Design and Implementation of Hand Gesture Detection System Using HM Model for Sign Language Recognition Development

Sharmin Akter Milu[1], Azmath Fathima[2], Tanmay Talukder[3], Inzamamul Islam[4*], Md. Ismail Siddiqi Emon[5]

[1]Department of Computer Science & Telecommunication Engineering (CSTE), Noakhali Science & Technology University, Noakhali, Bangladesh
[2]School of Computer, Mathematical & Natural Sciences (SCMNS), Morgan State University, Baltimore, MD, USA
[3]Department of Physics, Morgan State University, Baltimore, MD, USA
[4]Department of ECE, Morgan State University, Baltimore, MD, USA
[5]Department of Computer Science, Morgan State University, Baltimore, MD, USA
Email: *inzy.2018@gmail.com

Abstract

Gesture detection is the primary and most significant step for sign language detection and sign language is the communication medium for people with speaking and hearing disabilities. This paper presents a novel method for dynamic hand gesture detection using Hidden Markov Models (HMMs) where we detect different English alphabet letters by tracing hand movements. The process involves skin color-based segmentation for hand isolation in video frames, followed by morphological operations to enhance image trajectories. Our system employs hand tracking and trajectory smoothing techniques, such as the Kalman filter, to monitor hand movements and refine gesture paths. Quantized sequences are then analyzed using the Baum-Welch Re-estimation Algorithm, an HMM-based approach. A maximum likelihood classifier is used to identify the most probable letter from the test sequences. Our method demonstrates significant improvements over traditional recognition techniques in real-time, automatic hand gesture recognition, particularly in its ability to distinguish complex gestures. The experimental results confirm the effectiveness of our approach in enhancing gesture-based sign language detection to alleviate the barrier between the deaf and hard-of-hearing community and general people.

Keywords

Hand Gesture Recognition System

1. Introduction

Sign language is an important part of communication for people with speaking and hearing disabilities. In the modern world, almost every spoken language has its own verified sign language which is parallel to the written language. Most hearing people do not understand sign language and learning it is not an easy process. As a result, there is still an undeniable barrier between the hearing impaired and hearing majority. With the improvement of image recognition and computer speed, we can easily adopt the translation and make the communication easy, convenient, and fast. Gesture detection is the primary and most significant step for sign language detection.

Gestures are a natural and expressive method of communication between humans and computers in virtual reality systems. Due to their increased expressiveness and freedom of movement, hand gestures are an essential component of non-verbal communication between humans and computer systems. Hand gesture recognition has potential applications in multiple fields, including human-computer interaction (HCI), machine vision, virtual reality, and industrial machine control. In recent years, Human-Computer Interaction (HCI) has been a crucial area of study due to the growing demand for intuitive and natural interactions between humans and machines. Hand gesture recognition has emerged as one of the most promising HCI methods due to its naturalness, usability, and non-invasiveness. Hand gesture recognition enables users to interact with computers by communicating their intentions through hand movements, without the need for additional hardware or software. Several applications, such as games, virtual reality, augmented reality, and industrial automation, can benefit from this way of interacting, which can make the user's experience much better and make these applications more efficient and effective. But accurate hand gesture recognition in real time is hard because of things like lighting, occlusions, and the fact that individuals' hands are different in sizes and shapes. To deal with the challenge and overcome it, researchers used various types of machine learning approaches. Among them, the Hidden Markov model showed up and gained great and considerable attention.

The Hidden Markov model (HMM) is a statistical model first proposed by Baum L. E. that employs a Markov process with unknown and hidden parameters [1]. In recent years, Hidden Markov Models (HMMs) have attracted considerable interest due to their successful applications in speech recognition [2] and handwriting recognition [3]. These applications have shown how well HMMs can capture the temporal dynamics of sequential data and make it easier to recognize patterns accurately. Consequently, researchers have begun investigating the use of HMMs in spatio-temporal pattern recognition and computer vision [4] [5]. Utilizing Baum-Welch and other re-estimation techniques, HMMs can estimate model parameters from observation sequences, which is primarily responsible for their widespread application. To get the most out of HMMs for pattern recognition tasks like gesture recognition and computer vision applica-

tions, there is a growing interest in researching advanced techniques for training these models on multiple observation sequences.

