International Journal of Electrical and Computer Engineering (IJECE)
Vol. 14, No. 4, August 2024, pp. 4636~4645
ISSN: 2088-8708, DOI: 10.11591/ijece.v14i4.pp4636-4645

An overview of hand gesture recognition based on computer vision

Rifa Tasfia[1], Zeratul Izzah Mohd Yusoh[2], Adria Binte Habib[3], Tousif Mohaimen[2]
[1]Department of Computer Science and Engineering, National University Bangladesh, Gazipur, Bangladesh
[2]Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia
[3]Department of Computer Science and Engineering, BRAC University, Dhaka, Bangladesh

Article Info

Article history:

Received Oct 9, 2023
Revised Feb 25, 2024
Accepted Apr 22, 2024

Keywords:

Computer vision
Convolutional neural network
Deep neural network
Difficulties in hand gesture recognition
Hand gesture recognition
Hand gesture recognition methodology

ABSTRACT

Hand gesture recognition emerges as one of the foremost sectors which has gone through several developments within pattern recognition. Numerous studies and research endeavors have explored methodologies grounded in computer vision within this domain. Despite extensive research endeavors, there is still a need for a more thorough evaluation of the efficiency of various methods in different environments along with the challenges encountered during the application of these methods. The focal point of this paper is the comparison of different research in the domain of vision-based hand gesture recognition. The objective is to find out the most prominent methods by reviewing efficiency. Concurrently, the paper delves into presenting potential solutions for challenges faced in different research. A comparative analysis particularly centered around traditional methods and convolutional neural networks like random forest, long short-term memory (LSTM), heatmap, and you only look once (YOLO). considering their efficacy. Where convolutional neural network-based algorithms performed best for recognizing the gestures and gave effective solutions for the challenges faced by the researchers. In essence, the findings of this review paper aim to contribute to future implementations and the discovery of more efficient approaches in the gesture recognition sector.

Corresponding Author:

Rifa Tasfia
Department of Computer Science and Engineering, National University Bangladesh
Gazipur, Bangladesh
Email: rifa.tasfia2806@gmail.com

1. INTRODUCTION

Human-computer interaction (HCI) aims to optimize interfaces between users and computers. Hand gesture recognition (HGR) within HCI extends its application beyond device control to include advanced artificial intelligence (AI) robots. This paper provides a comprehensive review of HGR methodologies and challenges, analyzing existing works. It highlights that intelligent algorithms prove more effective, emphasizing the need for further attention to address challenges faced by researchers.

HGR is a prominent aspect of pattern recognition within HCI. Researchers are advancing vision-based technologies for HGR, achieving higher accuracy. Experiments, notably in automatic sign language (ASL) identification, have demonstrated success, with study [1] reporting 94.32% accuracy. The exploration of HGR as an eloquent and evolving field is emphasized in [2]. Figure 1 illustrates its applications, spanning egocentric HGR, ASL recognition, PowerPoint controlling, interactive projector screen, virtual mouse, fingertip identification, and PC controlling, showcasing the versatility of HGR across different contexts.

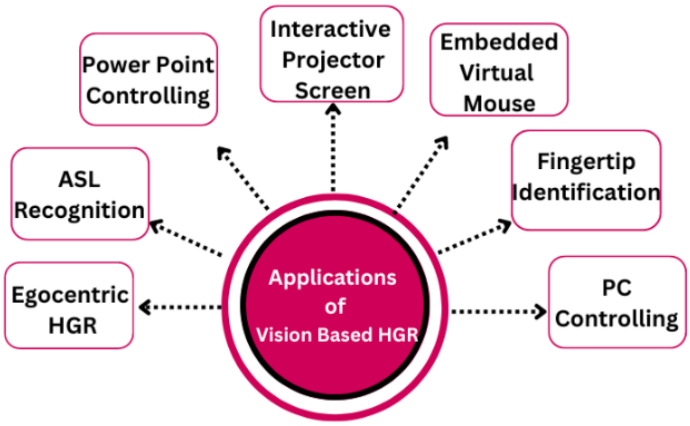

Figure 1. Applications of vision-based hand gesture recognition

HGR is a subset of pattern recognition under HCI. It has diverse approaches explained by researchers in [3], often employing vision-based technologies demonstrated in [4]. Vision-based HGR systems capture input through cameras connected to computers [4], and motion-based acquisition involves extracting still photos from motion videos for recognition [4]. Recent studies, such as [5], distinguishes between hand gesture recognition and evaluation, categorizing gestures as static and dynamic [6] and [7]. Workflow structures common across experiments encompass camera input, hand tracking and segmentation, feature extraction, classification and gesture recognition [6], [8] and [9]. Experiments, like those using the NVIDIA dataset [10], focus on dynamic gesture recognition. Unified learning in [11] categorizes gestures and fingertip recognition. Despite challenges, heatmap-based frameworks exhibit 99% accuracy with deep learning algorithms [4]. Deep learning algorithms excel in error reduction and complexity [8], [9] and [12], but vision-based acquisition faces challenges like complex backgrounds and lighting variations. In HCI, gesture recognition extends beyond hand movements, encompassing body and face gestures captured through cameras or sensors [2]. Habib *et al*. [9] and Fronteddu *et al*. [10] emphasize obtaining gestures through touch-based and touchless methods, with vision-based acquisition deemed most suitable [4]. Two fundamental types of gestures, static and dynamic, have been identified [2], [6], [7], and [13] with studies exploring effective recognition and classification algorithms. Static gestures, non-moveable gestures captured by cameras, achieve 93.09% accuracy [13]. For instance, Pinto *et al.* [14] used convolutional neural network (CNN) for static gestures, training an multilayer perception (MLP) neural network for skin color segmentation. Dynamic gestures involve continuous motion-based gestures, extensively explored by researchers [5], [10], [15] and [16]. ASL recognition demonstrated in [17]. Another unique technique reinforcement learning to classify EMG is applied to recognize hand gestures are demonstrated in [18]. This experiment got 90.49% accuracy for classification and 86.83% for recognizing hand gestures.

