# Image based Arabic Sign Language Recognition System

**Article** *in* International Journal of Advanced Computer Science and Applications · January 2018

DOI: 10.14569/IJACSA.2018.090327



**5 authors**, including:

Munerah H. Alzaidan
King Saud University
**1** PUBLICATION **48** CITATIONS
SEE PROFILE

Raghad Alghonaim
Imperial College London
**3** PUBLICATIONS **67** CITATIONS
SEE PROFILE

# Image based Arabic Sign Language Recognition System

Reema Alzohairi, Raghad Alghonaim, Waad Alshehri, Shahad Aloqeely,
Munera Alzaidan, Ouiem Bchir
Computer Science Department, College of Computer and Information Sciences,
King Saud University

***Abstract*—Through history, humans have used many ways of communication such as gesturing, sounds, drawing, writing, and speaking. However, deaf and speaking impaired people cannot use speaking to communicate with others, which may give them a sense of isolation within their societies. For those individuals, sign language is their principal way to communicate. However, most people (who can hear) do not know the sign language. In this paper, we aim to automatically recognize Arabic Sign Language (ArSL) alphabets using an image-based methodology. More specifically, various visual descriptors are investigated to build an accurate ArSL alphabet recognizer. The extracted visual descriptors are conveyed to One-Versus-All Support Vector Machine (SVM). The analysis of the results shows that Histograms of Oriented Gradients (HOG) descriptor outperforms the other considered descriptors. Thus, the ArSL gesture models that are learned by One-Versus-All SVM using HOG descriptors are deployed in the proposed system.**

***Keywords*—Component; Arabic sign language; image; visual descriptor; recognition**

## I. INTRODUCTION

Human communication has been evolving over time. Overages, humans have used petroglyphs, pictograms, ideograms, alphabet, sounds, signals, gestures as ways of communication. Nowadays, the dominant communication way relies on alphabet expression either orally, in writing, or as sign language. People suffering from hearing and/or speaking disorders cannot communicate orally with others. Moreover, they usually prove difficulties to learn how to write and read a text. Thus, sign language has emerged as an effective alternative to express their thoughts. According to World Health Organization over 5% of the world's population (360 million people) suffer from hearing impairment. Moreover, the World Federation of the Deaf stated that the number of deaf and hearing-impaired people among the Arab region exceeded 700,000 persons in 2008 [1].

Although many hearing-impaired people master sign language, few "normal" individuals understand and/or can use sign language. This affects the communication with deaf people and results in a kind of isolation between them and "normal" people world. This gap can be reduced using a system that allows the translation of sign language automatically to text and vice versa. Nowadays, many paradigm shifts in many technology fields have helped researchers to propose and implement systems targeting sign languages recognition [2]–[7]. Thus, several works on sign language recognition have been proposed for various sign languages, including American Sign Language, Korean Sign Language, Chinese Sign Language, etc. [8]. The proposed sign recognition systems rely on either image-based or sensor-based solutions.

Most of the sensor-based systems recognize gestures utilizing glove-based gadgets which provide information about the position and the shape of the hand [9]. However, these gadgets are cumbersome and generally have several links connected to a computer. This yields the need of utilizing non-intrusive, image-based methodologies for perceiving gestures [10]. Image-based systems have been proposed as an alternative solution that allows availability and naturalness. These image-based systems utilize image processing algorithms to recognize and track hand signs. This makes the process easier to the signer, since these systems do not require the impaired person to use any sensor. Moreover, they can be deployed to smart devices. Thus, due to the availability of cameras on the portable devices, image-based sign language recognition system can be used anytime and anywhere.

The key stone of any image-based system is feature (visual descriptor) extraction [11]. The role of these features is to translate the information perceived in the image to a numerical vector that can convey the appropriate information to the recognition system. Many features have been used and reported in the literature [11]–[13]. Some of them are general, describing either colors, textures, edges, or shapes of the content of the image [13]. Others, are application-dedicated and are designed for a specific application [14].

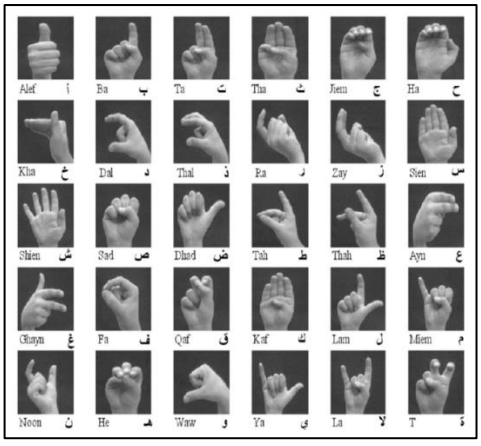


Fig. 1. The 30 gestures of the ArSL letters [14].

Recently, research on image-based recognition for ArSL have been reported [14]–[23]. ArSL include 30 gestures. As shown in Fig. 1, each gesture represents a specific hand orientation and finger positioning. In order to recognize these gestures, features have to be extracted. However, it is not straightforward to choose a feature that allows recognizing and segregating ArSL alphabet. This is due to the fact that ArSL has the characteristic of having several gestures that are very similar to each other like "Dal" and "Thal", "Ra", and "Zay", etc. In the literature [14]–[23], different types of features have been used. However, there is no empirical comparison that investigated which feature is suitable for Arabic letter recognition.

In this paper, we aim to design an ArSL recognition system that captures the ArSL alphabet gestures from an image in order to recognize automatically the 30 gestures displayed in Fig. 1. More specifically, we intend to investigate various features to build an ArSL alphabet recognizer.

This paper is organized as follows: Section I includes an introduction to the problem, Section II briefly explains Arabic sign language and compares it to other sign languages, Section III discusses the existence of related works, Section IV shows the proposed design of Arabic sign language recognizer, Section V discusses the result of running the experiments and how the system has been implemented, and Section VI concludes the paper.

## II. Arabic Sign Language

Sign language is the language that is used by hearing and speech impaired people to communicate using visual gestures and signs [24]. There are three kinds of image-based sign language recognition systems: alphabet, isolated word, and continuous sequences [23]. Usually, hearing and speech impaired communicate with others using words and continuous sequences, since it is faster than spelling each single word. However, if the desired word does not have a standard sign that represent it, signers use finger spelling. They spell out the word using gestures which have corresponding letters in the language alphabet. In this case, each letter is performed independently by a static posture [23]. Finger spelling gestures use a single-hand in some languages and two-hand gestures on others. For example, languages such as Australian, New Zealand and Scotland use two hands to represent the different alphabet [25].

Same as sign languages, finger spelling alphabet are not universal. Each language is characterized by a specific alphabet gestures. However, some languages share similar alphabet gestures. For instance, regardless of the unlikeness between Japanese and English orthography, Japanese Sign Language and American Sign Language (ASL) share a set of similar hand gestures. Also, the German and the Irish manual alphabet hand gestures are similar to the ASL ones. Similarly, French and Spanish alphabets share similar characteristics. Although the Russian language includes more alphabet to represent the Cyrillic ones, it has high similarities with the French and Spanish languages for the other gestures [25].

