INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 5, May 2026 | ISSN: 2456-4184 | IJNRD.ORG

The Comprehensive Review of Gesture Recognition System Using Convolutional Nueral Network

Dr. P. Janarthanan[1] and T. Shruthi[2]

[1,2]Department of Computer Science and Engineering, Sri Venkateswara College of Engineering, Sriperumbudur, TamilNadu, India.

ABSTRACT

This review examines advancements in vision-based hand gesture recognition systems documented in academic literature between 2014 and 2020. The primary objective is to assess developmental progress and identify areas requiring further investigation. Through systematic application of targeted keywords across reputable online databases, we retrieved and analyzed 98 relevant publications. Our findings indicate that vision-based hand gesture recognition represents an actively evolving research domain, with numerous studies contributing to dozens of annual publications in both journals and conference proceedings. The majority of examined works concentrate on three fundamental components: data collection methodologies, environmental conditions, and gesture representation techniques. Regarding system performance measured through recognition accuracy, signer-dependent approaches demonstrated accuracy ranging from 69% to 98%, averaging 88.8% across reviewed studies. Conversely, signer-independent systems achieved accuracy between 48% and 97%, with a mean value of 78.2%. The relatively limited progress in continuous gesture recognition suggests substantial research efforts remain necessary for developing practically applicable vision-based gesture recognition systems.

Keywords: Classification, Feature Extraction, Dynamic Hand Gesture Recognition, Sign Language Recognition, Vision-Based Hand Gesture, Recognition Accuracy.

1. INTRODUCTION

Non-verbal communication constitutes approximately 65% of human interaction, whereas verbal communication contributes merely 35% to our daily exchanges. Gestures encompass multiple categories including hand and arm movements (involving hand posture recognition, sign language interpretation, and entertainment applications), head and facial expressions (such as nodding, gaze direction, mouth movements during speech, winking), and full-body gestures.

Robust and accurate gesture recognition methods are essential for effective human-computer interaction (HCI). These recognition systems serve as alternatives to conventional HCI input devices including mice and keyboards. Hand gesture recognition stands among the most actively researched domains and represents one of the most significant areas within HCI, particularly valuable for applications requiring natural human-machine interaction.

The development of hand gesture recognition systems, particularly sign language applications, holds substantial importance in eliminating communication barriers between deaf communities and individuals unfamiliar with sign language. Technological solutions that automatically translate hand movements into written text or audible speech enable non-signing individuals to comprehend sign language, thereby reducing communication obstacles.

Vision-based hand gesture recognition systems find applications across diverse domains including communication, education, and rehabilitation tools. These systems can assist in situations where human interpreters are unavailable for sign language translation.

Hand gesture recognition presents considerable challenges for several reasons. First, systems must accommodate inputs that differ substantially from training data. For hand gesture recognition systems, unexpected inputs may include

environmental noise, variations among signers, language differences, and similar factors. Researchers typically impose environmental restrictions on signers to simplify segmentation and tracking challenges.

Another significant challenge involves managing transitional movements between consecutive signs, as identifying precise gesture boundaries proves difficult. A system's failure to detect boundaries between signs may result in inaccurate or poor recognition outcomes. Due to this complexity, researchers have devoted comparatively less attention to continuous sign language in vision-based hand gesture recognition, limiting real-world applicability.

Developing robust signer-independent hand gesture recognition systems—those usable by individuals not represented during training—represents an additional challenge. Such systems are highly desirable for practical applications as they accommodate diverse users without requiring individualized system training.

Although previous review papers have summarized existing hand gesture recognition research, these works typically examined overall progress broadly, encompassing both device-based and vision-based systems for sign language detection. Given that vision-based hand gesture recognition systems offer practical advantages for real-life applications, they must function effectively for any user across various environments. However, no previous review has specifically examined the extent of research progress toward vision-based hand gesture recognition systems and identified potential future directions. The present review addresses this gap by comprehensively examining historical and contemporary literature to evaluate progress in vision-based hand gesture recognition.

