Received: 23 July 2023 | *Revised:* 24 September 2023 | *Accepted:* 5 November 2023
DOI: 10.37917/ijeee.20.2.13 *Early View* | *December 2024*

Open Access

Iraqi Journal for Electrical and Electronic Engineering
Review Article

IJEEE University of Basrah College of Engineering

Advancements and Challenges in Hand Gesture Recognition: A Comprehensive Review

Bothina Kareem Murad[1], **Abbas H. Hassin Alasadi** *[1,2]
[1]College of Computer Science and Information Technology, University of Basrah, Basrah, Iraq.
[2]IEEE and ACIT member, Tel: 009647809835559, ORCID: (0000-0002-6627-4456) (orcid.org)

Correspondance
Abbas H. Hassin Alasadi
College of Computer Science and Information Technology,
University of Basrah, Basrah, Iraq.
Email: abbas.hassin@uobasrah.edu.iq or abbashh2002@gmail.com

Abstract

Hand gesture recognition is a quickly developing field with many uses in human-computer interaction, sign language recognition, virtual reality, gaming, and robotics. This paper reviews different ways to model hands, such as vision-based, sensor-based, and data glove-based techniques. It emphasizes the importance of accurate hand modeling and feature extraction for capturing and analyzing gestures. Key features like motion, depth, color, shape, and pixel values and their relevance in gesture recognition are discussed. Challenges faced in hand gesture recognition include lighting variations, complex backgrounds, noise, and real-time performance. Machine learning algorithms are used to classify and recognize gestures based on extracted features. The paper emphasizes the need for further research and advancements to improve hand gesture recognition systems' robustness, accuracy, and usability. This review offers valuable insights into the current state of hand gesture recognition, its applications, and its potential to revolutionize human-computer interaction and enable natural and intuitive interactions between humans and machines. In simpler terms, hand gesture recognition is a way for computers to understand what people are saying with their hands. It has many potential applications, such as allowing people to control computers without touching them or helping people with disabilities communicate. The paper reviews different ways to develop hand gesture recognition systems and discusses the challenges and opportunities in this area.

Keywords

Hand gesture recognition, HCI, Sign language recognition, Feature extraction, Machine learning algorithms.

I. Introduction

Human gesture recognition is a challenging and significant field in computer science, aiming to interpret human gestures using mathematical models.

Gestures are nonverbal communication methods using the body's movement to convey messages, originating from various parts of the human body, with the most common ones coming from the hand or face. In Webster's dictionary [1], it can find:

"A gesture is a movement usually of the body or limbs that expresses or emphasizes an idea, sentiment, or attitude." Gestures are essential for human communication, as they are the foundation of language and can convey various types of information. Without gestures, actions and movements are incomplete, lacking genuine feelings and thoughts. Pointing finger gestures can indicate time and space, indicating needs or wishes among people.

Hand gestures can be categorized into conversational, controlling, manipulated, and communication [2, 3]. Sign language is a crucial state of communication gestures used in vision algorithms for structural analysis. Analyzing pointing gestures for virtual identification is essential in Vision-Based Interface (VBI) research. Navigating gestures capture hand direction as 3D directional, while manipulated gestures enable natural

interaction with virtual objects, virtual assembly, and remote operation. Communicative gestures are crucial in human interaction and psychological research. Techniques based on vision capture motion techniques can aid this research [4].
Gestures can be categorized into static gestures based on hand shape and dynamic gestures influenced by hand movements [5]. In ref. [6], Pinto et al defined hand posture as a combination of position, orientation, and flexion observed at a specific time. Static hand gestures involve a single posture over time, allowing for interpretation through one or more hand images. Simple signs for this gesture include ”OK” or ”STOP.”
In a video signal, a dynamic hand gesture [7] is a series of postures connected by motion over a brief period. These gestures require identifying temporal context information and can be static or dynamic in certain situations. Sign language is a communication method using gestures and postures to communicate with deaf or dumb people. These gestures can be used in systems control, Augmented Reality, Gaming, robotics, and vision-based applications. Via imaging devices, a sign language recognition system interprets sign language into corresponding text. However, there are complexities in this process, as spoken languages change across different countries and regions, affecting the corresponding gestures [8].
Previous research shows that the deaf and dumb community will increase by 2050 compared to what it currently exists due to noise and other reasons, and with the increase in the number of the deaf and dumb community compared to the ignorance of ordinary people in sign language and the need of the two communities to communicate, recognition of sign language has become of great importance by using technological methods [9].

II. Hand Gesture Modeling

Modeling the hand is crucial for understanding posture and gestures as interfaces in Human-Computer Interaction (HCI) stages [10]. Hand modeling relies on kinematic structure for accurate techniques [11]. Gestures can be modeled spatially or temporally, focusing on posture characteristics in HCI application environments and dynamic hand gestures in time modeling. Spatial modeling for hand modeling in two- and three-dimensional spaces [12, 13].
Fig.1 displays four 2D shape types: shape, motion, colored marker, and deformable templates [12]. Geometric models are based on fingertips and palm features, while non-geometric models use features like silhouette, texture, color, contour, edges, image moments, and eigenvectors.
Deformable templates or flexible models provide a flexible level of the object shape change [14] to allow for a slight change in the hand's shape. The motion-based model can be implemented concerning color cues for hand tracing, color markers, or track and finger position estimation for hand shape modeling [15]. The 3D model will be discussed in appearance-based approaches in the section 3D model-based approach.

