# Hidden Markov Models for Silhouette Classification

Wael Abd-Almageed and Christopher Smith

Robotics, Artificial Intelligence and Vision Laboratory
Electrical and Computer Engineering Department
University of New Mexico
{wamageed, chsmith@eece.unm.edu}

**Abstract**

In this paper, a new technique for object classification from silhouettes is presented. Hidden Markov Models are used as a classification mechanism. Through a set of experiments, we show the validity of our approach and show its invariance under severe rotation conditions. Also, a comparison with other techniques that use Hidden Markov Models for object classification from silhouettes is presented.

## 1 Introduction

Object classification from silhouettes has a long history in pattern recognition and computer vision. It has a wide range of applications; for instance, robotic grasping [1], medical image processing [2] and content-based image retrieval [3]. Although silhouette recognition has been studied for several decades, this problem has no solution in the general case.

In this paper, we use Hidden Markov Models (HMMs) as a classification mechanism to solve the problem of object classification from silhouettes. HMMs have been used effectively to solving several computer vision and pattern recognition problems[4].

This paper is organized as follows. Section 2 presents a brief overview of the techniques used for object classification from silhouettes. In section 3, we discuss our proposed approach. Section 4 introduces some experimental results of applying our technique. In section 5, we compare the performance of our approach and other approaches that have used Hidden Markov Models for the same problem. Section 6 concludes the paper and introduces some ideas for future research.

## 2 Previous Work

The problem of object classification plays an important role in object recognition, matching and registration systems. In this section, we present a brief review of the different methods used to attack the problem of object classification from silhouettes. Several criteria can be used to categorize silhouette classification approaches. Belongie et. al. [5] categorized shape matching approaches based on their use of extracted silhouette features versus image brightness. Another categorization, due to Pavlidis [6], is based upon the shape of the boundary (or external) versus the global (or internal) shape of the object.

In this paper, we categorize silhouette classification approaches into several categories based on the classification method.

## 2.1 Statistical Methods

In statistical-based methods, the common theme is to operate the classifier in two phases: the *learning* phase and the *classification* phase. In the learning phase, we assume an underlying probability distribution for some feature(s) extracted from the shape/silhouette of the object and then estimate the parameters of the probability distribution. Estimating a different probability distribution for each individual class of objects yields decision boundaries in the feature space that separate patterns belonging to different classes. In the classification phase, the decision boundaries and/or the different probability distributions are used to classify a given silhouette. Hidden Markov Model (HMM)[4] and Support Vector Machines [7] are two of the most common statistical approaches for shape classification. In a HMM, for example, each class of objects is represented by a different HMM with a finite number of states. A supervised learning process is then performed to estimate the models' parameters. In the classification phase, the object being classified is tested against all models. For a more detailed review on statistical pattern recognition techniques, the reader is referred to [8].

## 2.2 Transform-Based Methods

The main idea in this class of methods is to transform the shape/silhouette of the object from the spatial-domain into a transform-domain (a transform-domain is a generalization of the frequency domain) which yields a set of coefficients that represents the shape/silhouette. This set of coefficients can be used as a feature vector to measure the similarity between different shape classes. The Fourier transform [9] is the most popular example of this category of methods. In this method, the boundary of the object is represented as a complex 1D signal, the Fourier transform is computed for the boundary signal, and then a limited number of Fourier coefficients (commonly known as Fourier descriptors) are compared to other sets of coefficients to classify the object. Wavelet transform [10] and Principal Component Analysis (PCA) [11] are other common transforms used for shape/silhouette classification.

## 2.3 Neural Networks-Based Methods

Neural Networks (NN) is one of the most commonly used tools for different classification and recognition tasks (e.g. face [12] and shape [13].) The NN is similar to the statistical approaches in that it operates in two different modes: a *learning* mode and a *classification* mode. In the learning mode, a set of features, extracted from different classes of objects, is shown to the network to adapt its parameters. In the classification mode, extracted feature(s) (similar to that used in the learning) are fed into the NN to classify it. Several variations of this approach have been proposed. These variations include supervised versus unsupervised learning and single-NN versus multiple-NN approaches. However, all variations follow the same idea described above.