For Arab countries, Arabic Sign Language (ArSL) is the official sign language for hearing and speech impaired [26]. It was officially launched in 2001 by the Arab Federation of the Deaf [26]. Although the Arabic Language is one of the most spoken languages in the world, ArSL is still in its evolutionary phases [27]. One of the largest issues that face ArSL is "Diglossia". In fact, in each country, the regional dialect is spoken rather than the written language [28]. Therefore, variant spoken dialects made variant ArSLs. They are as many as Arab countries but with many words in common and the same alphabet. The 30 Arabic alphabet gestures are represented in Fig. 1. There are also extra letters that have the same original gestures but with different rotation or additional motion. These are the different ways for writing "Alef". Fig. 2 displays these variants.

Fig. 2. ArSL variants.

ArSL is based on the letters shapes. Thus, it includes letters that are not similar to other languages letter representation. For example, Fig. 3 shows the American Sign Language (ASL) alphabet.

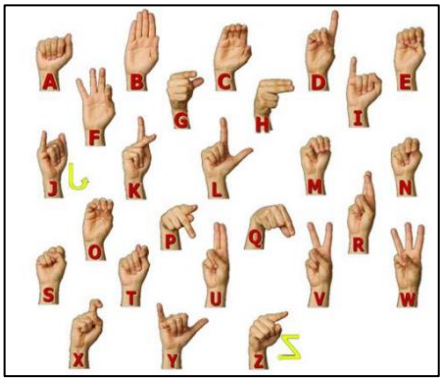

Fig. 3. ASL alphabet [29].

We notice from Fig. 3 that both ArSL and ASL letters are one-handed gestures. In addition to that, ArSL and ASL have some similar gestures (see Fig. 4). Some of them represent the same letter sound such as "Lam" and L (Fig. 4(a)), "Sad" and S (Fig. 4(b)), and "Ya" and Y (Fig. 4 (c)). On the other hand, there are other similar gestures for different letters sounds such as "Kaf" and B (Fig. 4(d)), "Ta" and u (Fig. 4(e)), and "Meem" and I (Fig. 4(f)).

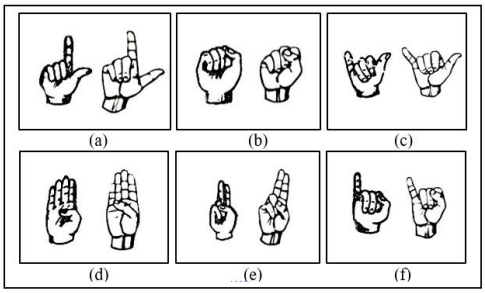


Fig. 4. Similar gestures between ASL and ArSL, (a)"Lam" and L, (b)"Sad" and S, (c)"Ya" and Y, (d)"Kaf" and B, (e)"Ta" and H, (f)"Meem" and I.

On the other hand, several ArSL letters are similarly gestured. This is a characteristic of ArSL since several Arabic letters are similarly shaped. For example, "Tah" and "Thah" are similar in the way that the index finger is raised and the rest of the fingers are facing the right side (Fig. 5(a)). Furthermore, "Ra" is similar to "Zay" but "Zay" has two curved fingers while "Ra" has only one curved finger (Fig. 5(b)). Similarly, the thumb and the index finger are curved in like a C shape in "Dal" and "Thal", yet the middle finger in "Thal" is curved too (Fig. 5(c)).

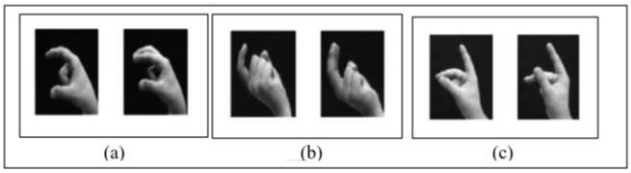


Fig. 5. Similar gestures in ArSL, (a)"Tah" and "Thah". (b)"Ra" and "Zay". (c)"Dal" and "Thal".

## III. Related Works

Recently, sign language recognition has become an active field of research [18]. Sign language recognition systems translate sign language gestures to the corresponding text or speech [30] in order to help in communicating with hearing and speech impaired people. These systems can be considered as HCI applications, where the computer would be able to identify those hand gestures and convert them to text or speech [14], [18]. They have been applied to different sign languages [18], [31], [32]. Sign language recognition systems are based on one of two ways to detect sign languages' gestures. They are sensor-based recognition systems and image-based recognition systems [14].

In sensor-based systems, sign language recognition is based on sensors that detect the hand's appearance. For this kind of system, two types are considered, which are the glove-based systems [33] and the Kinect-based systems [29]. Glove-based systems [33] use electromechanical devices to recognize hand gestures. Hearing and speech impaired signers are required to wear a glove that is linked to some sensors that gather information [34]. Although this technique can offer good results, it can be inconvenient to the signers [34]. For the second category, Kinect sensors are used to detect sign language gestures. Originally, these sensor devices were developed by Microsoft for their Xbox game as an input device to interact with video games without using any remote controllers [35]. Nowadays, the use of this device is expanding to include recognition systems like sign language recognition.

On the other hand, image-based systems use images or videos [32], [36], [37] along with image processing and machine learning techniques to recognize sign language gestures [34]. These systems fall into two categories. The first depends on using gloves containing visual markers to detect hand gestures, such as colored gloves [14]. However, this method prevents sign language recognition systems from being natural, where naturalness is expected from similar HCI systems [14]. The second category depends on images capturing hand gestures of the sign language [34]. When using these image-based recognition systems, hearing and speech impaired do not need to use any sensors or gloves with visual markers, which eliminates any inconvenience and makes the system more convenient [14]. The mentioned types of sign language recognition systems are shown in Fig. 6.

Image-Based Sign Language Recognition systems categorize the hand gesture from a 2D digital image by using image processing and machine learning techniques [34]. Such systems either recognize static or dynamic continuous gestures [38]. The images of ArSL shown in Fig. 1 are static gestures.

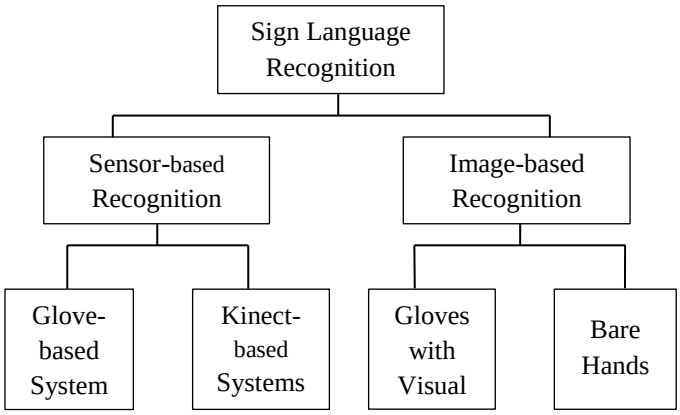


Fig. 6. Sign language recognition systems.

In the literature, various sign language recognition systems have been proposed. The authors in [32] have implemented a video-based continuous sign language recognition system. The system is based on continuous density hidden Markov models (HMM) [39]. It recognizes sign language sentences, based on a lexicon of 97 signs of German sign language (GSL). The system achieves an accuracy of 91.7%. Similarly, HMM [39] has been used by authors in [40]. The system recognizes Japanese sign language (JSL) words. This approach is video-based continuous recognition. Six visual descriptors were defined to recognize JSL, which are the flatness, the gravity center position, the area of the hand region, the direction of hand motion, the direction of the hand, and the number of protrusions. The system recognized 64 out of 65 words successfully. In [41] the authors have used a method to find the centroid for mapping the hand gesture of Sinhala Sign Language (SSL). The system recognizes image-based gestures of SSL words. A dataset of 15 Red-Green-Blue (RGB) image of gestures from 5 signers captured by a web camera has been used in this experiment. The system identified ten gestures with 100% accuracy, four gestures with 80% accuracy and one gesture with 60% accuracy. It recognized 92% of the 15 gestures.

