Investigating Hidden Markov Models' Capabilities in 2D Shape Classification

Manuele Bicego, *Student Member*, *IEEE*, and
Vittorio Murino, *Senior Member*, *IEEE*

Abstract—In this paper, Hidden Markov Models (HMMs) are investigated for the purpose of classifying planar shapes represented by their curvature coefficients. In the training phase, special attention is devoted to the initialization and model selection issues, which make the learning phase particularly effective. The results of tests on different data sets show that the proposed system is able to accurately classify objects that were translated, rotated, occluded, or deformed by shearing, also in the presence of noise.

Index Terms—Hidden Markov Models, 2D shape classification, model selection, probabilistic learning.

1 Introduction

Object recognition, shape analysis, and classification constitute important research areas in computer vision. Three-dimensional (3D) object recognition has been faced by a large number of approaches [1], many of which are based on the analysis of two-dimensional (2D) aspects of objects. In this context, a basic issue is surely the kind of object description to be used. In fact, shape analysis methods can be classified considering this issue in terms of boundary (or external) or global (or internal) algorithms. Typical examples of the former class are constituted by methods coding the object boundary like, for example, Fourier descriptors and chain code, whereas examples of the latter class are algorithms based on the medial axis extraction, or moment-based approaches [2]. In particular, object contours proved to be very effective in many applications, and different types of approaches have been proposed in the past years, each with different characteristics like robustness to noise and occlusions, invariance to translation, rotation and scale, computational requirements, and accuracy [2], [3].

In this context, this paper aims at investigating the capabilities of Hidden Markov Models (HMMs) [4] for 2D shape classification. Shapes are represented by contours and described by their curvature coefficients along the boundary [1]. Hidden Markov Models are a widespread approach to probabilistic sequence modeling and have been extensively employed in several applications in the last decade [4], [5], [6], [7], [8]. The use of HMMs for shape analysis has not been widely addressed. To the best of our knowledge, only a few papers have been found to exhibit some similarities to our approach [9], [10], [11], [12]. He and Kundu [9] were the first to employ HMMs for 2D shape classification. In their approach, contours were represented by autoregressive coefficients computed on segments extracted from the shape boundary. Results were quite interesting and presented as a function of the number of HMM states ranging from 2 to 6, using both stationary and nonstationary models. In [10], the 2D shape classification task was addressed by use of circular HMMs: This particular HMM topology allows one to achieve good classification accuracy despite scaling and deformations, and also presents useful characteristics for model training and testing. However, in both works, there is not an explicit and quantitative analysis of the noise or affine object transformations performances. Moreover, even if sensitivity to small occlusions is analyzed, shapes are always constrained to be closed contours. Another research study [11] addressed shape recognition by comparing HMMs with a syntactic modeling technique based on stochastic context-free grammars, but no original solutions were proposed for HMM design. Recently, an interesting approach was described in [12], in which Fourier spectral features were used to classify 2D shapes. A particular HMM topology was introduced in order to directly deal with these features but, also in this case, shapes were constrained to be closed and occluded and noisy views were not explicitly analyzed.

In this paper, we investigate the capabilities of HMMs in recognizing planar objects, showing their performances in the cases of translation, rotation, noise, occlusions, shearing transformations, and a combination of the above perturbations. It is worth noting that our approach does not rely on any specific HMM topology or a particular training algorithm; moreover, object shapes are *not* constrained to be closed, or represented by using a given number of symbols: our algorithm classifies any (closed or open) symbol string. The focus is on assessing the HMM potentialities in the case of shape analysis, not on presenting an actual object recognition system; for this reason, the segmentation issue is not considered in the paper. Special attention was instead devoted to the training of the HMMs, in particular, to the initialization of the learning session by using a Gaussian Mixture Model clustering approach. The initialization issue is crucial to the learning because of the local behavior of the standard procedure used to estimate the HMM parameters. Another practical but fundamental issue to be resolved when using HMMs is the determination of their structure, namely, the topology and the number of states. The choice of a good model structure is basic to the effectiveness of the learning process, but unfortunately, this issue is usually disregarded in the HMM literature. In our approach, a fast and reasonable model selection technique is applied in the initialization phase, following a Bayesian Inference Criterion (BIC) [13] approach.