## 2.4 Dynamic Programming

Dynamic programming is usually used for shape classification based on the shape boundary. The general idea is to represent the boundary of the unknown object and boundaries of the object prototypes (of the database) as strings. The problem then is reduced to that of the computing the distance between two given strings. The prototype that has the minimum distance with the unknown object is then considered to be the recognized class. In [14], Bunke and Buhler introduced a robust algorithm for 2D shape recognition using approximate string matching. The proposed algorithm was scale, rotation and translation invariant. They defined a cost function for the possible edit operations between the two strings (shapes) being compared. The minimum edit cost identifies the class to which the input shape belongs. Applying dynamic programming to shape/silhouette classification ranges from some simple, direct applications [14] to sophisticated applications [15], however, most of these applications follow the same theme as described before.

### 2.5 Hybrid Classifiers

The above approaches reflect the major shape/silhouette classification techniques. However, two or more of the approaches described above can be used in a hybrid classifier that fuses the results of each of the individual classifiers to obtain a robust classification. In [16] Abdelazim presented a fuzzy-neural approach for recognizing Arabic letters. Another example is [17], which uses a hybrid symbolic-neural approach for shape recognition.

### 2.6 Uncategorized Methods

Some other approaches have been proposed for solving the shape/silhouette classification problem that do not belong to any of the discussed categories and are not considered separate categories on their own. For example, in [18], Belongie and Malik introduced what they called the Shape Context for matching shapes. Also, some classification approaches adopted the invariant moments (sometimes classified as a statistical method) for classification [19]. Fuzzy sets have also been used for shape classification [20].

The above review is not a detailed review on shape/silhouette classification approaches. For more detailed reviews on the topic, the reader is referred to [8][21][22].

## 3 The Proposed Approach

**Silhouette Representation:** We represent the silhouette of an object by an ordered set of vectors $V = \{\overline{v}_1, \overline{v}_2, ..., \overline{v}_L\}$ in the 2D Euclidian space, where $L$ is the length of the object's silhouette. The external angle between any two successive vectors, $\overline{v}_i$ and $\overline{v}_{i+1}$, is then is defined by equation (1).

$$\theta_i = \cos^{-1} \frac{\overline{v}_i . \overline{v}_{i+1}}{|\overline{v}_i||\overline{v}_{i+1}|} \tag{1}$$

Because of the discrete nature of the image spatial domain, see figure 1, $\theta \in \Sigma = \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$. The set $\Sigma$ is called the alphabet set. Computing the external angle between all pair of successive vectors in the set $V$, yields an observation sequence $\Theta = \theta_1\theta_2...\theta_L$ that represents the contour of the object.

**Hidden Markov Model:** A discrete probability Hidden Markov Model can be characterized by a five-tuple $\lambda = \{S, \Sigma, A, B, \Pi\}$ as follows.

- $S = \{s_1, s_2, ..., s_N\}$ is the set of model states,
- $\Sigma = \{O_1, O_2, ..., O_M\}$ is the alphabet set,

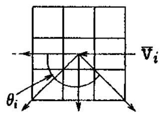


Figure 1: Discrete nature of the external angle

- $A = \{a_{ij}\}$ is the transition probability distribution where

$$a_{ij} = P[q_{t+1} = s_j | q_t = s_i], \qquad 1 \leq i, j \leq N, \tag{2}$$

and $q_t$ is the model state at time $t$.

- $B = \{b_{jk}\}$ is the observation probability distribution where

$$b_{jk} = P[O_k \; at \; t | q_t = s_j], \qquad 1 \leq j \leq N \quad and \quad 1 \leq k \leq M \quad and \tag{3}$$

- $\Pi = \{\pi_1, \pi_2, ..., \pi_N\}$ is the prior probability distribution where

$$\pi_i = P[q_1 = s_i], \qquad 1 \leq i \leq N. \tag{4}$$

In most of computer vision applications (e.g. face recognition,) the left-to-right HMM is sufficient to model the data. However, we use an ergodic (fully connected) HMM for modelling the silhouettes of the objects in the database. There are two main advantages of using such representation. First, an ergodic HMM does not require a starting feature that must be consistently extracted from the object's contour, because we are not constrained to start from the first state all of the time. The second advantage of using the ergodic HMM is that it tolerates greater amounts of partial occlusion. This is because, in the ergodic HMM, a transition is allowed from any state $s_i$ to any other state $s_j$ (i.e. all entries of the transition probability matrix are non-zero.)

