A Bag-of-Words Based Feature Extraction Scheme for American Sign Language Number Recognition from Hand Gesture Images

Rasel Ahmed Bhuiyan, Abdul Matin, Md. Shafiur Raihan Shafi, and Amit Kumar Kundu

***Abstract*—Human Computer Interaction (HCI) focuses on the interaction between humans and machines. An extensive list of applications exists for hand gesture recognition techniques, major candidates for HCI. The list covers various fields, one of which is sign language recognition. In this field, however, high accuracy and robustness are both needed; both present a major challenge. In addition, feature extraction from hand gesture images is a tough task because of the many parameters associated with them. This paper proposes an approach based on a bag-of-words (BoW) model for automatic recognition of American Sign Language (ASL) numbers. In this method, the first step is to obtain the set of representative vocabularies by applying a K-means clustering algorithm to a few randomly chosen images. Next, the vocabularies are used as bin centers for BoW histogram construction. The proposed histograms are shown to provide distinguishable features for classification of ASL numbers. For the purpose of classification, the K-nearest neighbors (kNN) classifier is employed utilizing the BoW histogram bin frequencies as features. For validation, very large experiments are done on two large ASL number-recognition datasets; the proposed method shows superior performance in classifying the numbers, achieving an F1 score of 99.92% in the Kaggle ASL numbers dataset.**

***Index Terms*—Human computer interaction (HCI), hand gesture recognition (HGR), American sign language (ASL), bag-of-words (BoW), kNN classifier.**

I. Introduction

The human computer interface (HCI) refers to the user inter-aces in a production or process-control system; it deals with the design, implementation, and assessment of new interfaces to improve the interaction between humans and machines [1], [2]. An efficient, robust, and customized interface can greatly reduce the gap between a human's mental model and the way a computer or robot accomplishes a given task. In recent industrial scenarios, gestures, hand and body poses, speech, and gaze are among the many natural interaction modes that can be used to design affordable user interfaces [3]. Among all the modes of natural interaction, gesturing serves as one of the most comfortable and expressive ways to conduct effective and meaningful communication between two individuals from different cultures; this is true even for physically challenged people, including those who are hearing-impaired or dumb [1]-[6]. However, very few people can understand hand gestures properly [1], so a communication gap exists that isolates those in the impaired community from the mass of people. Because of this, automatic hand gesture recognition (HGR) has become a major concern for researchers [7], [8]. It has been applied in some interesting fields, such as gaming, sign language recognition (SLR), and virtual reality [9]. For applications in the real world, automatic HGR faces significant challenges caused by its needs for accuracy and robustness [1], [2], [8]. There is a comprehensive and detailed analysis of existing research techniques for the recognition of sign language in [1]. It is accompanied by a discussion of the usual challenges for gesture recognition systems; the work aims to guide entry into (and to facilitate increasing efforts in) the SLR research field. Most studies of SLR are based on American Sign Language (ASL), Indian sign language, and Arabic sign language [1]. Various state-of-the-art HGR methods have been employed. Among them are methods based on hidden Markov models (HMMs), neural networks (NNs), and fuzzy logic; as they incorporate some complex processes [10] and varying parameters [11], their computational costs seem to be high [12]-[14]. A user valuation study, [7], conducted on 25 visually challenged people with the aim of enabling the impaired community to use hand gestures to interact with machines, has led the way to the proposal of an innovative dactylology. A quantitative rating analysis of the subjects' performances has led to the creation of an ideal collection of gestures. The same literature presents a module, accompanied by the proposed dactylology, aimed at recognizing dactylological symbols and enabling a writing support system. Kirsti and Thad [11], [15] explores the achievement of SLR through the use of HMMs, while Keshav [16] uses HMM and hand trajectory tracking techniques in an SLR system created to identify Roman numbers and Arabic alphabets. However, significant performance is not achieved when an SLR system uses a context-dependent HMM model only [17]. Sharmila [2] implements an approach that involves using different image processing and machine learning algorithms to recognize the ASL by means of hand gestures; this suggests the possibility of operating a system with no direct human touch. Gongfa [18] develops a method with moderate accuracy; based on a skeletonization algorithm and Convolutional Neural Network (CNN), it can reduce the effect of shooting angle and surroundings, both of which have a massive impact on recognition. Jayashree [4] proposes an approach that uses the minimum number of possible constraints and achieves a

Manuscript received December 11, 2019; revised May 23, 2020.

Rasel Ahmed Bhuiyan and Abdul Matin are with the Department of Computer Science and Engineering, Uttara University, Dhaka, Bangladesh (e-mail: rasel.cse@uttarauniversity.edu.bd, matin.cse.pust@gmail.com).