The authors in [42] used HMM [39] classifier. The system recognizes the vocabulary of GSL. The used dataset consists of a vocabulary of 152 signs of GSL performed by a single signer ten times each. The system achieved a recognition rate of 97.6%. The authors in [43] have proposed a novel recognition method based on Spatio-temporal visual modeling of sign language. It uses Support Vector Machines (SVMs) [44] to recognize Chinese Sign Language. Experimentation was conducted with 30 groups of the Chinese manual alphabet images. In [45] the authors have investigated the problem of recognizing words from a video. The words are finger spelled

using British Sign Language (BSL). A dataset of 1,000 low-quality web-cam videos of 100 words has been used in the experiment. The system achieved a word recognition accuracy of 98.9%.

A hand gesture recognition of Indian sign language (ISL) has been suggested in [46]. The system applies Histogram of Gradient Orientation (HOG) visual descriptors extraction approach. It is then converted to a neural network classification recognition purpose. The dataset consists of alphanumerical characters. They are collected using a simple web camera.

Recently, ArSL systems that recognizes static alphabet gestures have been proposed [14], [15], [18], [19], [21], [47]. These are image-based systems that do not rely on the use of sensors or colored gloves. In the following, we describe these approaches.

### A. Neuro-Fuzzy based Approach

The authors in [14] proposed an image-based recognition system for ArSL gestures. The system includes 6 phases. After the acquisition, images are filtered using 3×3 median filter to remove the noise and enhance the image quality. Next, the resulting images are segmented into two regions. One region is the gesture, and the other is the background. Segmentation is performed using iterative thresholding algorithm [48]. Then, the hand's direction and the center area are calculated. Border tracing algorithm was applied in order to detect the borders of the hand. Next, borders were smoothed by Gaussian filter [8] [9], to obtain continuous edges. Based on this information, a visual descriptor vector is extracted. It is scale, translation, and rotation invariant. The length of the vector is set to 30. Each entry is the distance between the center area and a point from the hand border. Not all border points are considered. In fact, 30 equidistant points lying from 90° before the orientation ax to 113° after it are selected.

To assure that the chosen visual descriptor is scale-invariant, normalization was applied by dividing each entry by the maximum vector length and then multiplying them by 100 to make the range from 0 to 100. The Adaptive Neuro-Fuzzy inference systems (ANFIS) which is a kind of artificial neural network [49] was used to recognize the different hand gestures. A fuzzy model is built for each of the 30 gestures. The process of fuzzy model identification is done using subtractive clustering algorithm [50] for determining the number of fuzzy rules, and the membership functions. Also, the least square estimate (LSE) method [49] is used for estimating the parameters. Moreover, the hybrid learning algorithm [51] which combines both Gradient descent [49] and LSE [49] was used in training the fuzzy model.

The dataset is acquired using a Computer-connected camera. It includes grayscale images of the 30 ArSL alphabet. The images were taken from 60 different people with different image sizes and orientations. Around 66% of the samples were used for the training set while the rest were used for the testing set. The experimental results were directly affected by the parameter of the cluster radius ($r_a$). Overfitting occurred when $r_a$ values were small. On the other hand, when $r_a$ values were huge, the training and testing results are not satisfactory. The best results are obtained with $r_a$ = 0.8. In this case, the system recognition rate reached 93.55%. However, some letters that are similar in their gestures were misclassified. These are "Ra" and "Zay", "Tah" and "Thah", and "Dal" and "Thal". Besides, "Ha" and "Sad" were misclassified too, although they are not similar visually.

### B. Adaptive Network Fuzzy Inference System based Approach

The authors in [21] developed a recognition system for ArSL alphabet gestures. It is an image-based system that does not rely on the use of sensors or colored gloves. After the acquisition, the images are pre-processed using a 3x3 median filter to reduce the noise. Then an iterative thresholding algorithm is applied in order to generate a binary image with black color as a background and white color for the hand region. The visual descriptors are extracted as in [14]. As described in Section III.A, in order to extract these visual descriptors, the hand direction, the coordinates of the hand area's centroid, and gesture boundary contour are computed. Since global boundary visual descriptors may not allow distinguishing alphabet with similar shapes, a hybrid approach based on both boundary and region information is used. The authors in [21] used k-means clustering technique to cluster the image into five regions. Then, the coordinates of hand centroid of each region are computed, and the distance between the global hand centroid and each region centroid is calculated. The length of the resulting visual descriptor vector is 35.

For the recognition stage, the authors build a fuzzy model for each class. Then an equivalent Adaptive Network-Based Fuzzy Inference System (ANFIS) [49] model is constructed and trained using a hybrid learning algorithm [51] which incorporates gradient descent method [49] and least-squares estimate [49] to estimate parameters values. The dataset used to experiment the system was collected using a camera connected to a computer.

A set of 1800 grayscale images for the 30 gestures was captured from different distances from the camera and different orientations. 1200 of the collected images were used as a training set while the other 600 were used as a testing set without cross-validation. The overall recognition rate of the system depends on the number of rules used in the ANFIS model. A 100% recognition rate was achieved when approximately 19 rules are used, and a 97.5% when approximately ten rules used. However, the authors in [21] are not using cross-validation. This may lead to the overfitting problem. In fact, while 100% accuracy is obtained for the data set when using 19 rules, the result can be different when using another dataset with a different number of rules.

### C. "ArSLAT: Arabic Sign Language Alphabets Translator"

The authors in [18] introduced an ArSL Alphabet Translator (ArSLAT) system. The proposed image-based system translates automatically hand gestures representing ArSL alphabet to text without using gloves or visual markers. ArSLAT system undergoes five phases, which are the pre-processing phase, followed by the best-frame detection phase, then the category detection phase, where the ArSL letters are categorized into three categories based on wrist direction to reduce processing time and increase accuracy of the system.

These three categories are wrist appearance from bottom-right, from bottom-left or the down-half. After the category detection phase comes the visual descriptor extraction phase and finally the classification phase.

In the visual descriptor extraction phase, the authors proposed to use visual descriptor vectors that are invariant with translation, scale, and rotation. Similarly as in [14] and in [21], in order to extract the visual descriptor vector. First, edge-detection is performed on all the images in the dataset. Then an orientation point is specified depending on the wrist's location. The visual descriptor vector is computed in such a way that each visual descriptor vector entry represents the distance between the orientation point and a point from the detected edge of the hand. Finally, the visual descriptor vectors are made scaling-invariant by dividing each visual descriptor element of the visual descriptor vector by the maximum value in that vector.

To recognize the alphabet, the system used two different classifiers, which are the minimum distance classifier and multilayer perceptron classifier. The minimum distance classifier (MDC) [52] classifies the visual descriptor vector of an unknown gesture image as the same class of the visual descriptor vector most similar to it from the training set. This similarity is computed based on the Euclidean distance between the two visual descriptor vectors. On the other hand, multilayer perceptron (MLP) classifier [53] is a Neural Network classifier. It learns and produces a model that determines the class of an unknown visual descriptor vector. It consists of a single input layer, a hidden layer, and a single output layer. The input layer is the input data, the hidden layer controls the classifier function, and the output layer returns the output. A dataset of 30 ArSL alphabets is collected. However, the authors limited the dataset to only 15 alphabets. As a result of experimenting only a subset of 15 letters, the accuracy of the system using MDC was 91.3%, while the accuracy of the system when using MLP classifier was 83.7%.

