IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 12, 2021, accepted November 15, 2021, date of publication November 19, 2021,
date of current version December 3, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3129650

A Review of the Hand Gesture Recognition System: Current Progress and Future Directions

NORAINI MOHAMED, (Graduate Student Member, IEEE), MUMTAZ BEGUM MUSTAFA, (Member, IEEE), AND NAZEAN JOMHARI

Department of Software Engineering, Faculty of Computer Science, and Information Technology, Universiti Malaya, 50603 Kuala Lumpur, Malaysia

Corresponding author: Noraini Mohamed (noraini.binti.mohamed@gmail.com)

This research was supported by University Malaya Research Grant (UMRG), Grant No.: RG284-14AFR and Fundamental Research Grant Scheme (FRGS), Grant No.: FP062-2020.

ABSTRACT This paper reviewed the sign language research in the vision-based hand gesture recognition system from 2014 to 2020. Its objective is to identify the progress and what needs more attention. We have extracted a total of 98 articles from well-known online databases using selected keywords. The review shows that the vision-based hand gesture recognition research is an active field of research, with many studies conducted, resulting in dozens of articles published annually in journals and conference proceedings. Most of the articles focus on three critical aspects of the vision-based hand gesture recognition system, namely: data acquisition, data environment, and hand gesture representation. We have also reviewed the performance of the vision-based hand gesture recognition system in terms of recognition accuracy. For the signer dependent, the recognition accuracy ranges from 69% to 98%, with an average of 88.8% among the selected studies. On the other hand, the signer independent's recognition accuracy reported in the selected studies ranges from 48% to 97%, with an average recognition accuracy of 78.2%. The lack in the progress of continuous gesture recognition could indicate that more work is needed towards a practical vision-based gesture recognition system.

INDEX TERMS Classification, feature extraction, dynamic hand gesture recognition, sign language recognition, vision-based hand gesture, recognition accuracy.

I. INTRODUCTION

Nonverbal communication is important in our life as it conveys about 65% of messages in comparison to verbal communication that contributes no more than 35% of our interactions [1]. Gestures can be categorized into hand and arm gestures (recognition of hand poses, sign languages, and entertainment applications), head and face gestures (such as nodding or shaking of the head, the direction of eye gaze, opening the mouth to speak, winking, and so on), and body gestures (involvement of full-body motion).

Effective human-computer interaction (HCI) requires gesture recognition methods that are robust and accurate. Such recognition systems are used to serve as an alternative for the commonly used HCI devices such as mouse, keyboard etc. [2]. Automatic recognition systems, such as hand gesture recognition, are among the most active research areas as well as one of the most significant for HCI [3]. Hand gesture recognition is particularly useful for applications that require natural human–machine interaction.

Developing hand gesture recognition systems such as sign language applications is extremely important to overcome the communication barrier with people that are unfamiliar with sign language. Technology that automatically translates hand motions into text or audible speech for a non-signing person to interpret can help to reduce this barrier.

The vision-based hand gesture recognition system is that can be applied in various applications such as communication, education, and rehabilitative tool. The system also can be used to assist in situations where a human interpreter may not be available for interpreting sign language.

The task of hand gesture recognition is very challenging for the following reasons. First, the ability of the system to handle inputs that vary considerably from the input used during the development stage. For hand gesture recognition systems, input that may not be considered during the development

The associate editor coordinating the review of this manuscript and approving it for publication was Zhenbao Liu.

stage includes environmental noise, signers' variability, language variability, and so on. This is because we will usually apply restrictions on the environment of the signers reduce the problems in the segmentation and tracking process [4].

Another challenge in hand gesture recognition development is handling the transition movements between two signs as it is difficult to distinguish the beginning and end of the hand gestures for each sign. The system's inability to identify the boundary between two signs may result in wrong or poor recognition. Due to this complexity, it appears that many researchers have placed less emphasis on continuous sign language for the vision-based hand gesture recognition, which has limited practicality for real-world applications [5].

The other challenge in this field is the development of robust signer-independent hand gesture recognition systems, e.g., a system that can be used by signers who are not represented during the training stage. Such systems are desirable for real-life applications as they can be used by a wide range of users without the need for every new user to be trained on the system [6].

Although the existing research on hand gesture recognition system had been summarized by several review papers, these papers had only examined the research progress in entirety. This includes the device-based and vision-based hand gesture recognition systems used for detecting sign language recognition. Since the vision-based hand gesture recognition system is practical for real-life applications, it must apply to any user in any environment. However, there had been no review which examined the extent of research made towards the development of the vision-based hand gesture recognition system and the possible future directions. As such, the current paper addresses that gap by reviewing current and past literature to examine the progress of vision-based hand gesture recognition system made thus far.

The remainder of this paper is organized as follows. Section II discusses the research background of this review. Section III discusses the research aim and approach that include the review aim, and the research questions applied. Section IV discusses the major findings of the review which addresses the research questions formulated. Section V concludes this paper.

II. BACKGROUND

Hand gesture recognition is an application that converts sign language hand gestures into outputs such as text or voice. It can be categorized by the way sign language hand gestures are captured by the system; the vision-based system (where the gestures are captured using one or more cameras), and device-based system (where a direct-measure device such as designated electronic gloves equipped with sensors are generally employed to connect the user with the system).

While the device-based systems are characterized by their efficiency, their real-life usage is limited due to the need of wearing the cumbersome device when interacting with the system. This issue, however, does not arise for the vision-based systems, allowing users to interact more naturally with the system [7]. In terms of applicability, it has a wider application in outdoor scenarios.

This easiness in use of the vision-based system was challenged by how it handles datasets made up of dynamic hand gestures in sign language, such isolated and continuous signs. According to [5], while most of the existing works focus on recognizing isolated gestures, their use in real-world applications is limited. Moreover, the development of hand gesture recognition using the vision-based system requires the use of more powerful feature extraction and discrimination methods [8].

The interest in gesture recognition has led to a large body of research, as has been noted in several review papers [3], [9]–[14]. Cheok *et al.* [3] reviewed the state-of-the-art technique used in recent hand gesture and sign language recognition research in areas such as data acquisition, pre-processing, segmentation, feature extraction, and classification. Wadhawan *et al.* [12] focused on academic literature published from 2007–2017. These papers were reviewed in six dimensions (data acquisition techniques, static/dynamic signs, signing mode, single/double handed signs, classification techniques, and recognition rates). More recently, Aloysius and Geetha [13] reviewed the vision-based continuous sign language recognition (CSLR) system, and Ratsgoo *et al.* [14] focused on the vision-based proposed models of sign language recognition.

It appears that the past works reviewed by researchers had left a gap; they had not examined the challenges and future direction of the vision-based hand gesture recognition system. Based on this, the current paper will address this gap by reviewing existing literature to identify the progress of research in vision-based hand gesture recognition systems for the present and for future directions.

III. RESEARCH AIMS AND APPROACH

This paper aims to review the current issues, progress, and potential future direction of the vision-based hand gesture recognition research. We have formulated two research questions for this purpose.

Research Question 1 (RQ1): What were the current issues and progress of the vision-based hand gesture recognition system in terms of data acquisition, data environment, and hand gesture representations?

To address this research question, we extracted articles related to the vision-based hand gesture recognition system from 2014 to 2020 in totality to identify the issues and solutions.

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

To answer this research question, we extract the performance of the vision-based hand gesture recognition system in term of recognition accuracy to identify the possible future directions in gesture recognition.

A. SEARCH METHODOLOGY

To carry out the search, we applied specific keywords. The search criteria included in the process of identifying articles that focused on the following keywords.

1. Sign language recognition,
2. Dynamic hand gestures recognition.

Our search covered popular databases such as:

1. Science Direct,
2. IEEE Explore Digital Library,
3. Springer Link,
4. Google Scholar.

To screen our initial search, we applied the following inclusion criteria:

- Publication date: between 2014 and 2020 inclusive.
- Search domain: science, technology, or computer science.
- Publication types: journals, proceedings, and transactions.
- Article type: full text and reviews.
- Subject: within the scope of hand gestures in sign language including isolated words, continuous sentences, and dynamic fingerspelling in the domain of the vision-based hand gesture recognition system.
- Language: English.

Additionally, we applied the following exclusion criteria:

- Studies that do not focus explicitly on the vision-based hand gesture recognition system in sign language.
- Studies that do not cover other forms of gesture recognition in sign language.
- Studies that discuss recognition of hand gestures in sign language as a side topic.
- Studies that reviewed the works of others.
- Studies that do not provide details of their experiments or experimental design.
- Full text of the paper is not available (physical and electronic forms).
- Opinions, viewpoints, keynotes, discussions, editorials, tutorials, comments, prefaces, anecdotal papers, and presentations in slide format, without any associated papers.

IV. FINDINGS OF THE REVIEW

Based on the search keywords we applied, 98 articles that satisfied the inclusion and exclusion criteria were selected. These articles were then closely examined by studying each article's abstract, methodology, discussion, and results. Table 1 shows the distribution of articles based on publication types, and the number of papers retrieved. Most of the papers were obtained from the IEEE Explore Digital Library.

A. RESEARCH QUESTION 1 (RQ1): WHAT WERE THE CURRENT ISSUES AND PROGRESS OF THE VISION- BASED HAND GESTURE RECOGNITION SYSTEM IN TERM OF - DATA ACQUISITION & DATA ENVIRONMENT, AND HAND GESTURE REPRESENTATIONS?

From the articles reviewed, most of the 98 articles reviewed had highlighted the issues and progress of data acquisition and data environment (n = 47), and hand gesture representation (n = 44).

TABLE 1. The distributions according to the publication type and the number of papers.