Md. Shafiur Raihan Shafi is with the Department of Computer Science and Engineering, Southeast University, Dhaka, Bangladesh (e-mail: shafiur.raihan@seu.edu.bd).

Amit Kumar Kundu was with the Department of Electrical and Electronics Engineering, Uttara University, Dhaka, Bangladesh (e-mail: amit31416@gmail.com).

doi: 10.18178/ijmlc.2021.11.1.1018

satisfactory detection rate; it identifies 26 different ASL alphabets in the presence of complex backgrounds that include varying lighting conditions, hand shapes and hand sizes. Antonakos [19] proposes a semi-supervised approach for classifying the extreme states from facial cues in sign language videos. Jaya [20] provides a feature extraction approach to identify ASL alphabets; using an SVM classifier, it is based on the DWT and F-ratio. Nasser [21] presents another gesture recognizer that uses bag-of-features and multiclass SVM techniques along with SIFT and K-means clustering. Ching [22] uses LMC for the recognition of ASL while Teak and Wenjinu [3], [5] proposes methods based on deep learning. Wenjinu [5] describes a unique approach to recognizing the ASL alphabet from depth images; it uses CNN along with multi-view augmentation and an inference fusion technique. Though this method outperforms the state-of-the-art methods with respect to some specific symbols, some of its technicalities lead to the misclassification of other signs and thence to the failure of recognition in some cases.

The objective of this paper is to develop an automatic scheme, based on bag-of-words (BoW) histogram features, for ASL number recognition. The first step is to extract a set of representative vocabularies by applying a K-means clustering algorithm to the pixel intensities of each of a few training images, randomly selected from each class of ASL numbers. At the feature extraction stage, these vocabularies are used as bin centers of the BoW-based histogram. For feature extraction, the pixel intensities of the image are mapped to the nearest vocabulary to construct the BoW-based histogram. The histogram bin frequencies are used as the feature vector. Finally, a supervised K-nearest neighbors (kNN) classifier is employed for classification. Experiments are done using a large number of captured images to evaluate the performance of the proposed method using a ten-fold cross-validation scheme.

The rest of the paper is organized as follows: Section II presents the proposed method of ASL number recognition in detail; in Section III, different experiments are performed to evaluate the performance of the proposed method; Section IV presents the concluding remarks.

II. Proposed BoW-Based Recognition Scheme for ASL Numbers

This section presents the proposed method in detail. It consists of five steps: preprocessing; BoW vocabulary extraction; feature extraction based on the BoW histogram; and classification with a kNN classifier.

A. Preprocessing

The many noisy background pixels in any image intended for ASL recognition may degrade the feature quality if they are included. To reduce noise and subtract the background, therefore, a preprocessing step is necessary. A captured back-ground image is subtracted from the given image and the resulting image filtered with a median filter to reduce noise. For simplicity, the filtered image is converted into a grayscale image, which is then submitted for feature extraction. Fig. 1 shows examples of the preprocessed grayscale images. The images in the first row represent the captured RGB images, those in the second row the corresponding grayscale images after background subtraction and noise reduction. Preprocessing enhances the hand gesture images and reduces background noise, so the step is expected to improve the feature quality for ASL number recognition.

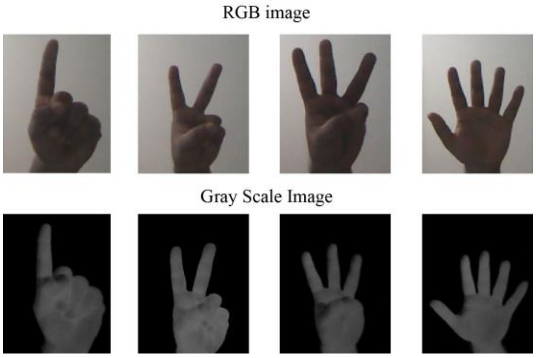

Fig. 1. Examples of preprocessed grayscale images for ASL number classification. The images in the first row represent the captured RGB images, those in the second row the corresponding grayscale images after background subtraction and noise reduction.