Having $C$ distinct object classes, we model each class by an individual DHMM. The models' parameters are estimated using the popular Baum-Welch algorithm [23]. The parameter estimation process is performed using different examples of the object's contour and is repeated for several iterations until convergence. During the classification phase, the object's contour being classified is tested against all DHMMs obtained previously and the probability that the object's contour is generated by each model, equation 5, is computed using the Viterbi algorithm [24].

$$P(\Theta_{Object} | \lambda_i), \qquad i = 1, 2, ..., C \tag{5}$$

In most HMM classifiers, the object is classified to the class that its HMM gives the maximum probability, i.e. $max_{\forall i}(P(\Theta_{Object}|\lambda_i)$. In our approach, we chose the class if

$$|max_{\forall i} P(\Theta_{Object}|\lambda_i) - P(\Theta_{Object}|\lambda_j)| < T, \qquad i, j = 1, 2, ..., C \tag{6}$$

where $T$ is a small threshold obtained experimentally. There are two main reasons for adopting such strategy. The first reason is that, for content-based image-retrieval applications, the objective usually is to obtain all objects that are similar to the query object, not just the most similar one. The second reason is that the parameter estimation process is an optimization process, which might not yield the optimum parameters of the HMMs. This will cause similar object classes to have similar models yielding classification inaccuracies, which is a problem for all object classification algorithms.

Table 1: Classification Statistics. Classification Ratio=93.3%

| Object 1 | Object 2 | Object 3 | Object 4 | Object 5 | Object 6 |
|---|---|---|---|---|---|
| 15/15 | 15/15 | 9/15 | 15/15 | 15/15 | 15/15 |

# 4 Experimental Results

To test our approach, we used a database of 90 images of 6 different objects. Figure 2 shows the different object classes in the database. A 5-state ($N = 5$) DHMM is obtained for each of the 6 object classes using the Baum-Welsh learning algorithm. We performed two experiments to test the robustness of our approach. In the first experiment, object images, from outside the training set, were classified using the approach described in section 3. Figure 3 shows the results of the experiment. In the second experiment, using the same DHMMs, we used a $90^o$ rotated version of the test images to test our approach. Unsurprisingly, the exact same classification results were obtained, as shown in figure 3. We repeated the same two experiments using different training and testing data sets and similar results were obtained. Table 1 quantifies the test results of our approach.

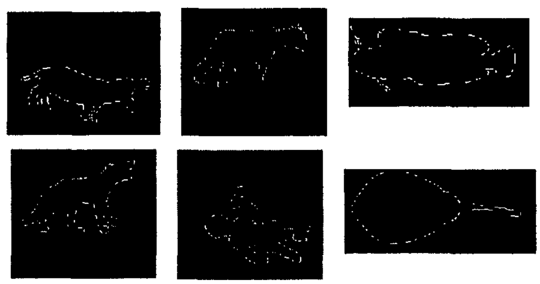

Figure 2: Object Classes

# 5 Performance Comparison

In [25], Arica and Vural used what they called circular hidden Markov model for shape classification. Excellent classification results were obtained. In their work, they used the 8-directional Freeman's chain code to encode the shape boundaries. The major disadvantage of this encoding scheme is its sensitivity to shape rotations. One solution to this problem is the use of the derivative of the chain code to encode the shape boundaries. The use of our proposed shape representation alleviates the problem of sensitivity to rotation, while using a smaller alphabet set (i.e. $B_{space} = \Theta(4N)$ in our approach versus $B_{space} = \Theta(7N)$ in [25].) Also, in [25], the authors use what is called *circular HMM* which is a modification of the Left-to-Right HMM that solves the problem of starting point. The disadvantage of this technique is, under moderate to large occlusion, the model will fail to classify correctly because of the zero transition probability if the required state transition is higher than a