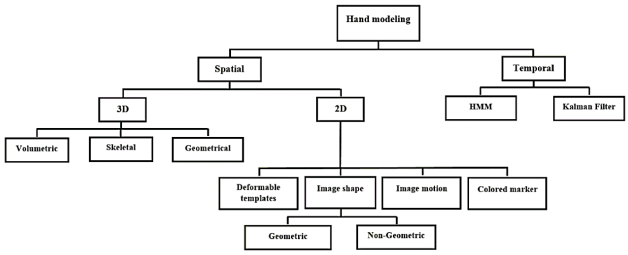

Fig. 1. Hand gesture modeling [16].

III. Hand Gesture Analysis Approaches

Identifying hand gestures involves obtaining data for specific tasks using Vision-Based Data-Gloves and Color-Marker techniques.

A. Vision-Based Approaches

Vision-based approaches use cameras or more to obtain human motion, allowing devices to interpret gesture properties like color and texture [17, 18]. However, challenges like lighting diversity, complex backgrounds, and clutter can arise, as well as time, speed, durability, and computation efficiency. These techniques differ in seven ways [19]:
1. Several cameras.
2. Response time and speed.
3. The physical features of the environment, such as the pace of movement and illumination.
4. What accessories and clothing are required for use?
5. Low-level features like histograms, silhouettes, edges, moments, and regions are utilized.
6. Decide between two- and three-dimensional representations.
7. Is the representation of time accurate?

B. Glove-Based Approaches

Sayre Glove was designed by the Electronic Visualization Lab in 1977, and it was the first data glove [20]. Researchers believe sign language influences gestures and can be used to create computer instructions. Data glove approaches use sensors to capture hand position, fingertips, acceleration, movement velocity, orientation, and motion, enabling accurate computation of finger palm and hand configuration coordinates [21]. Sensors struggle with easy computer connection due to physical user connection, high cost, and unsuitability for virtual reality environments, making them unsuitable for virtual real-

TABLE I.
A CONCISE COMPARISON OF HAND GESTURE RECOGNITION APPROACHES

Approach	Pros	Cons
Vision-based	Natural interaction, low cost	Lighting variations, complex backgrounds, clutter, time and computation requirements
Glove-based	Accurate computation of hand configuration	Expensive, difficult to connect to computers, unsuitable for virtual reality
Colored-markers	Simple and inexpensive	Limited natural interaction

ity environments [22]. This data-obtaining technique is often used in sign language [23] and Gaming [24]. Moore's Law predicts sensor size and affordability will increase in the future. Data gloves, including MIT, CyberGlove II, CyberGlove III, Fifth Dimension Sensor Glove Ultra, P5, and X-IST, are expected to become more prevalent.

C. Colored-Markers Approaches

The use of colored gloves for hand tracking and locating fingers and palms has been developed using markers and wool gloves [25]. These gloves extract geometric characteristics to outline the hand's shape. However, natural interaction between humans and computers remains insufficient, despite sensor or data glove advancements.
Table I shows a concise table of hand gesture recognition approaches.

IV. HAND GESTURE RECOGNITION APPROACHES

Techniques for gesture hand recognition rely on the bare hand and extract data, offering properties like simplicity and ease [26]. They provide direct connections with computers and can be categorized into two types.

A. Appearance-Based Approach

Appearance-based approaches design by extracting features from input hand images and comparing them to stored features [27]. This method is simpler and easier than three-dimensional models but can be affected by lighting and background objects. These patterns are part of a general pattern recognition problem consisting of three tasks: extracting features, classifying from labeled training samples, and classifying unknown samples.

B. 3D Model-based Approach

The hand's form is modeled and examined using the 3D model description [28], including the kinematic parameters required to project the three-dimensional model into a two-dimensional one. However, this method can lead to a loss of features. Various methods include volumetric, geometrical, and skeleton types [29]. Volumetric models deal with the three-dimensional and visual appearance of the human hand, while skeleton methods limit the set of parameters needed to form the hand's shape. Geometrical models mimic the efficiency of visual images but require several parameters and a time-consuming process. Geometric shapes, such as mesh polygons and cardboard models, can also be used to approach visual forms. However, this method has several disadvantages [30], including the need for initial parameters to be near the solution for each image, noise in the imaging phase, difficulty in extracting features, and incapability to deal with singularities caused by ambiguous views. In general, appearance-based detection in real-time is better than 3D-model-based methods, but it can represent a wide range of hand gestures.

V. FEATURE EXTRACTION

Feature extraction techniques collect data on gestures' position, direction, posture, and temporal progression. They process and analyze low-level data (pixel values) to produce high-level data like object contours. A manual algorithm is created to recognize and encode specific image features, such as texture, shape, and color. When an image is a collection of pixels, a manually defined algorithm is applied to obtain a feature vector that describes the image's contents. The resulting feature vectors are used as inputs for machine-learning models. Feature extraction reduces data dimensionality by encoding relevant information into a compressed representation and eliminating less discriminative data. The efficiency of gesture recognition relies heavily on feature extraction, making the most critical design decisions in hand motion, and gesture recognition is deciding which features to work with and how to extract them. The following subsections discuss popular feature types and computation techniques, including shape and motion features.

A. Motion

Using frame-to-frame comparisons, researchers use motion cues to detect hands, which are computationally efficient methods for finding foreground objects and detecting their motion and position. This method relies on assumptions like a static background, image pre-processing, and a stable camera. Binh et al [31] used the Kalman filter to predict the hand's position in one frame based on its observed position in the previous frame. Modern approaches combine motion information with other visual cues to enhance detection.

