European Conf. Computer Vision: Part III*, 2008.

[16] S. Petitjean, "A Survey of Methods for Recovering Quadrics in Triangle Meshes," *ACM Computing Surveys*, vol. 34, no. 2, pp. 211-262, 2002.

[17] L. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257-286, Feb. 1989.

[18] P. Smyth, "Clustering Sequences with Hidden Markov Models," *Advances in Neural Information Processing Systems*, pp. 648-654, 1997.

[19] J.W. Tangelder and R.C. Veltkamp, "A Survey of Content Based 3D Shape Retrieval Methods," *Proc. Int'l Conf. Shape Modeling and Applications*, 2004.

[20] O. van Kaick, H. Zhang, G. Hamarneh, and D. Cohen-Or, "A Survey on Shape Correspondence," *Proc. Conf. EuroGraphics: State-of-the-Art Report*, pp. 1-23, 2010.

[21] A. Zaharescu, E. Boyer, K. Varanasi, and R. Horaud, "Surface Feature Detection and Description with Applications to Mesh Matching," *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, pp. 373-380, 2009.