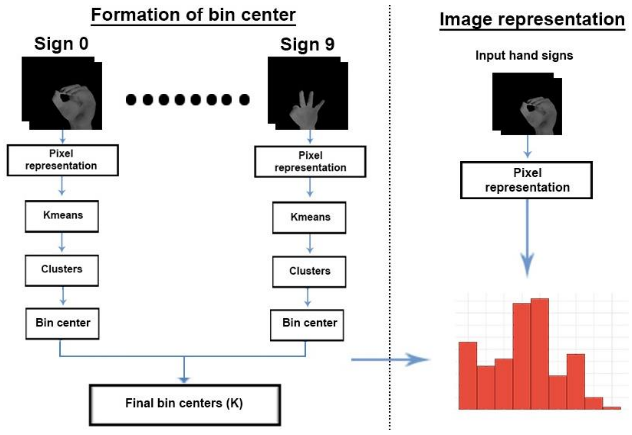

Fig. 2. Constructing a histogram based on the BoW model.

B. BoW Vocabulary Extraction

The efficient solution of a classification problem depends heavily on the quality of the extracted feature; extraction is therefore a challenging task for ASL number recognition, especially when a large dataset is involved. BoW-based feature extraction schemes are, however, widely popular for many pat-tern recognition problems [23], so this paper proposes such a scheme, using the pixel intensities for histogram construction. To begin, 10% images are randomly chosen for each class from the entire training set for BoW vocabulary formulation. Fig. 2 shows the procedure for BoW vocabulary construction.

The pixel intensities of the images are fed a K-means clustering algorithm [24] to obtain k cluster centers for each class. The k cluster centers are treated as the BoW vocabularies, $v_i = \{c_1^i, c_2^i, c_3^i, ..., c_k^i\}$ for the i-th considered class. The BoW vocabularies of each class are obtained using this method and then grouped together to form a BoW dictionary,

$$D = \bigcup_{i=1}^{N} Vi \tag{1}$$

where N is the total number of considered classes. Later, in both training and test phases, the vocabularies in D are considered as the histogram-bin centers for constructing a BoW-based histogram from an image.

C. Proposed BoW-Based Histogram Feature Extraction

After vocabulary or bin-center extraction, the next step is to obtain a BoW-based histogram from an image *I* for ASL number recognition. To do this, the pixel intensities of the image are mapped to the nearest vocabulary for histogram construction. First, the bin frequencies hm of all the bin centers ∈ *D* are initialized *t* zero. $h_m = 0$; $\forall_m \in D$. Then, the distances $d_m = \{dist(p_{ij}, c_m)\}$; $\forall_m \in D$ are calculated from the pixel intensity p_{ij} of image *I* for all the BoW histogram-bin centers. The *dist* function represents the Euclidean distance. The bin frequency of a bin center is increased by one if the bin center has a minimum distance from the pixel intensity of all bin centers. In this way, a histogram is created considering all the pixels in each sign image. Fig. 3 shows an example of BoW histogram construction for the ten classes considered. It is clear from the figure that the proposed BoW histograms are distinct for each of the ten classes under examination, a characteristic that can be considered as a strong feature for ASL number recognition. BoW histogram-bin frequencies are therefore used as main features for ASL number recognition in this paper. In this proposed BoW-based feature extraction scheme, pixel intensities of each grayscale image are used to extract the bin centers or for histogram construction. If the available sign images are in color, pixel intensity triplets could be used as local features for final histogram-based feature extraction.