B. Depth

Depth features calibrated from cameras or depth sensors like Light Detection and Ranging (LiDAR) [32] or Kinect [33] indicate a human face and nearest objects. However, due to coarse-grained and noisy depth, they are often paired with other image cues like color [34].

C. Color

Skin color [35] is a crucial image feature for detecting and tracking human hands, but it faces challenges in eliminating similar objects like the arm and face. To address this issue, people often wear long-sleeved shirts, restricting the colors of other objects. A skin color threshold is a broad range that accurately extracts the current skin color in an image. Fine-tuning this threshold is necessary to reduce skin color, as human skin conditions differ in shade, lighting, and pigmentation. Skin color includes both skin pixels and other skin pixels, and effective skin segmentation is necessary to eliminate false-positive pixels. Choosing the right color space for skin color models is crucial. Popular options include RGB, HSV, normalized RGB, YUV, and YCrCb. Chromatically based color components can enhance lighting changes and separate chromaticity and luminance components [36].

Color segmentation can be challenging due to background objects with color distributions similar to human skin. Background subtraction is the best approach but often relies on camera motion. Research has investigated dynamic correction methods for background models to compensate for these issues [37].

D. Shape

Hand detection in images relies on their distinct shape, collected by extracting the contours of the image's objects [38]. These contours reflect the hand's shape and are unaffected by skin color, viewpoint, or illumination. Edge detection-based contour extraction often produces many hands' edges and unrelated background objects, making complex post-processing methods crucial for increasing reliability. Edges are often combined with background subtraction motion cues and skin color. Local topological descriptors match the model to the image's edges, while a shape context descriptor, introduced by Zou et al [39], distinguishes the position of a specific point on a shape. The theory behind detection is that corresponding points on two unlike shapes would possess an ideally identical shape context.

E. Multi-cues

Many systems can merge information from multiple image cues to improve accuracy and speed. This involves combining appearance-based hands, body, or face detection with motion-based regions of interest (ROI) [40]. However, when one cue is removed, performance suffers. Multi-cue methods do not detect gestures based on color and shape, as they are performed sequentially rather than cooperatively. Real multi-cue methods face data fusion issues, similar to traditional sensor fusion.

F. Moment Invariants

Hu [41] proposed geometric moment invariants, which include scale, rotation, and translation invariants based on normalized central moments up to the third order. These invariant global features are commonly used in pattern recognition and image classification to define objects' unique shape features. However, high geometric moment values can cause numerical instabilities and noise sensitivity. Hu's moments were used to recognize hand gestures [42, 43].

G. Pixel Values

Numerous classification approaches have been developed to find hands based on texture and appearance in gray-level images. Mirehi et al [44] examined the appropriateness of various classification approaches for recognizing view-independent hand postures. Many methods use image samples to train classifiers to detect hands based on their appearances, with the fundamental presumption that hand appearance varies more between gestures than between people who make the same gesture. However, automatic feature selection remains a challenge. Boosting, a machine learning method, has shown strong hand and face detection results. Boosting involves linear grouping of imprecise or weak classifiers to create a highly precise or robust classifier [45].

H. 3D Model

Using powerful and inexpensive depth sensors, researchers can now recognize 3D scene details, background subtraction, and hand detection [46]. These sensors are more resistant to light changes, making them suitable for detecting hands in images. These methods can detect objects regardless of their view and require various image features. Kinematic hand models use line and point features to retrieve angles formed at hand joints. Hand postures can be estimated if the correspondences between the 3D model and observed features are fine-defined.

VI. Gesture Recognition Techniques

Various classifiers, such as SVM and specialized form classifiers, then process the extraction of features. The primary objective of hand gesture recognition is to understand the semantics transmitted by the hand's location, posture, or gesture [47]. Hand gestures can be classified as static or dynamic based on vision, with general classifiers detecting static gestures and Hidden Markov Models processing dynamic gestures. Learning algorithms can be categorized based on their

results, including supervised, unsupervised, semi-supervised, reinforcement, transduction, and learning to learn [48]. The chosen learning algorithm is largely determined by the hand gesture representation chosen.

Most available hand recognition technologies don't gather all the data or are static, making them inappropriate for many circumstances. Different gesture detection methods rely on strategies from modeling image processing, computer vision science, and pattern recognition. Popular methods for distinguishing between dynamic and static hand movements include:

A. Artificial Neural Networks

Artificial neural networks (ANNs) are non-linear algorithms that extract features through hidden layers and classify them by multilayer perceptrons [49]. These ANNs' approximate functions receive large numbers of unknown inputs. The nodes in ANNs are the core units of the brain, and links connect nodes with a weight representing a way to store data. ANNs have been used to address problems in speech recognition, Visual-Interpretation, and Roberts-Control. Most ANNs are implemented in sequential machines but can work in parallel machines, making them suitable for applications in these fields. Several researchers have used ANNs for gesture recognition, such as self-learning and big data processing. Some have used ANNs in classification processes for gestures, while others have used them to segment the hand. Maung [50] introduced a system for tracking hands and recognizing Myanmar-Alphabet-Language (MAL) by NNs, using an Adobe Photoshop filter to find the edges of the hand image. A system for recognizing static hand motions was developed by Stergiopoulou [51] utilizing SGONG Networks (Self-Growing and Self-Organized Neural Gas). Higher-level 2D features, such as angles, palm positions, and the number of lifted fingers, are extracted from hand topology to recognize static hand gestures.