Studies employ different data processing methods for gesture classification. For instance, Fronteddu *et al.* [10] condensed the methodology into hand gesture recognition and evaluation, while Aggarwal and Arora [7] subdivided it into data collection, data environment, and hand gesture recognition. Another approach outlined a five-step methodology: camera input, hand tracking and segmentation, feature extraction, classification and recognition, and gesture recognition [6]. Figure 2 provides a comprehensive depiction of these methodologies.

– Camera input: this involves obtaining images or videos using a camera connected to a computer, as detailed by authors in [4] using different cameras like red, green and blue (RGB), time of flight (TOF), Thermal, and night vision cameras.
– Hand tracking and segmentation: this means the process of partitioning an image to identify the region of interest (ROI), specifically focusing on the hand. Various segmentation strategies include skin color-based segmentation, region-based segmentation, edge-based segmentation, and Otsu thresholding. Sarma and Bhuyan [2] mention RGB, HSV, YCbCr, CMYK as methods for explicit boundary specification for skin tone.
– Gesture representation and feature extraction: the method of finding characteristics in gesture representation and extracting features. Sarma and Bhuyan [2] classified gestures are represented as model-based and appearance-based. Feature extraction techniques depend on the data. In static gestures, as highlighted by Yoon *et al.* [19], features encompass elements such as color, posture, and spatiotemporal patterns.

– Gesture classification and recognition: This step demonstrates classifying and recognizing gestures by comparing them with datasets (e.g., NVIDIA, DVS128, CNN). Gestures are processed through various algorithms and recognized using supervised or unsupervised methods. Common classifiers include unsupervised k-means, supervised k-nearest neighbor (K-NN), support vector machine (SVM), and artificial neural network (ANN) [2]. Real-time datasets with you only look once V3 (YOLOV3) (CNN-based) [2] and [17], and deep fusion networks [20] for recognizing gestures.

In this research, section 2 delves into the literature review, examining various works in the field, while section 3 conducts an analysis of algorithmic accuracy. In section 4, the challenges encountered in vision-based acquisition are elucidated. Section 5 encapsulates the summarized findings of this review, serving as a valuable resource for future exploration in this domain.

