

Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

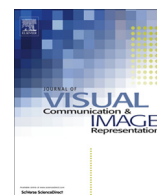
In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/authorsrights>



Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: [www.elsevier.com/locate/jvci](http://www.elsevier.com/locate/jvci)

# User-independent system for sign language finger spelling recognition



Djamila Dahmani, Slimane Larabi\*

Computer Science Department, University of Science and Technology Houari, Boumediene, Algiers, Algeria

## ARTICLE INFO

### Article history:

Received 14 May 2013

Accepted 28 December 2013

Available online 9 January 2014

### Keywords:

Hand posture

Shape recognition

Hand segmentation

Tchebichef moments

Hu moments

Sign language

Recognition

Classification

## ABSTRACT

We propose in this paper a framework for recognizing the sign language alphabet. To separate hand images from complex backgrounds, we use skin colour and texture attributes with neural networks. The recognition process is based on the combination of three shape descriptors: Discrete orthogonal Tchebichef moments applied on both internal and external outlines hand, Hu moments and a set of geometric features derived from the convex hull that encloses the hand shape taking into account the hand orientation.

The recognition is carried out using KNN and SVM classifiers. The proposed descriptors are combined in several sequential and parallel manners and applied on different datasets. The obtained results are compared to existing works.

© 2014 Elsevier Inc. All rights reserved.

## 1. Introduction

Sign consists of three main parts: Manual features involving gestures made with hands (employing hand shape and motion to convey meaning), Non-manual features such as facial expressions or body posture, which can both form part of a sign or modify the meaning of a manual sign, and finger spelling, where words are spelt out gesturally in the local verbal language [10].

To represent a hand gesture, four components are considered: The shape orientation, position and movement. Generally, hand gestures can be classified into two types: hand postures, which refer to the hand shape and orientation, and dynamic hand gestures, which are described in terms of the hand movements and position. Therefore, hand postures recognition is among the most important aspects of gesture recognition.

In this paper we propose a new hand-posture recognition system for sign language devoted to Arabic sign language alphabet. This system is user independent which does not use gloves or any other instruments or devices and operates in complex backgrounds. Hand segmentation is performed using skin color-texture attributes and neural networks. The system combines efficiently three appearance-based features selected to provide information regarding the hand shape and orientation. Discrete orthogonal Tchebichef moments applied on both internal and external outlines hand, Hu moments and a set of geometric features derived from the convex hull that encloses the hand shape.

The proposed geometric features take into account the hand orientation. The recognition is carried out using the KNN and SVM classifiers. The KNN classifier in spite of its simplicity, is shown to be successful when each class has many prototypes, even with a small number of training samples and then can deal with the user-independent recognition task. The SVM classifier outperforms the KNN classifier when the number of training samples increases.

Conducted experiments are based on two separate datasets. The first dataset consists of 30 hand signs from the Arabic sign language alphabet performed by 24 different subjects against simple background and 8 other subjects against complex backgrounds. The second is Jochen-Triesch static hand posture database. The system obtains good recognition rates for both datasets.

Section 2 is devoted to related works. Section 3 addresses the first task which consists to locate the hand in an image. In Section 4, we present some details concerning the extracted features from hand. The recognition process is presented in Section 5. The final sections are devoted to the experimental results and to a conclusion and future works.

## 2. Related works

Research on automatic sign language recognition (SLR) began approximately twenty years ago, particularly with American [37], Australian [25], and Korean sign languages [28]. Since then, many systems have been developed for Arabic [1,3], British [8], Chinese [16], French [4] and German [7,12].

Several user-independent hand posture recognition systems have been developed using many features:

\* Corresponding author. Fax: +213 21 24 76 07.

E-mail address: [slarabi@usthb.dz](mailto:slarabi@usthb.dz) (S. Larabi).

- Elastic graph matching, a recognition rate of 92.9% is obtained when classifying 10 hand postures [39],
- Modified census transform, a recognition rate of 89.9% was achieved using the Treisch database [24],
- Weighted eigenspace size functions, achieved a recognition rate of 85.1% for the Treisch data-set using the Triesch protocol [27],
- Hidden Markov models for signer-independent system of ArSL, recognizes a set of 30 isolated words with 90.6% accuracy [3].
- Weighted combination of geometric features and appearance-based features such as the area, gravity center [6], outline hand length and downscaled intensity directly extracted from images in the video frames [42].