The method Ismail et al [52] developed uses two recurrent NNs to recognize Arabic Sign Language. Segmentation was carried out using the Elman network, partially utilized independently of the recurrent and fully recurrent NNs. For Elman-NN and fully recurrent NN, respectively, the recognition rate was 89.67% and 95.1%.

B. Hidden Markov Models

In the mid-1990s, Hidden-Markov models (HMM) were introduced as a solution to the segmentation problem in recognition tasks. HMMs are non-deterministic and stochastic, with one or more arcs carrying the same value [53]. They are useful in speech and sign language recognition, as they can handle the limitations of Markov chains. Sarma and Bhuyan [54] introduced a hand gesture recognition system using real-time hand tracking and HMM for recognizing three-dimensional gestures.

HMMs have high accuracy and can define several gestures but require training and performance improvements. Each motion or posture must define the precise number of states. HMMs are the ideal choice when the types and quantity of gestures or postures are pre-defined. However, retraining may be more time-consuming when gestures and hand postures are determined during system development. The hidden Markov model requires significant training time, making knowing what is happening inside it difficult. Despite high precision, HMMs are considered the best method due to their high precision, reaching over 90% [55].

C. Fuzzy Clustering Algorithms

Patterns can belong to various data sets due to fuzzy-clustering algorithms' flexible grouping of the data [56]. Xingyan [57] unveiled a device that recognizes hand motions using a fuzzy c-means clustering technique in a mobile remote. The system performs pre-processing procedures such as removal of noise, removing undesired items, and threshold approaches. It also extracts the shape of the hand from camera images and converts them to HSV space color. The aspect ratio is the bounding box for the thirteen items that comprise the features vector, and grid cells are represented by grid cells. The FCM algorithm executes six motions with twenty language samples in various situations, including those with consistent lighting and complex backgrounds. The recognition accuracy rate of the system is 85.83

D. K-Nearest Neighbor

K-NN is a statistical technique for classifying objects in feature space using the nearest training instances. It uses locally approximating functions and postpones calculations to classification [58]. The object is categorized by a preponderance of votes from its neighbors, assigning it to the class with the most common k nearest neighbors. K-NN is easy to implement and understand, and its object property value can be assigned to the average of the k nearest neighbors' values for regression. The neighbors are selected from the objects group for which the right classification is defined.

E. Support Vector Machine

Vladimir Vapnik introduced the support vector machine (SVM), a nonlinear classifier that improves classification by mapping data to a high-dimensional space [59]. SVM classifiers are similar to neural networks, with a two-layer perceptron neural network equivalent. SVM models represent examples as space points separated by the widest possible distance and classify them according to their side of the gap. SVMs are designed as binary classifiers, but other methods treat multi-class problems as one optimization problem. Using SVMs as binary

TABLE II.
A COMPARISON OF MACHINE LEARNING ALGORITHMS FOR HAND GESTURE RECOGNITION.

Algorithm	Pros	Cons
Artificial Neural Networks	Can learn complex features and patterns, with high accuracy	Can be computationally expensive, difficult to train
Hidden Markov Models	Can handle sequential data, with high accuracy	Requires training, can be difficult to define the number of states
Fuzzy Clustering Algorithms	Flexible grouping of data can handle noisy data	Can be computationally expensive
K-nearest Neighbor	Simple to implement and understand, can handle numerical and categorical data	Sensitive to the choice of k, can be computationally expensive for large datasets
Support Vector Machines	High accuracy, robust to noise	Can be computationally expensive for large datasets
Dynamic Time Warping	Can handle sequential data with varying time scales, high accuracy	Can be computationally expensive for large datasets

classifiers has been suggested to break multi-class problems into multiple binary problems.

F. Dynamic Time Warping

Dynamic Time Warping (DTW) algorithm is widely used for finding the best alignment between two signals [60]. It computes the distance between point pairs and their related feature values, computes a cumulative distance matrix, and obtains the smallest costly path. This path reflects the perfect warp, reducing feature distance among synchronized points. DTW is used in data mining, speech recognition, and motion recognition. Prior work has mainly focused on algorithm acceleration, constraint analysis, algorithm approximation, and lower bounding methods.

Derivative DTW (DDTW) is a form of DTW that measures distances between first-order derivatives of points, focusing on shape characteristics [61]. Most of the work is one-dimensional.

Table II illustrates a concise table of machine learning algorithms for hand gesture recognition.

VII. DEEP LEARNING APPROACHES

Computer Vision (CV) applications have evolved from human activity recognition to speech recognition, image classification, and labeling [62]. Deep learning, a machine learning technology, has emerged as a successful solution since 2010. Researchers have transitioned from traditional handcrafted features to data-driven algorithms since 2010. Various learned-based feature methods, such as genetic programming and dictionary-based approaches, have been used for visual recognition tasks. Still, deep learning has significantly influenced computer vision in recent years.

Vision-based recognition algorithms include handcrafted features and learned-based features [63]. Handcrafted features resolve CV problems by collecting correct features from data. The pipeline typically includes four major stages; **Section 3.1** provides further details.

Unlike handcrafted methods, deep learning uses convolution neural networks to learn features automatically during training. This approach interprets problems in terms of conceptual hierarchy, with lower network concepts encoding simple representations and high-level layers constructing abstract concepts. This hierarchical learning process allows for the complete elimination of handcrafted feature extraction and allows convolution neural networks to function as end-to-end learners.

Traditional algorithms like Support Vector Machine, Random Forest, and Hidden Markow Model rely heavily on data representation, but handcrafted features often cause information loss. Deep learning algorithms have shown impressive results in CV challenges but require large amounts of data for accurate learning. Deep neural network training and parameterization also require significant computational resources and experimentation time [64].