The proposed approach is tested using three different databases in order to assess the robustness of the method to different object transformations such as translation, rotation, occlusion, noise, shearing, and combined perturbations. The resulting high performances make the presented method an interesting alternative to typical shape classification algorithms [2], [3].

The paper is organized as follows: A brief description of the HMM is provided in Section 2. The global description of the strategy used is presented in Section 3. Section 4 reports on experimental procedures and results. Finally, conclusions are drawn and future developments are envisaged in Section 5.

2 Hidden Markov Model

A discrete-time HMM is a probabilistic model that describes a random sequence $\mathbf{O} = O_1, O_2, \ldots, O_T$ as the indirect observation of an underlying (hidden) random sequence $\mathbf{Q} = Q_1, Q_2, \ldots, Q_T$, where this hidden process is Markovian, even though the observed process may not be so. For lack of space, HMMs are not exhaustively treated in this paper; we refer the reader to [4] for more details.

In summary, a discrete-time HMM λ is defined by the following elements [4]: a set $S = \{S_1, S_2, \cdots, S_k\}$ of (hidden) states; a transition matrix $\mathbf{A} = \{a_{ij}\}$, where $a_{ij} \geq 0$ represents the probability of going from state S_i to state S_j; an emission matrix $\mathbf{B} = \{b(o|S_j)\}$, indicating the probability of emission of symbol o from state S_j; an initial state probability distribution $\pi = \{\pi_i\}$, representing the probability of the first state $\pi_i = P[Q_1 = S_i]$.

• *The authors are with the Dipartimento di Informatica, Università di Verona, Ca' Vignal 2, Strada Le Grazie 15, 37134 Verona, Italy. E-mail: bicego@sci.univr.it and vittorio.murino@univr.it.*

Manuscript received 2 May 2002; revised 25 Mar. 2003; accepted 6 Aug. 2003.
Recommended for acceptance by A. Kundu.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 116470.

Given a sequence $\mathbf{O}$, there exists a well-established procedure able to determine the HMM parameters maximizing the probability $P(\mathbf{O}|\lambda)$. This technique, called the *Baum-Welch reestimation* procedure [14], is an instance of the well-known *Expectation-Maximization* (EM) algorithm [15] for Maximum Likelihood (ML) estimation. For a sequence O and an HMM λ, there is another standard recursive procedure able to compute the probability $P(\mathbf{O}|\lambda)$, and is called the *forward-backward* procedure [16].

In the classification stage, given an unknown sequence $\mathbf{O}$, the probability $P(\mathbf{O}|\lambda_i)$ is computed for each model λ_i, and $\mathbf{O}$ is classified as belonging to the class whose model shows the highest likelihood $P(\mathbf{O}|\lambda_i)$.

3 The Strategy

In this section, the proposed strategy is explained. After describing the object representation step, we detail the training and the classification procedures, focusing on the initialization and model selection issues.

3.1 Object Representation

Starting from a generic boundary point, each object is represented by the sequence of the curvature coefficients [1] computed at the contour points. This technique exhibits several attractive properties: First, the representation is invariant to object translation; second, object rotation is equivalent to a phase shift in the curvature signal. The scalar curvature value computed at each boundary point is rotationally invariant, but the object rotation implies a change in the initial point, so the curvature signal generally turns out to be shifted. The third and most important point lies in the fact that the curvature values can be computed for open contours, thus allowing one to deal with object occlusions. However, the main drawback of this method is sensitivity to contour noise, like any derivative operation (the derivation enhances the noise contribution). One possible solution is to apply a wide (i.e., with a large variance) Gaussian filter to the (X, Y) contour coordinates before curvature computation, reducing the noise impact on the signal estimation. Moreover, the use of Hidden Markov Models makes it possible to recover quite well from certain noisy situations, as can be seen in the following sections.

In our approach, the curvature is computed as follows: First, the contours are extracted by using the *Canny* edge detector (which implies a preliminary Gaussian filtering); the boundary is then approximated by segments of approximately fixed length d_L. Finally, the curvature value at point x is computed as the angle between the two consecutive segments intersecting at x. For a not occluded object, the initial point is the rightmost point lying on the horizontal line passing through the object centroid, following the boundary in a counterclockwise manner. If the object is occluded, the endpoint allowing the contour to be followed in an counterclockwise way is considered as the initial point.