### *D. Fourier-based Approach*

The authors in [19] proposed an image-based system that recognizes ArSL alphabet. The proposed method doesn't require the signers to wear gloves or any other marker devices to ease the hand segmentation. The system performs image preprocessing which consists in size normalization and skin detection. The size of the images is normalized to 150×150. Then, to detect the skin, images are converted from RGB to HSV, and the pixels values within a specific range are considered as skin.

After skin segmentation, the Fourier transform [54] is applied to the hand region. Then, based on the frequency information provided by the Fourier transformation, the Fourier descriptor (FD) [55] is extracted. The classifier that has been used in [19] is k-Nearest Neighbors algorithm (KNN) [52]. A total number of 180 images have been collected from 30 persons. Only six letters are considered. These are "Sad", "Zay", "Kaf", "Ba", "Lam", "Ya". In order to train the model, the authors in [19], used all the 180 images. As a result, the proposed system achieved a recognition accuracy of 90.55%. However, the number of letters is very limited. Also, since all the data is used for training, this will yield an over fitting problem. Another limitation of this approach is the range of the colors that have been used in skin detection. In fact, this range which is not specified in the paper [19], is not straightforward to set. In fact, skin color differs from one person to another and from one region to another. The choice of the range of skin color can yield another over fitting problem. Moreover, the parameter k of the KNN classifier [52] have not been specified.

### *E. Scale-Invariant Visual Descriptors Transform based Approach*

The authors in [34] propose an ArSL recognition system. The stages of the proposed recognition system are visual descriptor extraction using SIFT technique [56], visual descriptor vector's dimension reduction using LDA [57], and finally classification. The system uses the Scale-Invariant Features Transform (SIFT) [56] as visual descriptors. The SIFT algorithm [56] is used for visual descriptors extraction for its robustness against rotation, scaling, shifting, and transformation of the image. The SIFT algorithm [56] takes an input image and transforms it into a collection of local visual descriptor vectors. It extracts the most informative points of a given image, called key points. Since, visual descriptor vectors produced by SIFT [56] have high dimensions, Linear Discriminant Analysis (LDA) [57] is used to reduce their dimensions. Three classifiers are used in [34]. They are Support Vector Machine (SVM) [58], k-Nearest Neighbor (k-NN) [52], and minimum distance classifier [52].

The dataset used in the experiments is collected in Suez Canal University [34]. It is an ArSL database which includes 210 gray level ArSL images. Each image is centered and cropped to the size of 200×200. Thirty Arabic characters (seven images for each character) are represented in the database. The results of this experiment show that applying SVM classifier [58] achieved a better accuracy than the minimum distance [52] and k-NN classifiers [52]. The system has achieved an accuracy around 98.9%. We should mention here that the SIFT parameters have been investigated empirically. Moreover, different portions of training and testing samples have been tried in order to determine the optimal portion. These two facts may lead to an over-fitting problem. Besides, the system needs to be tested on a large dataset to check its scalability.

### *F. Pulse-Coupled Neural Network based Approach*

The authors in [47] introduced a new approach for image signature using a Pulse-Coupled Neural Network (PCNN) [59], [60] for ArSL alphabet recognition.

The recognition system used in [47] includes four main steps. First, the gesture image is put through first layer PCNN [59], [60], where image smoothing is applied to reduce noise. Second, the smoothed image is put through second layer PCNN for a certain number of times to output the image signature, also known as the global activity, which represents the time series that differentiates between the contents of the image. Third, visual descriptor extraction and selection is performed on the image signature using Discrete Fourier Transform (DFT). DFT maximum coefficients represent the visual descriptor vector since they represent the most informative signal. Finally, the visual descriptor vectors are

classified using Multi-Layer Perceptron (MLP) network [53]. The pulse-coupled neural network (PCNN) [59], [60] is a single layer network of neurons, where each neuron is associated with a pixel in the input image [17]. The dataset includes images of 28 ArSL gestures. Eight images are collected for each gesture. The system reaches a recognition rate of 90% when the size of the visual descriptor is equal to 3. However, this system considered only 28 of the 30 letters within a really small dataset. Therefore, this approach needs to be tested over a large data to check scalability and over-fitting.

In summary, different visual descriptors had been used in literature for Sign language recognition. The approaches in [14], [21] and [18] used application-dedicated visual descriptors. In fact, visual descriptor is based on the hand orientation, the hand center, and edges have been designed. Other approaches like in [19], [34], and [47] used general visual descriptors like Fourier descriptor [19], [47] and SIFT descriptor [34]. We also noticed that these approaches need pre-processing steps such as image segmentation and edge detection. We should also mention that some approaches like in [19], [47], [18], and [14] used a subset of the ArSL alphabet. Others, like in [34], [21] used a small data. This is due to the difficulty to recognize and segregate ArSL alphabet. In fact, ArSL has the characteristic of having several gestures that are very similar to each other like "Dal" and "Thal", "Ra", and "Zay", etc. In the literature, no study investigated or compared visual descriptors for ArSL recognition. In this project, we aim to empirically investigate existing visual descriptors in order to determine an appropriate one that will allow us to build an effective ArSL recognition system.

## IV. Image based Arabic Sign Language Recognizer

Visual descriptors play a significant role in any image-based recognition system and drastically affect its performance. They are intended to encode the image's visual characteristics into one or more numerical vectors in order to convey the image semantic contents to the machine learning component. Nevertheless, determining the most appropriate descriptor for a recognition system remains an open research challenge.

As reported in Section II, some ArSL alphabet gestures exhibit high similarity. For instance, as shown in Fig. 7, the gestures corresponding to the pairs "Tah" and "Thah", and "Dal" and "Thal" look almost the same. This makes determining the visual descriptor that is able to discriminate between similar gestures even more challenging. The aim of this project is to find a visual descriptor that allows differentiating between different ArSL gestures.

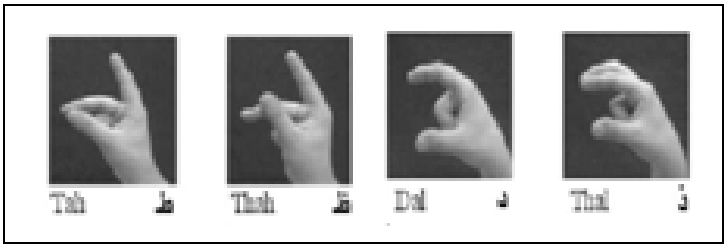


Fig. 7. Similar alphabet gesture example.

Various visual descriptors were proposed in the literature [61]-[63]. Namely, color, shape, and edge-based descriptors have been introduced. The color descriptors fail to extract relevant information for gesture recognition. In fact, the color of the hand, along with the background color, is irrelevant to the gesture characteristics. Moreover, the visual descriptors should not be sensitive to the color of the skin. Also, shape descriptors require prior processing before the extraction phase. Usually, the image needs to be segmented first in order to separate the region including the hand from the surrounding background. Moreover, for pairs of letters such as "Tah" and "Thah", and "Dal" and "Thal" shown in Fig. 7, the shape descriptor is not able to segregate between "Tah" and "Thah", or "Dal" and "Thal". This is because it does not yield information on the spatial position of the fingers. Thus, we do not intend to consider shape descriptors for our system.