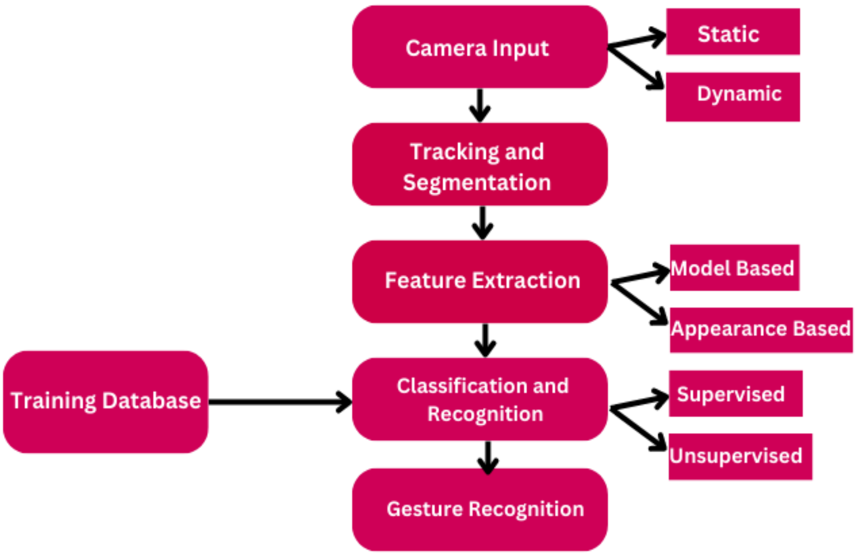

Figure 2. The common method of hand gesture recognition system

2. PREVIOUS WORKS

HGR has emerged as a dynamic field of research, witnessing diverse methodologies and algorithms aimed at enhancing gesture recognition accuracy. This comprehensive review delves into a multitude of studies, each contributing unique insights and innovations to the domain of HGR. The literature review encompasses an in-depth exploration of various studies, starting with the introduction of the hidden Markov model (HMM) in [21] and a low computational complexity model for HGR using RADAR in [22]. Notably, Kim *et al.* [23] proposes a novel approach utilizing machine learning (ML) and Wi-Fi for low-complexity gesture recognition, achieving commendable accuracy. On the basis of structural variety, Malima *et al.* [24] introduce an algorithm based on human skin color, utilizing red/green (R/G) intensities for segmentation and center of gravity (CoG) calculation. Further, Chanu *et al.* [25] have worked with both gloved based and vision based recognition using decision making algorithms and flex sensors for gloved based hands gestures. Do *et al.* [26] worked with depth based and skeletal based hand gestures and performed long short-term memory (LSTM) model on 3D data which gave excellent performance on dynamic hand gestures. A focus on automated hand gesture recognition using a deep convolutional neural network (DCNN) is presented in [27], where transfer learning and signer-dependent and signer-independent modes are explored across three datasets. The study in [28] leverages a 3D environment, using stereo vision, disparity maps, and centroid moments for gesture detection, achieving an 88% success rate. In [29], the focus shifts to Arabic sign language, employing ResNet50 and MobileNet V2 for gesture recognition. A low-cost radar sensor-based method is proposed in [30], achieving an impressive 97% accuracy through spectrogram analysis and CNN classification. Ray *et al.* [31], here worked with human activity recognition (HAR), where they reviewed different methodologies. They concluded that the idea of using state of art transfer learning methods that reduce difficulty and effort behind data collection. Joze *et al.* [32] introduce the multimodal transfer module (MMTM), applying it to four datasets in various domains. Depth-based 3D action recognition for hand gestures is explored in [33], utilizing 3D space voxelization and rank pooling, achieving an accuracy range from 82.4% to 93.5%. Wang *et al.* [34] implied bioinspired data with a fusion architecture of visual data and somatosensory data by stretchable strain sensors made from single walled carbon nano-tubes which got 100% accuracy. Gangrade and Bharti [35] applied depth-based 3D action recognition for hand gestures, utilizing 3D space voxelization and rank pooling to encode motion information. The study, using NTU RGB+D 120 and NTU RGB+D 60 datasets, achieved an accuracy range of 82.4% to 93.5%, showcasing effective recognition of 3D actions.

The investigation in study [36] delves into 3D-based hand gesture recognition using a multi-branch attention-based graph model, achieving accuracies of 94.12%, 92.00%, and 97.01% on benchmark datasets. sign language recognition is explored in [37], integrating CNN with classical non-intelligent techniques for a comprehensive gesture classification strategy. Wadhawan and Kumar [38] worked with Indian sign language (ISL). They dealt with robust modeling of static signs. They used deep learning (CNN). They got 99.72% on colored image and 99.09% on grayscale image.

He *et al.* [39] propose a method for predicting occupant thermal states through a fusion of infrared thermography, computer vision, and machine learning. The approach focuses on utilizing cheek, nose, and hand temperatures due to their resilience against blockage by hair, glasses, and clothing. The study employs geometrically defined sub-areas on the face and hands to measure the distribution of skin temperatures and leverages temperature differences within and between these areas to mitigate the impact of calibration drift in thermal infrared cameras. Tests conducted in both outdoor and indoor environments, consisting of extensive data on skin temperatures and subjective thermal sensations, demonstrate the effectiveness of random forest classification models. These models, utilizing absolute skin temperatures or intra- and inter-segment temperature differences, achieve high accuracy rates of 92%–96%.

This comprehensive review thoroughly explores the varied landscape of HGR, encompassing methodologies, algorithms, and innovations in diverse applications. The summarized insights offer a valuable resource for future exploration and development in the domain of pattern recognition through hand gesture recognition in computer vision. Table 1 provides a comparative analysis of selected experiments, enhancing the understanding of HGR advancements.

Table 1. A brief analysis of algorithms used in several experiments

Experiment no	Algorithms used	Experiment No	Algorithms used
[22]	Introduction of a low computational complexity model using RADAR for hand gesture recognition	[29]	Focus on Arabic sign language gesture recognition using ResNet50 and MobileNet V2 in [29]. Image preprocessing includes conversion to a three-channel image and the application of filters for quality enhancement
[23]	Proposal of a novel approach using ML and Wi-Fi to achieve low-complexity gesture recognition	[33]	Exploration of depth-based 3D action recognition using 3D space voxelization and rank pooling in [33], achieving an accuracy range from 82.4% to 93.5%.
[24]	Introduction of an algorithm based on human skin color using R/G intensities is deployed. Segmentation and CoG calculation are performed	[36]	Investigation into 3D-based hand gesture recognition using a multi-branch attention-based graph model in [36], achieving accuracies of 94.12%, 92.00%, and 97.01% on benchmark datasets.
[25]	Decision making algorithms were used for vision based real time data and flex sensors for gloved based data. Vision based recognition gave 100% accuracy.	[38]	Introduction of an innovative teleoperation method for controlling dexterous manipulators in space stations, relying on monocular hand motion capture and video-calling tools in [38].
[28]	Leveraging a 3D environment with stereo vision, disparity maps, and centroid moments for gesture detection in [28], achieving an 88% success rate.	[39]	Learning | Proposal of a method for predicting occupant thermal states through a fusion of infrared thermography, computer vision, and machine learning in [39]

3. ALGORITHMS AND ACCURACY OF HAND GESTURE RECOGNITION

Gesture recognition is broadly classified into supervised and unsupervised approaches. Supervised recognition involves comparing acquired images with pre-defined classes, and the recognition algorithm identifies the gesture accordingly. Mitra and Acharya [40] regard recognition as a sub-system of recognition frameworks, where a classifier organizes incoming gesture parameters into pre-defined classes (supervised) or based on their proximity (unsupervised). Various classifications are employed for both static and dynamic gestures, with researchers utilizing different supervised and unsupervised recognition methods.