The subsequent sections of this paper are structured as follows. Section 2 presents the research background. Section 3 discusses research objectives and methodology, including review aims and formulated research questions. Section 4 presents principal findings addressing the research questions. Section 5 concludes the paper.

2. BACKGROUND

Hand gesture recognition technology converts sign language hand movements into output formats such as text or voice. This technology can be classified based on gesture capture methodology: vision-based systems employing one or multiple cameras, and device-based systems utilizing direct-measurement devices such as sensor-equipped electronic gloves connecting users to the system.

While device-based systems offer efficiency advantages, their practical utility remains limited due to the requirement of wearing cumbersome equipment during system interaction. Vision-based systems avoid this limitation, enabling more natural user interaction and offering broader applicability in outdoor scenarios.

The user-friendliness of vision-based systems is counterbalanced by challenges in processing datasets containing dynamic hand gestures in sign language, including isolated and continuous signs. Research indicates that although most existing work focuses on isolated gesture recognition, such systems have limited real-world applicability. Furthermore, vision-based hand gesture recognition development necessitates more powerful feature extraction and discrimination methods.

Substantial research interest in gesture recognition has generated numerous review papers examining various aspects. Some reviews have surveyed state-of-the-art techniques in hand gesture and sign language recognition across data acquisition, preprocessing, segmentation, feature extraction, and classification. Others have focused on specific time periods or particular dimensions including data acquisition techniques, static versus dynamic signs, signing modes, one-handed versus two-handed signs, classification techniques, and recognition rates. Additional reviews have examined vision-based continuous sign language recognition systems or specific proposed models for sign language recognition.

3. RESEARCH AIMS AND APPROACH

This paper aims to analyze the current challenges, advancements, and possible future directions in vision-based hand gesture recognition research. To achieve this objective, two research questions were formulated.

Research Question 1 (RQ1): What are the current challenges and advancements in vision-based hand gesture recognition systems with respect to data acquisition, data environment, and hand gesture representation?

To address this question, research articles related to vision-based hand gesture recognition published between 2014 and 2020 were systematically collected and examined in order to identify existing issues and the solutions proposed by researchers.

Research Question 2 (RQ2): What is the performance of existing vision-based hand gesture recognition systems, and what are the possible future research directions?

To answer this question, the performance of various vision-based hand gesture recognition systems was analyzed based on recognition accuracy, enabling the identification of research gaps and future development opportunities in gesture recognition technology.

Search Methodology

A structured search methodology was adopted using specific keywords to identify relevant research articles. The search process focused on the following keywords:

1. Sign language recognition
2. Dynamic hand gesture recognition

The literature search was conducted using widely recognized academic databases, including:

1. ScienceDirect
2. IEEE Xplore Digital Library
3. SpringerLink
4. Google Scholar

To refine the search results, several inclusion criteria were applied:

- Publications released between 2014 and 2020
- Research within the domains of science, technology, and computer science
- Publication types including journals, conference proceedings, and transactions
- Full-text research and review articles
- Studies related to sign language hand gestures, including isolated words, continuous sentence recognition, and dynamic finger spelling in vision-based systems
- Articles published in English

In addition, the following exclusion criteria were considered to filter irrelevant studies:

- Studies not specifically focused on vision-based hand gesture recognition systems for sign language
- Research unrelated to sign language gesture recognition
- Studies where hand gesture recognition was only discussed as a secondary topic
- Review papers summarizing previous works without original contributions

- Papers lacking sufficient experimental details or methodology descriptions
- Articles whose full text was unavailable in either electronic or physical form
- Non-research materials such as opinions, editorials, keynote talks, tutorials, comments, anecdotal reports, discussion papers, and slide presentations without associated research papers.