VIII. HAND GESTURE RECOGNITION APPLICATIONS

Hand Gesture Recognition is crucial in various applications, including human-computer interaction, sign language recognition, virtual and augmented reality, and robotics. Characters use hand gestures in future films, highlighting how interaction systems can influence our interactions with computers if current systems do not allow such freedom.

A. Virtual and Augmented Reality Technologies

Virtual Reality (VR) is a digital technology that mimics a human's presence in a virtual setting by creating sounds, images, and feelings using headsets [65]. AR, or augmented reality, improves the real world, while VR completely replaces it with an artificial one. American teen Palmer Luckey built a VR headset prototype in 2010, which evolved into the Oculus Rift.

Rivals like the HTC Vivi, Sony PlayStation VR, Gear VR, and Google Cardboard headset have emerged, leading to developers working on virtual reality games and applications [66]. These technologies have various applications, including video games, art, medical applications, education, and flight training [67].

B. Recognition of Sign Language

Sign language is a primary language for individuals with speech and hearing impairments, utilizing nonverbal communication such as gestures. Sign languages are divided into three main parts: fingerspelling, sign language vocabulary at the word level and non-manual features [68]. Fingerspelling involves using gestures to spell words, while non-manual features involve facial expressions and body postures. Due to the limited understanding of sign language, hard-of-hearing individuals often require the assistance of a trained interpreter. However, hiring an interpreter can be costly and not always feasible. Using visual cues, a sign language recognition system could provide a low-cost, natural, and comfortable way to interact with deaf or hard-of-hearing individuals with impaired speech.

C. Robot and Ambient-Aided Living

Ambient Assisted Living (AAL) is a sub-field of Ambient Intelligence that integrates new technologies and social environments to improve life quality [69]. Vision-based assistive systems can benefit patients by observing their daily activities. As robots become more integrated into our daily lives, the challenge of communicating with them becomes more apparent. Hand gestures play a significant role in natural interaction, especially when a robot is designed to assist people in daily tasks.

Hand gestures can be used in rehabilitation treatment, controlling medical equipment, and assisting disabled individuals. Approaches for integrating hand gestures with physician-computer interfaces have been explored, with Gestix being a tracking system for hand gesture tracking in the operations room [70].

IX. Datasets Of Dynamic Gesture Recognition

Gesture recognition involves various datasets, including categories, scale, annotations type, sensors, and gesture domain. The Cambridge hand gestures datasets [71], a recent addition, consists of 900 RGB sequences from nine classes of gestures. Sheffield Kinect Gesture (SKIG) [72], a dynamic gesture datasets, has 2160 hand gesture sequences divided into ten classes. ChaLearn Gesture Challenge offers popular datasets like the ChaLearn LAP IsoGD, ConGD [73], and Multi-modal Gesture Datasets (MMGD) [74], which include gesture classes from nine domains, including Italian sign language, pantomime, and actions. The effort to recognize gestures inside a car is described in [75], which provides driver's hand gestures performed by eight different actors from one viewpoint against a plain background. Other datasets for sign language include RWTH-BOSTON-50, RWTH-400, NATOPS, and BIGHands [76]. The BIGHands datasets [77], designed for the posing of hands, is rich in hand pose variation and joint annotation but not explicitly reflective of gestures.

X. Evaluation of Model Performance

The following parameters are used for evaluating the classifier's performance [78]:

TN: is the number of true negatives.

TP: is the number of true positives.

FN: is the number of false negatives.

FP: is the number of false positives.

Precision: the classifier can avoid labeling a negative sample as positive. (See (1))

$$Precision = \frac{TP}{(TP+FP)} \tag{1}$$

Recall-Sensitivity: the classifier can find all positive samples as in (2).

$$Recall = \frac{TP}{(TP+FN)} \tag{2}$$

The F1-score is a weighted harmonic mean for recall and precision, with the highest value being one and the lowest value being 0, as in (3).

$$F1\_score = 2X\frac{recallxprecision}{recall + precision} \tag{3}$$

The support is the number of times in y-true every class appears.

Micro Average: expresses the function for computing the metric by considering all false negatives, true positives, and false positives (regardless of the dataset's predictions for each class). If you detect a class imbalance, a micro-average is preferred (i.e., you might have more samples of one class than others might).

Macro Average: states the function that calculates the metric independently for every class, and the average is returned without taking the percentage of every label in the datasets into account. (As a result, all classes are treated equally).

Weighted Average: states the function that computes f1 for each label in the datasets, and the average is returned based on the proportion in the datasets for every label.

XI. Conclusion

The review emphasizes the importance and challenges of hand gesture recognition in various fields, such as human-computer interaction, sign language recognition, virtual reality, gaming, and robotics. It explores different approaches, such as vision-based, sensor or data glove-based, and colored-marker techniques, and their advantages and limitations. Accurate hand modeling and feature extraction are crucial for capturing and analyzing hand gestures. Machine learning algorithms are essential for classifying and recognizing gestures based on extracted features. Challenges in hand gesture recognition include lighting variations, complex backgrounds, noise, and real-time performance. The review acknowledges the need for further research and advancements to improve hand gesture recognition systems' robustness, accuracy, and usability. The review provides valuable insights into the current state of hand gesture recognition, its applications, and the potential for enhancing human-computer interaction and communication between different communities.

Conflict of Interest

The authors have no conflict of relevant interest to this article.

References

[1] "Webster's dictionary accessed: 12-oct-2022."

