

Kurdish Sign Language Recognition System

Abdulla Dlashad, Fattah Alizadeh

Department of Computer Science and Engineering, University of Kurdistan Hewler, Erbil, Kurdistan Region - F.R. Iraq

*Corresponding author's email: f.ibrahim@ukh.edu.krd

Received: 01 February 2018

Accepted: 22 April 2018

Available online: 30 June 2018

ABSTRACT

Deaf people all around the world face difficulty to communicate with the others. Hence, they use their own language to communicate with each other. On the other hand, it is difficult for deaf people to get used to technological services such as websites, television, mobile applications, and so on. This project aims to design a prototype system for deaf people to help them to communicate with other people and computers without relying on human interpreters. The proposed system is for letter-based Kurdish Sign Language (KuSL) which has not been introduced before. The system would be a real-time system that takes actions immediately after detecting hand gestures. Three algorithms for detecting KuSL have been implemented and tested, two of them are well-known methods that have been implemented and tested by other researchers, and the third one has been introduced in this paper for the 1st time. The new algorithm is named Grid-based gesture descriptor. It turned out to be the best method for the recognition of Kurdish hand signs. Furthermore, the result of the algorithm was 67% accuracy of detecting hand gestures. Finally, the other well-known algorithms are named scale invariant feature transform and speeded-up robust features, and they responded with 42% of accuracy.

Keywords: Image Processing, Kurdish, Machine Learning, Sign Language

1. INTRODUCTION

Image processing applications are growing rapidly. Most of the daily life problems have been solved depending on computer vision applications such as optical character recognition, motion detection systems, surveillance CCTV cameras, facial recognition applications, (Satyanarayanan, 2001). More specifically, we will narrow down our concern into an interesting application that is very vital for people with special inabilities. The concern of this paper is an image processing application regarding Sign Language interpreter for deaf people in Kurdish society.

Sign language is a language that uses visible sign patterns instead of using textual alphabetic letters. These patterns could involve the movement of one hand, or two, or the

combination of a facial expression with hands. Each nation throughout the world has its own sign language. Hence, there are various kinds of sign languages among which the most popular one is American Sign Language (ASL) (Anon, 2015). More examples are British Sign Language (Anon, 2015), Australian Sign Language (Anon, 2015), Persian Sign Language (Karami et al., 2011), and Arabic Sign Language (Abdel-Fattah, 2005) (Anon, 2015). In addition to these languages, Kurdish society also has its own sign language that uses different hand gestures to represent words or phrases.

According to the World Health Organization statistics in 2015, 360 million people around the world suffer from hearing ability, that is, 5% of the world's population (Truong, et al., 2016). The statistic also states that 75% of the deaf community are unemployed because of the lack of communication bridge between typical and deaf people.

This project is aiming to develop a cross-platform webcam-based application that translates Kurdish Sign Language (KuSL) to the Kurdish-Arabic script.

Access this article online

DOI: 10.25079/ukhjse.v2n1y2018.pp1-6

E-ISSN: 2520-7792

Copyright © 2018 Dlashad and Alizadeh. Open Access journal with Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0).

2. RELATED WORKS

This section will discuss the previous works that have been done by other researchers. It will explain in brief the importance of implementing sign language recognition system. It will also demonstrate different types of algorithms, methods, devices, and techniques in designing the system for detecting and interpreting different sign languages to text or speech format.

One of the earliest researchers that have been on sign language goes back to 1998. Two students in IEEE computer society have done research regarding sentence-level continuous ASL (Weaver and Pentland, 1998). The research was basically developing two system prototypes for translating ASL to English Language. Furthermore, the system is a real-time system that is based on hidden Markov model (Eddy, 1996). The first system places a camera on a desk for the input of the data which are signer's hand gesture. Moreover, the second system is putting a camera on top of a cap to read the data from the top view. Moreover, both systems track hands gestures using true skin color method. Moreover, because the systems are real-time based, they need to be calibrated to get the accurate pixel values of x and y . The first system can be calibrated easily because the position of the camera is steady. Thus, the image segmentation would be an easy task, while the second system uses the top view of the nose as a calibration point because its position is fixed. As a final result, the first system got 92% of the accuracy of words and the second system got 97% of the accuracy of words.