On the other hand, texture descriptors [64] provide information on region homogeneity and the edges present in the image. In this paper, we investigate texture descriptors because they can capture ArSL gestures. More specifically, we intend to compare empirically five texture descriptors for the purpose of ArSL alphabet recognition. Namely, they are Histograms of Oriented Gradients (HOG) [66], Edge Histogram Descriptor (EHD) [65], Gray-Level Co-occurrence Matrix (GLCM) [61], Discrete Wavelet Texture Descriptor (DWT) [62], and Local Binary Pattern (LBP) [63].

The methodology of the proposed approach starts with extracting the visual descriptor i from the training images. Then, for each gesture, we build a model using one versus all SVM classifier [67]. In our case, we consider one class per ArSL alphabet gesture. This yields 30 classes. A model is learned for each gesture by training the classifier using one particular class against all the others.

The same ith descriptor is then extracted from the testing set of images. Using the 30 models built during the training phase, the testing alphabet is recognized. Finally, the performance of the recognition using the visual descriptor i is assessed using precision, recall, and accuracy.

This process is repeated for the five considered visual descriptors. Then, the results are compared to determine the most appropriate visual descriptor for ArSL alphabet recognition.

## V. Experiement

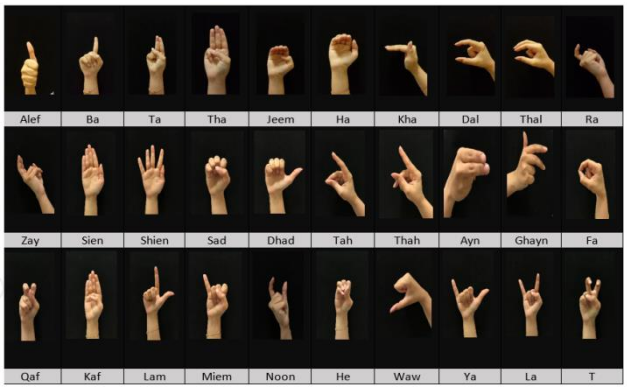


Fig. 8. A sample of ArSL alphabet.

The experiment was conducted using MATLAB. We captured the real images collection using different Smartphones and collected them with the help of 30 volunteers. Each volunteer gestured the 30 ArSL alphabets. Each alphabet is represented using a subset of 30 images from the original 900 photos. Fig. 8 shows a sample of ArSL images representing alphabet gestures. As it can be seen, all images have a uniform colored background. We proceeded with this choice in order to bypass the skin detection step, where we have to extract the hand region from the background before starting the gesture recognition phase.

First, we transform the 900 color images into gray level images. Then, we extract the five visual descriptors from the obtained images. These visual descriptors are used sequentially to represent the images and fed into the classifier individually. The recognition process will be conducted once for each visual descriptor. In order to avoid over-fitting, we set K to 10 for the K-fold cross validation training. We validate the discrimination power of each visual descriptor using the ground truth labels and the predicted categories obtained by the cross-validation.

When the HOG descriptor is provided as the input to the soft-margin SVM [44], the proposed ArSL system accuracy is 63.56 %. In order to further investigate this result, we display in Fig. 9 the per class performance. We notice that the performance varies from one letter to another. As can be seen, the letters Shien "ش" and Kha "خ" have an accuracy of 100%. However, Tha "ث" has an accuracy of 23.33 %. On the other hand, when using EHD descriptor [65], the proposed ArSL system accuracy is 42%. We display in Fig. 10 the per class performance. Similarly, we notice that the performance varies from one letter to another. As it can be seen, the letter Kha "خ" has an accuracy of 80%. However, Th "ت" has an accuracy of 6.67 %. Fig. 11 displays the per class performance when using LBP descriptor [63]. The proposed ArSL system accuracy is 9.78%. As can be seen, the letter Alef "أ" has an accuracy of 63.33%. However, many letters like Th "ت" and Ayn "ع" have an accuracy of 0 %. On the other hand, the proposed ArSL system accuracy is 8% when using DWT descriptor [62] as input to the one versus-all SVM [44]. In Fig. 12, we display the per class performance. As can be seen, the letter Dal "د" has an accuracy of 46.67%. However, many letters like Th "ت", Miem "م", and Ayn "ع" have an accuracy of 0 %. The worst accuracy result is obtained when using GLCM descriptor [61]. In fact, the proposed ArSL system accuracy is 2.89%. Fig. 13 displays the per class performance. We notice that the letter He "هـ" has an accuracy of 26.67%. However, many letters like Ta "ت", Miem "م", and Ayn "ع" have an accuracy of 0%.

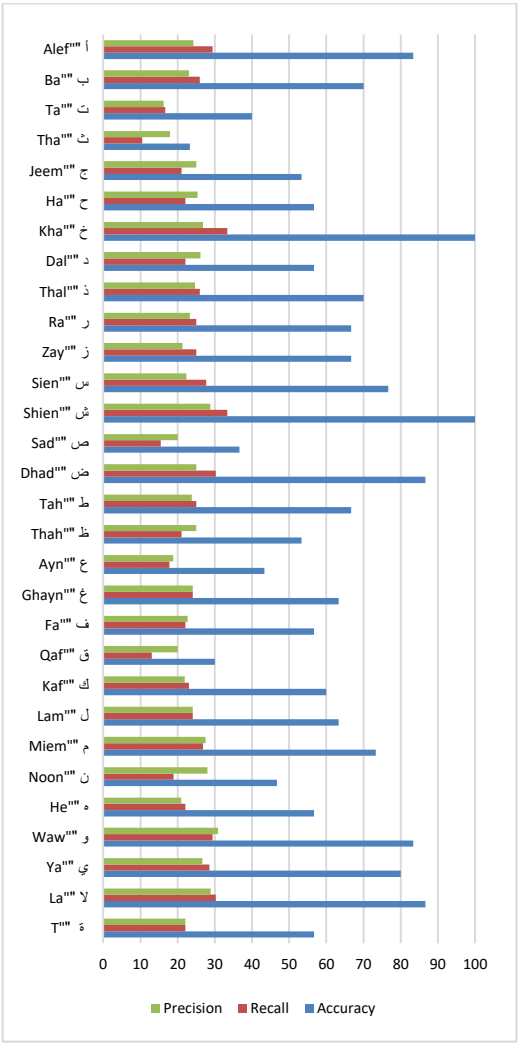


Fig. 9. The proposed ArSL system per class performance when using HOG descriptor.