In this work, a hidden Markov model (HMM)-based framework has been implemented for recognizing and detecting hand gestures.

2. Literature Review

Recognizing hand gestures is difficult due to the spatial-temporal variability of dynamic gestures. We observed various models that have been proposed [5] [6] [7] [8] [9] for this work such as Neural Network, Fuzzy Systems, and Hidden Markov Models (HMMs) [1] which completely differ from one another. As said before HMM received a wide application in handwriting recognition, character recognition and gesture recognition compared to other approaches due to their ability to model spatial-temporal time series.

However, there are challenges associated with the use of HMMs in hand gesture recognition. The spatial-temporal variability of dynamic gestures, such as differences in velocity, shape, duration, and integration, makes it challenging to accurately recognize them. Also, HMM-based methods that use local features can be affected by the sampling period and speed, which can lead to false recognition. To get around these problems, researchers have to look at the hand gesture's whole trajectory shape as well as how it moves at each time point. This will enhance recognition accuracy and decrease false recognition.

We observed some works available that employed HMM for hand gesture recognition.

Xiayan Wang *et al.* [10] propose a method for the automatic recognition of hand gestures using depth data. To recognize hand gestures, the proposed method combines techniques for feature extraction, feature selection, and classification. The extraction of features is based on shape and motion features, whereas a genetic algorithm is used to select the most relevant features for feature selection. Using a k-nearest neighbor (KNN) classifier, the classification is carried out. Nianjan Liu *et al.* [11] propose a method for offline recognition of cursive handwriting based on the recognition of discrete characters. In the proposed approach, the cursive script is broken up into discrete characters, and then a classifier is used to recognize each character on its own.

An automatic speaker recognition method based on discrete wavelet transform (DWT) and support vector machine (SVM) is presented by Shengluan Huang *et al.* [12]. The proposed method first uses discrete wavelet transform (DWT) to extract features from speech signals before moving on to support vector machines (SVM) for voice classification. The proposed method was tested with a collection of speech signals, and the outcomes demonstrate its ability to provide accurate recognition. When compared to other state-of-the-art methods, the proposed method is found to be more accurate at recognizing the input.

Nianjun Liu *et al.* [13] examine various training algorithms for Hidden Markov Models (HMMs) employed in letter based hand gesture recognition. The

purpose of this paper is to evaluate the performance of various HMM training algorithms in terms of recognition precision and computational efficiency. The study evaluates the performance of HMMs trained with two different algorithms, Baum-Welch and Viterbi using a dataset of letter hand gestures performed by various users. Different metrics, like precision, recall, and F1 score, are used to measure how well HMMs can recognize things. The study found that, when comparing the both algorithms, the Baum-Welch algorithm provided the highest recognition accuracy. However, the Viterbi algorithm is the most efficient in terms of computation. Additionally, the paper discusses the impact of various parameters, such as the number of hidden states and the size of the training dataset, on the performance of HMMs. Numerous studies have showcased the application of this methodology, and recent scientific advancements further underscore enhancements in our research [14]-[23].

3. Methodology

3.1. Dynamic Hand Gesture

In the realm of gesture recognition, a dynamic hand gesture embodies a spatio-temporal pattern characterized by four principal attributes: velocity, configuration, spatial positioning, and directional orientation. The kinematics of a hand gesture is expressible as a chronological array of spatial coordinates, centered around the centroid of the hand of the executing individual. This study chooses to abstract away from the morphological aspects of the hand, focusing instead on the locational dynamics of the gesture. Each instance of a dynamic hand gesture is thus distilled into a temporal series, defined as:

$$P_t = (x_t, y_t), (t = 1, 2, \cdots, T)$$

Here, T denotes the path length of the gesture, a parameter subject to variation across different gesture examples. In essence, such a gesture can be conceptualized as a sequential mapping from temporal progression to spatial coordinates. Figure 1 elucidates an instance of a dynamic hand gesture, illustrating its temporal unfolding and spatial projection onto a two-dimensional plane. Figure 1 shows these graphs demonstrate the hand gesture path using a parametric model, with visual elements to aid in understanding the sequence of the movement.