Digital database libraries	Keyword and hits		Total
	Sign language recognition	Dynamic hand gesture recognition	
Science Direct	10	3	13
IEEE Explorer Digital Library	52	13	65
Springer Link	4	-	4
Google Scholars	14	2	16
Total	80	18	98

1) ISSUES

a: DATA ACQUISITION AND DATA ENVIRONMENT

Table 2 depicts the issues that the existing works aim to solve. There are 47 articles that discuss the issues and progress on data acquisition and data environment. More than 80% of the 47 articles (39 articles) had been conducted in a restricted laboratory environment. Lim *et al.* [15], explain this by stating that the ideal background for gesture recognition should include only the signer with no background as the background clutter can affect the gesture recognition accuracy.

As such, almost all publicly available resources have been recorded under lab conditions for linguistic research purposes. Most of them share a common vocabulary size, the types/token ratio (TTR), and are signer/speaker dependent. This type of database, when trained, does not generalize very well because the structure of the signed sentences is often designed in advance, or it can only offer small variations. These can result in an over-fitted language model. Additionally, most self-recorded corpora consist of only a limited number of signers.

One of the key issues in hand gesture recognition is the environment the system needs to work. Uncontrolled environment, which refers to unanticipated conditions, such as the background colour where the system operates, is a challenge not resolved in current research. In developing a hand gesture recognition system within an uncontrolled environment, researchers are facing with the difficulty of separating objects in the background that are similar with the skin colour.

Hand gesture recognition needs to work with sign variations, such as non-restricted backgrounds, and different lighting conditions. Nonetheless, this is difficult to be addressed in practice, especially for vision-based systems because of the related constraints. Such constrain affects the performance of the image processing algorithms and these problems are yet to be solved [16], [17]. With recent research focusing on real-life applicability, the development of suitable datasets becoming more challenging as it should be larger, and closer to real-life signing. However, these take time to process and can be difficult to be replicated [18].

b: HAND GESTURE REPRESENTATIONS

As stated in [9], dynamic gesture representations can be classified into two types: isolated gestures or and continuous

gestures. As the focus of gesture recognition is toward the dynamic hand for sign language, we also include the works on fingerspelling in this review.

1. Isolated gestures – the signers perform one sign gesture at a time. For example, the isolated gesture of Chinese Sign Language for the word "WELCOME" in [19].
2. Continuous gestures – the signer perform continuously the signs. For example, continuous gestures of Indian Sign Language for the sentence "IT IS CLOSED TODAY" [20].
3. Fingerspelling – the act of spelling out the letters of the alphabet in a word using hand. For example, fingerspelling of American Sign Language for word "TULIP" in [21] and alphabet "Z" in [22].

The main problem that lies in hand gesture recognition is the issue of handling non-gesture movements, which often intersperse the sequence of hand gestures [9]. Some examples of non-gesture movements are movement epenthesis (ME), and coarticulation.

ME is the movement that occurs in continuous signs; it is not part of either of the signs. Moreover, it is not marked as to when the hands shifted to the starting position of the next sign. Aloysius and Geetha [13] state that ME does not have any information on the signs because there are no signs associated with them.

Coarticulation happens in sign language, where the current sign is affected by the preceding and succeeding signs. Yang and Sarkar [23] explain, the coarticulation effects occur over longer durations and at the same time affect different aspects of the sign, such as the hand shape, position, and movement. Because of this effect, the appearance of the end of the sign and the beginning of the next sign can be significantly different under different sentence contexts, making it difficult to recognize signs in sentences.

Gesture spotting is another critical issue in dynamic hand gesture recognition. Tanaka *et al.* [24] emphasize the significance of developing a method for spotting finger alphabets of words in sign language videos and displaying them on a screen to assist interpreters and the audience follow a presentation. As the hand shapes and movements used to create the letters in a finger alphabet are complicated, users may struggle to understand unfamiliar words spelled out by fingerspelling. They spot specific gestures expressed by fingerspelling in sign language video using temporal regularized canonical correlation analysis (TRCCA).

Due to hand segmentation issues, feature extraction faces restrictions on the signers' environment to achieve higher accuracy. Hand segmentation is a process in which images are partitioned into multiple distinct parts or objects. All subsequent processes in the hand gesture recognition system depended on the accuracy of the segmentation. If the data were missing due to inadequate segmentation, the accuracy of the system may reduce.

As a result, researchers usually place restrictions on background colour to avoid hand segmentation issues in the background [25]–[27]. Researchers also place several restrictions on the environment of the signers, such as wearing long sleeve clothing [28], [29] distance from the camera [30], uniform lighting [31], and using right hand gestures [32], [33]. Other researchers made use of coloured gloves to overcome the issue associated with skin colour, thereby making the process of segmentation easier [34]–[36].

c: SUMMARY

Overall, there are three major issues with the development of the gesture recognition system. The first is the data acquisition process that requires suitable devices for effective input of the gestures. The second challenge is the data environment in which the hand gesture needs to work in. From the review, we noted that more than 80% of the existing works had emphasized the restricted laboratory environment, which may have little similarity with the real world. Finally, the third challenge lies in the gestures posed by users, which are unique for everyone.

RQ1 focuses on the issues that limit the vision-based gesture recognition system to be more practical for real-life applications. When we deconstructed the issues of gesture recognition, we found that majority of the selected papers discussed issues related to data acquisition, data environment, and hand gesture representations.

2) PROGRESS

The progress in gesture recognition is discussed in terms of 1) data acquisition and data environment and 2) hand gesture representations which are depicted in Table 3.

a: DATA ACQUISITION METHOD

There are four types of vision-based approaches for capturing images or video of hand gestures with a video camera [3].

1. Single camera – use of one camera at a time, such as video camera, digital camera, Webcam, or smartphone camera.
2. Active techniques – use light projection to locate the hand and detect hand movement such as Microsoft Kinect camera and Leap Motion Controller
3. Invasive techniques – use body markers like wrist bands or color gloves.
4. Stereo camera – use multiple monocular cameras to capture images at the same time to provide depth information.

Our review indicated that about half (53%) of the articles use single cameras for the data acquisition process, as shown in FIGURE 1. However, in recent years, the focus of the vision-based hand gesture recognition research has moved to the integration of more in-depth information. Recent studies have found active techniques (39%) such as Microsoft Kinect and Leap Motion Controller considered for hand gesture recognition systems. The other approach is invasive techniques (8%). From the selected articles, not a single article has used the stereo camera for capturing hand gestures.

TABLE 2. Categorization of issues in selected articles based on RQ1.

Issues	Articles
Data acquisition and Data Environment	[16], [17], [18], [22], [25], [26], [27], [33], [34], [35], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58], [59], [60], [61], [62], [63], [64], [65], [66], [67], [68], [69], [70], [71], [72], [73].
Hand gesture representations	[19], [20], [21], [22], [24], [26], [32], [35], [41], [43], [61], [74], [75], [76], [77], [78], [79], [80], [81], [82], [83], [84], [85], [86], [87], [88], [89], [90], [91], [92], [93], [94], [95], [96], [97], [98], [99], [100], [101], [102], [103], [104], [105], [106].

i) SINGLE CAMERA

Single cameras appear to be the more common approach used to acquire hand gestures, particularly dynamic hand gestures [27], [37]–[39]. Digital or video cameras without additional sensor devices are favored by researchers for detecting hand regions as they are cost-effective and realistic [40], [41]. With the advancements in the technology of capturing devices and processing of high-quality images [25], as well as having higher mobility than other types of SLR systems [42], cameras are now capable of capturing gestures. According to Shohieb *et al.* [26], one of the benefits of using digital cameras is that they can create a database that considers different lighting and background conditions that can provide the flexibility for various research purposes.

Some researchers preferred webcams to collect data. Islam *et al.* [43] employ a low-cost color video using a webcam to address issues arising from the extraction of features and the detection of hand gestures captured using color videos. Mahmood *et al*. [44] used a Webcam (HP Pavilion dv6) to capture images for a real-time hand gesture recognition system.

Mobile video cameras may also be used in real-time to collect data. Kumar *et al.* [46] and Rao *et al*. [47] captured a video in selfie mode using a smartphone with a 5MP front camera attached to the end of a selfie stick. Takayama *et al* [48] used a smart-phones camera to record the videos of the signers. Athira *et al.* [41] developed a real-time signer independent system for recognizing gestures captured with mobile camera videos. In Uchil *et al*. [49] a real-time recognition system used a standard smartphone camera as an input to recognize medical terms.

Thongtawee *et al.* [45] stated that it is difficult to extract features from the boundary of binary images in alphabets gestures, thus proposing a 2D camera as it is fast enough to be used in real-time processing.

ii) ACTIVE TECHNIQUES

Microsoft Kinect (MK) has recently become a trend in hand gesture recognition. MK is a motion sensor that has been designed to track body movements. The use of cameras with MK for recognizing hand gestures is to make the system more reliable. MK can capture depth information [50]–[53]. Features extracted from MK are resilient to changes in illumination, size, and rotation [54]–[56].

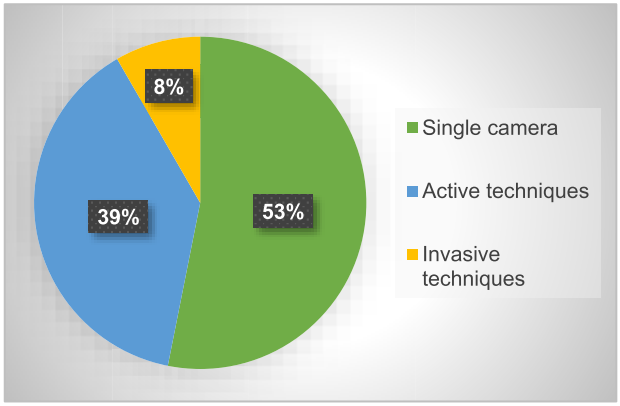

FIGURE 1. Breakdown of studies from 2014 to 2020 based on data environment types in hand gesture recognition systems.