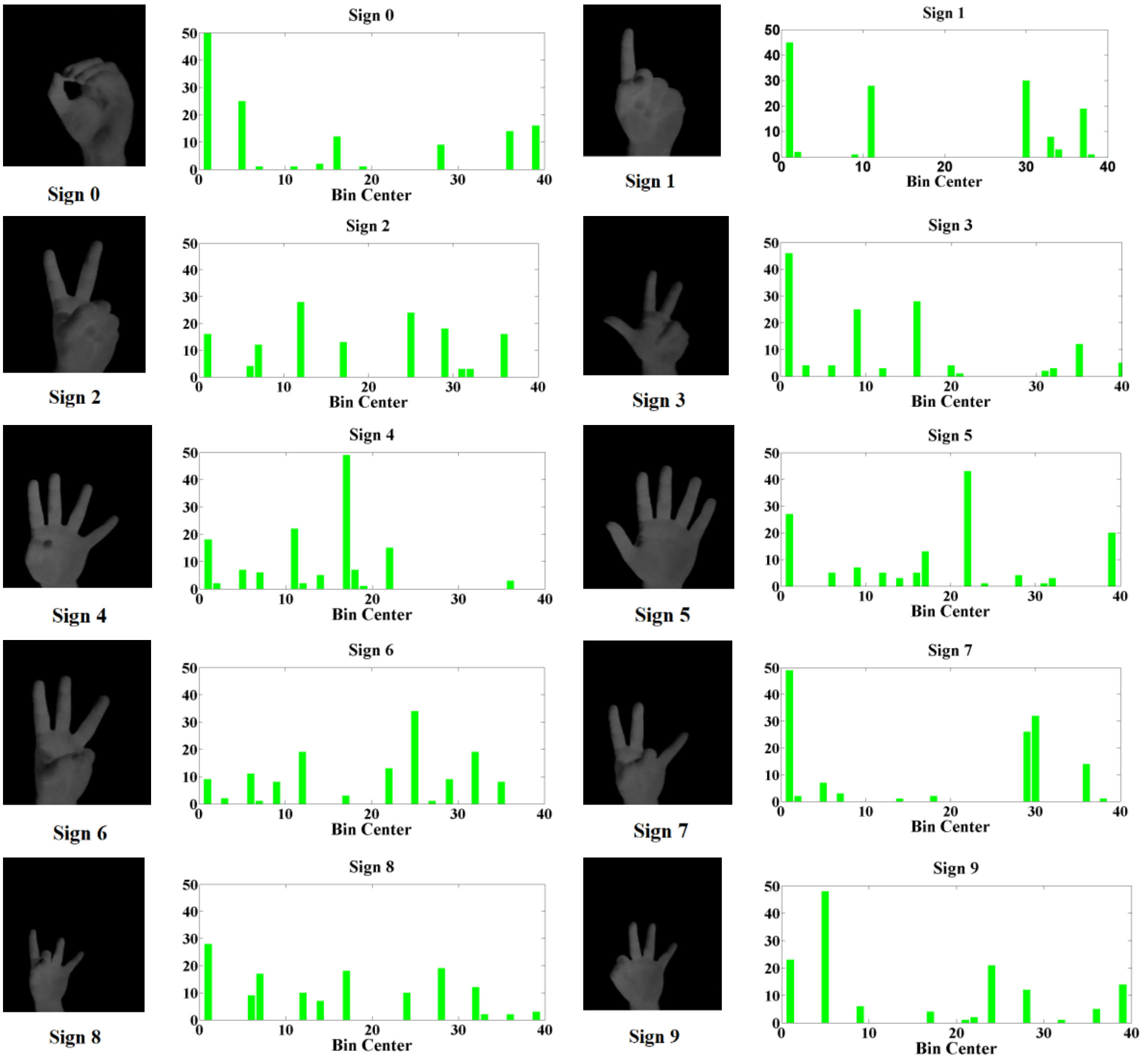

Fig. 3. BoW histograms for ASL number signs. The first and third columns present the images of ten ASL number signs; the second and fourth present the BoW histograms corresponding to each image.

D. Classification with a kNN Classifier

The kNN is a simple, non-parametric supervised classifier in common use [25], [26]. It predicts the class of a test feature, basing the prediction on a distance search to find the K number of the training neighbors nearest to the test feature. In general, the 'Cityblock', 'Cosine', 'Correlation', and 'Euclidean' distance functions are used to measure the distances from the new test feature to all the training samples in the feature space. The label that comprises the majority of the K nearest training samples is assigned to the test sample. To obtain a suitable value for K, various values of K are tried.

III. Results and Discussion

This section documents the performance of the proposed method and describes the dataset and performance measurement criteria.

A. Dataset

Two datasets, one publicly available and the other acquired, are used to validate the performance of the proposed method.

For the acquired dataset, 500 images are captured, using a webcam interfaced with MATLAB, for each ASL number class, thus providing 5,000 images for performance evaluation. To capture the images, an environment with a clear background is set up and suitably lit for recording videos with a webcam. The hand gestures are captured in a region of interest (ROI) that is set up in the video. Snapshots are taken from the video at three second intervals. Once the background frame has been captured by taking a snapshot with no gesture present at the ROI position, snapshots are taken with defined hand gestures (for example, ‘sign zero’) in the ROI position. In this way, 500 hand gesture images are taken for each of the ten ASL numbers, the image size being 250 × 274 pixels. After capturing all the images, the preprocessing step mentioned in Section II-A is employed to reduce the background noise and to enhance the hand gestures before feature extraction. For fair comparison, another dataset, the Kaggle ASL numbers dataset publicly available in [27], is also used for performance evaluation. In this dataset, 1000 images, size 30 × 30, are available for each class.

