A Survey of the Constraints Encountered in Dynamic Vision-Based Sign Language Hand Gesture Recognition

Ruth Wario(✉) and Casam Nyaga

University of the Free State, Bloemfontein, South Africa
wariord@ufs.ac.za

Abstract. Vision-based hand gesture recognition has received attention in the recent past and much research is being conducted on the topic. However, achieving a robust real time vision-based sign language hand gesture recognition system is still a challenge, because of various limitations (The term limitation in this study is used interchangeably to mean constraint or challenge in respect to the problems that can or are encountered in the process of implementing a vision-based hand gesture recognition system.). These limitations include multiple context and interpretations of gestures as well as well as complex non-rigid characteristics of the hand. This paper exposes the constraints encountered in the image acquisition via camera, image segmentation and tacking, feature extraction and gesture classification phase of vision-based sign language hand gesture recognition. It also highlights the various algorithms that have been used to address the problems. This paper will be useful to new as well as experienced researchers in this field. The paper is envisaged to act as a reference point for new researchers in vision-based hand gesture recognition in the journey towards achieving a robust system that is able to recognize full sign language.

Keywords: Algorithms · Constraints · Gesture recognition · Sign language

1 Introduction

Gestures form an important aspect in human communication, to the point that people gesture even in telephone conversations. Gesture recognition can be viewed as the ability of a computer based system to decode the meaning of a gesture [1]. Hand gesture recognition has many application areas for instance sign language recognition, robotic arm control and Human Vehicle Interaction (HVI) [2].

In this study, the main application area of interest was sign language recognition. Hand gesture recognition has demonstrated to be more convenient over other conventional methods of human computer interactions like mouse and key board [3]. There are two approaches to hand gesture recognition, namely data glove and vision-based [1]. The vision-based approach can be categorized as appearance-based methods and 3D hand model-based methods. Appearance-based methods are preferred in real-time performance, because it is less complex to perform image processing on a 2D image.

M. Antona and C. Stephanidis (Eds.): HCII 2019, LNCS 11573, pp. 373–382, 2019.
https://doi.org/10.1007/978-3-030-23563-5_30

The 3D hand model-based method provides a better description of hand features. However, as the 3D hand models are articulated, deformable objects with many degrees of freedom require a very large image database to cover all the characteristic shapes under different views. Matching the query image frames from video input with all images in the database is time-consuming and computationally expensive [4].

The vision-based approach is considered to provide a more natural and intuitive human computer interface [3]. However, hand gesture recognition has proved to be quite challenging due to the multiple context and interpretations of gestures amid other challenges like the complex non-rigid characteristics of the hand [5]. Sign language (SL) is also primarily grounded on spatial characteristics and iconicity characteristics. Hand parameters like the shape, motion of the hand, position in space as well as lips movement, and facial expressions are used to decode meaning of a sign [6].

Past research indicates that most research in sign language recognition is confined to a small subset of the whole sign language due to the constraints associated with vision-based hand gesture recognition [7]. This paper outlines the constraints associated with vision-based sign language hand gesture recognition.

2 Objective

The objectives of this study are to:

- Analyze the constraints in the hand tracking and segmentation phase of a vision-based sign language hand gesture recognition system.
- Analyze the constraints in the feature extraction phase of a vision-based sign language hand gesture recognition system.
- Analyze the constraints in the classification phase of a vision-based sign language hand gesture recognition system.

3 Methodology

In this study, a qualitative research design was employed through desktop research. The research comprised document analysis, which can be defined as an orderly process for reviewing or assessing printed and electronic documents [8]. Document analysis has been applied in many research studies to triangulate other methods, but can also be used singly in research [9]. It has been argued by [10] to be less time consuming, because it involves data selection as opposed to data collection and hence suitable for repeated reviews [10].

Desktop research, as guided by [11], has been successfully employed by [2, 12] and many other authors to bring out important conclusions; hence this method of data collection was used in this study. Twelve papers were reviewed in this study. The papers were searched using the google scholar search engine using key words matching the objectives.