[2] T. Zhang, Y. Ding, C. Hu, M. Zhang, W. Zhu, C. R. Bowen, Y. Han, and Y. Yang, "Self-powered stretchable sensor arrays exhibiting magnetoelasticity for real-time human–machine interaction," *Advanced Materials*, vol. 2203786, p. 2203786, 2022.

[3] F. A. Farid, N. Hashim, J. Abdullah, M. R. Bhuiyan, W. N. S. M. Isa, J. Uddin, M. A. Haque, and M. N. Husen, "A structured and methodological review on vision-based hand gesture recognition system," *Journal of Imaging*, vol. 8, no. 6, p. 153, 2022.

[4] M. G. A. J. P. Rawat, L. Kane and S. Sehgal, "A review on vision-based hand gesture recognition targeting rgb-depth sensors," *International Journal of Information Technology and Decision Making*, vol. 22, no. 01, pp. 115–156, 2023.

[5] S. Wu, Z. Li, S. Li, Q. Liu, and W. Wu, "An overview of gesture recognition," in *International Conference on Computer Application and Information Security (ICCAIS 2022)*, vol. 12609, pp. 600–606, SPIE, Mar 2023.

[6] R. F. Pinto Jr, C. D. Borges, A. M. Almeida, and I. C. Paula Jr, "Static hand gesture recognition based on convolutional neural networks," *Journal of Electrical and Computer Engineering*, vol. 2019, no. 1, p. 4167890, 2019.

[7] A. K. H. AlSaedi and A. H. H. AlAsadi, "A new hand gestures recognition system," *Indonesian journal of electrical engineering and computer science*, vol. 18, no. 1, pp. 49–55, 2020.

[8] P. Das, T. Ahmed, and M. F. Ali, "Static hand gesture recognition for american sign language using deep convolutional neural network," in *2020 IEEE Region 10 symposium (TENSYMP)*, pp. 1762–1765, IEEE, 2020.

[9] I. Papastratis, C. Chatzikonstantinou, D. Konstantinidis, K. Dimitropoulos, and P. Daras, "Artificial intelligence technologies for sign language," *Sensors*, vol. 21, no. 17, p. 5843, 2021.

[10] L. I. Khalaf, S. A. Aswad, S. R. Ahmed, B. Makki, and M. R. Ahmed, "Survey on recognition hand gesture by using data mining algorithms," in *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, pp. 1–4, IEEE, 2022.

[11] 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, 2020.

[12] D. V. Suma, "Computer vision for human-machine interaction-review," *Journal of Trends in Computer Science and Smart Technology*, vol. 1, no. 2, pp. 131–139, 2019.

[13] T. Vuletic, A. Duffy, L. Hay, C. McTeague, G. Campbell, and M. Grealy, "Systematic literature review of hand gestures in human-computer interaction interfaces," *International Journal of Human-Computer Studies*, vol. 129, pp. 74–94, 2019.

[14] H. Kawashima, *Active Appearance Models*, pp. 1–5. 2020.

[15] T. H. Tsai, C. C. Huang, and K. L. Zhang, "Design of hand gesture recognition system for human-computer interaction," *Multimedia tools and applications*, vol. 79, pp. 5989–6007, 2020.

[16] T. L. Dang, H. T. Nguyen, D. M. Dao, H. V. Nguyen, D. L. Luong, B. T. Nguyen, S. Kim, and N. Monet, "Shape: a dataset for hand gesture recognition," *Neural Computing and Applications*, vol. 34, pp. 21849–21862, Dec 2022.

[17] L. M. Dang, K. Min, H. Wang, M. J. Piran, C. H. Lee, and H. Moon, "Sensor-based and vision-based human activity recognition: A comprehensive survey," *Pattern Recognition*, vol. 108, Dec 2020.

[18] D. R. Beddiar, B. Nini, M. Sabokrou, and A. Hadid, "Vision-based human activity recognition: a survey," *Multimedia Tools and Applications*, vol. 79, pp. 30509–30555, Nov 2020.

[19] C. Z. Dong and F. N. Catbas, "A review of computer vision–based structural health monitoring at local and global levels," *Structural Health Monitoring*, vol. 20, pp. 692–743, Mar 2021.

[20] D. Dai, W. Zhuang, Y. Shen, L. Li, and H. Wang, "Design of intelligent mobile robot control system based on gesture recognition," in *Artificial Intelligence and Security: 6th International Conference, ICAIS 2020, Hohhot, China*, pp. 101–111, Springer Singapore, 2020.

[21] W. Lin, C. Li, and Y. Zhang, "Interactive application of data glove based on emotion recognition and judgment system," *Sensors*, vol. 22, Aug 2022.

[22] N. Magrofuoco, P. Roselli, and J. Vanderdonckt, "Two-dimensional stroke gesture recognition: A survey," *ACM Computing Surveys (CSUR)*, vol. 54, pp. 1–36, Jul 2021.

[23] N. Saleh, M. Farghaly, E. Elshaaer, and A. Mousa, "Smart glove-based gestures recognition system for arabic sign language," in *2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE)*, pp. 303–307, IEEE, Feb 2020.

[24] Q. Fu, J. Fu, J. Guo, S. Guo, and X. Li, "Gesture recognition based on bp neural network and data glove," in *2020 IEEE International Conference on Mechatronics and Automation (ICMA)*, pp. 1918–1922, IEEE, Oct 2020.

[25] R. Rastgoo, K. Kiani, and S. Escalera, "Sign language recognition: A deep survey," *Expert Systems with Applications*, vol. 164, 2021.