The Leap Motion Controller (LMC), on the other hand, can explicitly target the tracking of hand and finger movements in 3D digital format. It can also identify the joints of the fingers as well as trace their movements while gestures are being performed.

From the review, most of the studies had used a single depth sensor, either the MK or the LMC. Kumar *et al.* [58], [59] proposed a new multi-sensor fusion which is a combination of the MK and LMC. The overall recognition performance increases when MK and LMC are combined.

iii) INVASIVE TECHNIQUES

Several articles use invasive techniques like color gloves (which are not data gloves as they are not connected to any devices) for extracting hand gesture images. Single-colored gloves on each hand are used in [34], [60] and single-colored gloves with separated colors on each hand are used in [35], [61]. While [62] uses an optical camera and gloves with different color regions to identify hand and finger motions. Although the use of color gloves aided the image capturing process to some extent, [63] combined MK sensors with a pair of color gloves to extract three features-depth, motion, and color.

The environment in which the gesture recognition operates influences the solutions provided by the researchers. There are two types of environments in this field, which are controlled and uncontrolled environments. Less than 10% of the articles reviewed had looked at uncontrolled environments when focusing on hand gesture recognition.

TABLE 3. Categorization of progress in selected articles based on RQ1.

No.	References	Progress																					
		Hand gesture representation																		Data acquisition and data environment			
		Parameters					Feature extraction techniques						Classifiers										
		Hand shape	Hand orientation	Hand location	Hand motion	Others	Histogram of oriented gradient (HOG)	Principal component analysis (PCA)	Local binary pattern (LBP)	Zernike moments	Convolutional neural network (CNN)	Others	Support vector machine (SVM)	Hidden markov model (HMM)	K-Nearest neighbours (KNN)	Random forest (RF)	Distance metrics	Neural network	Deep neural network/ deep learning	Others	Single camera	Active technique	Invasive technique
1	[15]					✓		✓				✓					✓						
2	[16]					✓														✓	✓		
3	[17]	✓				✓						✓								✓	✓		
4	[18]		✓				✓													✓	✓		
5	[19]					✓						✓		✓									
6	[20]	✓				✓		✓				✓					✓						
7	[21]	✓	✓				✓	✓											✓				
8	[22]					✓						✓							✓			✓	
9	[24]		✓			✓						✓								✓			
10	[25]					✓						✓	✓		✓						✓		
11	[26]	✓			✓	✓														✓	✓		
12	[27]					✓						✓							✓		✓		
13	[28]		✓			✓						✓	✓	✓						✓			
14	[29]	✓	✓			✓						✓											
15	[30]					✓						✓								✓			
16	[31]	✓				✓						✓								✓			
17	[32]					✓						✓								✓			
18	[33]					✓						✓	✓									✓	
19	[34]					✓						✓						✓					✓
20	[35]											✓								✓			✓
21	[37]	✓				✓						✓				✓					✓		
22	[38]					✓						✓		✓							✓		
23	[39]	✓			✓	✓		✓			✓									✓	✓		
24	[40]					✓					✓							✓			✓		
25	[41]	✓	✓	✓	✓					✓		✓	✓								✓		
26	[42]		✓		✓	✓						✓					✓				✓		
27	[43]				✓						✓		✓								✓		
28	[44]					✓						✓						✓			✓		
29	[45]				✓	✓						✓						✓			✓		
30	[46]	✓				✓		✓										✓			✓		
31	[47]	✓				✓		✓				✓						✓			✓		
32	[48]	✓						✓				✓		✓							✓		
33	[49]				✓						✓							✓			✓		
34	[50]	✓				✓	✓					✓	✓									✓	
35	[51]	✓		✓	✓				✓			✓		✓								✓	
36	[52]				✓						✓					✓						✓	
37	[53]	✓					✓					✓								✓		✓	
38	[54]					✓						✓								✓		✓	
39	[55]	✓	✓					✓				✓	✓									✓	
40	[56]					✓						✓								✓		✓	
41	[57]					✓						✓						✓				✓	
42	[58]					✓						✓								✓		✓	
43	[59]					✓						✓		✓								✓	
44	[60]					✓						✓					✓						✓
45	[61]					✓					✓								✓				✓
46	[62]											✓		✓									✓
47	[63]				✓	✓						✓	✓										✓
48	[64]	✓	✓			✓	✓	✓												✓			✓
49	[65]	✓	✓				✓	✓														✓	
50	[67]	✓				✓						✓						✓			✓		
51	[66]		✓			✓						✓	✓								✓		
52	[68]	✓				✓	✓					✓						✓				✓	
53	[69]					✓						✓								✓		✓	
54	[70]					✓						✓								✓		✓	
55	[71]					✓						✓					✓					✓	
56	[72]					✓						✓							✓			✓	
57	[74]											✓								✓			
58	[75]	✓			✓						✓							✓					
59	[76]	✓			✓						✓							✓					
60	[77]				✓						✓								✓				
61	[78]				✓						✓								✓				
62	[79]	✓	✓	✓								✓	✓										
63	[80]					✓						✓	✓										
64	[81]		✓			✓	✓					✓	✓										
65	[82]	✓				✓		✓	✓					✓									
66	[83]	✓	✓									✓								✓			
67	[84]	✓				✓						✓								✓			
68	[85]					✓														✓			
69	[86]		✓				✓	✓											✓				
70	[87]			✓	✓	✓						✓								✓			
71	[88]				✓						✓							✓					
72	[89]			✓					✓									✓					
73	[90]	✓			✓						✓									✓			
74	[91]	✓	✓	✓						✓		✓								✓			

TABLE 3. ***(Continued.)*** **Categorization of progress in selected articles based on RQ1.**

75	[92]	✓				✓						✓						✓					
76	[93]				✓						✓								✓				
77	[94]	✓			✓	✓						✓								✓			
78	[95]	✓				✓	✓	✓				✓		✓									
79	[96]					✓						✓		✓									
80	[97]					✓						✓				✓							
81	[98]					✓						✓								✓			
82	[99]	✓	✓			✓	✓	✓				✓								✓			
83	[100]					✓						✓							✓				
84	[101]																			✓			
85	[102]					✓						✓			✓								
86	[103]				✓						✓							✓					
87	[104]					✓						✓						✓					
88	[105]					✓						✓								✓			
89	[107]					✓						✓						✓					
90	[108]				✓	✓						✓								✓			
91	[109]				✓						✓							✓					
92	[110]					✓						✓	✓		✓					✓			
93	[111]		✓		✓	✓						✓					✓						
94	[112]				✓	✓					✓							✓					
95	[113]					✓						✓	✓										
96	[114]				✓	✓						✓						✓					
97	[106]					✓						✓											
98	[73]											✓								✓			

b: DATA ENVIRONMENT

i) DATABASES

Several studies have revealed that the existing hand gesture recognition systems used private datasets [21], [75], [93], [95], [98]. These datasets were recorded for specific research purposes, and they differed significantly from the actual language used outside the research laboratory. Most publicly available datasets were limited in both quantity and quality, resulting in higher recognition error by the existing hand gesture recognition systems. This issue is further compounded by the lack of adequate training. In addition, the existing hand gesture recognition had only been recorded in a few languages as shown in Table 4. Consequently, researchers had to resort to using new datasets for developing the hand gesture recognition system.

While most of the selected articles focused on controlled environments, several works used datasets that ensure variability for addressing the issues associated with uncontrolled environments. In Kishore *et al.* [107], the dataset includes the occlusion of the signer's hands and head while performing the sign. Rokade and Doye [74] recorded videos of hand gestures with complex backgrounds. ElBadawy *et al.* [109] recorded gestures of two signers from different backgrounds and wearing different clothing. Teodoro *et al.* [111] created the dataset with uncontrolled lighting and a variety of backgrounds and clothing. The dataset in Sidig *et al*. [110] contains some variation in clothing, the use of both hands, and the distance between the camera and the signers, which adds to the variability in hand size. Kumar *et al.* [114] proposed a rotation and position invariant framework where all sign gestures were performed at three different rotational angles. Selfie sign language gestures were captured by Rao *et al.* [112] from five signers in different viewing angles and background settings.

Yang and Zhu [75], tackled the issues of the uncontrolled environment using publicly available video. Video images of the upper body were extracted from the Chinese sign language instructional video. On the other hand, works in [18], [39], [76]–[78], used publicly available databases for real-life continuous sign language recorded from the broadcasting news and weather forecast which are RWTH-PHOENIX-Weather, RWTH-PHOENIX-Weather-2012, and RWTH-PHOENIX-Weather-2014.

ii) FEATURE EXTRACTION

The effect of illumination under different lighting conditions can affect the skin and the image color. Kishore *et al.* [107] used an active contour model with intensity, color boundary, and shape information to extract signers from a clustered video background. Rokade and Doye [74] employ a hand segmentation algorithm that includes erosion operation, dilation operation, and conversion to YCbCr to distinguish regions of interest (the hands) from both uniform and non-uniform backgrounds as well as the rotation-invariant algorithm for feature extraction.