4 Technology Description

This paper was based on identification of the constraints associated with the implementation of a vision-based sign language hand gesture recognition system. Different authors have come up with different representations and terms of the phases that comprise a typical vision-based gesture recognition system. Below is Table 1, indicating some of the terms used.

Table 1. Vision-based hand gesture recognition system phases by different authors

Author	Phases				
[7]	Image acquisition from camera	Hand region segmentation	Hand detection and tracking	Hand posture recognition	Classified gesture (display as text or voice display as text or voice)
[13]	Capture image	Image preprocessing	Feature extraction	Gesture recognition system	Assign specific task
[14]	Image acquisition	Hand segmentation	Feature extraction	Gesture classification	Gesture recognition
[10]	Capture video	Hand tracking and segmentation	Feature extraction	Classification and recognition	Text application interface

As depicted in Table 1, the phases of a vision-based hand gesture recognition system are similar even though they represent different instances of different systems. The phase includes image acquisition, hand tracking and segmentation, feature extraction, classification and recognition. Below is a brief description of each phase and the constraints associated with the phase.

i. Image acquisition from camera

The first step in gesture recognition is to capture the gesture via a video camera, either attached to the computer or independent from the computer[1]. The constraints in this phase may be due to a number of factors. For instance, accuracy of gesture recognition may be affected by the following camera specifications: color range, resolution and accuracy, frame rate, lens characteristics and camera computer interface [5].

ii. Hand region segmentation

The main objective of the segmentation phase is to remove the background and noises, leaving only the Region of Interest (ROI), which is the only useful information

[1] Computer in this case refers to a desktop computer, tablet or even laptop computer that is used in the vision-based hand gesture recognition system.

in the image. This objective can be achieved in various ways like skin colour detection, hand shape features detection and background subtraction [3]. A Bayesian classifier, which a is supervised learning model, can be used for skin colour segmentation as well as an unsupervised model such as K-Mean clustering [3].

iii. Hand detection and tracking

Hand tracking is an important phase in gesture recognition and can be achieved through a number of algorithms. The algorithms return information such as the colour tracking, template matching, motion tracking and other cues, which can be returned in order to track the hand. These algorithms may include Kalman filtering, particle filtering, optical flow, camshaft, viola jones, and mean shift among others [3, 15].

In the tracking phase while using the skin color-based methods, the skin colour may vary from one person to another posing a major constraint. Hence the Hue Saturation and Value (HSV) and Yellow blue component and red component (YCBCr) colour models are used to give a better result than other models, because they separate luminance from chrominance components.

iv. Hand gesture classification and recognition

Classification of the gesture is also viewed as the point of recognition of the gesture, because it is the last step of a hand gesture recognition system. This phase involves matching the current gesture feature with stored features. The classification algorithms play an important role in the gesture recognition system as they determine the accuracy of the gesture. The speed of the classification algorithm is also important, especially for real time systems as speed is of the essence [9]. In this phase there are many algorithms, which can be applied. They can be categorized as mathematical model based algorithms such as Hidden Markov Model (HMM) and Finite State Machine (FSM), or as soft computing algorithms such as neural networks [3].

5 Result

Constraints as identified by different authors are summarized in Table 2.

Constraints arranged in the phase that they occur

(a) Constraints associated with image acquisition

Image acquisition is the first step in vision-based sign language hand gesture recognition. This is done via a camera attached to the system or attached on the system. Table 3 illustrates the constraints associated with image acquisition.