4. FINDINGS OF THE REVIEW

Through application of our search keywords and fulfillment of inclusion and exclusion criteria, 98 articles were selected for analysis. These articles underwent thorough examination of abstracts, methodologies, discussions, and results. Table 1 presents article distribution according to publication type and quantity retrieved. The IEEE Explore Digital Library contributed the majority of papers.

Table 1. Distribution according to publication type and number of papers

Digital Database Libraries	Keyword and Hits	Total
	Sign Language Recognition	Dynamic Hand Gesture Recognition
Science Direct	10	3
IEEE Explorer Digital Library	52	13
Springer Link	4	-
Google Scholars	14	2
Total	**80**	**18**

A. Research Question 1 (RQ1): Current Challenges and Progress in Vision-Based Hand Gesture Recognition Systems

Among the 98 reviewed articles, the majority highlighted challenges and progress related to data acquisition and environmental conditions (n=47) as well as hand gesture representation (n=44).

1) Challenges

a. Data Acquisition and Environmental Conditions

Table 2 presents the challenges addressed by existing research. Among 47 articles discussing data acquisition and environmental conditions, over 80% (39 articles) were conducted within restricted laboratory environments. Researchers explain that ideal gesture recognition backgrounds should contain only the signer without extraneous elements, as background clutter adversely affects recognition accuracy.

Consequently, nearly all publicly available resources have been recorded under laboratory conditions for linguistic research purposes. These resources typically share common vocabulary sizes, type/token ratios, and signer or speaker dependence. Such databases, when used for training, demonstrate poor generalization because signed sentence structures are often pre-designed or offer limited variations, potentially resulting in overfitted language models. Additionally, most self-recorded corpora include only limited numbers of signers.

Hand gesture recognition must accommodate sign variations including unrestricted backgrounds and varying lighting conditions. However, practical implementation of such capabilities proves difficult, particularly for vision-based systems due to associated constraints affecting image processing algorithm performance. These problems remain unsolved. As contemporary research increasingly focuses on real-life applicability, developing suitable datasets becomes more challenging, requiring larger scale and closer approximation to real-world signing scenarios. However, such datasets require extensive processing time and may prove difficult to replicate.

b. Hand Gesture Representations

Dynamic gesture representations can be classified into three types: isolated gestures, continuous gestures, and fingerspelling.

Isolated gestures involve signers performing single sign gestures sequentially. **Continuous gestures** involve signers performing signs in uninterrupted sequences. **Fingerspelling** involves spelling alphabet letters of words using hand movements.

The primary problem in hand gesture recognition involves managing non-gesture movements that frequently intersperse hand gesture sequences. Examples include movement epenthesis (ME) and coarticulation.

Movement epenthesis refers to transitional movements occurring between continuous signs that do not belong to either adjacent sign. Additionally, the precise moment when hands shift toward the next sign's starting position remains unmarked. Movement epenthesis carries no sign information as no signs are associated with these transitional movements.

Coarticulation occurs in sign language when the current sign is influenced by preceding and following signs. Coarticulation effects extend over longer durations while simultaneously affecting various sign aspects including hand shape, position, and movement. Due to this effect, sign endpoints and beginnings can appear substantially different across varying sentence contexts, making sign recognition within sentences difficult.

Gesture spotting represents another critical issue in dynamic hand gesture recognition. Researchers emphasize developing methods to identify finger alphabet words within sign language videos and display them on screens to assist interpreters and audiences in following presentations. Since hand shapes and movements for finger alphabet letters are complex, users may struggle to understand unfamiliar fingerspelled words. Researchers have employed temporal regularized canonical correlation analysis to identify specific fingerspelled gestures in sign language videos.

Due to hand segmentation issues, feature extraction faces restrictions on signer environments to achieve higher accuracy. Hand segmentation partitions images into distinct parts or objects. All subsequent hand gesture recognition system processes depend on segmentation accuracy. If data loss occurs due to inadequate segmentation, system accuracy may decrease.