[26] S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, pp. 311–324, Apr 2007.

[27] A. Dzedzickis, A. Kaklauskas, and V. Bucinskas, "Human emotion recognition: Review of sensors and methods," *Sensors*, vol. 20, Jan 2020.

[28] A. K. Al-Saedi and A. H. Al-Asadi, "Survey of hand gesture recognition systems," in *Journal of Physics: Conference Series*, vol. 1294, IOP Publishing, Sep 2019.

[29] L. I. Yang, J. Huang, T. I. Feng, W. A. Hong-An, and D. A. Guo-Zhong, "Gesture interaction in virtual reality," *Virtual Reality and amp; Intelligent Hardware*, vol. 1, pp. 84–112, Feb 2019.

[30] A. Mujahid, M. J. Awan, A. Yasin, M. A. Mohammed, R. Damaševičius, R. Maskeliūnas, and K. H. Abdulkareem, "Real-time hand gesture recognition based on deep learning yolov3 model," *Applied Sciences*, vol. 11, May 2021.

[31] N. D. Binh, E. Shuichi, and T. Ejima, "Real-time hand tracking and gesture recognition system," in *Proc. GVIP*, pp. 19–21, Dec 2005.

[32] J. Zhao, A. Lyons, A. C. Ulku, H. Defienne, D. Faccio, and E. Charbon, "Light detection and ranging with entangled photons," *Optics Express*, vol. 30, no. 3, pp. 3675–3683, 2022.

[33] B. Açış and S. Güney, "Classification of human movements by using kinect sensor," *Biomedical Signal Processing and Control*, vol. 81, 2023.

[34] S. Machado, V. Mercier, and N. Chiaruttini, "Limeseg: a coarse-grained lipid membrane simulation for 3d image segmentation," *BMC bioinformatics*, vol. 20, pp. 1–2, 2019.

[35] Y. Huang and J. Yang, "A multi-scale descriptor for real-time rgb-d hand gesture recognition," *Pattern Recognition Letters*, vol. 144, pp. 97–104, 2021.

[36] M. Singh, I. V. Tewari, and L. Sheth, *Skin-Colour-Based Hand Segmentation Techniques*, pp. 1–26. IGI Global, 2022.

[37] X. Larriva-Novo, C. Sánchez-Zas, V. A. Villagrá, M. Vega-Barbas, and D. Rivera, "An approach for the application of a dynamic multi-class classifier for network intrusion detection systems," *Electronics*, vol. 9, no. 11, 2020.

[38] Y. Zhou, H. Guo, L. Ma, Z. Zhang, and M. Skitmore, "Image-based onsite object recognition for automatic crane lifting tasks," *Automation in Construction*, vol. 123, 2021.

[39] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proceedings of the IEEE*, 2023.

[40] B. Noh, H. Park, S. Lee, and S. H. Nam, "Vision-based pedestrian's crossing risky behavior extraction and analysis for intelligent mobility safety system," *Sensors*, vol. 22, no. 9, 2022.

[41] M. K. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, 1962.

[42] S. Katoch, V. Singh, and U. S. Tiwary, "Indian sign language recognition system using surf with svm and cnn," *Array*, vol. 14, 2022.

[43] Z. Ren, F. Fang, N. Yan, and Y. Wu, "State of the art in defect detection based on machine vision," *International Journal of Precision Engineering and Manufacturing-Green Technology*, vol. 9, no. 2, pp. 661–691, 2022.

[44] N. Mirehi, M. Tahmasbi, and A. T. Targhi, "Hand gesture recognition using topological features," *Multimedia Tools and Applications*, vol. 78, pp. 13361–13386, 2019.

[45] M. Wagh and P. K. Nanda, "Decision-theoretic rough sets based automated scheme for object and background classification in unevenly illuminated images," *Applied Soft Computing*, vol. 119, 2022.

[46] W. Chen, C. Yu, C. Tu, Z. Lyu, J. Tang, S. Ou, Y. Fu, and Z. Xue, "A survey on hand pose estimation with wearable sensors and computer-vision-based methods," *Sensors*, vol. 20, no. 4, p. 1074, 2020.

[47] J. Qi, K. Xu, and X. Ding, "Approach to hand posture recognition based on hand shape features for a human-robot interaction," *Complex & Intelligent Systems*, 2021.

[48] 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.

[49] A. Thakur and A. Konde, "Fundamentals of neural networks," *International Journal for Research in Applied Science and Engineering Technology*, vol. 9, pp. 407–426, 2021.

[50] T. H. Maung, "Real-time hand tracking and gesture recognition system using neural networks," *International Journal of Computer and Information Engineering*, vol. 3, no. 2, pp. 315–319, 2009.

[51] E. Stergiopoulou and N. Papamarkos, "Hand gesture recognition using a neural network shape fitting technique," *Engineering Applications of Artificial Intelligence*, vol. 22, no. 8, pp. 1141–1158, 2009.

[52] M. H. Ismail, S. A. Dawwd, and F. H. Ali, "Dynamic hand gesture recognition of arabic sign language using deep convolutional neural networks," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 25, pp. 952–962, 2022.

[53] N. Rajawat, N. Gupta, and S. Lalwani, "A comprehensive review of hidden markov model applications in predicting human mobility patterns," *International Journal of Swarm Intelligence*, vol. 6, no. 1, pp. 24–47, 2021.

[54] 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, 2021.