3.2. Implementation Procedure

For the work hand gesture was recorded in a video stream. This recorded video is composed in multiple frames. Further skin color segmentation was done. Further to smooth the image and remove noise morphological operation carried out as a part preprocessing. After this image moments were calculated to find the hand is centroid in each frame. For each movement (35) or trajectory orientation was calculated which was quantized later on. Our system flow chart can be observed from Figure 2 & Figure 3.

3D Plot of Dynamic Hand Gesture Path

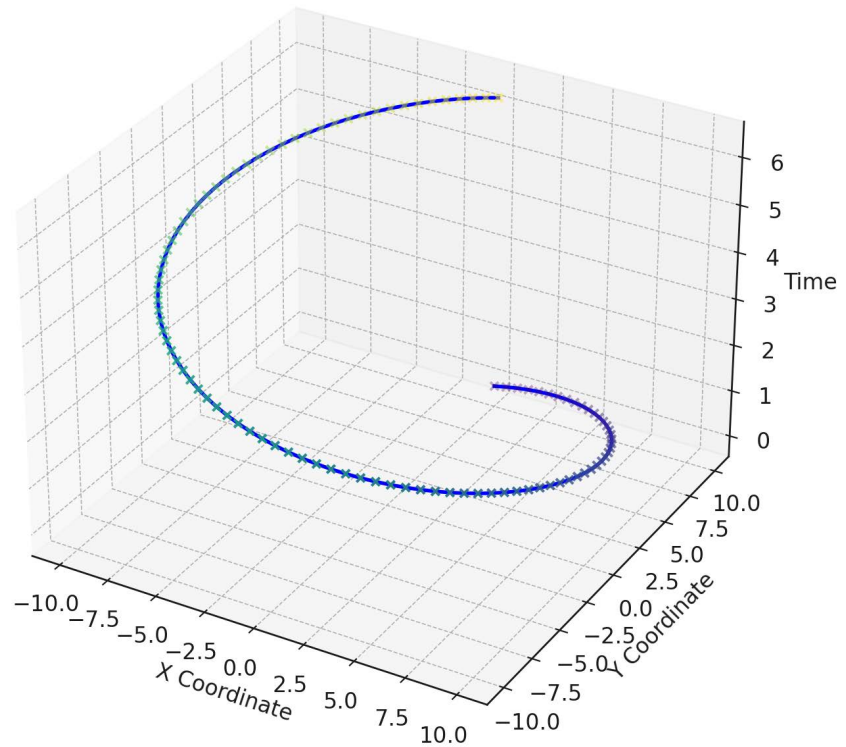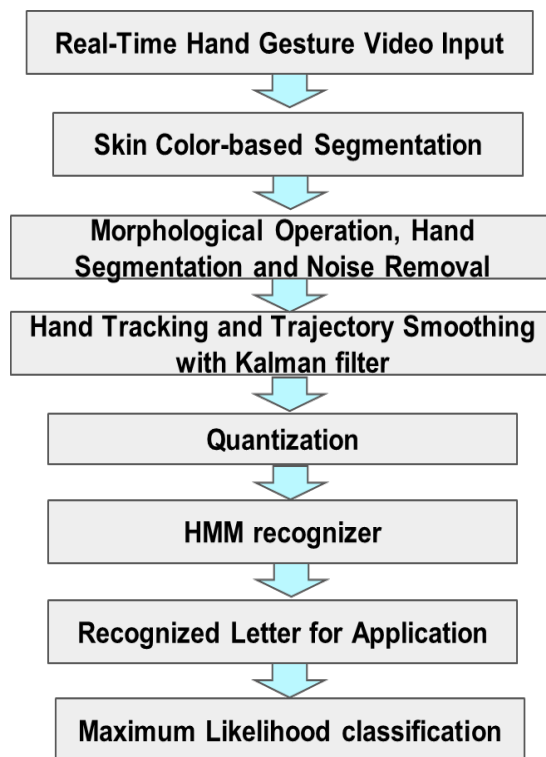**Figure 1.** 3D plot of dynamic hand gesture path.
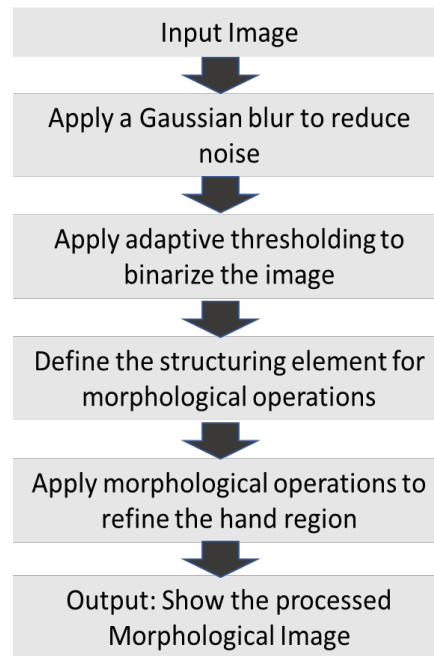