Consequently, researchers typically impose background color restrictions to avoid hand segmentation issues. Additional environmental restrictions include requiring long-sleeved clothing, specifying camera distance, maintaining uniform lighting, and using only right-hand gestures. Some researchers utilize colored gloves to overcome skin color-related issues, thereby simplifying segmentation processes.

c. Summary of Challenges

Three major challenges exist in gesture recognition system development. First, data acquisition requires appropriate devices for effective gesture input. Second, environmental conditions where hand gestures must function pose challenges. Our review revealed that over 80% of existing research emphasized restricted

laboratory environments bearing little resemblance to real-world conditions. Third, user-specific gesture variations present unique challenges for every individual.

RQ1 focused on challenges limiting vision-based gesture recognition system practicality for real-life applications. Upon deconstructing gesture recognition challenges, we found that most selected papers addressed issues related to data acquisition, environmental conditions, and hand gesture representations.

2) Progress

Progress in gesture recognition is discussed regarding: (1) data acquisition and environmental conditions, and (2) hand gesture representations.

a. Data Acquisition Methods

Four types of vision-based approaches exist for capturing hand gesture images or videos using video cameras:

1. **Single camera** - Using one camera at a time (video camera, digital camera, webcam, or smartphone camera)
2. **Active techniques** - Using light projection for hand location and movement detection (Microsoft Kinect camera, Leap Motion Controller)
3. **Invasive techniques** - Using body markers such as wristbands or colored gloves
4. **Stereo camera** - Using multiple monocular cameras simultaneously for depth information

Our review indicated that approximately 53% of articles employed single cameras for data acquisition. However, recent years have seen vision-based hand gesture recognition research shift toward integrating more detailed depth information. Contemporary studies have adopted active techniques (39%) including Microsoft Kinect and Leap Motion Controller for hand gesture recognition systems. Invasive techniques accounted for 8% of approaches. No selected articles utilized stereo cameras for hand gesture capture.

b. Hand Gesture Representations

Regarding hand gesture representations, most works focused on recognizing isolated gestures (67%) compared to continuous gesture dynamic recognition (21%). Only 12% of works addressed fingerspelling words and alphabets.

i) Isolated Gestures

ii) Continuous Gestures

iii) Fingerspelling

c. Feature Extraction Techniques

Prominent feature extraction techniques include Histogram of Oriented Gradients (HOG), Convolutional Neural Networks (CNN), and Principal Component Analysis (PCA).

B. Research Question 2 (RQ2): Performance of Existing Vision-Based Hand Gesture Recognition Systems and Future Directions

1) Performance of Sign Language Recognition Systems in Various Settings

Regarding signer-dependent approaches, recognition accuracy ranged from 69% to 98%, averaging 88.8% across selected studies. Signer-independent recognition accuracy ranged from 48% to 97%, averaging 78.2%.

Concerning input devices, single cameras achieved average signer-dependent accuracy of 88% and signer-independent accuracy of 79%. Active techniques achieved average signer-dependent accuracy of 89.6% and signer-independent accuracy of 77.2%. For invasive techniques, only one signer-independent recognition result was presented at 88%.

Regarding environmental conditions, restricted environments yielded average signer-dependent accuracy of 88% and signer-independent accuracy of 77%. For uncontrolled environments (with only three articles reporting results), average signer-dependent accuracy reached 98% and signer-independent accuracy reached 90%. Although not providing conclusive evidence, hand gesture research in uncontrolled environments shows promising results.

Concerning hand gesture representation types, isolated gestures achieved average signer-dependent accuracy of 92% and signer-independent accuracy of 77%. Continuous gestures achieved average signer-dependent accuracy of 84% and signer-independent accuracy of 82%. Fingerspelling achieved average signer-dependent accuracy of 81% and signer-independent accuracy of 71%.

2) Future Directions

a. Databases

Numerous articles postulate future directions where gesture databases will expand regarding gesture quantity, number of individuals represented, and language coverage. Developing databases for diverse sign languages—multilingual databases usable across future research—represents an important future direction.