3.2 Training

The obtained curvature representation is then used to train a continuous HMM, where the emission probability of each state is represented by a one-dimensional Gaussian function. Training is performed using the standard Baum-Welch reestimation method, which is stopped at likelihood convergence. As described in greater detail in the following, each HMM is carefully initialized and the number of states is roughly estimated by using a BIC (Bayesian Inference Criterion) approach linked to the initialization step. At the end of the training phase, we have one HMM λ_i for each object $\mathbf{O}_i$.

3.2.1 Initialization of a Hidden Markov Model

The standard algorithm used to estimate the HMM parameters (i.e., Baum Welch reestimation), starting from some initial estimate, converges to the nearest local maximum of the likelihood function. Therefore, the initialization process heavily affects the resulting model estimate, as the likelihood function is highly multimodal.

In this paper, a Gaussian Mixture Model (GMM) [17] clustering is used to initialize the emission matrix of the HMM before training. In greater detail, the initialization phase first considers the sequence of curvature coefficients as a set of scalar values (no matter in which order the coefficients appear); these values are then grouped into k clusters by following a GMM clustering approach, i.e., fitting the data by using k Gaussian distributions. The parameter k is the number of states of the HMM, and the Gaussian parameters are estimated by an EM-based [15], [18] method. Finally, the mean and variance of each cluster are used to initialize the Gaussian of each state, with a direct correspondence between clusters and states.

3.2.2 Model Selection

A practical but fundamental issue to be resolved when using an HMM is the determination of its structure, namely, the topology and the number of states. Although some special-purpose approaches have been proposed (e.g., [19], [20], [21], [22]), the typical solution is the use of some heuristic or general-purpose model selection methods, not adequately tuned for HMMs. The latter class of methods performs several training phases, choosing the HMM that maximizes a predefined criterion. In order to reduce the computational load, in our approach, the model selection issue is addressed in the initialization phase. In particular, the choice of the model is determined by a model selection analysis of the GMM clustering, i.e., the mixture that best fits the data is chosen. The number of states of the HMM is therefore set as the number of Gaussians of the best mixture: as a result, only one HMM training session is needed. Let us call this approach BOI (BIC On Initialization). It is worth noting that this model selection scheme determines the model that best fits the *unrolled* sequence: in this respect, this is a coarse model selection scheme as only the curvature values are considered and not the order in which they appear. Nevertheless, this is quite a reasonable assumption, which considers a shape as being made up of approximating segments with nearly similar curvatures, each group of segments being assigned to a single state. The dynamics of the sequence, i.e., the way in which these segments are ordered, is thus encoded into the transition matrix. To choose the GMM model that best fits the data, the Bayesian Inference Criterion (BIC) [13] is adopted, which is defined as:

$$BIC(M_k) = \log P(X|\hat{M}_k) - \frac{1}{2}|\hat{M}_k|\log(N), \quad (1)$$

where X is the data set (of cardinality N) to be modeled, $\{M_k\}$ $(k_{min} \leq k \leq k_{max})$ are the candidate models, $\hat{M}_k$ is the Maximum Likelihood estimate of the model M_k, and $|\hat{M}_k|$ is the number of free parameters of the model M_k. The strategy selects the model M_k for which the BIC criterion (1) is maximized.

To sum up, in our approach, the structure of the HMM is determined as follows: For $k_{min} \leq k \leq k_{max}$, the values of the unrolled sequences are clustered by different GMM clusterings $\mathcal{G}_k$, each with k Gaussians, evaluating the BIC criterion $\mathrm{BIC}(\mathcal{G}_k)$ for each modeling. Then, the $\mathcal{G}_{\hat{k}}$ maximizing the criterion is chosen and used to determine the number of HMM states ($\hat{k}$) and to initialize it. The training process is then carried out by using the standard Baum-Welch procedure.

