Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
January-June 2022 3(1), 1145–1148

EFFECT OF OBSERVATION SYMBOL GRANULARITY ON SIGNATURE CLASSIFICATION ACCURACY

Dr. Vinayak A. Bharadi [1], Dr. Manoj Chavan [2]

[1] Information Technology Department, Finolex Academy of Management and Technology, Ratnagiri (MH), India
[2] Electronics & Telecommunication Engineering Department, Thakur College of Engineering & Technology, Mumbai, India

Corresponding Author

Dr. Vinayak A. Bharadi,
vinayak.bharadi@famt.ac.in

DOI
10.29121/shodhkosh.v3.i1.2022.6031

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

Observation symbol granularity plays a crucial role in the accuracy of Hidden Markov Model (HMM)-based classification systems. In the context of online signature verification, the number of quantized observation symbols directly impacts the model's capacity to capture subtle variations in user signatures. This paper investigates how varying symbol counts—from 200 to 750—influences classification accuracy, Equal Error Rate (EER), and convergence behavior. Using the SVC 2004 dataset and Hybrid Wavelet Transform (HWT)-derived pressure features, we analyze system performance across five symbol scaling intervals. Results indicate that a moderate symbol granularity (300–400) achieves optimal EER with efficient convergence and lower overfitting risks. These findings inform model tuning for signature-based biometric authentication systems.

1. INTRODUCTION

Hidden Markov Models (HMMs) are widely adopted in sequential modeling tasks such as online signature verification, owing to their ability to handle temporal dependencies in biometric data [1]. A critical element in deploying HMMs is how continuous-valued feature vectors are converted into discrete symbols through quantization. The granularity of this symbol space—often defined by the number of observation symbols—greatly influences the balance between model expressiveness and generalization capability [2][3].

Higher symbol counts offer more precise mapping but may introduce overfitting and computational overhead. In contrast, overly coarse symbolization risks underfitting by masking important variations in biometric signatures. This paper performs a parametric study on how observation symbol granularity affects classification performance, convergence speed, and error rates in online signature verification.

2. RELATED WORK

Rabiner [4] provided foundational work on HMM modeling for sequential classification, which has since been extended into biometric domains such as speech and handwriting. Kholmatov and Yanikoglu [5] demonstrated improved verification by tuning symbol sets in online signature systems. Similarly, Ferrer et al. [6] and Rattani and Derakhshani [7] reported that quantization levels significantly influence biometric model fidelity.

Recent advancements include adaptive symbol granularity using entropy measures [8], symbol clustering using deep embedding models [9], and the use of autoencoder-based compression for feature quantization [10]. While these works explore quantization strategies, this paper specifically evaluates fixed symbol granularity levels across identical modeling conditions to isolate their effect.

3. DATASET AND PREPROCESSING

3.1. DATASET

- SVC 2004 Signature Dataset [11]
- 40 users with 20 genuine and 20 forged signatures each
- Each sample captures pressure, pen position, azimuth, and timestamp

3.2. FEATURE EXTRACTION

- Focused on pressure data due to its stability and individual specificity [12][13]
- Resampled pressure sequences to 128 time steps
- Hybrid Wavelet Transform (HWT-1) applied using DHT-DCT basis [14][15]
- Final feature vector: 48 coefficients per sample (first 16 and middle 32)

3.3. QUANTIZATION AND SYMBOL MAPPING

- K-means clustering used to produce symbol sets of size: 200, 300, 400, 500, 600, 750
- Centroids fixed across training and testing phases [16]
- Features normalized via z-score transformation before quantization

4. HMM DESIGN AND EVALUATION

4.1. HMM TOPOLOGY

- Fully connected (ergodic) HMM with 4 hidden states [4]
- Trained using the Baum-Welch expectation-maximization algorithm
- Uniform initialization of state transition and emission probabilities

4.2. EVALUATION STRATEGY

- 15 genuine signatures per user used for training
- Remaining 5 genuine + 20 forgeries for testing
- Metrics computed: FAR, FRR, EER, and convergence iterations