To accelerate database capture and development, devices such as 3D cameras and Kinect sensors will see increased application in future research. Expanding databases containing gestures from varied environments proves vital as future work aims to address environmental challenges. Databases must also cover important future areas including multilingual data, real-world data, and multi-signer databases.

b. Hand Gesture Representations

Various features can serve hand gesture recognition, including segmented signer hand shapes (primary information sources for interpreting specific signs), motion information, location, and gesture orientation. Most past studies utilized handshape for Sign Language Recognition, while hand motion remained least utilized.

c. Other Possible Future Directions

Future gesture recognition directions will likely cover several areas:

First, future research must expand current feature sets to recognize more gestures (including two-handed gestures and facial cues). Future gesture systems must address coarticulation due to extremely rapid hand movements, largely solvable through advanced data acquisition methods.

Second, computational cost will become a consideration during camera device development. Reducing computational cost enables shorter system development time and reduced learning time using advanced machine learning and unsupervised training.

Third, smart and wearable devices will receive increasing attention as data acquisition tools.

3) Summary

Regarding data acquisition, most data were collected using single cameras in restricted environments. Databases for hand gesture recognition system development employ limited numbers of standard signs that may not include sign variations for potential real-life applications. Moreover, many existing works did not address the possibility of large-size sign language databases. Restricted environment nature tends to constrain data collection choices, indirectly hindering existing hand gesture recognition system capabilities, particularly for inputs beyond restricted environments. Nevertheless, restricted environments permit examination of different solution effectiveness.

5. CONCLUSION AND FUTURE WORK

This paper examined challenges, progress, and potential future directions in vision-based hand gesture recognition systems over a seven-year period. Nearly every reviewed article highlighted the importance of data acquisition, features, and training data environments. Notably, the majority of databases employed in hand gesture recognition research originated from restricted environments, signaling the need for sign language databases that are less restrictive and incorporate diverse environmental conditions. This review concludes that preparing vision-based gesture recognition systems for real-life applications requires increased attention to uncontrolled environment settings, as such conditions provide researchers opportunities to improve system capabilities for recognizing hand gestures across any environmental context.

REFERENCES

[1] P. K. Pisharady and M. Saerbeck, ‘‘Recent methods and databases in vision-based hand gesture recognition: A review,’’ Comput. Vis. Image Understand., vol. 141, pp. 152–165, Dec. 2015, doi: 10.1016/j.cviu.2015.08.004.

[2] M. Yasen and S. Jusoh, ‘‘A systematic review on hand gesture recognition techniques, challenges and applications,’’ PeerJ Comput. Sci., vol. 5, p. e218, Sep. 2019.

[3] M. J. Cheok, Z. Omar, and M. H. Jaward, ‘‘A review of hand gesture and sign language recognition techniques,’’ Int. J. Mach. Learn. Cybern., vol. 10, no. 1, pp. 131–153, Jan. 2017, doi: 10.1007/s13042-017-0705-5.

[4] S. Kausar and M. Y. Javed, ‘‘A survey on sign language recognition,’’ in Proc. Frontiers Inf. Technol., 2011, pp. 95–98.

[5] H. Cooper, B. Holt, and R. Bowden, ‘‘Sign language recognition,’’ in Visual Analysis of Humans. London, U.K.: Springer, 2011, pp. 539–562.

[6] G. Fang, W. Gao, and D. Zhao, ‘‘Large vocabulary sign language recognition based on fuzzy decision trees,’’ IEEE Trans. Syst., Man, Cybern. A, Syst. Humans, vol. 34, no. 3, pp. 305–314, May 2004.

[7] M. Mohandes, M. Deriche, U. Johar, and S. Ilyas, ‘‘A signer-independent Arabic sign language recognition system using face detection, geometric features, and a hidden Markov model,’’ Comput. Electr. Eng., vol. 38, no. 2, pp. 422–433, 2012.