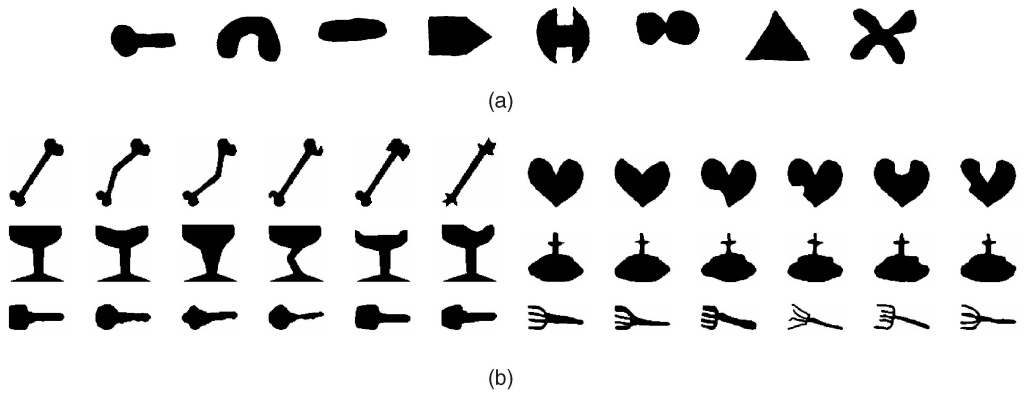

Fig. 1. Shapes used for the testing procedure: (a) the first set and (b) some shapes of the second data set.

3.3 Classification

Given an unknown sequence $\mathbf{O}$, the classification process is a standard Maximum Likelihood rule that computes, for each model λ_i, the probability $P(\mathbf{O}|\lambda_i)$. Then, the object $\mathbf{O}$ is assigned to the class C_ℓ whose model shows the highest likelihood, i.e., $\ell = \arg\max_i P(\mathbf{O}|\lambda_i)$.

4 Results and Discussion

The proposed method was tested by using three sets of shapes. The first data set, for which extensive testing was carried out, was employed by He and Kundu in [9], and is shown in Fig. 1a. Subsequently, the system was validated by using two other test sets: part of the object set used by Sebastian et al. in [23] (six classes, each containing 12 object instances—some examples are given in Fig. 1b) and the set composed of the 21 letters of the Italian alphabet (Helvetica font). One HMM for each object was trained, following the strategy proposed in Section 3. It is worth noting that each HMM is trained using the *only* object model present in the data set (Fig. 1), so that the following results were obtained by training a single shape. Invariance to rotation, occlusion, noise, shearing, and a combination of these transformations were tested, whereas invariance to translation was automatically managed by the curvature representation. Invariance to scale could be addressed by using the normalized curvature signal, which results in an oversampled or subsampled signal. HMMs proved to be robust also to this kind of degradation, as shown in [24]. Some examples of perturbations are presented in Fig. 2; the system was able to perfectly recognize the resulting deformed shapes.

4.1 Rotation

First of all, note that the rotation of an object causes only a shift in the curvature signal as, even if the curvature is rotationally invariant at each boundary point, the sequence of coefficients depends on the starting point which, in general, changes when the object is rotated. To test the invariance of our method, each object was rotated 10 times by an angle randomly chosen from 0 to 2π. The resulting classification accuracy was 100 percent, i.e., the HMM was able to exactly recognize the rotated objects.

4.2 Occlusion

An object is occluded by considering only a fragment of the object boundary, by removing an object part starting from a point randomly chosen (see Fig. 2c). It should be noted that the random choice of the initial point is important to assess the robustness of the method to this type of deformation. Given an open part of the contour, i.e., a fragment of the original one, the curvature is calculated and results in a string that is actually a substring of the original curvature signal. The trial was performed occluding each object 100 times, per level of occlusion, so a large number of possible object occlusions was considered and statistically significant results were obtained. The occluded part differed each time, starting from a randomly chosen initial point, and the degree of occlusion varied from 5 percent to 50 percent (only one half of the whole object was visible). The obtained accuracies were considerably high: even when 35 percent of each object was occluded, our technique was able to correctly classify all the fragments, and with only an half of the object visible the performances decreased to 95 percent (see Table 1a). These results are particularly interesting, considering that occlusion is one of the most severe problems in object recognition.