For instance, Sarma and Bhuyan [2] discusses the use of unsupervised k-means, a parametric classifier, and supervised K-NN algorithms with multi-dimensional information. Additionally, semi-supervised or self-trained algorithms are prevalent. Fronteddu *et al.* [10] utilized pre-trained NVIDIA datasets along with self-trained data, working with 27 classes collected in (*.avi*) format. Researchers here focused on virtual reality (VR), identifying gestures and fingertip positions using pre-trained and semi-trained datasets. OpenCV tools were employed incorporating deep learning through the VGG16 model, using CNN architecture for pretraining. Figure 3 and Table 2 in the paper present the accuracy of algorithms on different types of data applied in the experiment.

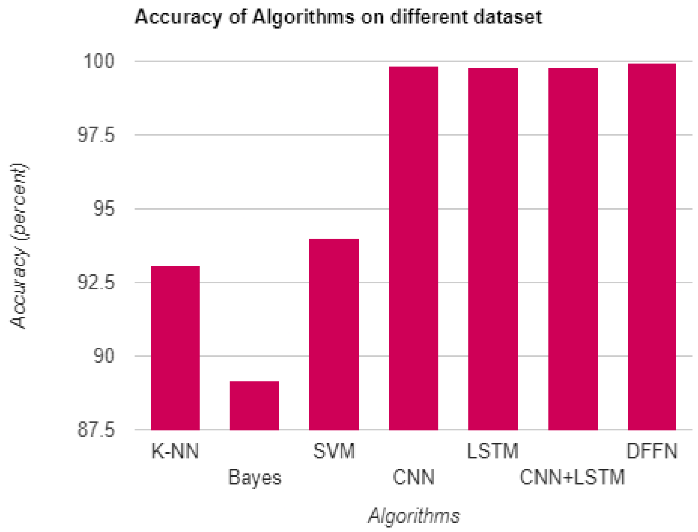

Figure 3. This bar graph here shows accuracy of different algorithms performed in Table 1 on different sets of data

Table 2. Accuracy of algorithms performed on different types of data

Algorithms applied	Accuracy
K-NN model	93.06%
Bayes model	89.15%
SVM model	94%
CNN model	99.82%
LSTM model	99.81%
CNN+LSTM	99.80%
DFFN	99.93%

DFFN: Deep feature fusion network

In [20], a gloved-based sensor system was utilized, combining sensor-based and vision-based HGR. A 3D model built on sensors determined the degrees of each figure, with Euler angles collected by a magnetic gyroscope. LSTM recognized gesture classification, and a deep feature fusion network extracted features. While comparing this paper, got the accuracy level the most for CNN based algorithms about 93.93%. A graphical representation of CNN based algorithms on the same type of data is showed in Figure 4. Table 3 in the paper presents the accuracy of algorithms on different types of data applied in the experiment.

Table 3. Accuracy of algorithms on the same type of data

Models	Learning rate	Images	Accuracy
VGG16	0.001	216	85.68%
SGD	0.1		77.98%
SSD	0.0001		82.00%
YOLOV3	0.001		97.68%

Pansare *et al.* [13] worked on different types of CNN based algorithms and obtained different levels of accuracy. A brief discussion of all these approaches is presented in Table 4. This will contribute to the future studies to find out the efficiency of CNN based approaches in vision based HGR system. Accuracy in this experiment has a range starting from 89.54% to 99.81%.

Table 4. Accuracy and methodology of algorithms [13]

Experiment	Algorithm used	Method used	Accuracy
Power point control	histogram of oriented gradients (HOG) feature extraction and K-NN classification	After image acquisition a skin detection algorithm is used then resized the image and background has removed. The binary image is acquired with the help of HOG feature extraction finally the image is classified with K-NN classification.	Accuracy drop has been found due to lack of brightness.
User dependent gesture recognition on android handheld devices	The distance metric algorithm was trained by large margin NN (LMNN) and K-NN classification	Data is recorded with accelerometer sensor readings which worked on feature extraction by using HOD- histogram of directions. This model is trained with the Distance metric algorithm and classified with the K-NN method.	99.81%
Embedded virtual mouse system	Convex-hull algorithm	After gaining the input skin detection algorithm is used to classify pixels and then raster scan is used for pixel labeling finally convex-hull algorithm is applied for finding the gesture.	92%
Android-based American sign language recognition system	HOG feature extraction & SVM	A phone camera is used for input taking the skin segmentation is carried out using the YCbCr system. Then HOG algorithm is used for feature extraction. Output is gained by trained dataset using SVM	89.54%
Hand gesture feature extraction using deep convolutional network	DCNN and multi-class support vector machines (MCSVMs)	Used for recognizing American sign language. DCNN is used for extracting efficient hand features. Inputs are collected using a webcam. The image is converted into a Grayscale image and classification was acquired by MCSVMs.	94.57%
High-speed gesture recognition	YOLOV2 & Spatial Refinement	This model used YOLOV2 model for detecting objects. This experiment used the spatial refinement module to change YOLOV2 by removing the last max-pooling layer and sixth convolutional layer. Light YOLO is a combination of YOLOV2 and the spatial refinement module.	98.06%