Table 2. Constraints as identified by different authors

Reference	Contributions (constraints identified)
[16]	• View point dependence constraint - vision-based hand gesture recognition systems require that the users have to position their hand facing the camera, which is a challenge to many and it compromises the naturalness of the system • Gesture intraclass variability constraint - a sign language gesture suffers from uniqueness of sign language dialect and it is also not possible to perform the same gesture the same way even by the same individual • Gesture start and stop detection constraint - most hand gesture recognition systems rely on classification based on frames, hence when multiple gestures are performed it may be challenging to detect when a gesture begins and ends, which may lead to inaccuracy of detection • Gesture context challenge - sign language hand gestures like any other language have a grammar context and in most cases, the hand gestures are performed with other cues like face expressions and lip reading. This makes it a very complex problem to solve • Use of both hands for hand gestures constraint - many sign languages use both hands in the process of performing hand gestures, which causes hand tracking and detection challenges
[5]	• There are many types of hand gestures, which have different meanings and are performed differently • The complexity of the hand and its ability to move in different directions according to its degrees of freedom makes recognizing it a challenge • The computer vision discipline is still not well understood by many people and the cameras used are also not up to the task, since they also suffer environmental challenges • The paper contributes to recognizing the challenges in the stages, which they are likely to manifest in a gesture recognition system
[18]	In this paper, the constraints that were identified include, occlusion, varying position of the person performing the gesture, loss of depth information by the camera, special temporal nature of the hand and difference in the signer's speed, and the co-articulation constraint not being able to know when the next gesture begins or when the previous one ends
[1]	In this paper, the constraints that were identified include, complex background, speed in real time tracking, feature selection and co-articulation

(b) Constraints in the hand tracking and segmentation phase of a vision-based sign language hand gesture recognition system

The main constraint in hand tracking is brought about by the ability of the hand to move in different directions depending on its 27 degrees of freedom. This constraint is referred to, by most researchers, as rotation. Other constraints in this phase include variation in the speed of hand gestures [3], variation in skin colour, illumination variation, background complexity, and occlusion. Table 4 below outlines the constraints associated with tracking and segmentation of hand gestures.

Table 3. Constraints associated with image acquisition [5]

Step	Constraint
Image acquisition	Camera specifications (colour range, resolution and accuracy, frame rate, lens characteristics, camera computer interface) and 3D image acquisition (depth accuracy, synchronization)

Table 4. Constraints associated with tracking and segmentation [5]

Step	Constraint
Segmentation	Illumination variation, complex background and dynamic background
Gesture detection	Hand articulation, occlusion

(c) Constraints in the feature extraction phase of a vision-based sign language hand gesture recognition system

The most notable constraints in this phase include rotation, scale and translation. Rotation constraintarises when the hand region is rotated in any direction in the scene. Scale constraint arises, because of the different sizes of people's hands making the gestures. The translation problem is the variation of hand positions in different images, which leads to erroneous representation of the features [19]. Table 5 indicates the constraints that can be encountered in the feature extraction phase of a vision-based sign language hand gesture recognition system.

Table 5. Constraints in the feature extraction phase [5]

Feature type	Examples	Constraint
Histogram-based	• Histogram of gradient (HoG) features	Complex background and image noise affect performance of the algorithm
Transform domain	•Fourier descriptor • Discrete Cosine Transform (DCT) descriptor • Wavelet descriptor	Challenge in differentiating gestures
Mixture of features	• Combined features	Compatibility may lower the recognition rate
Moments	• Geometric moments • Orthogonal moments	Moments cannot handle occlusion well
Curve fitting based	• Curvature scale space	Sensitive to distortion in the boundary

(d) constraints in the classification and recognition phase of a vision-based sign language hand gesture recognition system

An appropriate classifier identifies gesture features and categorizes them into either predefined classes (supervised) or by their similarity (unsupervised) [20]. Some of the limitations encountered in this phase include large data sets for classifier training in some algorithms, computational complexity, selection of optimum parameters and recognition of unknown gestures. Below is Table 6 outlining the constraints likely to be encountered in the classification phase.