B. Performance Measurement Criteria

Classification produces four recognition types for the signed images. An image belonging to one class may be misclassified as belonging to another, creating a false positive recognition (*Fp*) of that class, while an image belonging to another class may be misclassified as belonging to that class, creating a false negative (*Fn*) recognition of that class. When the class of a considered image is accurately predicted, the recognition is defined as a true positive (*Tp*) for the considered class and as a true negative (*Tn*) for all other classes. The standard performance measures accuracy, precision, recall, and F1 score are used to evaluate the performance of the proposed algorithm, which can be easily classified from the confusion matrix using the equations provided below:

$$\text{Accuracy}(Ac)=\frac{T_p+T_n}{T_p+F_p+F_n+T_n} \tag{2}$$

$$\text{Precision}(Pr)=\frac{T_p}{T_p+F_p} \tag{3}$$

$$\text{Recall}(Re)=\frac{T_p}{T_p+F_n} \tag{4}$$

$$\text{F}_1\text{Score}(FS)=\frac{2\times(Pr\times Re)}{Pr+Re} \tag{5}$$

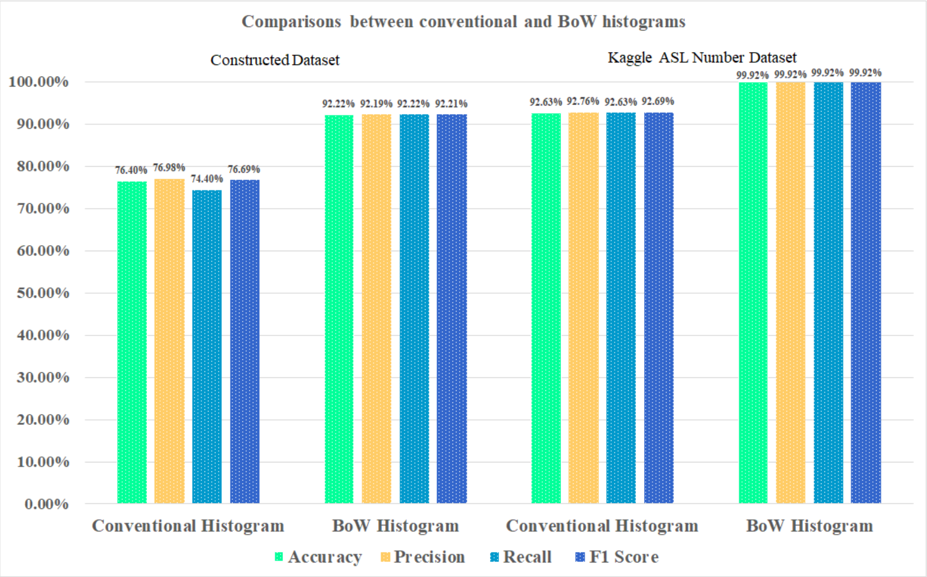

Fig. 4. Comparisons between conventional and BoW histograms.

TABLE I: PERFORMANCE COMPARISON OF THE PROPOSED METHOD USING VARIOUS CLUSTERING SIZE

Cluster Size	Bin Size	Performance (%)							
		Constructed Dataset				Kaggle ASL Numbers Dataset			
		Ac	Pr	Re	FS	Ac	Pr	Re	FS
4	40	87.04	87.04	87.04	87.04	99.56	99.56	99.56	99.56
8	80	90.58	90.58	90.58	90.58	99.85	99.85	99.85	99.85
16	160	92.22	92.19	92.22	92.21	99.92	99.92	99.92	99.92
32	320	91.80	91.75	91.80	91.78	99.98	99.98	99.98	99.98
64	640	91.74	91.68	91.74	91.71	99.98	99.98	99.98	99.98
128	1280	92.14	92.09	92.14	92.11	100	100	100	100