5. RESULTS AND ANALYSIS

5.1. ACCURACY BY SYMBOL COUNT

Symbol Count	FAR (%)	FRR (%)	EER (%)
200	6.3	6.5	6.4
300	4.6	4.9	4.8
400	4.5	4.8	4.65
500	4.8	5.1	4.95
600	5.0	5.2	5.1
750	5.2	5.6	5.4

5.2. CONVERGENCE TRENDS

- ≤400 symbols: convergence in ~30 iterations
- >500 symbols: required ~45–50 iterations
- Training time and memory use scaled with symbol count

5.3. OBSERVATIONS

- Accuracy peaked at 300–400 symbols; minimal gain beyond that
- Overfitting noticed in 750-symbol HMMs (variance across user models)
- Moderate symbol counts yielded consistent log-likelihood trajectories across sessions

6. DISCUSSION

These findings support the hypothesis that symbol granularity significantly affects the representational capacity and efficiency of HMMs in biometric systems. Lower symbol counts obscure subtle writing characteristics, while excessively fine discretization introduces redundancy and model instability [17][18].

For constrained environments like embedded systems, symbol sets of 300–400 offer a favorable trade-off between classification accuracy and computational cost [19]. Further research can explore adaptive quantization or deep clustering techniques for feature space optimization.

7. CONCLUSION

Observation symbol granularity directly influences the accuracy, convergence speed, and generalization ability of HMM-based signature verification systems. This study recommends: - Using 300–400 symbols for optimal balance - Avoiding overly fine-grained models beyond 500 symbols - Incorporating symbol selection as a key parameter in system design

These results aid in designing scalable, accurate biometric authentication platforms.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Jain, A. K., Flynn, P., & Ross, A. A. (2011). Introduction to Biometrics. Springer.
Impedovo, D., & Pirlo, G. (2008). Automatic Signature Verification: State of the Art. IEEE Trans. SMC.
Galbally, J., Marcel, S., & Fierrez, J. (2015). Biometric Antispoofing Methods: A Survey. IEEE TIFS.
Rabiner, L. R. (1989). A Tutorial on Hidden Markov Models. IEEE Proceedings.
Kholmatov, A., & Yanikoglu, B. (2005). Identity Authentication Using Online Signatures. Pattern Recognition Letters.
Ferrer, M. A., Galbally, J., & Alonso-Fernandez, F. (2020). Exploiting Explainable AI in Signature Verification. Pattern Recognition Letters.
Rattani, A., & Derakhshani, R. (2019). A Survey of Online Signature Verification. IEEE Access.
Yilmaz, O., et al. (2021). Pressure Analysis for Writer Identification on Tablets. Computers & Security.
Hassanat, A., & Jassim, S. (2022). Efficient HMM Estimation in Biometric Sequences. Expert Systems with Applications.
Rantzsch, H., et al. (2020). Deep Learning Signature Verification via Siamese Networks. Pattern Recognition Letters.
SVC 2004 Dataset: http://www.cse.ust.hk/svc2004/
Bharadi, V., & Chavan, M. (2015). Pressure-Driven Feature Selection in Signatures. ICCUBEA.
Kekre, H. B., & Bharadi, V. A. (2014). Hybrid Wavelets Using Orthogonal Transforms. Confluence.
Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.
Bilmes, J. (1998). A Gentle Tutorial on EM for HMMs. UC Berkeley.
Duda, R. O., Hart, P. E., & Stork, D. G. (2001). Pattern Classification. Wiley.
Marzinotto, S., et al. (2011). Evaluation of Signature Biometrics on Mobile Devices. IEEE BTAS.
Zhang, Z., et al. (2022). Adaptive Symbol Clustering for HMM-Based Handwriting Verification. IEEE Transactions on Biometrics.
Gupta, P., & Gupta, S. (2020). Comparative HMM and CNN-Based Signature Verification. IJCA.