One of the most trending methods of translating sign language to speech or text is using wearable gloves. Two undergraduate students; Thomas Pryor and Navid Azodi at the University of Washington have proposed and designed the system. They invented a brilliant way to make it possible for deaf and normal people to communicate. In addition, they won Lemelson-MIT Student Prize in 2016 (Wanshel, 2016). Azodi points out that "Our purpose for developing these gloves was to provide an easy-to-use bridge between native speakers of ASL and the rest of the world" (Wanshel, 2016). Moreover, they have named the pair of gloves SignAloud. The implementation of creating these gloves is not published. However, they stated that the gloves have some sensors that record the movement of the hands and transmit these movements to a computer using a Bluetooth device. Then, the computer program will read these signals and translate them to English text and speech. Finally, one of the good features about this device is that it translated both alphabet sign language and compound signs that represent words.

Despite all the researches on sign language, there are also commercial products that have been designed to help normal people and deaf people communicate easily. This could be evidence of the importance of doing researches that could help in developing sign language recognition system. Motion Savvy (Stemper, n.d.) is one of the most advanced systems that work as an interface between deaf and Normal people. It is basically a tablet-sized device that is designed to help the deaf and hearing people to communicate. Furthermore, the tabled is designed to interpret ASL to the English language.

3. PROPOSED WORK

The proposed system can be generally classified into the following steps: The first step is to detect the hand shape from the input image of video frames and applying some preprocessing techniques such as noise reduction, image thresholding, and background extraction. The second step is the identification of the region of interest (ROI) area using some feature extraction algorithms. The system will recognize and distinguish each hand sign according to their features. The following flowchart is the representation of the system procedures in a general view [Figure 1].

The final step is basically to compare the extracted features from the input image with the available hand gestures in the database and matching them. If the comparison conditions were met, the system will print the equivalent Kurdish-Arabic

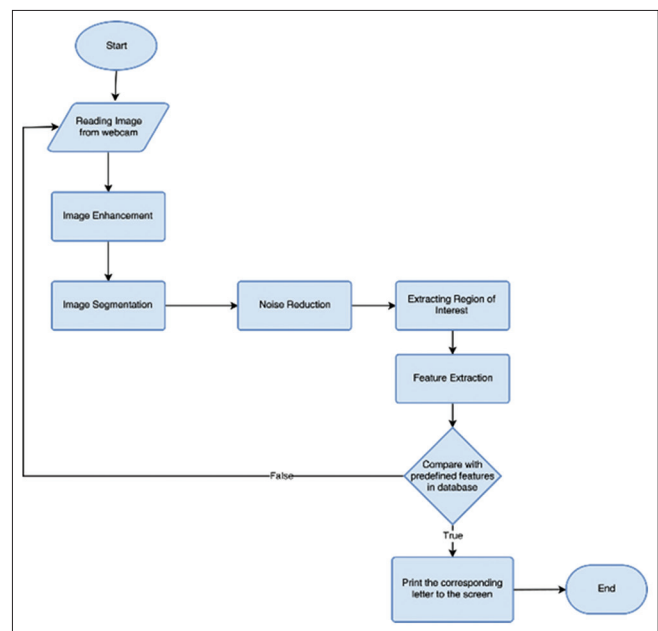


Figure 1. Flowchart of the system

script on the designated place. The system goes through some basic processes to completely detect and recognize the Kurdish Hand Signs.

3.1. Data Collection

The data that have been collected in this work is the image of Kurdish hand signs through the webcam of the laptop. The whole information about KuSL is indexed in the ministry of labor and social affairs in Kurdistan. In addition, the environment that is used in the development of KuSL recognition system is MacOS Sierra of Macbook Pro 13" with 720p resolution that is 1280×720 pixels.