Figure 2. System flowchart.

Figure 3. Morphological analysis.

Further Kalman filter was employed to track hand and trajectory smoothing. The Kalman filtering algorithm is made to figure out where an object is most likely to be in the current frame based on where it was in the previous frame. It accomplishes this by searching the neighboring area of the predicted target location. If the search area contains a target, the algorithm advances to the next frame for processing. The Kalman filter's prediction and update features are key to its effective functioning. Input video divided into 35 individual frames. Kalman filter applied on a full set of frame Kalman filter processed it and returned Observation Matrix Transition Matrix Emission Matrix. This values further used for Baum-Welch re-estimation process.

Quantization or vector quantization is typically employed for compressing audio (voice) and image data. However, it also has voice recognition and pattern recognition applications. Quantization was utilized to reduce the collection of continuous features to a single discrete representation in this study. This is a useful representation for discrete HMM. After smoothing the data, the orientation between consecutive points was calculated using an angle range of 0 to 360 degrees. This 360 degree range was divided by 20 quantizations and we obtained 18 different directional quantized data.

3.3. HMM Application

We define a gesture letter as a sequence of directional angles which are the observation symbols. Each letter is mapped to one hidden Markov model. We adopted the traditional Baum-Welch [1] to train the Hidden Markov Models over a range of model structures. To do that we defined Number of States, Observation Matrix, Transmission Probability Matrix, Emission Matrix which were

Fed to the data Baum-welch Re Estimation formulae. Determine Forward probability from observation sequence, start probability, Transition probably. Determine Backward probability from observation sequence, num of state, Transition probably, Emission matrix. The Baum-welch Re Estimation formulae can be observed from below-

$$\tilde{a}_{ij} = \frac{\Sigma_k W_k \Sigma_t^T \underline{k_1} a_i^k a_{ij} b_j \left(O_{t+1}^{(k)} \right) \beta_{t+1}^{(k)}(j)}{\Sigma_k W_k \Sigma_t^T \underline{k_1} a_i^k (i) \beta_t^{(k)}(i)}$$

$$\tilde{b}_{ij} = \frac{\Sigma_k W_k \Sigma_{O_t(k)=v_j} \alpha_t^k (i) \beta_t^{(k)}(i)}{\Sigma_k W_k \Sigma_t^T k_1 \alpha_i^k (i) \beta_t^{(k)}(i)}$$

Further maximum likelihood was calculated to predict the alphabet. Maximum Likelihood classifier consists of a log likelihood method which takes the input class means and class covariance. From which we predict the alphabet. We can obtain it from the following equation.

$$\hat{\ell}(\theta;x) = \frac{1}{n}\sum_{i=1}^{n} \ln f(x_i|\theta)$$

4. Results and Discussions

In the context of advancing human-computer interaction (HCI) methodologies, our research introduces a specialized interface centered on hand gesture recognition. This interface integrates seamlessly with standard webcams affixed to personal computers, which are instrumental in real-time acquisition of hand gesture imagery. Optimal performance of these webcams is contingent upon two critical specifications: a minimum frame rate of 25 frames per second and a capture resolution threshold of 640×480 pixels. The system is primarily designed for indoor environments, characterized by consistent background settings and stable lighting conditions.