ElBadawy *et al.* [109] used a scoring algorithm based on canny edge detection to select several frames as input to the developed system, which then uses 3D CNN to extract spatial temporal features. Teodoro *et al.* [111] compared five implemented skin segmentation methods using the WEKA implementations of Machine Learning algorithms with Rotation Forest algorithm that performs skin pixel classification. Sidig *et al.* [110] applied Fourier, Hartley, and Log-Gabor transforms to the accumulated image—after processing skeleton data from Kinect with affine transformation. Kumar *et al.* [114], extracted three distinct features from 3D segmented data: angular features, velocity, and curvature features. Yang and Zhu [75] detected the center of hands and captured the upper body images around the hand using Haar feature, skin color detection, and Hue-Saturation Value (HSV) color space.

iii) CLASSIFIERS

In general, researchers in gesture recognition focused on identifying suitable classifier(s) [28], [50], [52], [66], [79]–[81], [110], [113]. The classifiers' ability to discriminate a particular sign in any environment is crucial. Many of the existing works for controlled environment tend to

TABLE 4. Summary of the existing datasets use in the selected articles.

	Public dataset			Private dataset		
	IS	CO	FS	IS	CO	FS
American SL	✓		✓	✓	✓	✓
Arabic SL	✓	✓		✓		
Indian SL				✓	✓	✓
Chinese SL				✓	✓	
Japanese				✓		✓
Italian SL	✓					
German SL	✓	✓				
Turkish SL	✓	✓				
Argentine SL	✓					
Danish SL	✓					
New Zealand SL	✓					
Dutch SL		✓				
Brazilian SL	✓					
Others				✓	✓	

IS=Isolated gesture, CO=Continuous gesture, FS=Fingerspelling

apply the Support Vector Machine (SVM) as a classifier [14], [15], [18], [28], [30], [33], [63], [70], [71], [76], [81], [88], [90], [99].

The most commonly used classifier in uncontrolled environments is the Convolutional Neural Network (CNN) [75], [109], [112]. ElBadawy *et al.* [109] developed the system using 3D CNN. The system used depth map data and achieved more than 90% accuracy. Yang and Zhu [75] had used the CNN for Chinese Sign Language and achieve an accuracy of 99%. The three convolutional layers CNN model extracts the upper body images from videos directly, and it is also able to recognize the gestures in the images. On the other hand, Rao *et al.* [112] reported that the average recognition rate of the proposed four convolution layers CNN model is 92.88 %, a rate that is higher when compared with the other state-of-the-art classifiers.

While CNN is frequently used in uncontrolled environments, other classifiers had also been used in existing works. Kishore *et al.* [107] used the Artificial Neural Networks (ANN) to recognize gestures from video frames of signers in complex and variable backgrounds. The neural network can achieve a recognition rate of 93.63%. Sidig *et al.* [110] utilized three classifiers in uncontrolled environments; namely K-Nearest Neighbours (KNN), SVM, and Multi-Layer Perceptron (MLP). Among the three, SVM has the highest recognition accuracy at 98.8%. Other classifiers applied to uncontrolled environments include Random Forest (RF) [111], and the Hidden Markov Model (HMM) [114].

c: HAND GESTURE REPRESENTATIONS

From Figure 2, most works focused on recognizing isolated gestures at 67% compared to dynamic recognition for continuous gestures at only 21%. Only 12% of the works used fingerspelling words and alphabets.

The research in [19], [26], [89] focus on isolated gestures and continuous gestures. While [32], [41], [83] focused on developing hand gesture recognition systems for isolated gestures and fingerspelling. Others, as will be discussed further below, focus on a single type of hand gesture representation.

i) ISOLATED GESTURES

The review shows that most of the sign language datasets are for isolated gestures. There are several public datasets available for use including RWTH-BOSTON-50 database [15], The ASL Lexicon Video Dataset [91], and MSR Gesture3D dataset [55] for American Sign Language, SignsWorld Atlas database [26] for Arabic Sign Language, A3LIS-147 Database [28] and ChaLearn 2013 dataset [51], [80], [99] Italian Sign Language, Danish Sign Language database [76], [78], New Zealand Sign Language database [76], [78], LSA64 for Argentine Sign Language [35], Bosphorus Sign Dataset for Turkish Sign Language [104], and RPPDI dynamic gesture dataset for Brazilian Sign Language [73], [106] as shown in Table 4.

For private isolated gestures datasets, most researchers used self-generating datasets for various languages such as the Indian Sign Language, Arabic Sign Language, and Chinese Sign Language. With the growing interest in hand gesture recognition systems, many new isolated gesture datasets have been created, including Brazilian Sign Language [80], Persian Sign Language [38], Kurdish Sign Language [44], Korean Sign Language [93], Malaysia Sign Language [56], Mexican Sign Language [54], Indonesian Sign Language [60], [102], Filipino Sign Language [34], and Turkish Sign Language [104].

ii) CONTINUOUS GESTURES

For continuous gestures datasets, the largest publicly available datasets in the German Sign Language such as the RWTH-PHOENIX-Weather and SIGNUM used in [18], RWTH- PHOENIX-Weather-2012 [39], RWTH- PHOENIX-Weather-2014 [76]–[78], [90] and RWTH- PHOENIX-Weather-2014-multi-signer [39], [88]. Others datasets include the SignsWorld Atlas database [26] for Arabic Sign Language and HospiSign database for Turkish Sign Language [105].

There are many private continuous gestures datasets for Indian Sign Language and American Sign Language. Work by [20] and [29] used 10 Indian sign language sentences, which consist of two, three, and four types of gestures per sentence. On the other hand, [17], [31] used 50 signs and [67] 58 signs in one sentence to recognize the continuous gestures. For the American Sign Language, [85] uses a dataset that contain 25 signs. Work in [108] uses a small database with 5 sentences from two signers. Zafrulla *et al.* [97] used 92 different sentences that consist of a vocabulary of 10 noun signs and 5 classifier predicates (actions that describe how a pair of nouns interact). Other continuous gestures include Chinese Sign Language [19], Bangladesh Sign Language [89], and Lao Sign Language [30].

Elakkiya and Selvamani [85] deal with ME issues in continuous sign sentences for real-time hand detection in

uncontrolled environments. Choudhury *et al.* [115] use the height of the hand trajectory as a feature to detect ME. For efficient recognition, the signs obtained after eliminating the ME frames from the input sign sequence are recognized using a combination of spatial and temporal features.

iii) FINGERSPELLING

Some research has applied fingerspelling datasets, as most of the selected works emphasize static images. We discovered three works that used fingerspelling in sequence for American Sign language: [74] created a dataset of 22 fingerspelling words and [21], [86] created a dataset of 300 fingerspelling words. Work by [84] created continuous digits for Pakistan Sign Language. While [24] dataset comprises 8 fingerspelling words from different viewpoints, developed for Japanese Sign Language. In [32], a database for Latin with 28 fingerspelling words was developed.

Rokade and Doye [74] attempted keyframe detection algorithms on fingerspelling in sequence to distinguish the gesture and non-gesture frames in near real-time. Athira *et al.* [41] presented a method for removing coarticulation in dynamic fingerspelling alphabets. In this step, gesture spotting is used to determine the start and endpoints of a gesture pattern, and coarticulation is detected using the gradient of acceleration approach.

iv) FEATURE EXTRACTION TECHNIQUES

One of the solutions in hand gesture recognition is the features and the feature extraction techniques. Some of the more prominent techniques are Histogram of Oriented Gradient (HOG), Convolutional Neural Network (CNN), and Principal Component Analysis (PCA).

HOG is a feature descriptor that is often used to extract the appearance or the shape of an object. The HOG feature descriptor counts the occurrences of gradient orientation in localized portions of an image. There are several articles from the selected papers that extract features for handshape using HOG. Jiang *et al*. [53] extract a representative sign feature vector composed of normalized hand trajectories and HOG feature, which represents hand shapes as HOG can well describe the shape and appearance of an object and adapt to illumination variation or complex background. Huang *et al.* [68] found that both the change of hand shape and the trajectory of body movement are two of the most important features to describe a sign motion. They extract and combine two kinds of features: HOG from hand-shape image and coordinate locations of joints (trajectory) to train the GMM-HMM model. Wang *et al.* [95] adopted the HOG feature extracted from the hand region, which is segmented using a self-adaptive skin model and depth constraint. The self-adaptive skin model is initialized by the skin of the human face and updated by the skin of the detected human hands in previous frames. Since the dimension of the original HOG is too high, PCA is applied for dimensionality reduction and to retain only the most salient dimensions.

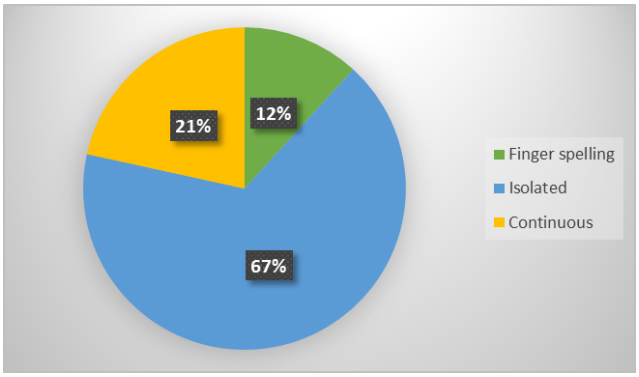

FIGURE 2. Breakdown of studies from 2014 to 2020 focusing on hand gesture representations.

He *et al.* [99] first adopt Histogram of Oriented Displacement (HOD) and Relative Distance Features (RDF) by using validation hidden markov model (VHMM) to describe the sign trajectory. As for hand shape feature, they consider HOG in local hand regions.