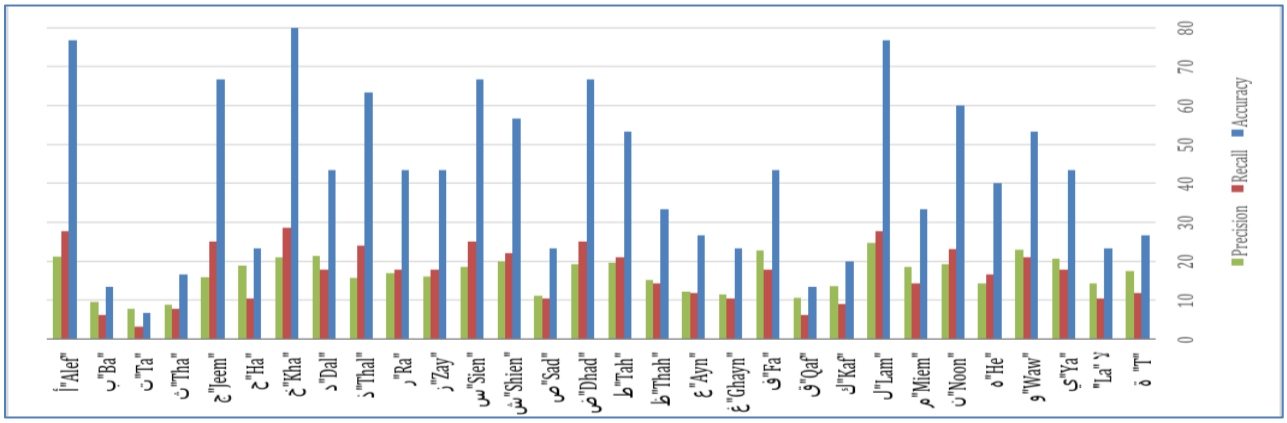


Fig. 10. The proposed ArSL system per class performance when using EHD descriptor.

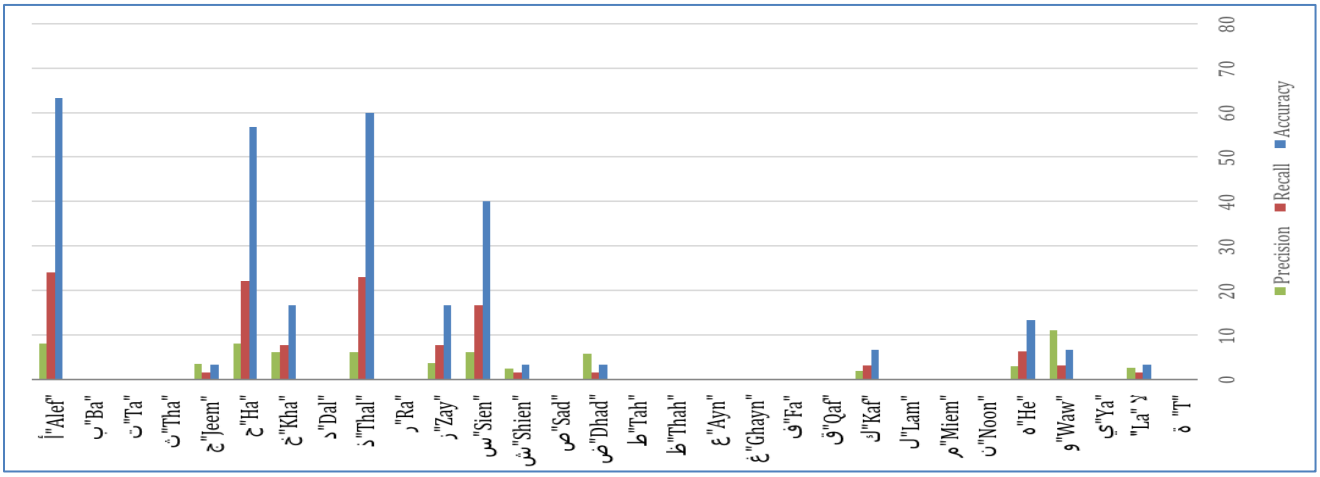


Fig. 11. The proposed ArSL system per class performance when using LBP descriptor.

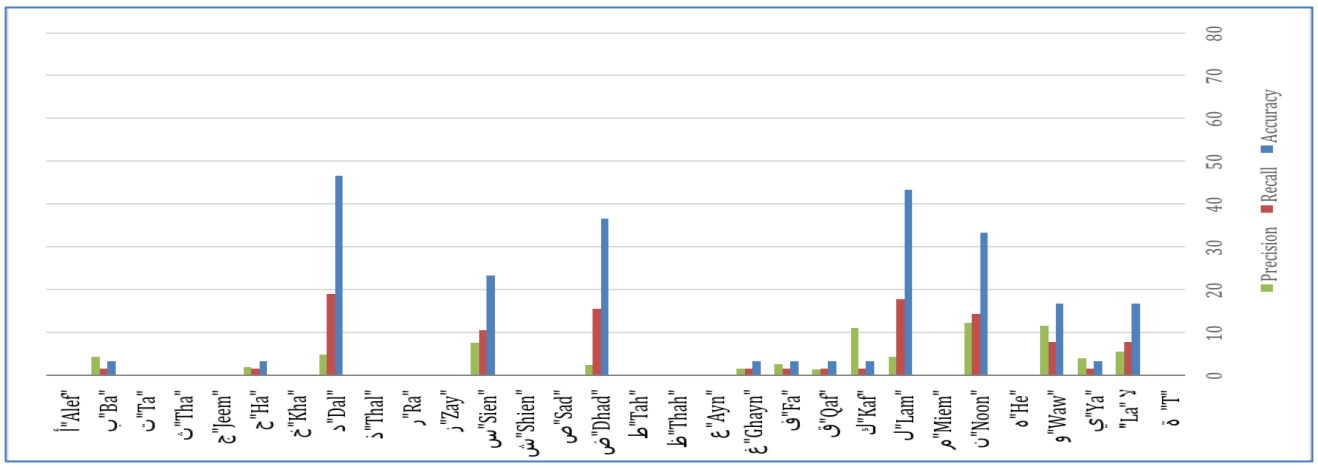


Fig. 12. The proposed ArSL system per class performance when using DWT descriptor.

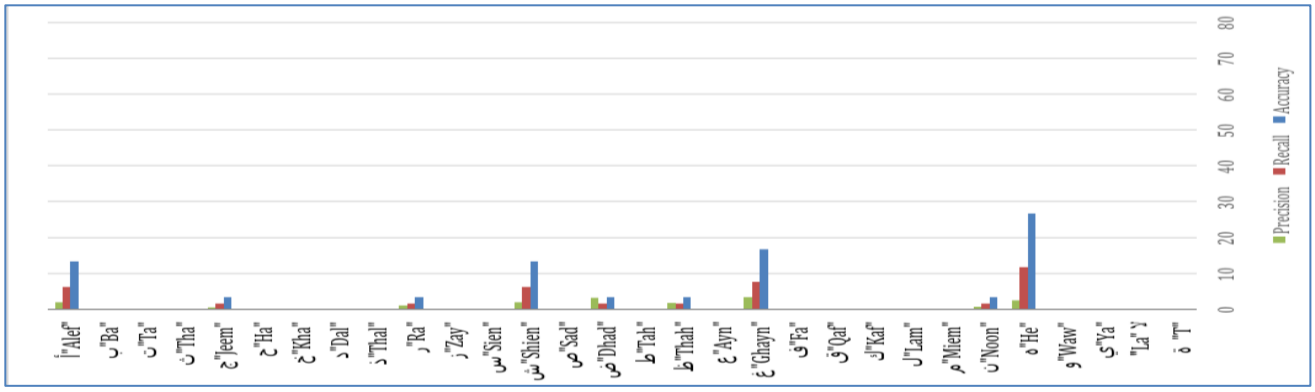


Fig. 13. The proposed ArSL system per class performance when using GLCM descriptor.

The results show that the HOG descriptor [66] achieved the highest performance followed by the EHD descriptor [65]. In fact, based on the achieved accuracy per letter (refer to Fig. 9 to 13), the HOG descriptor gives the highest accuracy for 27 letters. The three remaining letters which are ن "Noon", ل "Lam", and ج "Jeem" are best recognized when using EHD descriptor.