The interface's recognition capabilities encompass an array of hand gestures, involving diverse poses and movements. We have programmed the system to recognize and accurately track seven distinct hand gestures, which include:

1) Tracing three consecutive circles in a horizontal sequence via aerial hand movements.

2) Simulating the action of drawing a question mark in the air with one's hand.

3) Drawing three sequential circles vertically through air-bound hand maneuvers.

4) Elevating the hand in a straight vertical trajectory.

5) Conducting a left-to-right hand waving motion.

6) Executing a right-to-left hand waving sequence.

7) Forming an exclamation mark in the air using hand gestures.

For the quantification of local orientational features, we have empirically chosen 18 as the optimal number of codewords.

In this study, the optimal number of states for each gesture within the Hidden

Markov Model (HMM) was empirically determined through experimental analysis. Specifically, it was observed that for gestures "I" and "R", increasing the state count beyond 12 did not significantly improve recognition accuracy. For the remaining gestures "P", "S", and "Z", setting the state count to 10 was found to be optimal. Hence, this configuration of states was applied in all further experiments.

Our research involved a comprehensive collection of gesture trajectory samples. Over 900 samples for each gesture type were acquired from five individuals, forming the dataset for the training phase. In the validation phase, more than 340 samples for each gesture were obtained from a different group of eight participants. These results are systematically presented in Table 1 and Figure 4 shows the accuracy comparison between our and traditional methods. It can be seen that the proposed method can greatly improve the recognition process.

The analysis of this data highlighted a marked improvement in the recognition of complex gestures, particularly "I" and "R". The challenge in differentiating between these two gestures stemmed from their temporal similarities when analyzed using only local features. However, the algorithm we developed effectively addresses this issue, demonstrating a high degree of accuracy in distinguishing between gestures that exhibit closely related temporal patterns. For actual recognition examples we can observe Figures 4-8 where we showed the original image and morphological image after recognition.

Table 1. Hand gesture recognition results.

Gestures	Test sets' numbers	Our method (%)	Traditional method (%)
I	340	94	93
P	460	92	90
R	430	87	87
S	454	88	87
Z	439	91	90
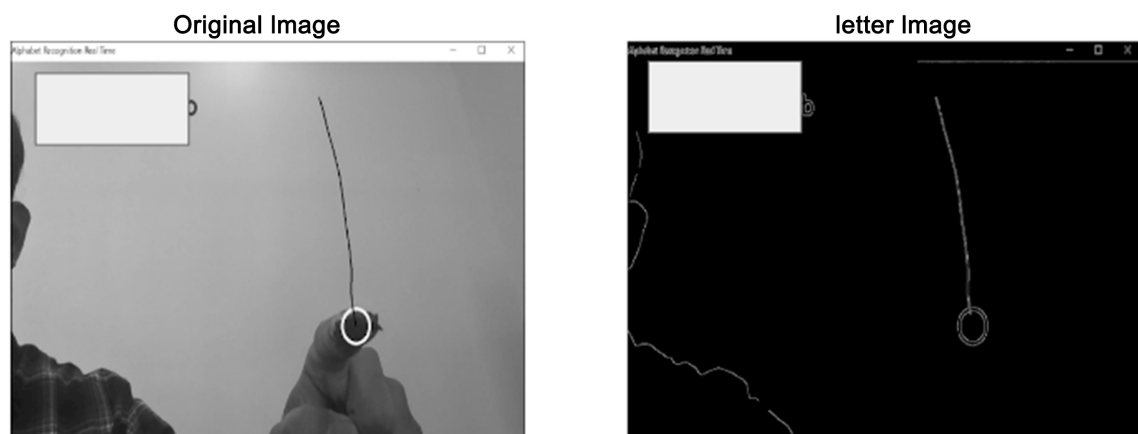
Figure 4. I gesture recognition and morphological image.