4.3 Noise

We tried to assess the robustness of our approach in noisy situations. To this end, two synthetic noise schemes are proposed. First, a Gaussian noise with zero mean and variance σ^2 ranging from 1 to 5, is added to the (X, Y) coordinates of an object. Shapes are not much affected by this kind of noise, and the resulting accuracy is 100 percent, especially owing to the Gaussian filter applied before calculating the curvature, as it is able to remove completely the effects of this kind of noise. The second type of noise scheme is adopted to degrade the object shapes more heavily

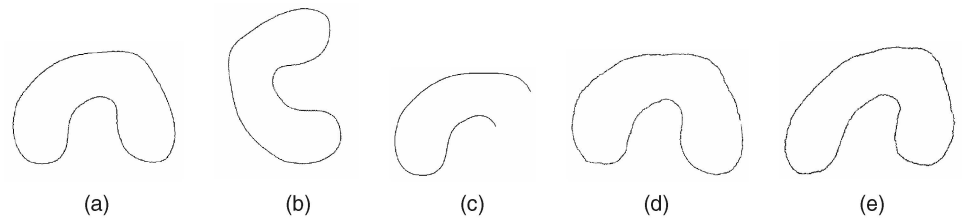

Fig. 2. Some examples of perturbations of the boundary of a shape; the system's recognition of these resulting shapes was perfect: (a) original, (b) rotated ($90°$), (c) occluded (35 percent), (d) noisy ($\sigma_d^2 = 0.35$), and (e) sheared shape ($\tau = 50°, \phi = 40°$).

TABLE 1
Classification Accuracies Obtained in: (a) Occlusion Experiments at Different Occlusion Levels and (b) Noisy Experiments at Different Noise Levels

Occlusion	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
Accuracy	100%	100%	100%	100%	100%	100%	100%	97.5%	96.2%	95%

(a)

σ_d^2	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85
Accuracy	100%	100%	100%	100%	97.5%	91.3%	83.8%	80%	71.3%

(b)

(see Fig. 2d). It is obtained by adding Gaussian noise to the *differential* signal, which results from computing, at each boundary point, the difference between the coordinates of the point and those of the following one. Subsequently, a zero-mean Gaussian noise is added to this difference-code; finally, the coordinates' values are recomputed starting from the prestored initial point. The test set was obtained by adding noise to each object 10 times, and the resulting accuracy values are presented in Table 1b, using the noise level (variance σ_d^2) as the varying parameter. As one can notice, the results are quite satisfactory, showing that HMMs can reduce the intrinsic curvature sensitivity to noise.

4.4 Combined Transformations

After assessing the robustness of the proposed method to single degradations, some experiments were carried out to evaluate the algorithm performances in the cases of combined transformations, i.e., 1) rotation and occlusion and 2) rotation, occlusion, and added noise. Occluded and rotated objects were obtained by rotating the objects by a random angle (between 0 and 2π) and considering fragments of their contours. From the results presented in Table 2a, it can be noticed that the accuracies are very high also in this case. A more difficult situation occurred when objects were first rotated by a random angle, then occluded, and finally degraded by the second type of noise described in Section 4.3. Results are given in Table 2b. Also, in this case, the accuracy values are satisfactory, but rapidly decreasing with increasing noise level.

4.5 Shearing

Finally, the robustness of our approach to shearing transformations was assessed. This experiment was characterized by a higher degree of complexity than those of the previous tests in that it involved actual strong deformations of the objects considered (see Fig. 2e). The shearing transformation was obtained by considering a 2D shape as a plane in a 3D space, and by varying its *tilt* and *slant* angles. The tilt τ of a planar surface is defined as the angle between the surface normal projected on the image plane and the reference x axis, whereas the slant ϕ is the angle between the surface normal and the line of sight [25]. The resulting transformed surface was then orthonormally projected on the original (X, Y) plane to get the usual fronto-parallel view (see Fig. 2e). The experiment was performed by adding synthetic noise to the sheared shapes, using the second type of noise described in Section 4.3, which degrades the shapes more heavily. The applied noise level was $\sigma_d^2 = 0.35$, a medium noise level, and each object was randomly affected by noise 10 times. The averaged results are given in Table 3. From these results, we can notice that sensitivity of our approach to tilt and slant changes is very different. The variation in the slant angle results in a severe distortion of the object appearance, whereas the variation in the tilt angle could be roughly considered as a kind of rotation of the slant-derived transformation. Our approach is very robust to shape rotations, so the performance level is mostly driven by slant variations. The results presented in Table 3 demonstrate that our approach is truly robust against shearing: only for large slant values, corresponding to severe distortions of the testing objects, do the classification accuracies decrease significantly, remaining more than two to three times above the random classification level. This is more remarkable considering that the sheared shapes are also affected by noise.