Zheng and Liang [55] proposed a discriminating descriptor called 3D motion map-based pyramid histograms of oriented gradient (M-PHOG) for depth-based human gesture recognition. The 3D motion map is generated through the entire depth video sequence to encode additional motion information from three projected orthogonal planes. By adding pyramid representation, the HOG descriptor is extended to M-PHOG which can characterize local shapes at different spatial grid sizes for gesture recognition.

Zhang *et al.* [64] proposed a new feature called enhanced shape context (eSC) to represent the spatial and temporal information feature for trajectory. In addition, they utilize the HOG feature for hand shape representation to describe the hands in video and PCA to reduce the dimension. Other works that use HOG are [18], [21], [50], [65], [81], [86].

PCA is a dimensionality reduction technique that has four main parts: feature covariance, eigen decomposition, principal component transformation, and choosing components in terms of variance. [21], [86] used the PCA for extracting Histogram of Gradient (HOG) features using multiple spatial grids (4×4, 8×8, and 16×16), while Wang *et al.* [95] used PCA for exacting the HOG feature from hand region, which is segmented by using self-adaptive skin model and depth constraint. Other researchers that used PCA for extracting HOG are [55], [64], [65]. Ahmed and Aly [82] used PCA for describing the texture and the shape of sign language images, while Tripathi and Nandi [20] applied PCA for extracting the Orientation Histogram (OH) and reducing the feature dimension. Rao and Kishore [47], have applied PCA for extracting the hand and head contour energies features for classification computed from discrete cosine transform.

CNN is one of the most widely used techniques in deep learning architectures. The main advantage of CNN over its predecessors is that it detects the important features automatically without any human supervision. A CNN model has two parts: feature extraction and classification. The convolution

pooling layers extract the features while the fully connected layers serve as a classifier. Islam *et al.* [43] propose a feature extraction process using CNN that consists of one or more fully connected convolutional layers as a standard multilayer neural network. Also, CNN has several dynamic parameters to train up the machine easily.

Camgoz *et al.* [78] used CNNs that take images as inputs and extract spatial features to extract Bidirectional Long Short-Term Memory Layers (BLSTM) temporally model the spatial features extracted by the CNNs. Cui *et al.* [88] used Recurrent CNN with temporal convolution and pooling for spatial and local temporal feature extraction, while Koller *et al.* [90], used hand shape-CNN as feature extractor for an additional GMM-HMM sign model. ElBadawy *et al.* [109] applied the 3D CNN for extracting spatial-temporal features and motion information encoded in multiple contiguous frames. Similarly, Liang *et al.* [103] applied 3D-CNNs to extract spatial and temporal features from video streams, and the motion information is captured by noting the variation in depth between each pair of consecutive frames. Other works that used CNN for feature extractions are [39], [40], [49], [52], [61], [75], [76], [112].

Another technique is the Local Binary Pattern (LBP), which is an efficient method used for texture feature extraction. This method is very popular for face detection and pattern recognition approaches. The LBP operator transforms an image into an array or image of integer labels describing the small-scale appearance of the image. Ahmed and Aly [82] employed the LBP to describe the texture and the shape of sign language images, while Santa *et al.* [89] used LBP for feature extraction. The LBP values on the hand region are calculated and 256- dimensional LBP histogram is generated from a hand region of interest (ROI) [51].

Other techniques of feature extraction from the selected articles are Zernike moments [41], [91], Radon transform [74], Convolutional self-organizing Map (CSOM) [92], scale-invariant feature transform (SIFT) [80], temporal accumulative features (TAF) [104], edge orientation histograms (EOH) [87], convexity approach [106] and convex invariant position based on RANSAC (CIPBR) algorithm [73].

B. RESEARCH QUESTION 2 (RQ2): WHAT ARE PERFORMANCES OF THE EXISTING VISION BASED HAND GESTURE RECOGNITION SYSTEMS AND THE POSSIBLE FUTURE DIRECTIONS?

1) PERFORMANCE OF THE SIGN LANGUAGE RECONGITION SYSTEM IN VARIOUS SETTING

Table 5 depicts the reported results in terms of recognition accuracy by the selected articles. For the signer dependent, the recognition accuracy ranges from 69% to 98%, with an average of 88.8% recognition accuracy among the selected studies. On the other hand, the signer independent's recognition accuracy reported in the selected studies ranges from 48% to 97%, with an average accuracy of 78.2%.

The recognition accuracy based on the input devices is presented in Table 6. For single cameras, the average recognition accuracy for signer dependent is 88%, and for signer independent is 79%. For active techniques, the average recognition accuracy for signer dependent is 89.6%, and for signer independent is 77.2%. On the other hand, for the invasive technique, only one result was presented for signer independent recognition, which is 88%.

We also looked at the recognition accuracy based on the data environment as shown in Table 7. For restricted environment, the average recognition accuracy for signer dependent is 88%, and for signer independent is 77%. However, for the uncontrolled environment (with only three articles reporting the results), the average recognition accuracy for signer dependent is 98%, and for signer independent is 90%. Though it does not provide conclusive evidence, it can be observed that the research on hand gestures in an uncontrolled environment shows promising results.

Table 8 shows the recognition accuracy for hand gesture recognition. For isolated gestures, the average recognition accuracy for signer dependent is 92%, and for signer independent is 77%. For continuous gestures, the average recognition accuracy for signer dependent is 84%, and for signer independent is 82%. On the other hand, for the fingerspelling, the average recognition accuracy for signer dependent is 81%, and for signer independent is 71%.

2) FUTURE DIRECTIONS

a: DATABASES

Many of the articles have postulated a future direction in which the gesture database will be bigger in term of number of gestures, number of people in the database and the coverage of the language [21], [35], [40], [51], [53], [60], [61], [64], [69], [75], [87], [97], [99]. It was highlighted in [105] that one of the important future directions in gesture recognition is the development of databases for many different sign languages i.e. a multilingual database which can be used in many research in the future.

To increase the rapid capturing and development of databases, devices such as 3D cameras and Kinects will be increasingly applied in future research [22], [41], [58], [59], [100], [111]. The increase in the database that contains gestures in many environments is vital as many of the future works will be aiming to solve this issue. The database also needs to cover many areas that will be important in the future such as multilingual database [105], real-world data [24], and multi-signers database [95].

b: HAND GESTURE REPRESENTATIONS

A variety of features can be used for hand gesture recognition, such as the shape of the segmented signer's hands (which is the main source of information for interpreting a specific sign), the motion information, the location, and the orientation of the gestures. Most past studies had used the handshape for Sign Language Recognition (SLR), while hand motion was the least used in hand gesture recognition.

TABLE 5. Recognition rate for signer dependent and signer independent vision-based hand gesture recognition system from the selected articles.

	Researcher	Data acquisition			Data environment		Hand gesture representation			Recognition rate	
		Single camera	Active techniques	Invasive techniques	Restricted	Uncontrolled	Isolated gestures	Continuous gestures	Fingerspelling	Signer dependent (%)	Signer independent (%)
1.	[28]	✓			✓		✓			-	48.06
2.	[38]	✓			✓		✓			95.30	78.00
3.	[109]	✓				✓	✓			98.00	85.00
4.	[42]	✓				✓	✓			-	97.00
5.	[41]*	✓			✓		✓			-	89.00
6.	[92]	✓			✓		✓				89.50
7.	[53]		✓		✓		✓			97.45	92.36
8.	[96]		✓		✓		✓			95.13	92.50
9.	[95]		✓		✓		✓			94.00	83.60
10	[81]		✓		✓		✓			92.40	70.90
11	[113]*		✓		✓		✓			-	88.00
12	[100]		✓		✓		✓			-	86.20 63.30
13	[52]		✓		✓		✓			69.20	65.80
14	[65]		✓		✓		✓			-	55.57
15	[105]		✓		✓		✓			-	57.40 65.80 97.50
16	[57]			✓			✓				88.00
17	[67]	✓			✓			✓		-	89.48
18	[31]	✓			✓			✓		-	92.49
19	[87]	✓			✓			✓		84.82	58.94
20	[113]*		✓		✓			✓		-	85.20
21	[86]	✓			✓				✓	69.70	39.40
22	[21]	✓			✓				✓	92.4	83.80
23	[41]*	✓			✓				✓	-	91.00

Feature extraction of dynamic signs appeared to be more challenging than static gestures. Dynamic signs can be sub-divided into a set of basic movements which can be analyzed in higher recognition layers, with the support of a powerful grammar model. A clear advantage of dividing a gesture into basic units is that it is possible to represent a huge range of sign gestures. This scheme can also skip any undetected or erroneous units though using a powerful grammar model can be computationally expensive.

Several authors have estimated that the future research in gesture recognition will be focusing on 3D gesture and the use of non-manual features [29], [58], [59], [71], [77], [84]. On top of that, greater emphasis should be made on dynamic gesture recognition [44], [45], as well as the continuous sign language [17], [25], [53], [71], [74].

c: OTHER POSSIBLE FUTURE DIRECTIONS IN GESTURE RECOGNITION

The future direction for gesture recognition will likely cover several areas as follows:

First, the future for this research lies in the need to expand the current feature set to be able to recognize more gestures

TABLE 6. Input method and recognition accuracy.

Input method	Recognition accuracy		
	Min	Max	Average
Single camera	SD = 70 SI= 39	SD = 98 SI = 97	SD = 88 SI = 79
Active techniques	SD = 69 SI = 56	SD = 97 SI = 97	SD = 90 SI = 77

SD=Signer Dependent, SI=Signer Independent

TABLE 7. Data environment and recognition accuracy.