Table 6. Constraints in the classification phase [5]

Classifier	Constraint
Dynamic time wrapping (DTW)	• Requires a huge training data set • Not computationally effective for large gesture vocabulary size
K nearest neighbour (k-XN)	• Computationally intensive for a large dataset • Performance degrades as the dimensionality of the feature space increases
Deep networks	• Difficult to get an optimized solution for non-convex and nonlinear systems • Needs a huge training data set
Finite state machine (FSM)	• Complex to manage large states, • The state transition conditions are rigid
Artificial neural networks (ANN)	• Difficult to set parameters (e.g. the optimal number of nodes, hidden layers, sigmoid functions) • Training is computationally intensive and requires a large set of training data for obtaining acceptable performance • It acts like a 'black box,' and hence, it is difficult to identify errors in a complex network
Conditional random fields (CRF)	• High computational complexity during training makes it difficult to re-train the model when new training gesture sequences become available • CRFs cannot recognize totally unknown gestures, i.e., gestures that are not present in the training dataset
k-means	• Dependent on initial cluster centre values • Sensitive to outliers • Does not perform well in the presence of non-globular clusters • Selecting an appropriate value of *k* is challenging
Support vector machines (SVM)	• Right kernel function selection is challenging • Computationally expensive • SVM is a binary classifier; and hence, multi-class classification requires multiple pairwise classifications • Multi-class SVMs based on single optimization are difficult to implement
Hidden Markov Model (HMM)	• The number of states and the structure of the HMM must be predefined • Statistical nature of an HMM precludes a rapid training phase. Well-aligned data segments are required to train an HMM • The stationarity assumption in HMM may not hold true for a complete gestural action

(*continued*)

Table 6. (*continued*)

Classifier	Constraint
Mean-shift	• The algorithm is computationally complex • Data needs to be sufficiently dense with a discernible gradient to locate the cluster centres • Susceptible to outliers or data points located between natural clusters

The constraints can also be categorized by the cause. The three causes include the hand itself, the system and equipment in use and environmental factors, as indicated in Fig. 1.

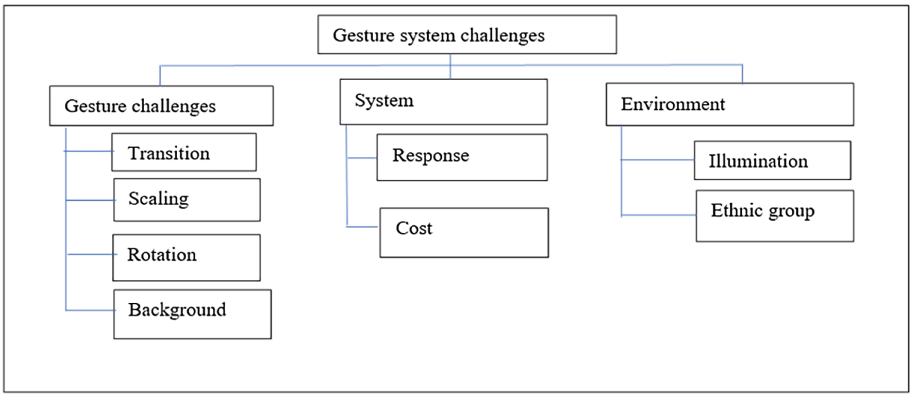

Fig. 1. Pictorial representation of gesture challenges or constraints [14]

6 Business Benefits

This study highlights the constraints encountered in vison-based system implementation in a logical way since the constraints are presented in the phases they are mostly likely to occur. This will help researchers and gesture recognition system developers to easily identify the constraints they want to address using new or a combination of existing algorithms. This can be beneficial in many hand gesture recognition application areas like robot control, game applications sign language recognition, amongst others.

The results of this study can provide a basis for a better sign language hand gesture recognition system capable of full sign language interpretation. Sign language interpretation systems are beneficial for communication, because they assist hearing impaired individuals to understand the non-hearing impaired and *vice versa*. Vision-based sign language interpretation systems enable communication in a natural way without the need for a human interpreter, hence they are likely to be more cost efficient. The vision-based gesture recognition interpretation systems can be deployed as

software applications on mobile phones, computers, laptops and even tablets. This can facilitate communication for hearing impaired individuals in public facilities like banks, airports, churches and schools.