4.6 Results on Other Data Sets

To increase the statistical significance of the results, the method was also tested on two other sets of shapes. The first was used in

TABLE 2
Classification Accuracies for Combined Transformations: (a) Occluded and Rotated Objects at Different Occlusion Levels and (b) Occluded, Rotated, and Noisy Objects at Different Occlusion and Noise Levels

Occlusion	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
Classification	100%	100%	100%	100%	98.8%	98.8	95%	94%	85%	87.5%

(a)

Occlusion	Noise $\sigma_d^2 = 0.1$	Noise $\sigma_d^2 = 0.3$	Noise $\sigma_d^2 = 0.5$
10%	100.00%	97.50%	87.50%
20%	98.75%	93.75%	80.00%
30%	98.75%	90.00%	80.00%
40%	93.75%	87.50%	77.50%
50%	86.25%	83.75%	75.00%

(b)

TABLE 3
Classification Accuracies for Noisy and Sheared Objects at Different Tilt and Slant Angles

		slant ϕ								
		0°	10°	20°	30°	40°	50°	60°	70°	80°
tilt τ	0°	98.8%	95.0%	100%	97.5%	97.5%	95.0%	65.0%	32.5%	27.5%
	10°	95.0%	98.8%	100%	100%	100%	93.8%	52.5%	37.5%	30.0%
	20°	100%	100%	100%	100%	96.3%	96.3%	51.3%	37.5%	37.5%
	30°	100%	100%	98.8%	100%	98.8%	88.8%	51.3%	48.8%	37.5%
	40°	97.5%	98.8%	100%	100%	100%	77.5%	52.5%	50.0%	37.5%
	50°	98.8%	98.8%	98.8%	100%	100%	75.0%	48.8%	37.5%	33.8%
	60°	97.5%	100%	98.8%	98.8%	98.8%	71.3%	46.3%	37.5%	31.3%
	70°	98.8%	98.8%	98.8%	98.8%	92.5%	70.0%	48.8%	40.0%	22.5%
	80°	96.3%	100%	98.8%	98.8%	82.5%	56.3%	43.8%	40.0%	17.5%
	90°	98.8%	98.8%	97.5%	97.5%	75.0%	53.8%	37.5%	38.8%	22.5%
	100°	96.3%	98.8%	97.5%	92.5%	71.3%	48.8%	33.8%	40.0%	22.5%
	110°	97.5%	96.3%	100%	91.3%	78.8%	43.8%	35.0%	37.5%	23.8%
	120°	100%	98.8%	100%	87.5%	80.0%	50.0%	36.3%	37.5%	26.3%
	130°	98.8%	98.8%	97.5%	92.5%	82.5%	63.8%	38.8%	37.5%	25.0%
	140°	100%	97.5%	98.8%	93.8%	83.8%	67.5%	50.0%	33.8%	25.0%
	150°	98.8%	96.3%	97.5%	96.3%	88.8%	72.5%	50.0%	35.0%	25.0%
	160°	98.8%	98.8%	100%	100%	97.5%	86.3%	52.5%	32.5%	25.0%
	170°	100%	97.5%	100%	100%	98.8%	90.0%	65.0%	36.3%	25.0%

[23], and was characterized by 12 deformed object instances for each class (six in all). In this case, instead of training one HMM for each shape class, one HMM was trained for each instance, resulting in 72 HMMs. Accuracy was computed by using the Leave One Out error scheme [26] and assigning an unclassified object to the class of the object whose model showed the maximum likelihood. The results were equal to 100 percent, thus confirming that the proposed approach was robust and accurate also for this set. Moreover, we evaluated the performances in the presence of occlusions using the same procedure as described in Section 4.2; the results are presented in Table 4. Also, these results are very satisfactory, even though not so good as on the He and Kundu database. In particular, as expected, the classification accuracies increased as the occlusion level decreased. By analyzing the errors on each specific class, we noted that the objects of the second class (glasses) were near perfectly classified at all occlusion levels. The class for which the results were worst was the sixth one (keys). Nevertheless, it is also worth noting that the instances of this class represented the same object only semantically, but the related shapes were indeed very different. Yet, the method's performances did not degrade too much.