## VI. Conclusion and Future Works

A Sign language recognition system allows hearing and speech impaired people to communicate and facilitates their societal integration. ArSL is the official sign language for the Arab world. Despite its similarity to other international sign languages, ArSL includes alphabet representations that are exclusive to Arabic language. This makes non-ArSL sign

language recognition systems inapplicable to ArSL. ArSL recognition gets even more challenging due to the highly similar gestures representing Arabic letters. State-of-the-art ArSL recognition systems rely either on a sensor-based or an image-based approach to identify ArSL alphabet. However, image-based systems proved to be more effective because of their flexibility, portability, and friendly use. In fact, they can be deployed using smart devices incorporating digital camera(s), and could be used everywhere. After investigating existing visual descriptors for ArSL alphabet recognition systems, we proposed a new ArSL recognition system. The proposed system consists of extracting the HOG descriptor that is conveyed to a one versus all soft-margin SVM [58]. The resultant system succeeds in recognizing 63.5% of Arabic Alphabet gestures.

As future work, we intend to investigate kernel SVM [68] in order to further enhance the performance of the proposed system. In fact, Kernel SVM [68] allows mapping the features to a new space where it exhibits linear patterns. Thus, the ArSL gestures will be linearly separable in the new space. Besides, since some features achieve a better accuracy in recognizing certain letters than others, we intend to assign a relevance feature weight with respect to each gesture.

## REFERENCES

[1] C. Allen, "Global survey report WFD interim regional Secretariat for the Arab region (WFD RSAR) global education Pre-Planning project on the human rights of deaf people," 2008.

[2] M. F. Tolba and A. S. Elons, "Recent developments in sign language recognition systems," 2013 8th Int. Conf. Comput. Eng. Syst. ICCES, pp. 26–28, Nov. 2013.

[3] V. I. Pavlovic, R. Sharma, and T. S. Huang, "Visual interpretation of hand gestures for human-computer interaction: A review," IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 7, pp. 677–695, Jul. 1997.

[4] S. S. Fels and G. E. Hinton, "Glove-talk: A neural network interface between a data-glove and a speech synthesizer," IEEE Trans. Neural Netw., vol. 4, no. 1, pp. 2–8, 1993.

[5] D. J. Sturman and D. Zeltzer, "A survey of glove-based input," IEEE Comput. Graph. Appl., vol. 14, no. 1, pp. 30–39, Jan. 1994.

[6] D. L. Quam, "Gesture Recognition with a Dataglove," Proc. IEEE Natl. Aerosp. Electron. Conf. NAECON 90, vol. 2, pp. 755–760, May 1990.

[7] J. Eisenstein, S. Ghandeharizadeh, L. Huang, C. Shahabi, G. Shanbhag, and R. Zimmermann, "Analysis of clustering techniques to detect hand signs," Proc. 2001 Int. Symp. Intell. Multimed. Video Speech Process. ISIMP 2001 IEEE Cat No01EX489, pp. 259–262, May 2001.

[8] H. Brashear, V. Henderson, K. Park, H. Hamilton, S. Lee, and T. Starner, "American Sign Language Recognition in Game Development for Deaf Children," Oct. 2006.

[9] K. Mahesh, S. Mahishi, S. R, N. S, and V. Pujari, "Finger Detection for Sign Language Recognition," Mar. 2009.

[10] S.-K. Kang, K.-Y. Chung, K.-W. Rim, and J.-H. Lee, "Development of Real-Time Gesture Recognition System Using Visual Interaction" 2011.

[11] S. Ding, H. Zhu, W. Jia, and C. Su, "A survey on feature extraction for pattern recognition," Artif Intell Rev, vol. 37, pp. 169–180, 2012.

[12] B. Manjunath, P. Salembier, and T. Sikora, "Introduction to MPEG 7: Multi- media content description language," 2002.

[13] M. Nixon and A. Aguado, "Feature Extraction and Image Processing for Computer Vision," 2012.

[14] O. Al-Jarrah and A. Halawani, "Recognition of gestures in arabic sign language using neuro-fuzzy systems," Artif. Intell., vol. 133, no. 1–2, pp. 117–138, Dec. 2001.

[15] K. Assaleh and M. Al-Rousan, "Recognition of Arabic sign language alphabet using polynomial classifiers," vol. 2005, pp. 2136–2145, 2005.

[16] "Error detection and correction approach for arabic sign language recognition," 2012 Seventh Int. Conf. Comput. Eng. Syst. ICCES, pp. 117–123, Nov. 2012.

[17] M. F. Tolba, A. Samir, and M. Aboul-Ela, "Arabic sign language continuous sentences recognition using PCNN and graph matching," Neural Comput. Appl., vol. 23, no. 3–4, pp. 999–1010, Aug. 2012.

[18] N. El-Bendary, H. M. Zawbaa, M. S. Daoud, A. E. Hassanien, and K. Nakamatsu, "ArSLAT: Arabic sign language Alphabets translator," 2010 Int. Conf. Comput. Inf. Syst. Ind. Manag. Appl. CISIM, 2010.

[19] N. R. Albelwi and Y. M. Alginahi, "Real-time arabic sign language (arsl) recognition," 2012, pp. 497–501.

[20] M. F. Tolba, A. Samir, and M. Abul-Ela, "A proposed graph matching technique for Arabic sign language continuous sentences recognition," 8th IEEE Int. Conf. Inform. Syst. INFOS, pp. 14–20, 2012.

[21] O. Al-Jarrah and F. A. Al-Omari, "IMPROVING GESTURE RECOGNITION IN THE ARABIC SIGN LANGUAGE USING TEXTURE ANALYSIS," Appl. Artif. Intell., vol. 21, no. 1, pp. 11–33, Jan. 2007.

[22] A. A.A, A. Elsayed, and H. Hamdy, "Arabic sign language (ArSL) recognition system using HMM," Int. J. Adv. Comput. Sci. Appl., vol. 2, no. 11, 2011.

[23] M. Mohandes, M. Deriche, and J. Liu, "Image-based and sensor-based approaches to Arabic sign language recognition," vol. 44, no. 4, pp. 551–557, 2014.

[24] "Sign language - definition," Oxford Dictionaries. [Online]. Available: en.oxforddictionaries.com/definition/sign_language. [Accessed: 04-Oct-2016].

[25] J. D. Schein and D. A. Stewart, Language in motion: Exploring the nature of sign. Gallaudet University Press, 1995.

[26] N. Cleaver, "WASLI country reports issue 1-November 2005," presented at the WASLI 2007, Spain, 2007, p. 2.

[27] M. A. Abdel-Fattah, "Arabic sign language: A perspective," J. Deaf Stud. Deaf Educ., vol. 10, no. 2, pp. 212–221, 2005.

[28] Cambridge University, Sign Languages. Cambridge University Press, 2010.

[29] A. Agarwal and M. K. Thakur, "Sign language recognition using Microsoft Kinect," presented at the 2013 Sixth International Conference on Contemporary Computing (IC3), 2013.

[30] A. K. Sahoo, S. Gouri, K. Mishra, and R. Kumar, "SIGN LANGUAGE RECOGNITION: STATE OF THE ART," vol. 9, no. 2, 2014.

[31] T. Starner, J. Weaver, and A. Pentland, "Real-time American sign language recognition using desk and wearable computer based video" IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 12, pp. 1371–1375, 1998.