3.2. Image Training and Reading

The very first stage of developing the proposed system is basically to extract the appropriate features and store the relevant shape descriptors of each hand sign into the database. Then, the new hand sign will be fed into the system through a computer webcam, as a query. The pre-processing steps are again will be applied to the query image to improve the quality of the image and to extract the relevant features more easily. Figure 2 shows the original image that has been taken from computer webcam under normal lighting condition and low detail background.



Figure 2. Original image captured from webcam



Figure 3. Smoothed Image

3.3. Image Enhancement

Image enhancement is one of the important stages of the development of this system. One of the image enhancement techniques that are used is essentially applying a Gaussian filter to the input image (Deng and Cahill, n.d.). This filter will smooth the image to make it ready for the next stages. The purpose of applying this filter is to smooth the image and to remove some noises, especially those of appears in the background, so, it will make it easier to detect the hand gesture. Figure 3 shows a sample of a blurred image.

3.4. Image Segmentation

Image segmentation plays an important role in almost every image processing application (Haralick and Shapiro, 1985). Hence, the segmentation section of this project is the color space conversion from RGB to YCbCr. It has been widely used in applications that are related to human skin detection (Albiol, et al., 2001). In the new color space, the skin color, including hand shapes, will be segmented and detected more accurately. In the new YCbCr color space, Y is for luminance, Cb is for blue difference, and Cr is for red difference. Once we blurred out the image using a Gaussian filter, we need to convert the smoothed image to YCbCr. Figure 4 shows the output image of YCbCr color space and binary image of the result.



Figure 4. YCbCr color space and thresholded output

3.5. Noise Reduction

Noise reduction is an important stage in this system that helps in resulting a more accurate detection of hand gesture. The output image of the previous stage is the binary image of skin region. Without applying noise removing techniques, we will be having lots of noise and holes in the image. Hence, to overcome this issue, we need to apply some morphological operations on the

binary image before passing it to other stages. Morphological operators, including dilation and erosion, are simply applied on binary images. They can also be applied to grayscale images. Figure 5 shows the input image of thresholded hand shape and the output for removed noise image.

3.6. Extract ROI

After smoothing and noise-reduction of the input image, it is now ready to go under the important stage of ROI extraction. Instead of comparing the resulted image to those of the database, the ROI (only hand shape) should be extracted. The process of ROI extraction can be done by calculating contours of the segmented binary image. Then, returning the biggest contour of the frame since the hand is the only contour in the frame. Once we have done that, the boundary box around the hand contour will be calculated and is considered as the ROI. Figure 6 illustrates the extracted ROI of a binary image and its equivalent region of the original hand image.

3.7. Feature Extraction

Feature extraction is the final and the most vital stage of the system. There are various types or feature extraction methods that can be utilized for different applications such as Haar-like features (Pham and Cham, 2007), BRIEF (Colender et al., 2010), scale invariant feature transform (SIFT) (Lowe, 2004), speeded-up robust features (SURF) (Bay et al., 2008), and so much more. Each one of these algorithms is working the best for some specific applications and in some specific environments. In the current work, we have used three algorithms for feature extraction. They are as follows:

- SIFT
- SURF
- Grid-based gesture descriptor

3.7.1. SIFT

SIFT feature extractor is a powerful algorithm which is widely used for academic purposes for extracting good features of an object in an image. In addition, it is scale invariant, rotation invariant, illumination invariant, and viewport invariant (Lowe, 2004). It uses Laplacian of Gaussian (LoG) to find scale space. Furthermore, SIFT needs to go through major stages to generate sets of features for a particular image. The stages are scale-space extrema detection, key-point localization, orientation assignment, and finally key-point descriptor (Lowe, 2004).

3.7.2. SURF

This feature detector is designed, in 2006, to overcome the performance issue of SIFT detector (Bay et al., 2008). The algorithm is almost similar to SIFT, while it uses LoG with box filter for finding scale space. The result of using this algorithm was not successful, and it gives an inaccurate comparison.

