

HIDDEN MARKOV MODEL BASED WEIGHTED LIKELIHOOD DISCRIMINANT FOR MINIMUM ERROR SHAPE CLASSIFICATION

Ninad Thakoor and Sungyong Jung

Jean Gao

Electrical Engineering
The University of Texas at Arlington
Arlington, TX 76019

Computer Science and Engineering
The University of Texas at Arlington
Arlington, TX 76019

ABSTRACT

The goal of this communication is to present a weighted likelihood discriminant for minimum error shape classification. Different from traditional Maximum Likelihood (ML) methods in which classification is carried out based on probabilities from independent individual class models as is the case for general hidden Markov model (HMM) methods, our proposed method utilizes information from all classes to minimize classification error. Proposed approach uses a Hidden Markov Model as a curvature feature based 2D shape descriptor. In this contribution we present a Generalized Probabilistic Descent (GPD) method to weight the curvature likelihoods to achieve a discriminant function with minimum classification error. In contrast with other approaches, a weighted likelihood discriminant function is introduced. We believe that our sound theory based implementation reduces classification error by combining hidden Markov model with generalized probabilistic descent theory. We show comparative results obtained with our approach and classic maximum-likelihood calculation for fighter planes in terms of classification accuracies.

1. INTRODUCTION

Object recognition is a classic problem in image processing, computer vision, and database retrieval. Among others, object recognition based on shape is widely used. First step towards the design of a shape classifier is feature extraction. Shapes can be represented by their contours or regions [1]. Curvature, chain codes, Fourier descriptors, etc. are contour based descriptors while medial axis transform, Zernike moments, etc. are region based features. Contour based descriptors are widely used as they preserve the local information which is important in classification of complex shapes.

Feature extraction is followed by shape matching. In recent years, dynamic programming (DP) based shape matching is being increasingly applied [2], [3], [4], [5]. DP approaches are able to match the shapes part by part rather than point by point, and are robust to deformation and occlusion. Hidden Markov Models (HMMs) are also being explored as one of the possible shape modeling and classification frameworks [6], [7], [8], [9], [10]. Apart from having all the properties of DP based matching, HMM provides a probabilistic framework for training and classification.

The current HMM approaches apply maximum likelihood (ML) as their classification criterion. Due to good generalization property of HMM, applying ML criterion to similar shapes does not provide good classification. Also, ML criterion is evaluated using information from only one class and does not take advantage of information from the other classes. Generally shapes can be

discriminated using only parts of the boundaries rather than comparing whole boundary. ML does not provide such mechanism.

To overcome these shortcomings, we propose a weighted likelihood discriminant for shape classification. The weighting scheme emulates comparison of parts of shape rather than the whole shape. The weights are estimated by applying Generalized Probabilistic Descent (GPD) method. Unlike ML criterion, GPD uses information from all the classes to estimate the weights. As GPD method is designed to minimize the classification error, the proposed classifier gives good classification performance with similar shapes. This paper is organized as follows: The description phase of the proposed method is discussed in Section 2, while Section 3 formulates discriminative training with GPD. Experimental results are presented in Section 4 and the paper ends with the conclusions and suggestions for further research in Section 5.

2. SHAPE DESCRIPTION WITH HMM

Before proceeding to the detailed topology of HMM, we introduce the terminology used in the rest of the paper.

1. S , set of states. $S = \{S_1, S_2, \dots, S_N\}$, where N is number of states. State of HMM at instance t is denoted by q_t .
2. A , state transition probability distribution. $A = \{a_{ij}\}$, a_{ij} denotes the probability of changing the state from S_i to S_j .

$$a_{ij} = P[q_{t+1} = S_j | q_t = S_i], \quad 1 \leq i, j \leq N. \quad (1)$$

3. B , observation symbol probability distribution. $B = \{b_j(o)\}$, $b_j(o)$ gives probability of observing the symbol o in state S_j at instance t .

$$b_j(o) = P[o \text{ at } t | q_t = S_j], \quad 1 \leq j \leq N. \quad (2)$$

4. π , initial state distribution. $\pi = \{\pi_i\}$, π_i gives probability of HMM being in state S_i at instance $t = 1$.

$$\pi_i = P[q_1 = S_i], \quad 1 \leq i \leq N. \quad (3)$$

If C_j is j^{th} shape class where $j = 1, 2, \dots, M$ and M is total number of classes then for convenience, HMM for C_j can be compactly denoted as,

$$\lambda_j = (A, B, \pi). \quad (4)$$

An in depth description about HMM can be found in [11].