A final test was carried out using the 21 letters of the Italian alphabet, Helvetica font (some letters fell into the same class: for example, "p" and "d," since the former is only a rotation of the latter). Also, in this case, the system reached a 100 percent classification accuracy, thus confirming the appropriateness of the proposed approach.

4.7 Significance of the Classification Scheme

The typical HMM classification scheme is a *Maximum Likelihood* approach, where one unknown object **O** is assigned to the class C_ℓ whose model shows the highest likelihood. This criterion is not able *per se* to provide a reliability measure of the classification decision: the "winning" class is represented by the maximum value, neglecting how "far" the second classified is. Obviously, the more distant the second one, the more reliable the decision. In this paper, a reliability measure has been defined, that is based on the log-likelihood distance between the first two choices of the classifier. This distance has been computed and averaged over the experiments regarding occlusion and noise degradation, and the expected behavior has been confirmed. This distance decreases the difficulty of the task increases (for example, when one increases the occlusion level), yet it remains at a satisfactory level. Two considerations can be derived from this analysis: the first is that our approach is robust, but, as expected, its robustness decreases with increasing task difficulty; the second is that it seems relatively simple to obtain a rejection rule by merely thresholding the likelihood difference between the first two choices of the algorithm: if a classification is not sufficiently reliable, it can be rejected.

TABLE 4
Classification Accuracies in the Presence of Occlusions for the Second Object Set Employed for Testing at Different Occlusion Levels

Occlusion	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
Accuracies	99.7%	99%	97.9%	97.4%	94.4%	94%	89.7%	88.1%	84.4%	82.2%

4.8 Discussion

In general, our approach proved to be very powerful in classifying shapes with a large class of deformations, especially occlusions, which constitute one of the hardest problems in object recognition. Nevertheless, the proposed method presents some weaknesses and limitations. First, like every contour-based method, it cannot manage all possible shapes, e.g., it cannot be applied for objects with holes. Second, since it deals with curvatures, it is sensitive to noise, i.e., its performances may be affected by rugged contours. This problem may indeed be attenuated by a proper use of the Gaussian filter applied before curvature computation. A further limitation of our approach is that it is not suited to distinguish very similar objects with only slight differences between them. Actually, the main characteristic and power of the Hidden Markov Modeling methodology lies in its ability to "generalize" well, i.e., to capture the essential features of a shape so as to recognize an object even if its aspect is not precisely the same of the model. Another problem may rise in case of occlusions. Although our experiments have demonstrated a certain resilience of the method, even in case of large occlusions, lower performances may occur when a part of the occluding object is part of the object contour to be classified. In this case, the performances of our approach obviously depend on the form and the extension of the part of the occluding object enclosed in the considered contour. In principle, our approach could also recognize the object, but with a lower likelihood. However, we have not addressed such a case since this aspect is to be considered as part of the segmentation process rather than a classification issue, which is the core of the present paper.

5 Conclusions

In this paper, the HMM behavior in the classification of planar shapes has been investigated. One HMM has been trained for each shape represented by curvature coefficients, paying particular attention to the HMM initialization and to the model selection issues during the learning process. Testing objects have been fed to all the trained HMMs, and each object has been classified as belonging to the class whose model provided the maximum likelihood. Tests on three different data sets have demonstrated that the proposed system is able to recognize objects that are modified instances of their original shape after rotation, occlusion, shearing, and noise degradations. The system has shown to be particularly robust to all these kinds of alterations, despite the intrinsic curvature sensitivity. This proves that the Hidden Markov Model methodology, with particular care to the training procedure, represents a powerful approach to shape classification.

References

[1] O. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, 2003.

[2] S. Loncaric, "A Survey of Shape Analysis Techniques," *Pattern Recognition*, vol. 31, no. 8, pp. 983-1001, 1998.