3.7.3. Grid-based gesture descriptor

The main contribution of the work is related to the new shape descriptor which is a simple algorithm and requires less computational complexity comparing to the other two algorithms. It is basically dividing the input image into $N \times N$ submatrices. Before performing the division, the algorithm should resize the original image and query image to have the same size. We can perform this by applying bitwise operations to the mask and input binary contour. Once the image is masked and the division is done, we need to count the number of white pixels (in the binary image) for each submatrix and calculate the average difference between the query image and those of in the database. Then, the comparison will be done among the images with a threshold of 0.2. In another word, it will consider a submatrix as truly-matched for any difference $<20\%$. The image of the maximum similarities is considered as the correct output.

The next section will show the performance results of the three utilized methods.

4. RESULTS

This section will demonstrate the test result of each algorithm of feature extraction. Furthermore, the test data are taken

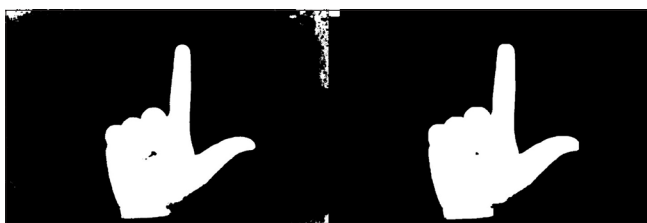


Figure 5. Removed noise output

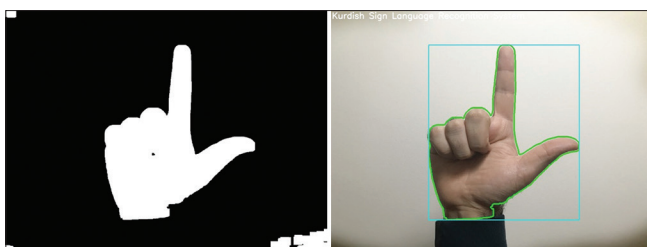


Figure 6. Extracted ROI and drwand contours and bounding box around it.

Table 1. SIFT test result

ا	ى	0	17%
ب	ى	0	7%
ح	م	0	6%
د	نا	0	2%
ه	ى	0	17%
و	م	1	0%
ز	ن	1	0%
س	ى	0	5%
ش	س	1	0%
نا	ش	1	0%
ى	ى	0	1%
	ى	1	0%
	Result	42%	4.6%

SIFT: Scale invariant feature transform

Table 2. SURF test result

Alphabets	Read by SURF	Result	Error rate
ا	ى	0	15%
ب	ى	0	19%
ح	ع	0	4%
د	ع	1	0%
ه	ى	0	8%
و	ى	0	6%
ز	ن	1	0%
س	ى	0	27%
ش	س	1	0%
نا	ش	1	0%
ى	ى	0	5%
	ى	1	0%
	Result	42%	7.0%

from one set of samples of 12 Kurdish signs. The samples are taken from normal conditions of lighting, viewport, and rotation. Moreover, the general formula for finding error rate is as follows:

$$1 - \frac{\text{Similarity Score of query image}}{\text{Similarity Score of correct image}} * 100 \% \quad (1)$$

4.1. SIFT Algorithm

Table 1 shows the test result for the SIFT algorithm. It is clearly shown that most of the signs are detected inaccurately. For example, the letter “ب” is recognized as “ى” while it is not true. In addition, the error rate indicates the accuracy of matching of every single sign.

4.2. SURF Algorithm

Table 2 shows the test result for the SIFT algorithm. It is clearly shown that most of the signs are detected inaccurately. For example, the letter “ب” is recognized as “ى” while it is not true. In addition, the error rate indicates the accuracy of matching of every single sign.

Table 3. GRIDDING algorithm test result

Alphabets	Read by GRIDDING	Result	Error rate
ا	ا	1	0%
ب	ب	1	0%
ح	ح	1	0%
د	د	1	0%
ه	ه	0	0%
و	م	1	2%
ز	ق	0	0%
س	ق	1	22%
ش	س	1	0%
نا	ى	0	4%
ى	ق	0	3%
	ى	1	0%
	Result	67%	2.6%

4.3. GRIDDING Algorithm

The proposed shape descriptor in this work is outperforming the other two algorithms, in terms of both accuracy and error rate [Table 3]. In another word, it has a lower error rate for the signs and has more correct outputs. As shown in Table 3, the average accuracy of the algorithm is 67%, and the error rate is 2.6%.