For the approach proposed in this paper, the description phase employs HMM topology proposed by Bicego and Murino [7]. The

curvature of the shape is used as the descriptor. The shape is filtered with large variance Gaussian filter to reduce the effect of noise in curvature estimation. The filtered shape is normalized to a fixed length to simplify comparison and its major eigen-axis is aligned horizontally to achieve an invariant starting point. Let the aligned shape be indicated by $D = \{D_n\}$ and $D_n = (x_n, y_n)$ for $1 \leq n \leq T$, where T is the normalized length of the shape, and D_n indicates coordinates of n^{th} point of the shape. Finally, approximate curvature at each point is calculated as the turn angle at that point. The turn angle θ_n at point D_n is defined as,

$$\theta_n = \arctan \frac{y_n - y_{n-1}}{x_n - x_{n-1}} - \arctan \frac{y_n - y_{n+1}}{x_n - x_{n+1}} \quad (5)$$

The turn angle θ_n is treated as observation O_n for the HMM model. Shape class C_j is modeled by a N -state ergodic HMM λ_j^N and observation symbol probability distribution, i.e., b_j of each state is modeled as one-dimensional Gaussian distribution. Gaussian Mixture Model (GMM) [12] for N clusters estimated from unrolled values of curvature is used to initialize observation distribution B . Baum-Welch algorithm is then applied to estimate the parameters of the HMM $\lambda_j = (A, B, \pi)$. Optimum N for the HMM is selected using Bayesian Inference Criterion (BIC). In [7] BIC is applied to GMM to select optimal k , but this gives optimal N for GMM and not for HMM. In proposed approach, BIC is applied to HMM to ensure proper model selection. For HMM topology discussed, BIC can be written as,

$$\text{BIC}(\lambda_j^N) = \log P(O|\lambda_j^N) - \frac{N^2 + 2N - 1}{2} \log(T). \quad (6)$$

Model with N states, i.e., λ_j^N is selected to maximize $\text{BIC}(\lambda_j^N)$.

ML training approach described in this section utilizes information from only one class to build the models. Though other approaches, like Maximum Mutual Information (MMI), Generalized Probabilistic Descent (GPD), which use information of all the classes, have been proposed for model training. However, classification performance of properly designed and ML trained HMM cannot be improved significantly with MMI or GPD training of HMM. Therefore in our paper, we will stay with optimally designed HMM and make our contributions in designing robust discriminant functions with minimum error. Hence Section 3.

3. DISCRIMINANT FUNCTION FORMULATION AND TRAINING

In this section, we formulate a minimum error classifier with weighted likelihood discriminant function. The weights introduced in the discriminant function are trained with GPD method. A detailed review of GPD method can be found in [13].

3.1. Discriminant Function Selection

Consider observation sequence to be classified, $O = O_1 O_2 \dots O_T$. After modeling this sequence with j^{th} class HMM $\lambda_j = (A, B, \pi)$ and solving optimal path problem, optimum path is given by $Q^* = q_1^* q_2^* \dots q_T^*$ in which q_t is the state of HMM at instance t . The probability of observation sequence O given the state sequence Q^* and model λ_j is given by,

$$P(O|Q^*, \lambda_j) = b_{q_1^*}(O_1) \cdot b_{q_2^*}(O_2) \dots b_{q_T^*}(O_T). \quad (7)$$

Probability of state sequence Q^* is given by,

$$P(Q^*|\lambda_j) = \pi_{q_1^*} \cdot a_{q_1^* q_2^*} \dots a_{q_{T-1}^* q_T^*}. \quad (8)$$

Then probability of the both occurring simultaneously is given by,

$$\begin{aligned} P(O, Q^*|\lambda_j) &= P(O|Q^*, \lambda_j) \cdot P(Q^*|\lambda_j) \\ &= \pi_{q_1^*} b_{q_1^*}(O_1) a_{q_1^* q_2^*} b_{q_2^*}(O_2) \\ &\quad \dots a_{q_{T-1}^* q_T^*} b_{q_T^*}(O_T). \end{aligned} \quad (9)$$

Let Υ be defined as,

$$\Upsilon_{t,j} = \begin{cases} \log \{ \pi_{q_1^*} b_{q_1^*}(O_1) \}, & t = 1; \\ \log \{ a_{q_{t-1}^* q_t^*} b_{q_t^*}(O_t) \}, & 2 \leq t \leq T. \end{cases} \quad (10)$$

Therefore, Eqn.(9) can be expressed as,

$$\log P(O, Q^*|\lambda_j) = \sum_{t=1}^T \Upsilon_{t,j}. \quad (11)$$

Eq.(11) can be used as discriminant for classification of the observation sequence O . This function gives equal importance to every point of the shape in shape classifications. Hence, we introduce a new discriminant function which weights the curvature likelihood of shape points according to their importance in classification. The new discriminant function, g_j is given by,

$$g_j = \sum_{t=1}^T w_j(t) \cdot \Upsilon_{t,j}, \quad (12)$$

where w_j is weighting function for class C_j . w_j provides additional discrimination among the classes. These weights will be tuned by applying GPD method to minimize the classification error.