7 Conclusion

In this paper, the phases of a typical vison-based sign language hand gesture recognition system are identified. The constraints that can be encountered in each stage of a vision-based hand gesture recognition system are outlined. It is evident from the literature that the challenges begin right from the first phase, which is image acquisition where camera resolution and quality can affect the gesture recognition rate. Background noise and lighting also pose serious constraints.

These constraints coupled with many others as mentioned in this paper have resulted in development of many algorithms. Each of these algorithms has its strengths and weaknesses. Hence the choice of the algorithm to use for sign language application may vary from one researcher to another. Further work needs to be done in order to find better solutions to overcome the constraints.

References

1. Choudhury, A., Kumar, A. Kumar, K.: A review on vision-based hand gesture recognition and applications (2015)
2. Micheni, E., Murumba, J.: The role of ICT in electoral processes: case of Kenya (2018)
3. Zhu, Y., Yang, Z., Yuan, B.: Vision based hand gesture recognition (2013)
4. Chen, Q., Georganas, N., Petriu, E.: Real-time vision-based hand gesture recognition using haar-like features. In: IEEE Instrumentation and Measurement Technology Conference IMTC (2007)
5. Chakraborty, B., Sarma, D.. Bhuyan, M., Macdorman, K.: Review of constraints on vision-based gesture recognition for human – computer interaction (2018)
6. Braffort, A.: Research on computer science and sign language: ethical aspects. In: Wachsmuth, I., Sowa, T. (eds.) GW 2001. LNCS (LNAI), vol. 2298, pp. 1–8. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-47873-6_1
7. Bhuyan, P., Ghoah, D.: A framework for hand gesture recognition with application to sign language (2006)
8. Corbin, J., Strauss, A.: Basics of qualitative research: techniques and procedures for developing grounded theory (2008)
9. Bowen, G.: Document analysis as a qualitative research document analysis as a qualitative research method. Qual. Res. J. **9**(2), 27–40 (2017)
10. Ghotkar, A.: Study of vision based hand gesture recognition using (2014)
11. McLeod, S.: Qualitative vs quantitative data simply psychology (2017)
12. Gamundani, A., Nekare, I.: A review of new trends in cyber attacks: a zoom into distributed database systems (2018)
13. Ahmed, T., Bernier, O., Viallet, J.: A neural network based real time hand gesture recognition system (2012)
14. Darwish, S., Madbouly, M., Khorsheed, M.: Hand gesture recognition for sign language: a new higher order fuzzy HMM approach. Hand **1**, 18565 (2016)

15. Ghotkar, S., Kharate, G.: Study of vision based hand gesture recognition using indian sign language. Int. J. Smart Sens. Intell. Syst. **7**(1), 96–115 (2014)
16. Zabulis, X., Baltzakis, H., Argyros, A.: Vision-based hand gesture recognition for human-computer interaction. Gesture, 1–56 (2009)
17. Wachs, J., Kölsch, M., Stern, H., Edan, Y.: Vision-based hand-gesture applications. Commun. ACM **54**(2), 60 (2011)
18. Bauer, B., Karl-Friedrich, K.: Towards an automatic sign language recognition system using subunits. In: Wachsmuth, I., Sowa, T. (eds.) GW 2001. LNCS (LNAI), vol. 2298, pp. 64–75. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-47873-6_7
19. Simei, A., Wysoski, G., Marcus V., Susumu, K.: A rotation invariant approach on static-gesture recognition using boundary histograms and neural networks. In: IEEE 9th International Conference on Neural Information Processing (2002)
20. Mitra, S., Acharya, T.: Gesture recognition: a survey. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **37**(3), 311–324 (2007)