[55] S. Mandal, Z. Li, T. Chatterjee, K. Khanna, K. Montoya, L. Dai, C. Petersen, L. Li, M. Tewari, A. Johnson-Buck, and N. G. Walter, "Direct kinetic fingerprinting for high-accuracy single-molecule counting of diverse disease biomarkers," *Accounts of Chemical Research*, vol. 54, no. 2, pp. 388–402, 2020.

[56] J. Arora, K. Khatter, and M. Tushir, "Fuzzy c-means clustering strategies: A review of distance measures," in *Software Engineering: Proceedings of CSI 2015*, pp. 153–162, 2019.

[57] R. S. Gaikwad and L. S. Admuthe, "A review of various sign language recognition techniques," in *Modeling, Simulation, and Optimization: Proceedings of CoMSO 2021*, pp. 111–126, jun 2022.

[58] K. Taunk, S. De, S. Verma, and A. Swetapadma, "A brief review of the nearest neighbor algorithm for learning and classification," in *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, pp. 1255–1260, IEEE, may 2019.

[59] S. Ghosh, A. Dasgupta, and A. Swetapadma, "A study on support vector machine-based linear and non-linear pattern classification," in *2019 International Conference on Intelligent Sustainable Systems (ICISS)*, pp. 24–28, IEEE, feb 2019.

[60] M. Yu, J. Jia, C. Xue, G. Yan, Y. Guo, and Y. Liu, "A review of sign language recognition research," *Journal of Intelligent & Fuzzy Systems*, vol. 43, no. 4, pp. 3879–3898, 2022.

[61] U. Moser and D. Schramm, "Multivariate dynamic time warping in automotive applications: A review," *Intelligent Data Analysis*, vol. 23, no. 3, pp. 535–553, 2019.

[62] D. Bhatt, C. Patel, H. Talsania, J. Patel, R. Vaghela, S. Pandya, K. Modi, and H. Ghayvat, "Cnn variants for

computer vision: History, architecture, application, challenges, and future scope," *Electronics*, vol. 10, no. 20, 2021.

[63] M. A. Khan, M. Mittal, L. M. Goyal, and S. Roy, "A deep survey on supervised learning based human detection and activity classification methods," *Multimedia Tools and Applications*, vol. 80, pp. 27867–27923, jul 2021.

[64] W. Chen, Q. Sun, X. Chen, G. Xie, H. Wu, and C. Xu, "Deep learning methods for heart sound classification: A systematic review," *Entropy*, vol. 23, no. 6, 2021.

[65] B. Xie, H. Liu, R. Alghofaili, Y. Zhang, Y. Jiang, F. D. Lobo, C. Li, W. Li, H. Huang, M. Akdere, and C. Mousas, "A review of virtual reality skill training applications," *Frontiers in Virtual Reality*, vol. 2, apr 2021.

[66] C. Lewis and F. C. Harris Jr, "An overview of virtual reality," in *Proceedings of 31st International Conference*, vol. 88, pp. 71–81, nov 2022.

[67] A. Rizzo, S. Koenig, and B. Lange, "Clinical virtual reality: The state of the science," in *APA Handbook of neuropsychology, Volume 2: Neuroscience and neuro methods*, vol. 2, pp. 473–491, 2023.

[68] N. B. Ibrahim, H. H. Zayed, and M. M. Selim, "Advances, challenges, and opportunities in continuous sign language recognition," *Journal of Engineering and Applied Sciences*, vol. 15, no. 5, pp. 1205–1227, 2020.

[69] J. Wachs, H. Stern, Y. Edan, M. Gillam, C. Feied, M. Smith, and J. Handler, "A hand gesture sterile tool for browsing mri images in the or," *Journal of the American Medical Informatics Association*, vol. 15, pp. 321–323, may 2008.

[70] Z. Hosseinaee, M. Le, K. Bell, and P. H. Reza, "Towards non-contact photoacoustic imaging," *Photoacoustics*, vol. 20, dec 2020.

[71] Y. Zhang, S. Q. Xie, H. Wang, and Z. Zhang, "Data analytics in steady-state visual evoked potential-based brain-computer interface: A review," *IEEE Sensors Journal*, vol. 21, pp. 1124–1138, aug 2020.

[72] M. B. Shaikh and D. Chai, "Rgb-d data-based action recognition: A review," *Sensors*, vol. 21, jun 2021.

[73] J. Wan, Y. Zhao, S. Zhou, I. Guyon, S. Escalera, and S. Z. Li, "Chalearn looking at people rgb-d isolated and continuous datasets for gesture recognition," in *Proceedings of the IEEE Conference on computer vision and pattern recognition workshops*, pp. 56–64, 2016.

[74] S. e. a. Escalera, "Chalearn multi-modal gesture recognition 2013: grand challenge and workshop summary," in *Proceedings of the 15th ACM on International conference on multimodal interaction*, pp. 365–368, dec 2013.

[75] P. e. a. Molchanov, "Online detection and classification of dynamic hand gestures with recurrent 3d convolutional neural network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4207–4215, 2016.

[76] V. e. a. Athitsos, "The american sign language lexicon video dataset," in *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1–8, IEEE, jun 2008.

[77] S. e. a. Yuan, "Bighand2.2m benchmark: Hand pose dataset and state-of-the-art analysis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4866–4874, 2017.

[78] E. P. Costa, A. C. Lorena, A. C. Carvalho, and A. A. Freitas, "A review of performance evaluation measures for hierarchical classifiers," in *Evaluation methods for machine learning II: papers from the AAAI-2007 Workshop*, vol. AAAI Technical Report WS-07-05, pp. 1–6, 2007.