Original Image   Morpholgy applied Image
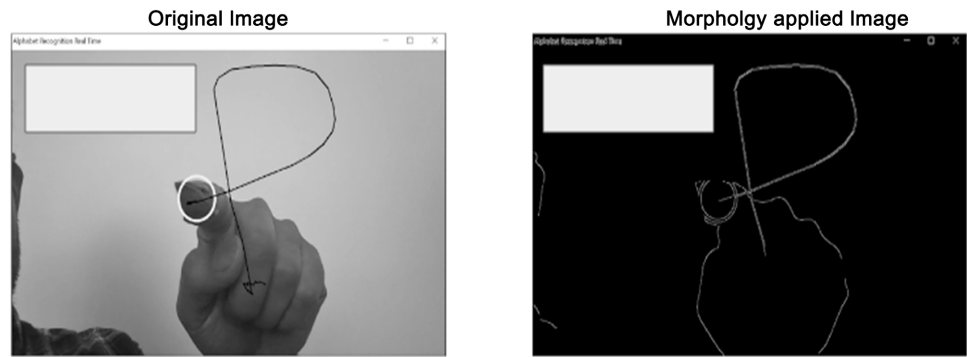
Figure 5. P gesture recognition and morphological image.

Original Image   Marpholgied Image
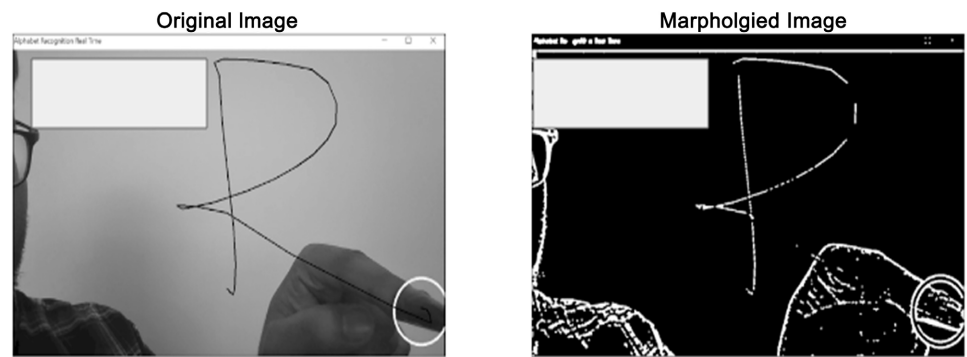
Figure 6. R gesture recognition and morphological image.

Original Image   Morpholgied Image

Figure 7. S gesture recognition and morphological image.

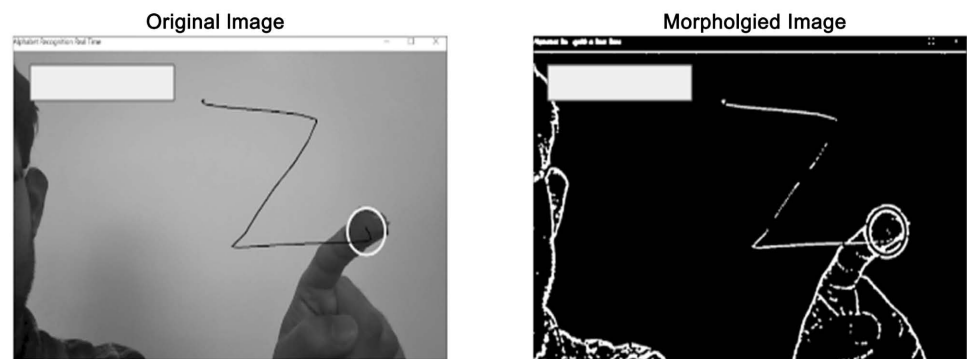Original Image   Morpholgied Image

Figure 8. Z gesture recognition and morphological image.