Data environment	Recognition accuracy		
	Min	Max	Average
Restricted	SD = 70 SI = 39	SD = 98 SI = 97	SD = 88 SI = 77
Uncontrolled	SD = 98 SI = 85	SD = 98 SI = 97	SD = 98 SI = 90

SD=Signer Dependent, SI=Signer Independent

TABLE 8. Hand gesture representation and recognition accuracy.

Hand gesture representation	Recognition accuracy		
	Min	Max	Average
Isolated gestures	SD = 70 SI = 48	SD = 98 SI = 97	SD = 92 SI = 77
Continuous gestures	SD = 84 SI = 59	SD = 84 SI = 92	SD = 84 SI = 82
Fingerspelling	SD = 70 SI = 39	SD = 92 SI = 91	SD = 81 SI = 71

SD=Signer Dependent, SI=Signer Independent

(like those involving two hands or facial cues). The future gesture needs to deal with coarticulation [66] due to the extremely fast movements of the hand and this can be largely solved with an advanced data acquisition method. Second, the computational cost problem will be one of the issues to consider when developing camera devices [15], [91]. Reducing the computational cost means that the system can be developed in much a shorter time and reduce the learning time using advanced machine learning and unsupervised training. Third, another area that will be concentrated in the future is the use of smart and wearable devices [47], [49], [94] as the data acquisition tool.

3) SUMMARY

With regards to the data acquisition, most data were collected using single cameras in a restricted environment. The database used for the development of the hand gesture recognition system use limited numbers of standard signs which may not normally include the variation of signs for possible real-life applications. Moreover, many of the existing works did not look at the possibility of having a large size database for sign language. In addition, the nature of the restricted environment tends to constrict the choices made for data collection, which indirectly hinders the ability of the existing hand gesture recognition system, particularly in handing inputs that are beyond the restricted environments. Nevertheless, we must also understand that the use of restricted environments permits us to examine the effectiveness of the different solutions.

The emphasis of the vision-based gesture recognition that was performed by the present paper was motivated by the fact that most of us own smartphones that have built-in cameras. As smartphone technology had improved tremendously, so has the ability of these cameras in capturing high quality images. Thus, integrating the vision-based gesture recognition system into an existing ecosystem of smartphone cameras would be an effective way to bring the system to be utilized for real-life applications. This explains why most of the existing works used digital cameras as the primary method for data collection.

Further, most of the existing studies had also focused on recognizing isolated gestures, which have limited use. Only 20% of the existing works had focused on continuous gestures. The lack in the progress of looking at continuous gesture recognition could indicate that a lot more work is needed towards a practical vision-based gesture recognition system. The lack of using continuous gestures could indicate the complexity of recognizing the gestures as well as the inability of the existing solutions to detect these gestures with acceptable levels of accuracy.

V. CONCLUSION

This paper had looked at the issues, progress, and possible future direction of the vision-based hand gesture recognition system over a period of seven years. It appears that almost every article we reviewed had highlighted the importance of data acquisition, features, and the environment of the training data. It was also noted that the majority of the databases used in hand gesture recognition research were those from a restricted environment, thereby signaling the need for sign language databases to be less restrictive and contain different environments. This paper thus concludes that to make the vision-based gesture recognition system ready for real-life application, more attention needs to be focused on the uncontrolled environment setting as it can provide researchers the opportunity to improve the ability of the system in recognizing hand gestures in any form of environment.

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.
[23] R. Yang and S. Sarkar, "Detecting coarticulation in sign language using conditional random fields," in *Proc. 18th Int. Conf. Pattern Recognit. (ICPR)*, vol. 2, Jan. 2006, pp. 108–112.
[24] S. Tanaka, A. Okazaki, N. Kato, H. Hino, and K. Fukui, "Spotting fingerspelled words from sign language video by temporally regularized canonical component analysis," in *Proc. IEEE Int. Conf. Identity, Secur. Behav. Anal. (ISBA)*, Feb. 2016, pp. 1–7.
[25] N. Singh, N. Baranwal, and G. C. Nandi, "Implementation and evaluation of DWT and MFCC based ISL gesture recognition," in *Proc. 9th Int. Conf. Ind. Inf. Syst. (ICIIS)*, Dec. 2014, pp. 1–7.
[26] S. M. Shohieb, H. K. Elminir, and A. M. Riad, "Signsworld atlas: A benchmark Arabic sign language database," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 27, pp. 68–76, Jan. 2015, doi: 10.1016/j.jksuci.2014.03.011.
[27] B. Mocialov, G. Turner, K. Lohan, and H. Hastie, "Towards continuous sign language recognition with deep learning," in *Proc. Workshop Creating Meaning Robot. Assistants*, 2017, pp. 1–5.
[28] M. Fagiani, E. Principi, S. Squartini, and F. Piazza, "Signer independent isolated Italian sign recognition based on hidden Markov models," *Pattern Anal. Appl.*, vol. 18, no. 2, pp. 385–402, May 2015, doi: 10.1007/s10044-014-0400-z.
[29] N. Baranwal, K. Tripathi, and G. C. Nandi, "Possibility theory based continuous Indian sign language gesture recognition," in *Proc. IEEE Region 10 Conf.*, Nov. 2015, pp. 1–5.
[30] V. Sombandith, A. Walairacht, and S. Walairacht, "Recognition of lao sentence sign language using Kinect sensor," in *Proc. 14th Int. Conf. Electr. Eng./Electron., Comput., Telecommun. Inf. Technol.* , Jun. 2017, pp. 656–659.
[31] P. V. V. Kishore, D. A. Kumar, and M. Manikanta, "Continuous sign language recognition from tracking and shape features using fuzzy inference engine," in *Proc. Int. Conf. Wireless Commun., Signal Process. Netw. (WiSPNET)*, Mar. 2016, pp. 2165–2170.
[32] P. Kumar, R. Saini, S. K. Behera, D. P. Dogra, and P. P. Roy, "Real-time recognition of sign language gestures and air-writing using leap motion," in *Proc. 15th IAPR Int. Conf. Mach. Vis. Appl. (MVA)*, May 2017, pp. 157–160.
[33] P. T. Hai, H. C. Thinh, B. Van Phuc, and H. H. Kha, "Automatic feature extraction for Vietnamese sign language recognition using support vector machine," in *Proc. 2nd Int. Conf. Recent Adv. Signal Process., Telecommun. Comput. (SigTelCom)*, Jan. 2018, pp. 146–151.
[34] J. R. Balbin, D. A. Padilla, F. S. Caluyo, J. C. Fausto, C. C. Hortinela, C. O. Manlises, C. K. S. Bernardino, E. G. Finones, and L. T. Ventura, "Sign language word translator using neural networks for the aurally impaired as a tool for communication," in *Proc. 6th IEEE Int. Conf. Control Syst., Comput. Eng. (ICCSCE)*, 2016, pp. 425–429.
[35] F. Ronchetti, F. Quiroga, and L. Lanzarini, "LSA64: An Argentinian sign language dataset," in *Proc. Cong. Argentino Ciencias Comput. (CACIC)*, 2016, pp. 794–803.
[36] D. Konstantinidis, K. Dimittropoulos, and P. Daras, "A deep learning approach for analyzing video and skeletal features in sign language recognition," in *Proc. Int. Conf. Imag. Syst. Techn.*, Oct. 2018, pp. 1–6.
[37] M. Shi, Y. Huang, Z. Hu, and Q. Dai, "Dynamic sign language recognition algorithm using weighted gesture units," *J. Inf. Comput. Sci.*, vol. 12, no. 15, pp. 5611–5621, 2015.
[38] S. G. Azar and H. Seyedarabi, "Continuous hidden Markov model based dynamic Persian sign language recognition," in *Proc. 24th Iranian Conf. Electr. Eng. (ICEE)*, May 2016, pp. 1107–1112.
[39] O. Koller, S. Zargaran, H. Ney, and R. Bowden, "Deep sign: Hybrid CNN-HMM for continuous sign language recognition," in *Proc. Brit. Mach. Vis. Conf.*, 2016, pp. 136.1–136.12.
[40] Y. Ji, S. Kim, and K.-B. Lee, "Sign language learning system with image sampling and convolutional neural network," in *Proc. 1st IEEE Int. Conf. Robot. Comput. (IRC)*, Apr. 2017, pp. 371–375.
[41] P. K. Athira, C. J. Sruthi, and A. Lijiya, "A signer independent sign language recognition with co-articulation elimination from live videos: An Indian scenario," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 15, pp. 1–10, May 2019, doi: 10.1016/j.jksuci.2019.05.002.
[42] N. B. Ibrahim, M. M. Selim, and H. H. Zayed, "An automatic Arabic sign language recognition system (ArSLRS)," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 30, no. 4, pp. 470–477, 2017, doi: 10.1016/j.jksuci.2017.09.007.
[43] M. R. Islam, U. K. Mitu, R. A. Bhuiyan, and J. Shin, "Hand gesture feature extraction using deep convolutional neural network for recognizing American sign language," in *Proc. 4th Int. Conf. Frontiers Signal Process. (ICFSP)*, Sep. 2018, pp. 115–119.
[44] M. R. Mahmood, A. M. Abdulazeez, and Z. Orman, "Dynamic hand gesture recognition system for Kurdish sign language using two lines of features," in *Proc. Int. Conf. Adv. Sci. Eng. (ICOASE)*, Oct. 2018, pp. 42–47.
[45] A. Thongtawee, O. Pinsanoh, and Y. Kitjaidure, "A novel feature extraction for American sign language recognition using webcam," in *Proc. 11st Biomed. Eng. Int. Conf. (BMEiCON)*, Nov. 2018, pp. 1–5.
[46] D. A. Kumar, P. V. V. Kishore, A. S. C. S. Sastry, and P. R. G. Swamy, "Selfie continuous sign language recognition using neural network," in *Proc. IEEE Annu. India Conf. (INDICON)*, Dec. 2016, pp. 1–6.
[47] G. A. Rao and P. V. V. Kishore, "Selfie video based continuous Indian sign language recognition system," *Ain Shams Eng. J.*, vol. 9, no. 4, pp. 1929–1939, Dec. 2018, doi: 10.1016/j.asej.2016.10.013.
[48] N. Takayama and H. Takahashi, "Sign words annotation assistance using Japanese sign language words recognition," in *Proc. Int. Conf. Cyberworlds (CW)*, Oct. 2018, pp. 221–228.
[49] A. P. Uchil, S. Jha, and B. G. Sudha, "Vision based deep learning approach for dynamic Indian sign language recognition in healthcare," in *Proc. Int. Conf. Comput. Vis. Bio Inspired Comput.*, 2019, pp. 371–383.