5. CONCLUSION AND FUTURE WORKS

This work aimed to develop a system to aid deaf people to understand the sign languages using different algorithms to translate the signs to proper letters. The education of Kurdish deaf people needs to be developed more and more. One of the advancements is designing a system to enable them to communicate with other people.

The challenging point of this work was implementing feature extraction methods. Three algorithms have been implemented, and their performance has been tested. The test result of the proposed algorithm turned out to have 67% of the accuracy of recognizing sign hands while the other two well-known algorithms (SIFT and SURF) responded with the accuracy of 42%. As a future work, we can suggest to test the system under various environmental conditions such as low light, background noise, and different kinds of skin colors. Furthermore, utilizing larger set of training images and using learning algorithms can improve the quality of the recognized letters.

REFERENCES

- Abdel-Fattah, M. A. (2005). Arabic sign language: A perspective. *Journal of Deaf Studies and Deaf Education*, 10(2), 212-221.
- Al, S. (2017). *SIFT: Theory and Practice*. Available from: <http://>

- www.aishack.in/tutorials/sift-scale-invariant-feature-transform-introduction/. [Last accessed on 2017 Jun 19].
- Albiol, A., Torres, L. & Del, E. J. (2001). *Optimum Color Spaces for Skin Detection*. 2001 International Conference, 1, 122-124.
- Anon. (2015). *Sign Language and Hearing Aids*. Available from: <http://www.disabled-world.com/disability/types/hearing/communication>. [Last accessed on 2017 Jan 08].
- Bay, H., Ess, A., Tuytelaars, T. & Van Gool, L. (2008). Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3), 346-359.
- Colender, M., Lepetit, V., Strech, C. & Fua, P. (2010). Brief: Binary robust independent elementary features. *Computer Vision–ECCV* 6314, 778-792.
- Deng, G. & Cahill, L. (n.d.). *An adaptive Gaussian Filter for Noise Reduction and Edge Detection*. 1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference.
- Eddy, S. R. (1996). Hidden Markov Models. *Current Opinion in Structural Biology*, 6(3), 361-365.
- Haralic, R. M. & Shapiro, L. G. (1985). Image segmentation techniques. *Computer Vision, Graphics, and Image Processing*, 29(1), 100-132.
- Karami, A., Zanj, B. & Sarkaleh, A.K. (2011). Persian sign language (PSL) recognition using wavelet transform and neural network. *Expert Systems with Applications*, 38(3), 2661-2667.
- Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91-110.
- Pham, M. T. & Cham, T.J. (2007). *Fast Training and Selection of Haar Features using Statistics in Boosting-Based Face Detection*. In: Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. p. 1-7.
- Reed, T. R. & Du Buf, J. M. H. (1993). A review of recent texture segmentation and feature extraction techniques. *CVGIP: Image Understanding*, 57(3), 359-372.
- Satyanarayanan, M. (2001). Pervasive computing: Vision and challenges. *IEEE Personal Communications*, 8(4), 10-17.
- Stemper, J. (n.d.). *MotionSavvy UNI: 1st Sign Language to Voice System*. Available from: <https://www.indiegogo.com/projects/motionsavvy-uni-1st-sign-language-to-voice-system#>. [Last accessed on 2017 Jan 21].
- Truong, V. N., Yang, C. K. & Tran, Q. V. (2016). *A Translator for American Sign Language to Text and Speech*.
- Wanshel, E. (2016). *Students Invented Gloves That Can Translate Sign Language into Speech and Text*. Available from: http://www.huffingtonpost.com/entry/navid-azodi-and-thomas-pryor-signaloud-gloves-translate-american-sign-language-into-speech-text_us_571fb38ae4b0f309bae06d. [Last accessed on 2017 Jan 12].
- Weaver, J. & Pentland, A. (1998). Real-time American sign language recognition using desk and wearable computer based video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, 1371-1375.