3.2. Weighting Functions

Weighting function at individual observation can be estimated by applying GPD to current formulation. But due to the large number of parameters (equal to T), the convergence of GPD will be slower and will need large number of observation sequences for training. As mentioned in Section 1, to discriminate between similar shapes, comparison between parts of their contour is sufficient. As a result, shape can be weighted part by part instead of weighting it pointwise. Following this intuitive idea, weighting functions are chosen to be windows which can adapt their position, spread and height. Any smooth window function can be selected. Our approach uses weighting function given in Eqn.(13), which is sum of S Gaussian shaped windows.

$$w_j(t) = \sum_{i=1}^S p_{i,j} \cdot e^{-\frac{(t-\mu_{i,j})^2}{s_{i,j}^2}}. \quad (13)$$

Parameter $p_{i,j}$ governs the height, $\mu_{i,j}$ controls the position, while $s_{i,j}$ determines spread of i^{th} window of j^{th} class. In this case, we have only $3S$ parameters to estimate. The discriminant function can now be written as,

$$g_j = \sum_{t=1}^T \sum_{i=1}^S p_{i,j} \cdot e^{-\frac{(t-\mu_{i,j})^2}{s_{i,j}^2}} \cdot \Upsilon_{t,j}. \quad (14)$$

In the next subsection GPD method is applied to above formulation.

3.3. GPD Algorithm

To complete the formulation of GPD, we introduce misclassification measure for observation sequence of j^{th} class as,

$$d_j = -g_j + \frac{1}{\eta} \log \left(\frac{1}{M-1} \sum_{k, k \neq j} e^{\eta \cdot g_k} \right) \quad (15)$$

and corresponding cost function as,

$$l_j = \frac{1}{1 + e^{-\xi \cdot d_j}} \quad (16)$$

The probabilistic descent re-estimation rule for parameters Λ is given as,

$$\Lambda_{n+1} = \Lambda_n - \varepsilon U \nabla l_j. \quad (17)$$

For the proposed method, U is chosen to be identity matrix. The re-estimation rules in iteration n , for i^{th} window parameters of k^{th} class when C_j is the correct class are given by,

$$p_{i,k}^{n+1} = p_{i,k}^n - \varepsilon_p \cdot \frac{\partial l_j}{\partial p_{i,k}}, \quad (18)$$

$$\mu_{i,k}^{n+1} = \mu_{i,k}^n - \varepsilon_\mu \cdot \frac{\partial l_j}{\partial \mu_{i,k}}, \quad (19)$$

$$s_{i,k}^{n+1} = s_{i,k}^n - \varepsilon_s \cdot \frac{\partial l_j}{\partial s_{i,k}}, \quad (20)$$

for $1 \leq i \leq S$, $1 \leq k \leq M$.

Partial derivatives appearing in Eqns.(18)-(20) can be calculated by chain rule as,

$$\frac{\partial l_j}{\partial p_{i,k}} = \frac{\partial l_j}{\partial d_j} \cdot \frac{\partial d_j}{\partial g_k} \cdot \frac{\partial g_k}{\partial p_{i,k}}, \quad (21)$$

$$\frac{\partial l_j}{\partial \mu_{i,k}} = \frac{\partial l_j}{\partial d_j} \cdot \frac{\partial d_j}{\partial g_k} \cdot \frac{\partial g_k}{\partial \mu_{i,k}}, \quad (22)$$

$$\frac{\partial l_j}{\partial s_{i,k}} = \frac{\partial l_j}{\partial d_j} \cdot \frac{\partial d_j}{\partial g_k} \cdot \frac{\partial g_k}{\partial s_{i,k}}, \quad (23)$$

where,

$$\frac{\partial l_j}{\partial d_j} = \frac{\xi e^{-\xi \cdot d_j}}{(1 + e^{-\xi \cdot d_j})^2}, \quad (24)$$

$$\frac{\partial d_j}{\partial g_k} = \begin{cases} -1, & j = k; \\ \frac{e^{\eta \cdot g_j}}{\sum_{k' \neq k} e^{\eta \cdot g_{k'}}}, & j \neq k. \end{cases} \quad (25)$$