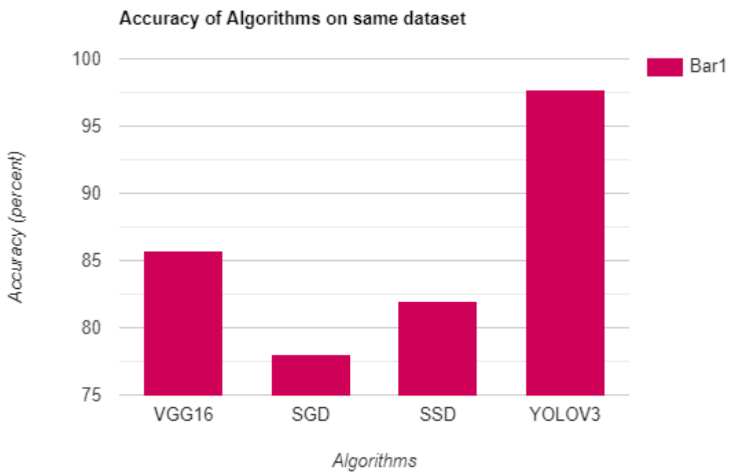

Figure 4. This bar graph here shows accuracy of different algorithms performed in Table 3 on same set of data

4. CHALLENGES IN HAND GESTURE RECOGNITION

The realm of HGR systems has garnered extensive attention in research, with diverse methodologies and algorithms addressing unique challenges. Notably, in [4] and [13], complexities such as lighting variations, background intricacies, and processing time constraints were documented, underscoring the multifaceted nature of HGR.

Difficulties faced in re-identifying postures by authors of [41] is solved by using consensus based algorithms. Complexities outlined in [42] pertained to interconnected links and joints, acknowledging variability based on datasets and algorithms. Addressing unmodeled variables, [43] proposed the frame reduction technique and optical flow for high-resolution image and moving object handling. Short comes like interference factors in gesture recognizing in [44] was solved by using Deep Learning algorithms which improved the quality of sEMG pattern recognition. Authors in [45] proposed occlusion solutions involving video frame extraction and foreground object segmentation using MATLAB. In [46], researchers addressed the constraints of HGR by examining the shortcomings in the vision-based HGR system. They focused on integrating diverse textual and interpretative aspects of gestures, along with handling complex non-rigid hand characteristics. The researchers introduced various categories and subcategories of methods, leveraging neural networks such as convolutional neural network (CNN) and deep neural network (DNN) achieving accuracy levels ranging from 68% to 97%. Liao *et al.* [47] worked with occlusion problem. To overcome this problem, they used single shot MultiBox detection (SSD). As an outcome the MobileNets is selected as the frontend network and the MobileNets-SSD network is improved. In study [48], researchers worked with the limitations of the HGR system, specifically addressing the challenges associated with occlusion. The researchers conducted experiments using both RGB and RGB-D cameras, providing a comprehensive review through quantitative and qualitative comparisons of various algorithms. Their findings led them to the conclusion that prediction algorithms could serve as a promising solution to solve the identified challenges in HGR systems. Singha *et al.* [49] encountered background intricacies in hand gesture recognition, devising a two-stage HGR system. The initial stage was developed with three-frame differencing and then tracked by Kanade-Lucas-Tomasi feature tracker and finally a fusion model of ANN, SVM and K-NN model worked for recognition. Muneeb *et al.* [50] found several difficulties like tilting, rotation and acceleration. They used several feature extraction algorithms and a random forest classifier which achieved 94% of accuracy.

Despite these challenges, CNN algorithms emerged as efficient solutions, providing superior results and faster processing across diverse HGR problems. However, there remains a gap in understanding and exploring the full spectrum of difficulties in HGR, warranting further research in this domain. The summarized insights from this diverse body of literature offer a comprehensive overview, paving the way for future exploration in hand gesture recognition in computer vision. In Table 5 a brief comparison between these experiments is presented.

Table 5. Challenges and solutions in hand gesture recognition

Experiment	Methods/Platform used	Challenges	Solutions
[41]	Re-identification algorithm	Recognizing same postures in different camera views	Consensus based algorithm.
[46]	Multiple static technologies	Comprising multiple text and interpretations of gesture along with complex non-rigid hand characteristics	CNN and DNN.
[47]	Unrecognizable data	Occlusion problem was seen	The SSD in TensorFlow platform.
[48]	RGB, RGB-D cameras for acquisition	Occlusion (size, shape and speed variation)	Prediction algorithm.
[49]	Multiple dynamic HGR techniques	Complicated background, change in illumination, occlusion	Fusion model of (CNN, SVM and K-NN)
[50]	Sensor embedded hand glove	Tilting, rotation and algorithm.	Several feature extraction algorithm and random forest classifier.

5. CONCLUSION

The paper delves into the realm of vision-based HGR, a vital aspect of HCI. It scrutinizes various working methodologies such as color-based, appearance-based, motion-based, skeleton-based, depth-based, 3D model-based, and deep learning-based approaches, showcasing their diverse applications in different sectors. With a primary focus on creating interfaces in the domain of HGR, the paper extensively reviews numerous researchers' contributions. It concludes that CNN-based algorithms notably outperform classical non-intelligent methods, while fusion-based algorithms also achieve success. The review identifies challenges faced by researchers, highlighting CNN-based solutions for resolution, occlusion, and low-quality image problems. Despite significant advancements, the paper underscores the need for further research to address challenges in hand gesture recognition systematically. It provides a foundational overview of vision-based HGR methodologies, comparing algorithms based on accuracy levels and offering mathematical representations, making it a valuable resource for comprehending pattern recognition through hand gesture recognition in computer vision.

REFERENCES

[1] M. M. Islam, S. Siddiqua, and J. Afnan, "Real time hand gesture recognition using different algorithms based on American sign language," in *2017 IEEE International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, 2017, pp. 1–6, doi: 10.1109/ICIVPR.2017.7890854.