[3] P. Suetens, P. Fua, and A. Hanson, "Computational Strategies for Object Recognition," *ACM Compuiting Surveys*, vol. 24, no. 1, pp. 5-61, 1992.

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

[5] J. Hu, M. Brown, and W. Turin, "HMM Based Online Handwriting Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 10, pp. 1039-1045, oct. 1996.

[6] R. Hughey and A. Krogh, "Hidden Markov Model for Sequence Analysis: Extension and Analysis of the Basic Method," *Computer Applications in the Biosciences*, vol. 12, pp. 95-107, 1996.

[7] S. Eickeler, A. Kosmala, and G. Rigoll, "Hidden Markov Model Based Continuous Online Gesture Recognition," *Proc. IEEE Int'l Conf. Pattern Recognition*, vol. 2, pp. 1206-1208, 1998.

[8] T. Jebara and A. Pentland, "Action Reaction Learning: Automatic Visual Analysis and Synthesis of Interactive Behavior," *Proc. Int'l Conf. Computer Vision Systems*, 1999.

[9] Y. He and A. Kundu, "2-D Shape Classification Using Hidden Markov Model," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 13, no. 11, pp. 1172-1184, Nov. 1991.

[10] N. Arica and F. Yarman-Vural, "A Shape Descriptor Based on Circular Hidden Markov Model," *proc. IEEE Int'l Conf. Pattern Recognition*, vol. 1, pp. 924-927, 2000.

[11] A. Fred, J. Marques, and P. Jorge, "Hidden Markov Models vs. Syntactic Modeling in Object Recognition," *Proc. IEEE Int'l Conf. Image Processing*, vol. 1, pp. 893-896, 1997.

[12] J. Cai and Z.-Q. Liu, "Hidden Markov Models with Spectral Features for 2D Shape Recognition," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 23, no. 12, pp. 1454-1458, Dec. 2001.

[13] G. Schwarz, "Estimating the Dimension of a Model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461-464, 1978.

[14] L. Baum, T. Petrie, G. Soules, and N. Weiss, "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains," *Annals of Math. Statistics*, vol. 41, no. 1, pp. 164-171, 1970.

[15] A. Dempster, N. Laird, and D. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Royal Statistical Soc. B*, vol. 39, pp. 1-38, 1977.

[16] L. Baum, "An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of Markov Processes," *Inequality*, vol. 3, pp. 1-8, 1970.

[17] G. McLachlan and D. Peel, *Finite Mixture Models*. New York: John Wiley & Sons, 2000.

[18] C. Wu, "On the Convergence Properties of the EM Algorithm," *The Annals of Statistics*, vol. 11, no. 1, pp. 95-103, 1983.

[19] A. Stolcke and S. Omohundro, "Hidden Markov Model Induction by Bayesian Model Merging," *Advances in Neural Information Processing Systems*, S. Hanson, J. Cowan, and C. Giles, eds., pp. 11-18, vol. 5, Morgan Kaufmann, 1993.

[20] M. Brand, "An Entropic Estimator for Structure Discovery," *Advances in Neural Information Processing Systems*, MIT Press, M. Kearns, S. Solla, and D. Cohn, eds., vol. 11, 1999.

[21] M. Bicego, A. Dovier, and V. Murino, "Designing the Minimal Structure of Hidden Markov Models by Bisimulation," *Energy Minimization Methods in Computer Vision and Pattern Recognition*, M. Figueiredo, J. Zerubia, and A. Jain, eds., Springer, pp. 75-90, 2001.

[22] M. Bicego, V. Murino, and M. Figueiredo, "A Sequential Pruning Strategy for the Selection of the Number of States in Hidden Markov Models," *Pattern Recognition Letters*, vol. 24, nos. 9-10, pp. 1395-1407, 2003.

[23] T. Sebastian, P. Klein, and B. Kimia, "Recognition of Shapes by Editing Shock Graphs," *Proc. Int'l Conf. Computer Vision*, pp. 755-762, 2001.

[24] M. Bicego and V. Murino, "2D Shape Recognition by Hidden Markov Models," *Proc. IEEE Int'l Conf. Image Analysis and Processing*, pp. 20-24, 2001.

[25] L. Shapiro and G. Stockman, *Computer Vision*. Prentice Hall, 2001.

[26] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. Academic Press, 1999.