The Arabic sign language (ArSL) alphabet has not received sufficient attention and only few research studies have been conducted [1,2,5,13,38]. All of them were user dependent. Al-Jarrah and Halawani used contour-based features to recognize 30 signs from Arabic sign language with a reported accuracy of 93.55%. Al-Roussan and Hussain, developed a system to recognise isolated signs for 28 letters from ArSL using coloured gloves for data collection and adaptive neuro-fuzzy inference systems (ANTFS) to achieve a recognition rate of 88%. In a similar work the recognition rate was increased to 93.41% using polynomial networks [5]. Tolba et al. [38] proposed a recognition system based on PCNN image signatures. This system achieved a recognition rate of 90.4%, but no database was specified. Later, Elons et al. increased the recognition rate to 92% by adding a weighting factor [13].

Moments are useful tools allowing the description of shape characteristics. They have been commonly used for pattern recognition.

For static gesture signs recognition, Gu and Su [18] used the Zernike and pseudo-Zernike moment features introduced by Chang et al. [9] which are invariant to rotation. More recently, for view-independent gesture classification, geometric and orthogonal moments were evaluated and the results indicated that Krawtchouk and Tchebichef moments are more robust in achieving view-point and user-independent gesture recognition [34].

Although the Tchebichef moments have been successfully used in pattern recognition and image reconstruction, they have not been used in sign language recognition. In addition Tchebichef moments have been computed using the binary or grey-scale images.

We investigated in this paper what can be inferred from moments and some geometric features applied on the outline and internal contours of the hand for the recognition of hand posture.

The main contributions of our method are:

- The first one, we propose a new approach for hand segmentation based on the skin colour-texture characteristics and Multi-Layered Perceptron (MLP) NN. This method is able to accurately characterize skin region and segment hand images under different lighting conditions from complex backgrounds.
- The second one, unlike to state of the art methods we use the external and internal edges of the hand shape. The Tchebichef moments are computed in our system from the internal and external hand edges. We show that hand shape representation using internal and external edges leads to better hand posture recognition.
- The third one concerns the proposition of geometric features that give information about finger tips and take into account the hand orientation. A large variation in orientation may give a posture a different meaning in sign language. In our case, we use the normalized farthest and closest distances from the centre of gravity of the hand shape.
- The last one concerns the combination of the three shape descriptors. This combination is motivated by the expectation that classification errors can be reduced if a set of classifiers

rather a single classifier is used for a given task [29]. We will show that the shape descriptors used in this paper give complementary informations about the hand shape.

### 3. Hand localisation in image

We assume that coloured images of hand are taken against complex backgrounds. Our aim is to locate the hand region in the image. Existing skin color segmentation techniques are not effective when the illumination conditions vary rapidly or when background colors are near to those of wood or red colors.

However, Ilea et al. found that a mixture of the colour and texture attributes allows describing the image content accurately [23]. We propose in this paper a new skin segmentation approach based on the skin colour-texture attributes, and MLP neural network for hand segmentation.

#### 3.1. Skin color modeling

The colour space used is one of the most important component in color segmentation techniques. The orthogonal colour spaces ( $Y C_b C_r$ ,  $YIQ$ ,  $YUV$ ,  $YES$ ) reduce the redundancy present in RGB color channels and represent the color with statistically independent components [26].  $Y C_b C_r$  which represents the colour as luminance  $Y$  and  $C_b, C_r$  the blue and red chrominance is chosen.

#### 3.2. Skin texture modeling

Texture is an important feature used for identifying regions of interest in image. Grey-Level Co-occurrences Matrices (GLCM) is one of the most widely used approaches for the texture feature extraction and analysis [19]. Each value  $x$  at the coordinate  $(i, j)$  in a Co-occurrence matrix  $P_{d,\theta}$  represents the frequency of the grey levels  $i$  and  $j$  separated by a given distance  $d$  and in a given direction  $\theta$ . Formally for an  $N \times M$  image  $f$  the normalized GLCM  $P_{d,\theta}$  is given by:

$$P_{d,\theta}(i, j) = \frac{\text{Card}\{(n, m) : f(n, m) = i, f(n + d\cos\theta, m + d\sin\theta) = j\}}{N \times M} \quad (1)$$

where Card is the cardinality of the set. The taken values of GLCM directions of analysis are: horizontal  $\theta = 0^\circ$ , vertical ( $\theta = 90^\circ$ ), right diagonal ( $\theta = 45^\circ$ ) and left diagonal ( $\theta = 135^\circ$ ).