5. Conclusion

This study presents a breakthrough in hand gesture recognition for sign language detection, employing Hidden Markov Models (HMMs) to accurately interpret hand movements. Utilizing standard webcams, our system effectively captures and processes hand gestures corresponding to five English alphabet letters ("I", "P", "R", "S", "Z"). Key to our methodology is the optimized use of states within the HMM, tailored specifically for each gesture, which significantly enhances recognition accuracy. Our comprehensive dataset, with over 900 training samples and 340 validation samples, demonstrates the system's robustness. A notable achievement is the system's ability to discern between gestures with similar temporal features, particularly the "I" and "R" gestures, a task that posed a significant challenge when relying solely on local features. The integration of skin color-based segmentation, morphological operations, and advanced hand tracking technologies, such as the Kalman filter, addresses common issues in gesture recognition like variable lighting conditions, occlusions, and individual hand differences. The application of the Baum-Welch Re-estimation Algorithm within the HMM framework, coupled with a maximum likelihood classifier, has proven critical in the system's ability to distinguish complex hand gestures in real-time. Our results, as outlined in Table 1 and accompanying figures, not only confirm the effectiveness of our method but also establish its superiority over traditional hand recognition approaches. This advancement has significant implications for enhancing interactive experiences across various applications, from virtual reality to industrial automation along with sign language detection.

6. Future Work

Future work could focus on expanding the system's application in more diverse and challenging environments, further cementing its utility in next-generation HCI systems to employ it in sign language detection. In particular, our future plan is to utilize this gesture recognition-based system to detect regional sign language specially for our own Bengali sign language.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Baum, L.E. and Petrie, T. (1966) Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *The Annals of Mathematical Statistics*, **37**, 1554-1563. https://doi.org/10.1214/aoms/1177699147

[2] Rabiner, L.R. and Juang, B.H. (1993) Fundamentals of Speech Recognition. Prentice Hall, New Jersey.

[3] Lee, J.J. and Kim, J.H. (2001) Data-Driven Design of HMM Topology for Online Handwriting Recognition. *International Journal of Pattern Recognition and Artifical Intelligence*, **15**, 107-121. https://doi.org/10.1142/S0218001401000769

[4] Wilson, A.D. and Bobick, A.F. (2001) Hidden Markov Models for Modeling and Recognizing Gesture under Variation. *International Journal of Pattern Recognition and Artificial Intelligence*, **15**, 123-160. https://doi.org/10.1142/S0218001401000812

[5] Nefian, A. and Hayes, M. (1998) Hidden Markov Models for Face Recognition. *Proceedings of the International Conference on Acoustics*, *Speech*, *and Signal Processing* (*ICASSP*98), Seattle, 12-15 May 1998, 2721-2724.

[6] Xu, D.Y. (2006) A Neural Network approach for Hand Gesture Recognition in Virtual Reality Driving Training System of SPG. *Proceedings of the* 18*th International Conference on Pattern Recognition* (*ICPR* '06), Hong Kong, 20-24 August 2006, 519-522. https://doi.org/10.1109/ICPR.2006.109

[7] Hsieh, C.-C., Liou, D.-H. and Lee, D. (2010) A Real Time Hand Gesture Recognition System Using Motion History Image. 2010 2*nd International Conference on Signal Processing Systems*, Dalian, 5-7 July 2010, V2-394-V2-398. https://doi.org/10.1109/ICSPS.2010.5555462

[8] Holden, E., Owens, R. and Roy, G. (1996) Hand Movement Classification Using an Adaptive Fuzzy Expert System. *International Journal of Expert Systems*, **9**, 465-480.

[9] Elmezain, M., Al-Hamadi, A. and Michaelis, B. (2008) Real-Time Capable System for Hand Gesture Recognition Using Hidden Markov Models in Stereo Color Image Sequences. *Journal of WSCG*, **16**, 65-72.

[10] Wang, X.Y., Xia, M., Cai, H.W., Gao, Y. and Cattani, C. (2012) Hidden-Markov-Models-Based Dynamic Hand Gesture Recognition. *Mathematical Problems in Engineering*, **2012**, Article ID: 986134. https://doi.org/10.1155/2012/986134

[11] Liu, N.J., Lovell, B.C., Kootsookos, P.J. and Davis, R.I.A. (2004) Model Structure Selection & Training Algorithms for an HMM Gesture Recognition System. *Ninth International Workshop on Frontiers in Handwriting Recognition*, Tokyo, 26-29 October 2004, 100-105.