[32] B. Bauer and H. Hienz, "Relevant Features for Video-Based Continuous Sign Language Recognition," presented at the Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000, pp. 64–75.

[33] N. Tubaiz, T. Shanableh, and K. Assaleh, "Glove-based continuous arabic sign language recognition in user-dependent mode," IEEE Trans. Hum.-Mach. Syst., vol. 45, no. 4, pp. 526–533, Aug. 2015.

[34] A. Tharwat, T. Gaber, A. E. Hassenian, M. K. Shahin, and B. Refaat, "SIFT-Based Arabic Sign Language Recognition System," vol. 334, Nov. 2014.

[35] J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with Microsoft Kinect sensor: A review," IEEE Trans. Cybern., vol. 43, no. 5, pp. 1318–1334, Oct. 2013.

[36] M. AL-Rousan, K. Assaleh, and A. Tala'a, "Video-based signer-independent arabic sign language recognition using hidden Markov models," Appl. Soft Comput., vol. 9, no. 3, pp. 990–999, Jun. 2009.

[37] T. Shanableh, K. Assaleh, and M. Al-Rousan, "Spatio-Temporal feature-extraction techniques for isolated gesture recognition in arabic sign language," IEEE Trans. Syst. Man Cybern. Part B Cybern., vol. 37, no. 3, pp. 641–650, Jun. 2007.

[38] S. Kausar and Y. Javed, "A Survey on Sign Language Recognition," 2011.

[39] P. Dymarski, "Hidden Markov models, theory and applications," 19-Apr-2011.[Online].Available:http://www.intechopen.com/books/hidden-markov-models-theory-and-applications. [Accessed: 18-Nov-2016].

[40] T. Nobuhiko, S. Nobutaka, and S. Yoshiaki, "Extraction of Hand Features for Recognition of Sign Language Words."

[41] H. C. M. Herath, W. A. L. V. Kumari, W. A. P. . Senevirathne, and M. . Dissanayake, "IMAGE BASED SIGN LANGUAGE RECOGNITION SYSTEM FOR SINHALA SIGN LANGUAGE," SAITM Res. Symp. Eng. Adv., pp. 107–110, 2013.

[42] J. Zieren and K.-F. Kraiss, "NON-INTRUSIVE SIGN LANGUAGE RECOGNITION FOR HUMAN-COMPUTER INTERACTION," 2000.

[43] Q. Yang, "Chinese sign language recognition based on video sequence appearance modeling," 2010, pp. 1537–1542.

[44] C. Cortes and V. Vapnik, "Support-vector networks," vol. 20, no. 3, pp. 273–297, 1995.

[45] S. Liwicki and M. Everingham, "Automatic recognition of fingerspelled words in British sign language," 2009, pp. 50–57.

[46] N. V. Tavari and A. Deorankar, "Indian Sign Language Recognition based on Histograms of Oriented Gradient," vol. 5, 2014.

[47] M. Fahmy Tolba, M. Saied Abdel-Wahab, M. Aboul-Ela, and A. Samir, "Image signature improving by PCNN for arabic sign language recognition," vol. Volume 1, no. 1, pp. 1–6, Jan. 2010.

[48] R. C. Jain, R. Kasturi, and B. G. Schunck, Machine vision. New York: McGraw Hill Higher Education, 1995.

[49] J.-S. R. Jang, "ANFIS:Adaptive-network-based fuzzy inference system," IEEE Trans. Syst. Man Cybern., vol. 23, no. 3, pp. 665–685, 1993.

[50] A. Priyono, M. Ridwan, A. J. Alias, R. A. O. K. Rahmat, A. Hassan, and M. A. Mohd. Ali, "Generation of fuzzy rules with Subtractive clustering," J. Teknol., vol. 43, no. 1, 2005.

[51] D. . Loganathan and K. . Girija, "Hybrid Learning For Adaptive Neuro Fuzzy Inference System," Res. Inven. Int. J. Eng. Sci., vol. 2, no. 11, 2013.

[52] B. Thuraisingham, M. Awad, L. Wang, M. Awad, and L. Khan, Design and implementation of data mining tools. Boca Raton, FL: Auerbach Publishers, 2009.

[53] M. Kantardzic, Data mining: Concepts, Models, Methods, and Algorithms, 2nd ed. Oxford: Wiley-Blackwell (an imprint of John Wiley & Sons Ltd), 2011.

[54] S. Bochner and K. Chandrasekharan, Fourier Transforms.(AM-19), vol. 19. Princeton University Press, 2016.

[55] R. L. Cosgriff, Identification of shape. 1960.

[56] D. G. Lowe, "Distinctive image features from scale-invariant Keypoints," Int. J. Comput. Vis., vol. 60, no. 2, pp. 91–110, Jan. 2004.

[57] J. Lepš, P. Šmilauer, J. Leps, and P. Smilauer, Multivariate analysis of ecological data using CANOCO. Cambridge: Cambridge University Press, 2003.

[58] S. Abe and A. Shigeo, Support vector machines for pattern classification, 2nd ed. London: Springer-Verlag New York, 2010.

[59] R. Eckhorn, H. J. Reitboeck, M. Arndt, and P. Dicke, "Feature linking via Synchronization among distributed assemblies: Simulations of results from cat visual cortex," Neural Comput., vol. 2, no. 3, pp. 293–307, Sep. 1990.

[60] R. C. Mureşan, "Pattern recognition using pulse-coupled neural networks and discrete Fourier transforms," Neurocomputing, vol. 51, pp. 487–493, Apr. 2003.

[61] M. Partio, B. Cramariuc, M. Gabbouj, and A. Visa, "ROCK TEXTURE RETRIEVAL USING GRAY LEVEL CO-OCCURRENCE MATRIX," Proceeding 5th Nord. Signal Process. Symp., 2002.

[62] D. F. Long, D. H. Zhang, and P. D. Dagan Feng, "Fundamentals of Content-Based Image Retrieval," in Multimedia Information Retrieval and Management, D. Dagan Feng, W.-C. Siu, and H.-J. Zhang, Eds. pp. 1–26.

[63] M. Pietikäinen, A. Hadid, G. Zhao, and T. Ahonen, "Local Binary Patterns for Still Images," in Computer Vision Using Local Binary Patterns, pp. 13–47.

[64] T. Sikora, "The MPEG-7 Visual Standard for Content Description--An overview," IEEE Trans. CITCUITS Syst. VIDEO Technol., vol. 11, no. 6, pp. 696–702, Jun. 2001.

[65] Y. M. R. Ro, M. K. Kim, H. K. K. Kang, B. S. M. Manjunath, and J. K. Kim, "MPEG-7 homogeneous texture Descriptor," ETRI J., vol. 23, no. 2, pp. 41–51, Jun. 2001.

[66] F. Suard, A. Rakotomamonjy, A. Bensrhair, and A. Broggi, "Pedestrian Detection using Infrared images and Histograms of Oriented Gradients," presented at the Intelligent Vehicles Symposium, 2006.

[67] A. Statnikov, C. F. Aliferis, D. P. Hardin, and er Statnikov, A gentle introduction to support vector machines in Biomedicine: Volume 1: Theory and methods. Singapore, Singapore: World Scientific Publishing Co Pte, 2011.

[68] B. Schölkopf, C. J. C. Burges, and A. J. Smola, Advances in Kernel Methods: Support Vector Learning. Cambridge, MA: MIT Press, 1999.