Fourteen features were extracted by Haralick et al. [19] from the GLCM to characterize texture. In this paper we select four discriminate features to skin-texture detection:

$$\text{Energy} = \sum_i \sum_j P_{d,\theta}^2(i, j) \quad (2)$$

$$\text{Contrast} = \sum_i \sum_j (i - j)^2 P_{d,\theta}^2(i, j) \quad (3)$$

$$\text{Correlation} = \sum_i \sum_j (i - \mu_i)(j - \mu_j) P_{d,\theta}^2(i, j) / (\sigma_i \sigma_j) \quad (4)$$

where  $\mu_i, \mu_j$  are the means and  $\sigma_i, \sigma_j$  are the standard deviations of the partial probability density functions, obtained by summing up the rows or the columns of the matrix  $P_{d,\theta}$  respectively.

$$\text{LocalHomogeneity} = \sum_i \sum_j P_{d,\theta} / (1 + (i - j)^2) \quad (5)$$

The skin-color features  $Y, C_b, C_r$  and skin-texture features: Energy, Contrast, Correlation and Local Homogeneity are then used to train a Multi-Layered Perceptron (MLP) for skin pixel classification.

### 3.3. Skin-pixel classification

MLPs have been successfully applied in many pattern recognition areas due to their ability to learn input–output relationships. In this work, the proposed MLP is trained with 515520 pixels. The performance was tested using the Mean-Squared Error (MSE) which is defined as the average squared error between the network outputs and target outputs. After exploring the algorithm for different pairs  $(d, \theta)$  for calculating the GLCM, we found that the values  $d = 1$  and  $\theta = 0$  are the most suitable for our hand segmentation task. After running many simulations we have chosen the MLP NN with 7 input nodes, one hidden layer with 6 neurons and one output neuron which contain the value 1 if the pixel is skin-pixel and 0 otherwise.

There are some images where non-skin pixels are falsely detected due to the effect of lighting conditions. We assume that the hand is the largest skin-object in the image, and we find the connected components to identify the region of the hand. Fig. 1 shows some segmentation results where the hand region is separated from complex backgrounds under different light conditions including a wood background.

## 4. Features extraction from hand

The proposed method for hand posture recognition is based on features extracted from hand in image and their use for sign recognition using the KNN and SVM classifiers. Three shape descriptors are extracted and combined: Tchebichef moments, Hu moments and geometric features.

### 4.1. The Tchebichef orthogonal moments

#### 4.1.1. A summary

Moments are scalar quantities used to characterize a function and to describe its significant features. Mathematically, moments are “projections” of function onto polynomial basis [15] and are able to represent global features of image and widely used in image processing and pattern recognition.

The orthogonal moments are moments to an orthogonal polynomial basis. There is a group of orthogonal polynomials defined directly on series of points and therefore are especially suitable for digital images [15]. Among those moments, the discrete Tchebichef orthogonal moments introduced by Mukundan et al. [32] are good signal descriptors and sufficient to provide discriminate power for pattern recognition [41].

The discrete Tchebichef moments  $T_{pq}$  of the order  $(p + q)$  in an  $N \times N$  discrete-space image are defined as follows:

$$T_{pq} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} t'_p(x) t'_q(y) f(x, y) \quad (6)$$

where  $t'_p(x), t'_q(y)$  are the normalised Tchebichef polynomials defined as:

$$t'_p(x) = t_p(x) / \sqrt{\rho(p, N)} \quad (7)$$

$$t'_q(x) = t_q(x) / \sqrt{\rho(q, N)} \quad (8)$$

in which:

$$\rho(p, N) = N(1 - 1/N^2) \dots (1 - p^2/N^2) / (2p + 1) \quad (9)$$

The discrete Tchebichef polynomials are defined as follows [8]:

$$t_p(x) = (1 - N) {}_3F_2(-p, -x, 1 + p, 1 - N, 1) \quad (10)$$

in which  ${}_3F_2$  is the hyper geometric function defined as:

$${}_3F_2(a_1, a_2, a_3, b_1, b_2, z) = \sum_{v=0}^{\infty} ((a_1)_v (a_2)_v (a_3)_v z^v) / (b_1)_v (b_2)_v v! \quad (11)$$

and  $a_v$  is the Pochhammer symbol given by the following relationship:

$$a_v = a(a + 1)(a + 2) \dots (a + v - 1) \quad (12)$$

#### 4.1.2. Usefulness for hand sign recognition

Many algorithms have been proposed for the computation of moments in images using either binary images, boundary of binary object or grey-level images [9,18,22,27,34]. However, in the all proposed works, the internal edges have not been considered yet in spite of its usefulness to avoid ambiguity of sign recognition due to the added information provided. Indeed, some letter signs have similar external contours but their internal contours are different (see Fig. 2).

Furthermore, it was shown that Tchebichef moments are quite stable compared to the other discrete orthogonal moments and perform the best for the rougher images [36]. This motivated the computation of Tchebichef moments of the internal and external edges of the hand shape. To compute the moments using both contours, the function  $f(x, y)$  in the formula (6) takes values 1 for each pixel of the internal or external contours and the value 0 otherwise.