[8] S. C. W. Ong, S. Ranganath, and Y. V. Venkatesh, ‘‘Understanding gestures with systematic variations in movement dynamics,’’ Pattern Recognit., vol. 39, no. 9, pp. 1633–1648, Sep. 2006.

[9] B. K. Chakraborty, D. Sarma, M. K. Bhuyan, and K. F. MacDorman, ‘‘Review of constraints on vision-based gesture recognition for human– computer interaction,’’ IET Comput. Vis., vol. 12, no. 1, pp. 3–15, Feb. 2018, doi: 10.1049/iet-cvi.2017.0052.

[10] S. S. Rautaray and A. Agrawal, ‘‘Vision based hand gesture recognition for human computer interaction: A survey,’’ Artif. Intell. Rev., vol. 43, no. 1, pp. 1–54, Jan. 2012, doi: 10.1007/s10462-012-9356-9.

[11] M. A. Moni and A. B. M. S. Ali, ‘‘HMM based hand gesture recognition: A review on techniques and approaches,’’ in Proc. 2nd IEEE Int. Conf. Comput. Sci. Inf. Technol., 2009, pp. 433–437.

[12] A. Wadhawan and P. Kumar, ‘‘Sign language recognition systems: A decade systematic literature review,’’ Arch. Comput. Methods Eng., vol. 28, pp. 785–813, May 2021, doi: 10.1007/s11831-019-09384-2.

[13] N. Aloysius and M. Geetha, ‘‘Understanding vision-based continuous sign language recognition,’’ Multimedia Tools Appl., vol. 79, nos. 31–32, pp. 22177–22209, Aug. 2020, doi: 10.1007/s11042-020-08961-z.

[14] R. Rastgoo, K. Kiani, and S. Escalera, ‘‘Sign language recognition: A deep survey,’’ Expert Syst. Appl., vol. 164, Feb. 2021, Art. no. 113794, doi: 10.1016/j.eswa.2020.113794.

[15] K. M. Lim, A. W. C. Tan, and S. C. Tan, ‘‘A feature covariance matrix with serial particle filter for isolated sign language recognition,’’ Expert Syst. Appl., vol. 54, pp. 208–218, Jul. 2016, doi: 10.1016/j.eswa.2016.01.047.

[16] W. Ahmed, K. Chanda, and S. Mitra, ‘‘Vision based hand gesture recognition using dynamic time warping for Indian sign language,’’ in Proc. Int. Conf. Inf. Sci. (ICIS), Aug. 2016, pp. 120–125.

[17] M. V. D. Prasad, P. V. V. Kishore, D. A. Kumar, and C. R. Prasad, ‘‘Fuzzy classifier for continuous sign language recognition from tracking and shape features,’’ Indian J. Sci. Technol., vol. 9, no. 30, pp. 1–9, Aug. 2016, doi: 10.17485/ijst/2016/v9i30/98726.

[18] O. Koller, J. Forster, and H. Ney, ‘‘Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers,’’ Comput. Vis. Image Understand., vol. 141, pp. 108–125, Dec. 2015, doi: 10.1016/j.cviu.2015.09.013.

[19] W. Yang, J. Tao, and Z. Ye, ‘‘Continuous sign language recognition using level building based on fast hidden Markov model,’’ Pattern Recognit. Lett., vol. 78, pp. 28–35, Jul. 2016, doi: 10.1016/j.patrec.2016.03.030.

[20] K. Tripathi and N. B. G. C. Nandi, ‘‘Continuous Indian sign language gesture recognition and sentence formation,’’ Proc. Comput. Sci., vol. 54, pp. 523–531, Jan. 2015.

[21] T. Kim, J. Keane, W. Wang, H. Tang, and J. Riggle, ‘‘Lexicon-free fingerspelling recognition from video: Data, models, and signer adaptation,’’ Comput. Speech Lang., vol. 46, pp. 209–232, Nov. 2017, doi: 10.1016/j.csl.2017.05.009.

[22] 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, Oct. 2018, doi: 10.3390/s18103554.