$$\frac{\partial g_k}{\partial p_{i,k}} = \sum_{t=1}^T e^{-\frac{(t-\mu_{i,k})^2}{s_{i,k}^2}} \Upsilon_{t,k}, \quad (26)$$

$$\frac{\partial g_k}{\partial \mu_{i,k}} = \sum_{t=1}^T \frac{2p_{i,k}(t-\mu_{i,k})e^{-\frac{(t-\mu_{i,k})^2}{s_{i,k}^2}}}{s_{i,k}^2} \Upsilon_{t,k}, \quad (27)$$

$$\frac{\partial g_k}{\partial s_{i,k}} = \sum_{t=1}^T \frac{2p_{i,k}(t-\mu_{i,k})^2 e^{-\frac{(t-\mu_{i,k})^2}{s_{i,k}^2}}}{s_{i,k}^3} \Upsilon_{t,k}. \quad (28)$$

Note that in above formulation $\Upsilon_{t,k}$ is treated to be a constant as the HMM parameters are not affected by the change in Λ .

Table 1. Classification accuracy in %

Class	HMM-ML	HMM-WtL
Mirage	79.42	99.33
Eurofighter	78.14	98.66
F-14 Closed	79.28	98.66
F-14 Open	52.57	99.33
Harrier	78.28	100.00
F-22	68.71	100.00
F-15	62.57	100.00

4. EXPERIMENTAL RESULTS

The proposed classification scheme was verified with fighter aeroplane shapes. These fighter aeroplanes include Mirage, Eurofighter, F-14, Harrier, F-22 and F-15 [14]. Since F-14 has two possible shapes, one when its wings closed and another when its wings opened, total number of shape classes are seven. Shape database was created by taking digital pictures of die-cast replica models of these aeroplanes. 30 pictures for each class were captured at 640×480 resolution, and were segmented using Spedge and Medge [15] color image segmentation algorithm. Contours of the segmented planes were used for training and testing of the classifier. Figure 1 shows the extracted shapes for different classes.

Shapes were filtered with Gaussian filter (standard deviation = 10) and shape length was normalized to 512 points. The normalized shapes were split randomly into 15 training and 15 testing samples. For one of the training samples of each class, HMM was built as explained in Section 2. Optimum number of HMM states were selected by applying BIC to models with 3 to 10 states. Sum of 20 Gaussian windows was used for formulation and training of the discriminant function. The window parameters were initialized to spread the windows uniformly over the shape (refer Figure 2(a)). The training vectors were used to train the classifier with $\xi = 1$ and $\eta = 10$. Once the training was complete, testing samples were used to determine the classification performance.

Table 1 gives classification results for ML classification (HMM-ML) and GPD based for weighted likelihood classification (HMM-WtL). These results were averaged over 20 runs of the classifier design, each time with different combination of training samples. Figure 2 shows the initial and estimated weights for various classes for one of the experiments. Weights for all the classes are initialized as shown in Figure 2(a). After training, the weights modified to minimize the classification error can be seen in Figure 2(b) and (c) for classes Mirage and F-14 respectively.

5. CONCLUSION AND FUTURE WORK

In this paper, we proposed a weighted likelihood discriminant function for shape classification by the combination of generalized probabilistic descent theory and hidden Markov model. A training algorithm based on GPD method to estimate the optimal weights with minimum classification error was formulated. The performance of the proposed shape classification scheme was tested on shapes of seven fighter planes. The classification accuracy is found to be 99.43% which is much higher than 71.3% of ML discriminant.

Though the shapes used for training and testing of the classifier exhibit some deformation (due to varying view points) and noise (due to automatic segmentation), a comprehensive analysis

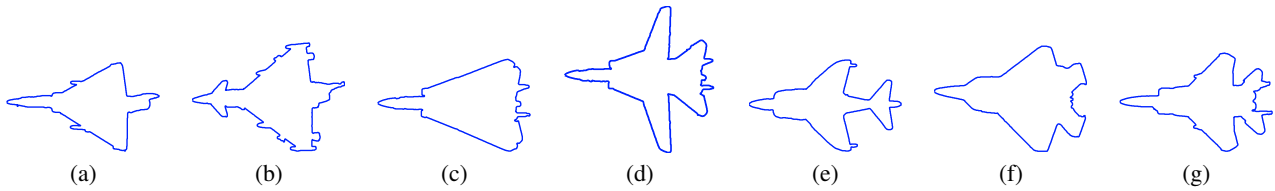


Fig. 1. Aeroplane shape classes: (a) Mirage, (b) Eurofighter, (c) F-14 wings closed, (d) F-14 wings opened, (e) Harrier, (f) F-22, (g) F-15.