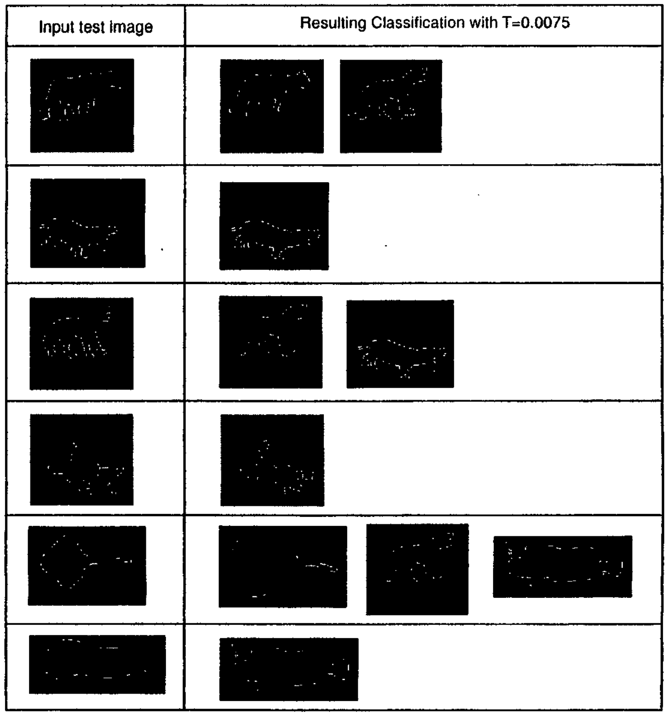


Figure 3: Object Classesifications

specified threshold ($N$ in their approach.) In our work, the use of ergodic HMM gives a finite non-zero probability for such state transitions if the object is partially occluded.

Another HMM-based shape classifier was proposed in [4]. In this work, the authors used a centroid radial scan to extract features from the object boundary. The extracted features are then models by autoregressive functions to obtain the observation sequences of the object. The computational complexity of using such a feature extractor ($\Theta(KL)$, $K$ and $L$ are defined in [4]) is much higher than computing the angle between each pair of vectors in our scheme ($\Theta(L)$). Also, in [4], the authors used a continuous probability HMM (CHMM.) The computational complexity of the classification phase of both approaches will be similar asymptotically ($\Theta(LN)$), if using an optimized CHMM. However, using a DHMM avoids the need of the quantization step needed to obtain such complexity.

# 6 Conclusion

A Hidden Markov Models-based object classification technique was presented. Compared to the previously presented approaches that adopted HMM as a classification mechanism, the new algorithm has two main advantages. The first advantage is that it is rotational invariant due to the extracted feature from the object's contour. The second advantage is that using DHMM makes the approach computationally less expensive compared to other approaches that use continuous probability HMM. The similarity criterion presented suits the different kinds of applications. Setting $T > 0$ is well suited to content-based image-retrieval applications, while setting $T = 0$ is well suited to the classification applications where only the best matching class is required.

The proposed approach was evaluated using a silhouette database of 90 images of 6 classes of objects. Accurate classification results were obtained. The approach shows robustness against rotation conditions in contrast to other HMM-based silhouette classifiers.

For future research, we would like to investigate using embedded HMMs for object classification. The idea is to model each of the contour features by a separate HMM instead of a state in our current model. This should yield more robust results specially when greater amounts of occlusion occur.

# References


[1] H. Kobayashi, K. Igawa, T. Bito, and K. Kikuchi, "3d object recognition and grasping for human support robotic system.," in *9th IEEE International Workshop on Robot and Human Interactive Communication. IEEE RO-MAN 2000, 27-29 Sept. 2000, Osaka, Japan.*

[2] E. Claridge, P.N. Hall, M. Keefe, and J.P. Allen, "Shape analysis for classification of malignant melanoma.," *Journal of Biomedical Engineering*, vol. 14, no. 3, pp. 229 – 34, May 1992.

[3] B. Gunsel and A.M. Tekalp, "Shape similarity matching for query-by-example.," *Pattern Recognition*, vol. 31, no. 7, pp. 931 – 44, July 1998.