[12] Huang, S.L. and Hong, J.X. (2011) Moving Object Tracking System Based on Camshift and Kalman Filter. 2011 *International Conference on Consumer Electronics*, *Communications and Networks* (*CECNet*), Xianning, 16-18 April 2011, 1423-1426. https://doi.org/10.1109/CECNET.2011.5769081

[13] Liu, N.J., Lovell, B.C. and Kootsookos, P.J. (2003) Evaluation of HMM Training Algorithms for Letter Hand Gesture Recognition. *Proceedings of the* 3*rd IEEE International Symposium on Signal Processing and Information Technology* (*IEEE Cat. No.* 03*EX*795), Darmstadt, 17 December 2003, 648-651.

[14] Nahar, L., Zabu, Z.A., Raihan, A., Islam, M.I. and Emon, M.I. (2022) A Comparative Selection of Best Activation Pair Layer in Convolution Neural Network for Sentence Classification Using Deep Learning Model. 2022 7*th International Conference on Communication and Electronics Systems* (*ICCES*), Coimbatore, 22-24 June 2022, 1398-1404. https://doi.org/10.1109/ICCES54183.2022.9835806

[15] Alam, S.U., Uddin, R., Alam, M.J., Raihan, A., Mahtab, S.S. and Bhowmik, S. (2022) Mathematical Modeling and Performance Evaluation of 3D Ferroelectric neGative Capacitance FinFET. *Modelling and Simulation in Engineering*, **2022**, Article ID: 8345513. https://doi.org/10.1155/2022/8345513

[16] Minhajul Alam, S.M., Barua, A., Raihan, A., Alam, M.J., Chakma, R., Mahtab, S.S. and Biswas, C. (2021) Design and Implementation of a Smart Helmet System for Underground Miner's Safety. In: Bindhu, V., Tavares, J.M.R.S., Boulogeorgos, A.A.A. and Vuppalapati, C., Eds., *International Conference on Communication, Computing and Electronics Systems*, Springer, Singapore, 301-311.

https://doi.org/10.1007/978-981-33-4909-4_22

[17] Alam, M.J., Mahbub, T., Rahman, M., Anonto, R.A., Raihan, A. and Mahtab, S.S. (2021) Effect of Anti Reflecting Coating (ARC) in GaAs Based P-I-N Solar Device with Gap as Window Layer. *AIP Conference Proceedings*, **2327**, Article ID: 020046. https://doi.org/10.1063/5.0040004

[18] Ali, F.B., Chakma, R., Raihan, A., Hasan, M., Khan, M.F., Alam, M.J., Akter, R. and Mahtab, S.S. (2021) Design and Implementation of a Time-Based Sun Tracking Solar System. *AIP Conference Proceedings*, **2327**, Article ID: 020024.

[19] Mahtab, S.S., Akter, R., Mahbub, T., Raihan, A., Anonto, R.A. and Alam, M.J. (2021) Impact Analysis of Anti-Reflection Coating on P-i-N Solar Device. In: Bindhu, V., Tavares, J.M.R.S., Boulogeorgos, A.A.A. and Vuppalapati, C., Eds., *International Conference on Communication, Computing and Electronics Systems*, Springer, Singapore, 279-287. https://doi.org/10.1007/978-981-33-4909-4_20

[20] Akter, R., Mahin, M.M., Mahbub, T., Nahreen, K., Raihan, A., Chakma, R., Alam, M.J. and Mahtab, S.S. (2021) Solar Cells: Varieties and Utilization—A Short Review. *AIP Conference Proceedings*, **2327**, Article ID: 020023. https://doi.org/10.1063/5.0039501

[21] Hossain, M.D.I., Akter, R., Hasan, M., Raihan, A., Khan, M.F., Ali, F.B., Chakma, R., Alam, M.J. and Mahtab, S.S. (2021) Generation of Electrical Energy Using Road Transport Pressure at Speed Breaker. *AIP Conference Proceedings*, **2327**, Article ID: 020025. https://doi.org/10.1063/5.0039999

[22] Raihan, A. (2023) Structural, Optical, and Electrical Properties of WSe2 Thin Films Synthesized by Magnetron Sputtering. Master's Thesis, Morgan State University, Baltimore.

[23] Raihan, A., Jyotsna, D., Ravinder, K. and Dereje, S. (2023) Controlled Growth and Transport Property of Bi_2Se_3 and WSe_2 Thin Film. *APS March Meeting Abstracts*, **2023**, SS01-007.