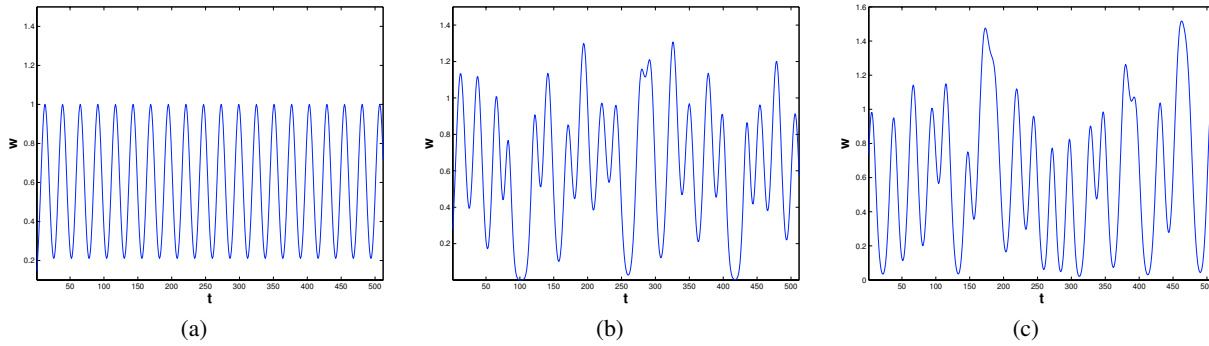


Fig. 2. Weights for various classes: (a) Initial weights for all classes, (b) Weights for Mirage after GPD, (c) Weights for F-14 open after GPD.

in presence of noise, occlusion and deformations needs to be evaluated by designing appropriate experiments. Currently, the number of weighting windows is selected manually and their parameters are initialized uniformly. A better initialization strategy like initializing them at curvature extremes can be employed.

6. REFERENCES

- [1] D. Zhang and G. Lu, "Review of shape representation and description technique," *Pattern Recognition*, vol. 37, no. 1, pp. 1–19, Jan. 2004.
- [2] E. Miliou and E. G. M. Petrakis, "Shape retrieval based on dynamic programming," *IEEE Transactions on Image Processing*, vol. 9, no. 1, pp. 141–147, Jan. 2000.
- [3] E. G. M. Petrakis, A. Diplaros, and E. Miliou, "Matching and retrieval of distorted and occluded shapes using dynamic programming," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 11, pp. 1501–1516, Nov. 2002.
- [4] 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. 1312–1328, Dec. 1999.
- [5] T. Adamek and N. E. O'Connor, "A multiscale representation method for nonrigid shapes with a single closed contour," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 5, pp. 742–753, May 2004.
- [6] N. Arica and F. Yarman-Vural, "A shape descriptor based on circular hidden markov model," *Proceedings of IEEE International Conference on Pattern Recognition*, vol. 1, pp. 924–927, 2000.
- [7] M. Bicego and V. Murino, "Investigating hidden markov models' capabilities in 2d shape classification," *IEEE Transactions on Pattern Recognition Machine Intelligence*, vol. 26, no. 2, pp. 281–286, Feb 2004.
- [8] J. Cai and Z.-Q. Liu, "Hidden markov models with spectral features for 2d shape recognition," *IEEE Transactions on Pattern Analysis Machine Intelligence*, vol. 23, no. 12, pp. 1454–1458, Dec. 2001.
- [9] Y. He and A. Kundu, "2-d shape classification using hidden markov model," *IEEE Transactions on Pattern Analysis Machine Intelligence*, vol. 13, no. 11, pp. 1172–1184, Nov. 1991.
- [10] A. Fred, J. Marques, and P. Jorge, "Hidden markov models vs. syntactic modeling in object recognition," *Proceedings of IEEE International Conference of Image Processing*, vol. 1, pp. 893–896, 1997.
- [11] L. R. Rabiner, "A tutorial on hidden markov models and selected application in speech recognition," *Proceedings of IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
- [12] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using gaussian mixture models," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 1, pp. 72–83, Jan. 1995.
- [13] S. Katagiri, B.-H. Juang, and C.-H. Lee, "Pattern recognition using a family of design algorithms based upon the generalized probabilistic descent method," *Proceedings of IEEE*, vol. 86, no. 11, pp. 2345–2373, Nov. 1998.
- [14] D. Rendall, *Jane's Aircraft Recognition Guide*, Harper Collins, 1999.
- [15] J. Gao, A. Kosaka, and A. Kak, "Interactive color image segmentation editor driven by active contour model," *Proceedings of International Conference on Image Processing*, vol. 3, pp. 245–249, Oct. 1999.