C. Performance of the Proposed ASL Number Recognition Scheme

To evaluate the performance of the proposed ASL number recognition method, the first step is to vary the cluster size in the K-means clustering algorithm in the range $k \in \{4, 8, 16, ..., 128\}$. Next, using the cluster centroids as bin centers, BoW histograms are constructed. The number of bin centers is therefore ten times the value of the chosen *k*. The supervised kNN classifier now performs the classification with $K = 1$. All the results are evaluated using a tenfold cross-validation scheme and appear in Table I, which shows that the best performance is achieved using the constructed dataset and a clustering size of 16. The results are as follows: accuracy is 92.22%; precision is 92.19%; recall is 92.22%; and the F1-score is 92.21%. Table II shows the confusion matrix, where ‘A’ represents the actual class and ‘P’ the predicted class. The computational complexity of the proposed algorithm is O(N(p ×q)m). The proposed algorithm takes 0.0123 seconds for feature extraction per sample and 0.02666 seconds for classification (System configuration is: Intel(R) Core(TM) i5-6200U CPU @ 2.30GHz 2.40Ghz, 8GB RAM, 64-bit OS). In the rest of the paper, therefore, results from the constructed dataset are reported using a clustering size of 16 unless otherwise specified. The best results for the Kaggle ASL dataset are 100% for all performance indices with k = 128. Next, the performance of the proposed BoW histogram is compared with that of the conventional histogram based approach. This involves using the gray pixel intensities of the sign images to construct the histogram for a bin size of 30, then choosing the bin centers by dividing the overall range of gray-level intensity into equal portions. The histogram bin frequencies are chosen as features in the kNN classifier with ‘cityblock’ distance. The results are shown in Fig. 4, which makes it clear that the proposed BoW histogram approach outperforms that of the conventional histogram. This paper therefore proposes the use of BoW histogram-based features for the recognition of signed numbers. Table III compares the performance of the proposed method for three supervised classifiers: artificial neural network (ANN); support vector machine (SVM), and K-nearest neighbors (kNN). The ANN classifier is implemented for the ‘trainscg’, ‘trainrp’, ‘trainbfg’, ‘trainlm’, and ‘traingd’ training functions and for a hidden node size of $N \in \{10, 20, 30, ..., 100\}$. The SVM classifier is implemented for multi-class classification using the ‘one versus all’ coding scheme with regularization parameter $C \in \{1, 2, 4, ..., 128\}$ and Gaussian Radial Basis Function (RBF) kernel parameter $\sigma \in \{1, 2, 4, ..., 128\}$. The kNN classifier is implemented for ‘Cityblock’, ‘Cosine’, ‘Cor-relation’, and ‘Euclidean’ distances with $K \in \{1, 2, 3, ..., 10\}$. Table III shows the best performance of each classifier, each with its best possible settings. The kNN classifier, with K = 1 and ‘cityblock’ distance, produces the best result for both datasets. Finally, Table IV compares the BoW method with two others, one proposed in [26] and the other by SK Dixit [26], who documented the use of features obtained using the Combined Orientation Histogram and Statistical (COHST) and Discrete Wavelet Transform (DWT) approaches for ASL number recognition. The statistical features in [26] are extracted from the entire image; this may degrade the feature quality when the image pixels of different classes are at the same intensity level. Note also that the DWT method extracts features only from low-frequency components. Table IV shows that the proposed BoW-based feature extraction scheme outperforms the methods proposed in [26], in terms of all performance.

TABLE II: CONFUSION MATRIX OF THE PROPOSED SCHEME

A P	0	1	2	3	4	5	6	7	8	9
0	478	0	0	2	2	9	2	2	1	4
1	2	477	1	0	0	0	0	0	20	0
2	0	0	477	2	2	0	2	16	1	0
3	0	0	1	485	3	1	2	1	0	7
4	2	0	1	1	452	7	17	11	6	3
5	3	0	0	2	5	487	0	0	0	3
6	0	0	7	2	31	0	422	35	2	1
7	3	2	28	3	13	0	32	409	4	6
8	1	19	2	7	9	2	3	9	448	0
9	6	0	0	8	2	6	1	1	0	476

TABLE III: COMPARISON AMONG CLASSIFIERS

Classifier	Classifier Settings	Performance (%)							
		Constructed Dataset				Kaggle ASL Numbers Dataset			
		Ac	Pr	Re	FS	Ac	Pr	Re	FS
ANN	trainscg’ function with N = 50	83.82	83.71	83.82	83.77	97.17	97.23	97.17	97.20
SVM	C = 16, σ = 16	90.78	92.60	90.78	91.68	99.88	99.88	99.88	99.88
kNN	**‘cityblock’ distance**	**92.22**	**92.19**	**92.22**	**92.21**	**99.92**	**99.92**	**99.92**	**99.92**

TABLE IV: COMPARISON WITH OTHER METHODS

Authors	Methods	Performance (%)							
		Constructed Dataset				Kaggle ASL Numbers Dataset			
		Ac	Pr	Re	FS	Ac	Pr	Re	FS
Thalange *et al.* [28]	COHST+Neural Network	89.44	89.13	89.44	89.29	96.96	96.17	96.96	96.56
Thalange *et al.* [28]	DWT+Neural Network	87.92	88.23	87.92	88.07	94.64	94.74	93.64	94.18
Proposed Method	**BoW+kNN**	**92.22**	**92.19**	**92.22**	**92.21**	**99.92**	**99.92**	**99.92**	**99.92**