[2] D. Sarma and M. K. Bhuyan, "Methods, databases and recent advancement of vision-based hand gesture recognition for HCI systems: a review," *SN Computer Science*, vol. 2, no. 6, Aug. 2021, doi: 10.1007/s42979-021-00827-x.

[3] G. R. S. Murthy and R. S. Jadon, "A review of vision based hand gestures recognition," *Computer Science and Application*, vol. 10, no. 5, pp. 990–1001, 2020, doi: 10.12677/csa.2020.105102.

[4] M. Oudah, A. Al-Naji, and J. Chahl, "Hand gesture recognition based on computer vision: a review of techniques," *Journal of Imaging*, vol. 6, no. 8, p. 73, Jul. 2020, doi: 10.3390/JIMAGING6080073.

[5] Y. Dong, J. Liu, and W. Yan, "Dynamic hand gesture recognition based on signals from specialized data glove and deep learning algorithms," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–14, 2021, doi: 10.1109/TIM.2021.3077967.

[6] J. S. Sonkusare, N. B. Chopade, R. Sor, and S. L. Tade, "A review on hand gesture recognition system," in *Proceedings - 1st International Conference on Computing, Communication, Control and Automation, ICCUBEA 2015*, Feb. 2015, pp. 790–794, doi: 10.1109/ICCUBEA.2015.158.

[7] K. Aggarwal and A. Arora, "An approach to control the PC with hand gesture recognition using computer vision technique," in *Proceedings of the 2022 9th International Conference on Computing for Sustainable Global Development, INDIACom 2022*, Mar. 2022, pp. 760–764, doi: 10.23919/INDIACom54597.2022.9763282.

[8] T. R. Gadekallu *et al.*, "Hand gesture recognition based on a Harris Hawks optimized convolution neural network," *Computers and Electrical Engineering*, vol. 100, May 2022, doi: 10.1016/j.compeleceng.2022.107836.

[9] A. B. Habib, F. Bin Ashraf, and A. Shakil, "Finding efficient machine learning model for hand gesture classification using EMG data," *2021 5th International Conference on Electrical Engineering and Information and Communication Technology, ICEEICT 2021*, Nov. 2021, doi: 10.1109/ICEEICT53905.2021.9667856.

[10] G. Fronteddu, S. Porcu, A. Floris, and L. Atzori, "A dynamic hand gesture recognition dataset for human-computer interfaces," *Computer Networks*, vol. 205, Mar. 2022, doi: 10.1016/j.comnet.2022.108781.

[11] M. M. Alam, M. T. Islam, and S. M. M. Rahman, "Unified learning approach for egocentric hand gesture recognition and fingertip detection," *Pattern Recognition*, vol. 121, Jan. 2022, doi: 10.1016/j.patcog.2021.108200.

[12] Y. Xing, G. Di Caterina, and J. Soraghan, "A new spiking convolutional recurrent neural network (SCRNN) with applications to event-based hand gesture recognition," *Frontiers in Neuroscience*, vol. 14, Nov. 2020, doi: 10.3389/fnins.2020.590164.

[13] J. Pansare, P. Gajanan Aochar, T. Salvi, J. Braganza, A. Tiwari, and D. Kesharkar, "Effective computer vision techniques for real-time hand gesture recognition and detection," *International Journal of Advances in Engineering and Management (IJAEM)*, vol. 3, 2021.

[14] R. F. Pinto, C. D. B. Borges, A. M. A. Almeida, and I. C. Paula, "Static hand gesture recognition based on convolutional neural networks," *Journal of Electrical and Computer Engineering*, vol. 2019, pp. 1–12, Oct. 2019, doi: 10.1155/2019/4167890.

[15] B. Ionescu, D. Coquin, P. Lambert, and V. Buzuloiu, "Dynamic hand gesture recognition using the skeleton of the hand," *Eurasip Journal on Applied Signal Processing*, vol. 2005, no. 13, pp. 2101–2109, Aug. 2005, doi: 10.1155/ASP.2005.2101.

[16] M. Ur Rehman *et al.*, "Dynamic hand gesture recognition using 3D-CNN and LSTM networks," *Computers, Materials and Continua*, vol. 70, no. 3, pp. 4675–4690, 2022, doi: 10.32604/cmc.2022.019586.

[17] A. Mujahid *et al.*, "Real-time hand gesture recognition based on deep learning YOLOv3 model," *Applied Sciences (Switzerland)*, vol. 11, no. 9, p. 4164, May 2021, doi: 10.3390/app11094164.

[18] J. P. Vásconez, L. I. Barona López, Á. L. Valdivieso Caraguay, and M. E. Benalcázar, "A comparison of EMG-based hand gesture recognition systems based on supervised and reinforcement learning," *Engineering Applications of Artificial Intelligence*, vol. 123, Aug. 2023, doi: 10.1016/j.engappai.2023.106327.

[19] H. S. Yoon, J. Soh, Y. J. Bae, and H. Seung Yang, "Hand gesture recognition using combined features of location, angle and velocity," *Pattern Recognition*, vol. 34, no. 7, pp. 1491–1501, Jan. 2001, doi: 10.1016/S0031-3203(00)00096-0.

[20] G. Yuan, X. Liu, Q. Yan, S. Qiao, Z. Wang, and L. Yuan, "Hand gesture recognition using deep feature fusion network based on wearable sensors," *IEEE Sensors Journal*, vol. 21, no. 1, pp. 539–547, 2021, doi: 10.1109/JSEN.2020.3014276.