[4] Y. He and A. Kundu, "Shape classification using hidden markov model.," in *ICASSP 91: 1991 International Conference on Acoustics, Speech and Signal Processing, 14-17 April 1991, Toronto, Ont., Canada.*

[5] S. Belongie, J. Malik, and J. Puzicha, "Matching shapes.," in *Eighth IEEE International Conference on Computer Vision, 7-14 July 2001, Vancouver, BC, Canada.*

[6] T. Pavlidis, "A review of algorithms for shape analysis.," *Computer Graphics and Image Processing*, vol. 7, no. 2, pp. 243 – 58, April 1978.

[7] V.N. Vapnik, "An overview of statistical learning theory.," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988 – 99, September 1999.

[8] A.K. Jain, P.W. Duin, and Jianchang Mao, "Statistical pattern recognition: a review.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4 – 37, January 2000.

[9] C.T. Zahn and R.Z.L. Roskies, "Fourier descriptors for plane closed curves.," *IEEE Transactions on Computers*, vol. c-21, no. 3, pp. 269 – 81, March 1972.

[10] Mahmoud Khalil and Mohamed Bayoumi, "A dyadic wavelet affine invariant funntion for 2d shape recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001.

[11] V. Gouaillier and L. Gagnon, "Ship silhouette recognition using principal components analysis.," *Proceedings of the SPIE - The International Society for Optical Engineering*, vol. 3164, pp. 59 – 69, 1997.

[12] S. Lauria and R. Mitchell, "Weightless neural networks for face recognition.," in *11th Italian Workshop on Neural Nets - WIRN Vietri-99, 20-22 May 1999, Salerno, Italy.*

[13] P.N. Suganthan, Eam Khwang Teoh, and D.P. Mital, "Hopfield network with constraint parameter adaptation for overlapped shape recognition.," *IEEE Transactions on Neural Networks*, vol. 10, no. 2, pp. 444 – 9, March 1999.

[14] H. Bunke and U. Buhler, "Applications of approximate string matching to 2d shape recognition.," *Pattern Recognition*, vol. 26, no. 12, pp. 1797 – 812, December 1993.

[15] Y. Gdalyahu and D. Weinshall, "Flexible syntactic matching of curves and its application to automatic hierarchical classification of silhouettes.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1312 – 28, December 1999.

[16] H.Y. Abdelazim, "A hybrid fuzzy-neural approach to the recognition of arabic script.," in *5th International Conference on Multi-lingual Computing (The New World Order and Information-Security and Welfare of Nations), 11-13 April 1996, Cambridge, UK.*

[17] Mon-Chu Chen and Yu-Tung Liu, "A symbolic/neural hybrid approach to multiple shape recognition.," in *28th Southeastern Symposium on System Theory, 31 March-2 April 1996, Baton Rouge, LA, USA.*

[18] S. Belongie and J. Malik, "Matching with shape contexts.," in *IEEE Workshop on Content-based Access of Image and Video Libraries, 12 June 2000, Hilton Head Island, SC, USA.*

[19] Y.S. Tawfik, A.M. Darwish, and S.I. Shaheen, "Energy matching based on deformable templates.," in *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99, 15-19 March 1999, Phoenix, AZ, USA.*

[20] B. Lazzerini and F. Marcelloni, "A fuzzy approach to 2d-shape recognition.," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 1, pp. 5 – 16, February 2001.

[21] R. Veltkamp and M. Hagedoorn, "State-of-the-art in shape matching," Tech. Rep., Utrecht University, the Netherlands, 1999.

[22] S. Loncaric, "A survey of shape analysis techniques," *Pattern Recognition*, vol. 31, no. 8, pp. 983–1001, 1998.

[23] L.E. Baum and G.R. Sell, "Growth functions of transformation of manifolds," *Pacific Journal of Mathematics*, vol. 73, no. 2, pp. 360–362, 1967.

[24] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Transactions on Information Theory*, vol. IT-13, pp. 260–269, April 1976.

[25] N. Arica and F.T.Y. Vural, "A shape descriptor based on circular hidden markov model.," in *15th International Conference on Pattern Recognition, 3-7 Sept. 2000, Barcelona, Spain.*