[50] Y. Chen and W. Zhang, "Research and implementation of sign language recognition method based on Kinect," in *Proc. 2nd IEEE Int. Conf. Comput. Commun. (ICCC)*, Oct. 2016, pp. 1947–1951.

[51] T. Kodama, T. Koyama, and T. Saitoh, "Kinect sensor based sign language word recognition by mutli-stream HMM," in *Proc. 56th Annu. Conf. Soc. Instrum. Control Eng. Jpn. (SICE)*, Sep. 2017, pp. 94–99.

[52] Y. Ye, Y. Tian, M. Huenerfauth, and J. Liu, "Recognizing American sign language gestures from within continuous videos," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 2145–2154.

[53] Y. Jiang, J. Tao, W. Ye, W. Wang, and Z. Ye, "An isolated sign language recognition system using RGB-D sensor with sparse coding," in *Proc. IEEE 17th Int. Conf. Comput. Sci. Eng.*, Dec. 2014, pp. 21–26.

[54] G. Garcia-Bautista, F. Trujillo-Romero, and S. O. Caballero-Morales, "Mexican sign language recognition using Kinect and data time warping algorithm," in *Proc. Int. Conf. Electron., Commun. Comput. (CONIELECOMP)*, 2017, pp. 1–5.

[55] L. Zheng and B. Liang, "Sign language recognition using depth images," in *Proc. 14th Int. Conf. Control, Autom., Robot. Vis. (ICARCV)*, Nov. 2016, pp. 1–6.

[56] M. B. A. Majid, J. B. M. Zain, and A. Hermawan, "Recognition of Malaysian sign language using skeleton data with neural network," in *Proc. Int. Conf. Sci. Inf. Technol. (ICSITech)*, Oct. 2015, pp. 231–236.

[57] A. S. Elons, M. Ahmed, H. Shedid, and M. F. Tolba, "Arabic sign language recognition using leap motion sensor," in *Proc. 9th Int. Conf. Comput. Eng. Syst. (ICCES)*, Dec. 2014, pp. 368–373.

[58] P. Kumar, H. Gauba, P. P. Roy, and D. P. Dogra, "A multimodal framework for sensor based sign language recognition," *Neurocomputing*, vol. 259, pp. 21–38, Oct. 2017, doi: 10.1016/j.neucom.2016.08.132.

[59] P. Kumar, H. Gauba, P. P. Roy, and D. P. Dogra, "Coupled HMM-based multi-sensor data fusion for sign language recognition," *Pattern Recognit. Lett.*, vol. 86, pp. 1–8, Jan. 2017, doi: 10.1016/j.patrec.2016.12.004.

[60] D. R. Kartika, S. T. R. Sigit, and S. T. Setiawardhana, "Sign language interpreter hand using optical-flow," in *Proc. Int. Seminar Appl. Technol. Inf. Commun.*, 2016, pp. 197–201, doi: 10.1109/ISEMANTIC.2016.7873837.

[61] D. Konstantinidis, K. Dimitropoulos, and P. Daras, "Sign language recognition based on hand and body skeletal data," in *Proc. Con., True Vis.-Capture, Transmiss. Display 3D Video (3DTV-CON)*, Jun. 2018, pp. 1–4.

[62] Y. Okayasu, T. Ozawa, M. Dahlan, H. Nishimura, and H. Tanaka, "Performance enhancement by combining visual clues to identify sign language motions," in *Proc. IEEE Pacific Rim Conf. Commun., Comput. Signal Process. (PACRIM)*, Aug. 2017, pp. 1–4.

[63] P. Usachokcharoen, Y. Washizawa, and K. Pasupa, "Sign language recognition with Microsoft Kinect's depth and colour sensors," in *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Oct. 2015, pp. 186–190.

[64] J. Zhang, W. Zhou, C. Xie, J. Pu, and H. Li, "Chinese sign language recognition with adaptive HMM," in *Proc. IEEE Int. Conf. Multimedia Expo. (ICME)*, Jul. 2016, pp. 1–6.

[65] M. Elpeltagy, M. Abdelwahab, M. E. Hussein, A. Shoukry, A. Shoala, and M. Galal, "Multi-modality-based Arabic sign language recognition," *IET Comput. Vis.*, vol. 12, no. 7, pp. 1031–1039, Oct. 2018, doi: 10.1049/iet-cvi.2017.0598.

[66] A. Kumar, K. Thankachan, and M. M. Dominic, "Sign language recognition," in *Proc. 3rd Int. Conf. Recent Adv. Inf. Technol. (RAIT)*, 2016, pp. 422–428.

[67] P. V. V. Kishore, M. V. D. Prasad, D. A. Kumar, and A. S. C. S. Sastry, "Optical flow hand tracking and active contour hand shape features for continuous sign language recognition with artificial neural networks," in *Proc. IEEE 6th Int. Conf. Adv. Comput. (IACC)*, Feb. 2016, pp. 346–351.

[68] J. Huang, W. Zhou, H. Li, and W. Li, "Sign language recognition using 3D convolutional neural networks," in *Proc. IEEE Int. Conf. Multimedia Expo. (ICME)*, Jun. 2015, pp. 1–6.

[69] S. Aliyu, M. Mohandes, M. Deriche, and S. Badran, "Arabie sign language recognition using the Microsoft Kinect," in *Proc. 13rd Int. Multi-Conf. Syst., Signals Devices (SSD)*, Mar. 2016, pp. 301–306.

[70] S. Aliyu, M. Mohandes, and M. Deriche, "Dual LMCs fusion for recognition of isolated Arabic sign language words," in *Proc. 14th Int. Multi-Conf. Syst., Signals Devices (SSD)*, Mar. 2017, pp. 611–614.

[71] A. Eqab and T. Shanableh, "Android mobile app for real-time bilateral Arabic sign language translation using leap motion controller," in *Proc. Int. Conf. Electr. Comput. Technol. Appl. (ICECTA)*, Nov. 2017, pp. 1–5.

[72] D. Avola, M. Bernardi, L. Cinque, G. L. Foresti, and C. Massaroni, "Exploiting recurrent neural networks and leap motion controller for the recognition of sign language and semaphoric hand gestures," *IEEE Trans. Multimedia*, vol. 21, no. 1, pp. 234–245, Jan. 2019, doi: 10.1109/TMM.2018.2856094.

[73] D. Santos, B. Fernandes, and B. Bezerra, "HAGR-D: A novel approach for gesture recognition with depth maps," *Sensors*, vol. 15, no. 11, pp. 28646–28664, Nov. 2015.

[74] R. S. Rokade and D. D. Doye, "Spelled sign word recognition using key frame," *IET Image Process.*, vol. 9, no. 5, pp. 381–388, May 2015, doi: 10.1049/iet-ipr.2012.0691.

[75] S. Yang and Q. Zhu, "Video-based Chinese sign language recognition using convolutional neural network," in *Proc. IEEE 9th Int. Conf. Commun. Softw. Netw. (ICCSN)*, May 2017, pp. 929–934.

[76] O. Koller, H. Ney, and R. Bowden, "Deep hand: How to train a CNN on 1 million hand images when your data is continuous and weakly labelled," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3793–3802.

[77] J. Pu, W. Zhou, and H. Li, "Dilated convolutional network with iterative optimization for continuous sign language recognition," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 885–891.

[78] N. C. Camgoz, S. Hadfield, O. Koller, and R. Bowden, "SubUNets: End-to-end hand shape and continuous sign language recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3075–3084.

[79] B. Pathak, A. S. Jalal, S. C. Agrawal, and C. Bhatnagar, "A framework for dynamic hand gesture recognition using key frames extraction," in *Proc. 5th Nat. Conf. Comput. Vis., Pattern Recognit., Image Process. Graph. (NCVPRIPG)*, Dec. 2015, pp. 1–4.

[80] E. Escobedo-Cardenas and G. Camara-Chavez, "A robust gesture recognition using hand local data and skeleton trajectory," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 1240–1244.

[81] X. Chai, H. Wang, F. Yin, and X. Chen, "Communication tool for the hard of hearings," in *Proc. Int. Conf. Affect. Comput. Intell. Interact.*, 2015, pp. 1–3.

[82] A. A. Ahmed and S. Aly, "Appearance-based Arabic sign language recognition using hidden Markov models," in *Proc. Int. Conf. Eng. Technol. (ICET)*, Apr. 2014, pp. 1–6.

[83] K. C. P. Carrera, A. P. R. Erise, E. M. V. Abrena, S. J. S. Colot, and R. E. Telentino, "Application of template matching algorithm for dynamic gesture recognition of American sign language finger spelling and hand gesture," *Asia Pacific J. Multidiscip. Res.*, vol. 2, no. 4, pp. 154–158, 2014.