[21] M. Elmezain, A. Al-Hamadi, J. Appenrodt, and B. Michaelis, "A hidden Markov model-based continuous gesture recognition system for hand motion trajectory," *Proceedings - International Conference on Pattern Recognition*, Dec. 2008, doi: 10.1109/icpr.2008.4761080.

[22] Y. Zhao, V. Sark, M. Krstic, and E. Grass, "Low computational complexity algorithm for hand gesture recognition using mmWave RADAR," in *Proceedings of the International Symposium on Wireless Communication Systems*, Oct. 2022, vol. 2022-Octob, doi: 10.1109/ISWCS56560.2022.9940253.

[23] J. S. Kim, W. S. Jeon, and D. G. Jeong, "WiFi-based low-complexity gesture recognition using categorization," in *IEEE Vehicular Technology Conference*, Jun. 2022, vol. 2022-June, doi: 10.1109/VTC2022-Spring54318.2022.9860660.

[24] A. Malima, E. Özgür, and M. Çetin, "A fast algorithm for vision-based hand gesture recognition for robot control," in *2006 IEEE 14th Signal Processing and Communications Applications Conference*, 2006, vol. 2006, doi: 10.1109/SIU.2006.1659822.

[25] O. R. Chanu, A. Pillai, S. Sinha, and P. Das, "Comparative study for vision based and data based hand gesture recognition techniquee," in *ICCT 2017 - International Conference on Intelligent Communication and Computational Techniques*, Dec. 2017, pp. 26–31, doi: 10.1109/INTELCCT.2017.8324015.

[26] N. T. Do, S. H. Kim, H. J. Yang, and G. S. Lee, "Robust hand shape features for dynamic hand gesture recognition using multi-level feature LSTM," *Applied Sciences (Switzerland)*, vol. 10, no. 18, p. 6293, Sep. 2020, doi: 10.3390/APP10186293.

[27] M. Al-Hammadi, G. Muhammad, W. Abdul, M. Alsulaiman, M. A. Bencherif, and M. A. Mekhtiche, "Hand gesture recognition for sign language using 3DCNN," *IEEE Access*, vol. 8, pp. 79491–79509, 2020, doi: 10.1109/ACCESS.2020.2990434.

[28] M. A. Laskar, A. J. Das, A. K. Talukdar, and K. K. Sarma, "Stereo vision-based hand gesture recognition under 3D environment," *Procedia Computer Science*, vol. 58, pp. 194–201, 2015, doi: 10.1016/j.procs.2015.08.053.

[29] A. Alnuaim, M. Zakariah, W. A. Hatamleh, H. Tarazi, V. Tripathi, and E. T. Amoatey, "Human-computer interaction with hand gesture recognition using ResNet and MobileNet," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–16, Mar. 2022, doi: 10.1155/2022/8777355.

[30] S. Franceschini *et al.*, "Hand gesture recognition via radar sensors and convolutional neural networks," in *IEEE National Radar Conference - Proceedings*, Sep. 2020, vol. 2020-Septe, doi: 10.1109/RadarConf2043947.2020.9266565.

[31] A. Ray, M. H. Kolekar, R. Balasubramanian, and A. Hafiane, "Transfer learning enhanced vision-based human activity recognition: a decade-long analysis," *International Journal of Information Management Data Insights*, vol. 3, no. 1, Apr. 2023, doi: 10.1016/j.jjimei.2022.100142.

[32] H. R. V. Joze, A. Shaban, M. L. Iuzzolino, and K. Koishida, "MMTM: multimodal transfer module for CNN fusion," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2020, pp. 13286–13296, doi: 10.1109/CVPR42600.2020.01330.

[33] Y. Wang *et al.*, "3DV: 3D dynamic voxel for action recognition in depth video," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2020, pp. 508–517, doi: 10.1109/CVPR42600.2020.00059.

[34] M. Wang *et al.*, "Gesture recognition using a bioinspired learning architecture that integrates visual data with somatosensory data from stretchable sensors," *Nature Electronics*, vol. 3, no. 9, pp. 563–570, Jun. 2020, doi: 10.1038/s41928-020-0422-z.

[35] J. Gangrade and J. Bharti, "Vision-based hand gesture recognition for Indian sign language using convolution neural network," *IETE Journal of Research*, vol. 69, no. 2, pp. 723–732, Nov. 2023, doi: 10.1080/03772063.2020.1838342.

[36] A. S. M. Miah, M. A. M. Hasan, and J. Shin, "Dynamic hand gesture recognition using multi-branch attention based graph and general deep learning model," *IEEE Access*, vol. 11, pp. 4703–4716, 2023, doi: 10.1109/ACCESS.2023.3235368.

[37] M. M. Damaneh, F. Mohanna, and P. Jafari, "Static hand gesture recognition in sign language based on convolutional neural network with feature extraction method using ORB descriptor and Gabor filter," *Expert Systems with Applications*, vol. 211, Art. no. 118559, Jan. 2023, doi: 10.1016/j.eswa.2022.118559.

[38] A. Wadhawan and P. Kumar, "Deep learning-based sign language recognition system for static signs," *Neural Computing and Applications*, vol. 32, no. 12, pp. 7957–7968, Jan. 2020, doi: 10.1007/s00521-019-04691-y.

[39] Y. He *et al.*, "Smart detection of indoor occupant thermal state via infrared thermography, computer vision, and machine learning," *Building and Environment*, vol. 228, p. 109811, Jan. 2023, doi: 10.1016/j.buildenv.2022.109811.