IV. CONCLUSION

An efficient ASL number recognition scheme is developed in this paper, using bag-of-words (BoW) histograms to ex-tract features. It is experimentally shown that the BoW-based features are very suitable for classifying the ASL numbers. For classification, kNN, the simplest and most widely used classifier, is employed. The performance of the proposed scheme is evaluated in terms of accuracy, precision, recall, and F1 score, using two large datasets. The results show that the proposed BoW-based histogram method outperforms methods using conventional histograms; when compared with two other methods, it is shown to outperform both in terms of all performance indices. The best performance is achieved using our constructed dataset with a clustering size of 16. The results are as follows: accuracy is 92.22%; precision is 92.19%; recall is 92.22%; and the F1-score is 92.21%, whereas for the Kaggle ASL dataset all performance indices are 100% with a k of 128. The proposed method is therefore expected to help deaf and dumb people by allowing them to communicate with others via intelligent devices; it is also expected to help develop the games industry and robotics industries use hand gestures. In future, the authors wish to extend the proposed BoW-based model for ASL number recognition using semi-supervised learning techniques with larger dataset where labeling of all images are quite challenging.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Rasel Ahmed Bhuiyan and Abdul Matin actively participated in the implementation of the proposed model and prepared the draft paper. Md. Shafiur Raihan Shafi and Amit Kumar Kundu prepared the manuscript and actively worked as a proofreader. During the review process, all of us work together to solve the reviewer's comments.

Acknowledgment

We would like to thanks Amit Kumar Kundu to guide us. He has supervised us in the right direction. His supervision encourages us to take the implementation further.

References

[1] M. J. Cheok, Z. Omar, and M. H. Jaward, "A review of hand gesture and sign language recognition techniques," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 1, pp. 131-153, 2019.

[2] S. Gaikwad, A. Shetty, A. Satam, M. Rathod, and P. Shah, "Recognition of American sign language using image processing and machine learning," *International Journal of Computer Science and Mobile Computing*, vol. 8, no. 3, pp. 352-357, 2019.

[3] T. W. Chong and B. G. Lee, "American sign language recognition using leap motion controller with machine learning approach," *Sensors*, vol. 18, no. 10, p. 3554, 2018.

[4] J. R. Pansare, S. H. Gawande, and M. Ingle, "Real-time static hand gesture recognition for American sign language (ASL) in complex background," *Journal of Signal and Information Processing*, vol. 3, no. 3, p. 364, 2012.

[5] W. Tao, M. C. Leu, and Z. Yin, "American sign language alphabet recognition using convolutional neural networks with multi-view augmentation and inference fusion," *Engineering Applications of Artificial Intelligence*, vol. 76, pp. 202-213, 2018.

[6] M. Mohandes, S. Aliyu, and M. Deriche, "Arabic sign language recognition using the leap mo-tion controller," in *Proc. IEEE 23rd International Symposium on Industrial Electronics (ISIE)*, 2014, pp. 960-965.

[7] G. Modanwal and S. Kishor, "Utilizing gestures to enable visually impaired for computer interaction," *CSI Transactions on ICT*, pp. 1-5, 2019.

[8] A. B. Rasel, K. T. Abdul, A. Akm, S. Jungpil, and I. Rashedul, "Reduction of gesture feature dimension for improving the hand gesture recognition performance of numerical sign language," in *Proc. 20th IEEE International Conference of Computer and Information Technology (ICCIT)*, 2017, pp. 1-6.

[9] R. Cui, L. Hu, and Z. Changshui, "Recurrent convolutional neural networks for continuous sign language recognition by staged optimization," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7361-7369.

[10] Y. Wu, "Thomas S h. Vision based gesture recognition: A review," in *Proc. International Gesture Workshop*, 1999, pp. 103-115.

[11] K. Grobel and A. Marcell, "Isolated sign language recognition using hidden markov models," in *Proc. IEEE International Conference on Systems, Man, and Cybernetics*, 1997, vol. 1, pp. 162-167.

[12] S. Saengsri, N. Vit, and A R. Chotirat, "Tfrs: Thai finger spelling sign language recognition system," in *Proc. 2nd International Conference on Digital Information and Communication Technology and Its Applications (DICTAP)*, 2012, pp. 457-462.

[13] P. S. Rajam and G. Balakrishnan, "Recognition of Tamil sign language alphabet using image processing to aid deaf-dumb people," *Procedia Engineering*, vol. 30, pp. 861-868, 2012.

[14] P. S. Rajam and G. Balakrishnan, "Real time Indian sign language recognition system to aid deaf-dumb people," in *Proc. IEEE 13th International Conference on Communication Technology*, 2011, pp. 737-742.

[15] T. Starner, J. Weaver, and A. Pentland, "Real-time American sign language recognition using desk and wearable computer based video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, pp. 1371-1375, 1998.