[84] M. Raees and S. Ullah, "Continuous number signs recognition," in *Proc. 12nd Int. Conf. Frontiers Inf. Technol.*, Dec. 2014, pp. 274–279.

[85] R. Elakkiya and K. Selvamani, "An active learning framework for human hand sign gestures and handling movement epenthesis using enhanced level building approach," *Proc. Comput. Sci.*, vol. 48, pp. 606–611, Jan. 2015.

[86] T. Kim, W. Wang, H. Tang, and K. Livescu, "Signer-independent fingerspelling recognition with deep neural network adaptation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 6160–6164.

[87] H. S. Nagendraswamy, B. M. Chethana Kumara, and R. L. Chinmayi, "Indian sign language recognition: An approach based on fuzzy-symbolic data," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2016, pp. 1006–1013.

[88] R. Cui, H. Liu, and C. Zhang, "Recurrent convolutional neural networks for continuous sign language recognition by staged optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1610–1618.

[89] U. Santa, F. Tazreen, and S. A. Chowdhury, "Bangladeshi hand sign language recognition from video," in *Proc. 20th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2017, pp. 1–4.

[90] O. Koller, S. Zargaran, H. Ney, and R. Bowden, "Deep sign: Enabling robust statistical continuous sign language recognition via hybrid CNN-HMMs," *Int. J. Comput. Vis.*, vol. 126, no. 12, pp. 1311–1325, Dec. 2018, doi: 10.1007/s11263-018-1121-3.

[91] S. Mathur and P. Sharma, "Sign language gesture recognition using Zernike moments and DTW," in *Proc. 5th Int. Conf. Signal Process. Integr. Netw. (SPIN)*, Feb. 2018, pp. 586–591.

[92] S. Aly and W. Aly, "DeepArSLR: A novel signer-independent deep learning framework for isolated Arabic sign language gestures recognition," *IEEE Access*, vol. 8, pp. 83199–83212, 2020, doi: 10.1109/ACCESS.2020.2990699.

[93] H. Park, J.-S. Lee, and J. Ko, "Achieving real-time sign language translation using a Smartphone's true depth images," in *Proc. Int. Conf. Commun. Syst. Netw. (COMSNETS)*, Jan. 2020, pp. 622–625.
[94] S. Shrenika and M. M. Bala, "Sign language recognition using template matching technique," in *Proc. Int. Conf. Comput. Sci., Eng. Appl. (ICCSEA)*, Mar. 2020, pp. 1–5.
[95] H. Wang, X. Chai, Y. Zhou, and X. Chen, "Fast sign language recognition benefited from low rank approximation," in *Proc. 11st IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG)*, May 2015, pp. 1–6.
[96] O. Amin, H. Said, A. Samy, and H. K. Mohammed, "HMM based automatic Arabic sign language translator using Kinect," in *Proc. 10th Int. Conf. Comput. Eng. Syst. (ICCES)*, Dec. 2015, pp. 389–392.
[97] Z. Zafrulla, H. Sahni, A. Bedri, P. Thukral, and T. Starner, "Hand detection in American sign language depth data using domain-driven random forest regression," in *Proc. 11st IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG)*, May 2015, pp. 1–7.
[98] D. Guo, W. Zhou, M. Wang, and H. Li, "Sign language recognition based on adaptive HMMS with data augmentation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 2876–2880.
[99] J. He, Z. Liu, and J. Zhang, "Chinese sign language recognition based on trajectory and hand shape features," in *Proc. Vis. Commun. Image Process. (VCIP)*, 2016, pp. 1–4.
[100] T. Liu, W. Zhou, and H. Li, "Sign language recognition with long short-term memory," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 2871–2875.
[101] S. S. Hazari, L. Alam, and N. A. Goni, "Designing a sign language translation system using Kinect motion sensor device," in *Proc. Int. Conf. Electr., Comput. Commun. Eng. (ECCE)*, Feb. 2017, pp. 344–349.
[102] W. N. Khotimah, N. Suciati, Y. E. Nugyasa, and R. Wijaya, "Dynamic Indonesian sign language recognition by using weighted K-Nearest neighbor," in *Proc. 11st Int. Conf. Inf. Commun. Technol. Syst. (ICTS)*, Oct. 2017, pp. 269–274.
[103] Z. Liang, S.-B. Liao, and B.-Z. Hu, "3D convolutional neural networks for dynamic sign language recognition," *Comput. J.*, vol. 61, no. 11, pp. 1724–1736, 2018, doi: 10.1093/comjnl/bxy049.
[104] A. A. Kindiroglu, O. Ozdemir, and L. Akarun, "Temporal accumulative features for sign language recognition," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 1288–1297.
[105] S. Tornay, M. Razavi, and M. Magimai.-Doss, "Towards multilingual sign language recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 6309–6313.
[106] P. Barros, N. T. Maciel-Junior, B. J. T. Fernandes, B. L. D. Bezerra, and S. M. M. Fernandes, "A dynamic gesture recognition and prediction system using the convexity approach," *Comput. Vis. Image Understand.*, vol. 155, pp. 139–149, Feb. 2017.
[107] P. V. V. Kishore, A. S. C. S. Sastry, and A. Kartheek, "Visual-verbal machine interpreter for sign language recognition under versatile video backgrounds," in *Proc. 1st Int. Conf. Netw. Soft Comput. (ICNSC)*, Aug. 2014, pp. 135–140.
[108] R. S. Rokade and D. D. Doye, "Spelled sentence recognition using radon transform," in *Proc. Sci. Inf. Conf.*, Aug. 2014, pp. 351–354.
[109] M. ElBadawy, A. S. Elons, H. A. Shedeed, and M. F. Tolba, "Arabic sign language recognition with 3D convolutional neural networks," in *Proc. 8th Int. Conf. Intell. Comput. Inf. Syst. (ICICIS)*, Dec. 2017, pp. 66–71.
[110] A. A. I. Sidig, H. Luqman, and S. A. Mahmoud, "Transform-based Arabic sign language recognition," *Proc. Comput. Sci.*, vol. 117, pp. 2–9, Nov. 2017, doi: 10.1016/j.procs.2017.10.087.
[111] B. T. Teodoro, J. Bernardes, and L. A. Digiampietri, "Skin color segmentation and levenshtein distance recognition of BSL signs in video," in *Proc. 30th SIBGRAPI Conf. Graph., Patterns Images*, Oct. 2017, pp. 95–102.
[112] G. A. Rao, K. Syamala, P. V. V. Kishore, and A. S. C. S. Sastry, "Deep convolutional neural networks for sign language recognition," in *Proc. Conf. Signal Process. Commun. Eng. Syst. (SPACES)*, Jan. 2018, pp. 194–197.
[113] J. Zhang, W. Zhou, and H. Li, "A new system for Chinese sign language recognition," in *Proc. IEEE China Summit Int. Conf. Signal Inf. Process. (ChinaSIP)*, Jul. 2015, pp. 534–538.
[114] P. Kumar, R. Saini, P. P. Roy, and D. P. Dogra, "A position and rotation invariant framework for sign language recognition (SLR) using Kinect," *Multimedia Tools Appl.*, vol. 77, no. 7, pp. 8823–8846, Apr. 2018, doi: 10.1007/s11042-017-4776-9.
[115] A. Choudhury, A. K. Talukdar, M. K. Bhuyan, and K. K. Sarma, "Movement epenthesis detection for continuous sign language recognition," *J. Intell. Syst.*, vol. 26, no. 3, pp. 471–481, Jul. 2017.

NORAINI MOHAMED (Graduate Student Member, IEEE) received the Bachelor of Education degree (Hons.) in information technology from Universiti Pendidikan Sultan Idris (UPSI), Perak, Malaysia, in 2006, and the master's degree in computer science from Universiti Malaya (UM), Kuala Lumpur, Malaysia, in 2011, where she is currently pursuing the Ph.D. degree with the Department of Software Engineering, Faculty of Computer Science. Then, she works as a Research Assistant for the High Impact Research Project, in speech recognition for both impaired and unimpaired speakers. Her research interests include hand gesture recognition, sign language recognition, automatic speech recognition, and deep learning methods.

MUMTAZ BEGUM MUSTAFA (Member, IEEE) received the B.Sc. degree in software engineering from Universiti Putra Malaysia (UPM), in 2002, and the M.Sc. degree in software engineering and the Ph.D. degree in computer science from Universiti Malaya (UM), in 2006 and 2012, respectively. She is currently an Associate Professor at UM. She has undertaken several Speech Synthesis research and holds grants from the Ministry of Higher Education. Her research and development of the HMM-based Malay speech synthesis system has won The Most Prestigious Award (MPA) for Excellent Research 2012 from MIMOS Bhd. the National Research and Development Centre in ICT and have won gold medals for several national level competitions. She has published several papers in prestigious speech conferences and journals. She has established network with several International Speech Research Laboratory in Japan and Singapore. She supervises a group of Ph.D. and Master Students working on speech synthesis, speech recognition, and speech signal processing. She has published her work in many of prestigious international journals. Her research interests include emotional speech synthesis and speech assistive tools for disabled individuals.

NAZEAN JOMHARI is currently a Senior Lecturer with the Department of Software Engineering, Faculty of Computer Science and Information Technology. She is also a member of the Centre of Quranic Research, Universiti Malaya, and a member of the Interaction Design Foundation and myHCI-UX Malaysia Chapter. She is also actively involved in Active Learning in Engineering Education (ALIEN) project under the Erasmus Plus. Her research interests include design and evaluation of interactive apps that help people with special needs, including autism, deaf, syndrome down, and older adult to gain spiritual knowledge.