[40] S. Mitra and T. Acharya, "Gesture recognition: a survey," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 37, no. 3, pp. 311–324, May 2007, doi: 10.1109/TSMCC.2007.893280.

[41] A. Barman and S. K. Shah, "SHaPE: a novel graph theoretic algorithm for making consensus-based decisions in person re-identification systems," in *Proceedings of the IEEE International Conference on Computer Vision*, Oct. 2017, vol. 2017-Octob, pp. 1124–1133, doi: 10.1109/ICCV.2017.127.

[42] A. S. Ghotkar and G. K. Kharate, "Study of vision based hand gesture recognition using indian sign language," *International Journal on Smart Sensing and Intelligent Systems*, vol. 7, no. 1, pp. 96–115, Jan. 2014, doi: 10.21307/ijssis-2017-647.

[43] N. O'Mahony *et al.*, "Deep learning vs. traditional computer vision," *Advances in Intelligent Systems and Computing*, vol. 943, pp. 128–144, 2020, doi: 10.1007/978-3-030-17795-9_10.

[44] W. Li, P. Shi, and H. Yu, "Gesture recognition using surface electromyography and deep learning for prostheses hand: state-of-the-art, challenges, and future," *Frontiers in Neuroscience*, vol. 15, Apr. 2021, doi: 10.3389/fnins.2021.621885.

[45] Y. Z. Cheong and W. J. Chew, "The application of image processing to solve occlusion issue in object tracking," *MATEC Web of Conferences*, vol. 152, 2018, doi: 10.1051/matecconf/201815203001.

[46] F. Al Farid *et al.*, "A structured and methodological review on vision-based hand gesture recognition system," *Journal of Imaging*, vol. 8, no. 6, May 2022, doi: 10.3390/jimaging8060153.

[47] S. Liao *et al.*, "Occlusion gesture recognition based on improved SSD," *Concurrency and Computation: Practice and Experience*, vol. 33, no. 6, Oct. 2021, doi: 10.1002/cpe.6063.

[48] P. K. Pisharady and M. Saerbeck, "Recent methods and databases in vision-based hand gesture recognition: a review," *Computer Vision and Image Understanding*, vol. 141, pp. 152–165, Dec. 2015, doi: 10.1016/j.cviu.2015.08.004.

[49] J. Singha, A. Roy, and R. H. Laskar, "Dynamic hand gesture recognition using vision-based approach for human–computer interaction," *Neural Computing and Applications*, vol. 29, no. 4, pp. 1129–1141, Aug. 2018, doi: 10.1007/s00521-016-2525-z.

[50] M. Muneeb, H. Rustam, and A. Jalal, "Automate appliances via gestures recognition for Elderly living assistance," *2023 4th International Conference on Advancements in Computational Sciences, ICACS 2023 - Proceedings*, Feb. 2023, doi: 10.1109/ICACS55311.2023.10089778.

BIOGRAPHIES OF AUTHORS

Rifa Tasfia received the B.Sc. degree in computer science and engineering from National University of Bangladesh, in 2020.She is currently working as a teacher in a school. She is involved in several research works on the ground of human computer interaction and computer vision. Her research interests include computer vision, pattern recognition, deep neural network, face recognition, hand gesture recognition and emotion detection. Rifa has contributed to a conference paper focusing on pattern recognition and she hopes to contribute more in this domain in future. She can be contacted at email: rifa.tasfia2806@gmail.com.

Zeratul Izzah Mohd Yusoh is a senior lecturer in the Department of Intelligent Computing and Analytics, Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM) where she has been a faculty member since 2003. She completed her Ph.D. in Queensland University of Technology, Australia. Her Ph.D. work is focusing on SaaS resource management system in cloud computing using evolutionary computation. She is also a visiting researcher at University of Kobe, Japan, where her work is focusing on autonomous driver's model optimization in emission test cycle. She did her master at University of Edinburgh, Scotland and her undergrad study is at Universiti Teknologi Malaysia (UTM), Johor. Zeratul is a certified IT Professional (Malaysia) and EMC data science Associate. Her research interests lie in the area of developing soft computing solutions to improve the systems' intelligence. Through her past research in Dyslexia area, she has won two gold medals, one silver medal, and one bronze medal at both international and national exhibitions. The product of the research has also been commercialized and applied for patent. In recent years, her research is focusing on intelligent analytics, evolutionary computation techniques, software as a service (SaaS) and cloud computing area. She can be contacted at email: zeratul@utem.edu.my.

Adria Binte Habib earned her bachelor of science in computer science from BRAC University. She currently holds the position of research assistant at BRAC University while pursuing her master's degree in the same field at the same institution. Focusing her research on areas such as data science, pattern recognition, biomedical imaging, and time series analysis, Adria hopes to become a data scientist. Presently, she is actively engaged in the development of a journal paper centered around biomedical imaging. So far Adria has contributed to four published conferences and authored three blog articles in various fields from basic mathematics to explanations of various algorithms. She can be contacted at email: binte.adria708@gmail.com.

Tousif Mohaimen is a dedicated computer science student specializing in artificial intelligence (AI) at the University Technical Malaysia Melaka. His research interests primarily revolve around the dynamic fields of computer vision and natural language processing (NLP). He is passionate about harnessing AI technologies to solve real-world problems and advance our understanding of human-computer interactions. With a commitment to academic excellence and a forward-thinking approach, he strives to contribute to the ever-evolving landscape of AI research. She can be contacted at email: tousifmohaimen@icloud.com.