[16] K. Sinha, R. Kumari, A. Priya, and P. Paul, "A computer vision based gesture recognition using hidden markov mode," *Innovations in Soft Computing and Information Technology*, pp. 55-67, 2019.

[17] C. Vogler and D. Metaxas, "Asl recognition based on a coupling between hmms and 3d motion analysis," in *Proc. Sixth International Conference on Computer Vision*, 1998, pp. 363-369.

[18] D. Jiang, G. Li, Y. Sun, J. Kong, and B. Tao, "Gesture recognition based on skeletonization algorithm and CNN with ASL database," *Multimedia Tools and Applications*, pp. 1-18, 2018.

[19] E. Antonakos, V. Pitsikalis, and P. Maragos, "Classification of extreme facial events in sign language videos," *EURASIP Journal on Image and Video Processing*, vol. 1, p. 14, Dec. 1, 2014.

[20] J. P. Sahoo, S. Ari, and D. K. Ghosh, "Hand gesture recognition using dwt and f-ratio based feature descriptor," *IET Image Processing*, vol. 12, no. 10, pp. 1780-1787, 2018.

[21] N. H. Dardas and N. D. Georganas, "Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 11, pp. 3592-3607, 2011.

[22] C. H. Chuan, E. Regina, and C. Guardino, "American sign language recognition using leap motion sensor," in *Proc. IEEE 13th International Conference on Machine Learning and Applications*, 2014, pp. 541-544.

[23] J. Wang, P. Liu, M. F. She, S. Nahavandi, and A. Kouzani, "Bag-of-words representation for biomedical time series classification," *Biomedical Signal Processing and Control*, vol. 8, no. 6, pp. 634-644, 2013.

[24] N. Dhanachandra, K. Manglem, and Y. J. Chanu, "Image segmentation using k-means clustering algorithm and subtractive clustering algorithm," *Procedia Computer Science*, vol. 54, pp. 764-771, 2015.

[25] Y. Zhang, G. Cao, B. Wang, and X. Li, "A novel ensemble method for k-nearest neighbor," *Pattern Recognition*, 85, pp. 13-25, 2019.

[26] Y. Rao, H. Yang *et al.*, "A generalized mean distance-based k-nearest neighbor classifier," *Expert Systems with Applications*, vol. 115, pp. 356-72, 2019.

[27] A. Gupta, *Kaggle ASL Numbers Dataset*, 2018.

[28] A. Thalange and S. Dixit, "Cohst and wavelet features based static Asl numbers recognition," *Procedia Computer Science*, vol. 92, pp. 455-460, 2016.

Rasel Ahmed Bhuiyan received the B.Sc. degree in computer science and engineering (CSE) from the University of Asia Pacific (UAP), Dhaka, Bangladesh in 2017. Currently, he has been serving as a lecturer in the Department of CSE at Uttara University (UU), Dhaka, Bangladesh. His research interest includes machine learning, computer vision, signal processing, and pattern recognition. He has several publications on the areas mentioned above. He has also been involved as a research assistant in Computer Vision and Pattern Recognition Lab (CVPR), UAP. Previously, he worked as a programmer at Adiva Graphics and Teaching Assistant at the University of Asia Pacific (UAP).

Abdul Matin has been serving as a senior lecturer in the Department of Computer Science and Engineering at Uttara University since 2017. His research interest lies in the area of computer vision (CV) and machine learning (ML), ranging from theory to design to implementation. He has several publications on ML and CV. He has collaborated actively with researchers in several other disciplines of computer science. Abdul completed his bachelor of science in engineering at Pabna University of Science and Technology, Bangladesh.

Md. Shafiur Raihan Shafi received the B.Sc. degree from the Department of Computer Science and Engineering (CSE), Ahsanullah University of Science and Technology, Bangladesh in 2016. He is currently pursuing his M.Sc. in CSE degree from Bangladesh University of Engineering and Technology, Bangladesh. He worked as a lecturer from 2017 to 2019 and as a senior lecturer in 2019 in the Department of CSE, Uttara University, Bangladesh. Currently he is working as a lecturer in the CSE Department, Southeast University, Bangladesh. His research interest lies in data science and machine learning.

Amit Kumar Kundu received both the B.Sc. and the M.Sc. degree from the Electrical and Electronic Engineering (EEE) Department, Bangladesh University of Engineering and Technology, Bangladesh. He is currently pursuing his Ph.D. degree from the University of Maryland, College Park, USA. He was a lecturer with the EEE Department, Uttara University, Bangladesh. His research interest lies in biomedical signal processing, and machine learning.