### 4.2. The Hu moments

#### 4.2.1. A summary

The Hu moments are a reformulation of the non-orthogonal centralised moments and are invariant to translation, scale change



Fig. 1. Some hand segmentation results using ArSL postures.

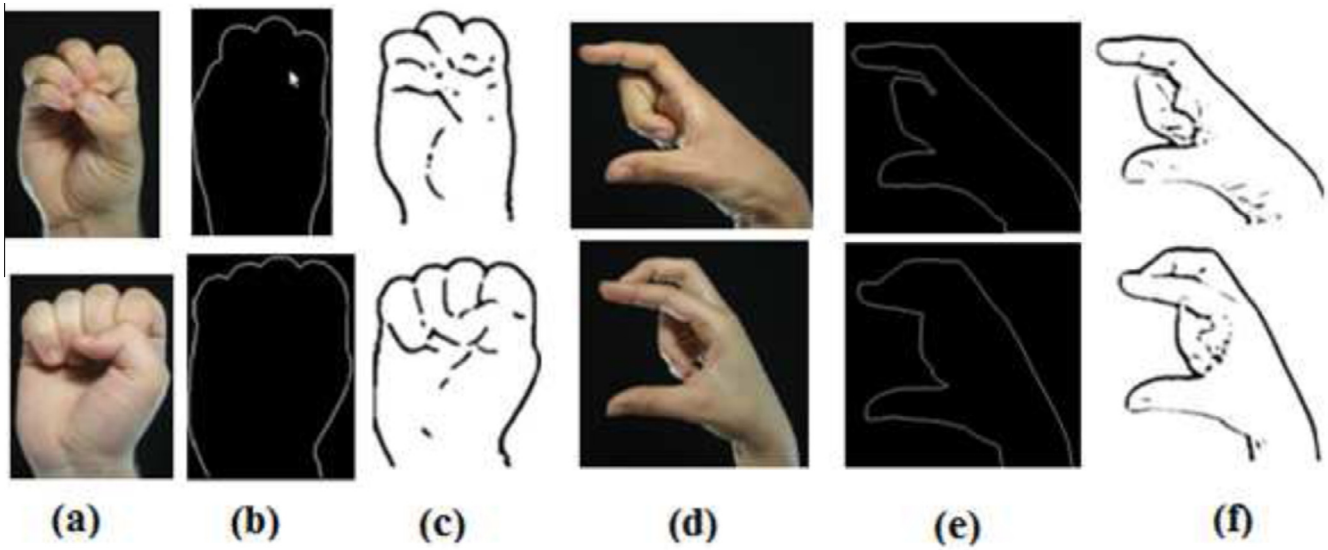


Fig. 2. (a,d) Images of gestures (He and Sad, Dal and Dhal) in Arabic sign language; (b,e) Outline contour; (c,f) Internal contours.

and rotation [21]. They are derived from the 2D moments and central geometric moments, using the algebraic invariants and defined as follows:

$$\begin{cases} \phi_1 = \mu_{2,0} + \mu_{0,2} \\ \phi_2 = \mu_{2,0} - \mu_{0,2} + 4\mu_{1,1}^2 \\ \phi_3 = (\mu_{3,0} - 3\mu_{1,2})^2 + (3\mu_{2,1} - \mu_{3,0})^2 \\ \phi_4 = (\mu_{3,0} + \mu_{1,2})^2 + (\mu_{2,1} + \mu_{0,3})^2 \\ \phi_5 = (\mu_{3,0} - 3\mu_{1,2})(\mu_{3,0} + \mu_{1,2})[(\mu_{1,2} + \mu_{3,0})^2 - 3(\mu_{2,1} + \mu_{0,3})^2] \\ \quad + (3\mu_{2,1} - \mu_{0,3})(\mu_{2,1} + \mu_{0,3})[3(\mu_{3,0} + \mu_{1,2})^2 - (\mu_{0,3} + \mu_{2,1}^2)] \\ \phi_6 = (\mu_{2,0} - \mu_{0,2})[(\mu_{3,0} + \mu_{1,2})^2 - (\mu_{2,1} + \mu_{0,3}^2)] \\ \quad + 4\mu_{1,1}(\mu_{3,0} + \mu_{1,2})(\mu_{0,3} + \mu_{2,1}) \\ \phi_7 = (3\mu_{2,1} - \mu_{0,3})(\mu_{3,0} + \mu_{1,2})[(\mu_{1,2} + \mu_{3,0}^2) - 3(\mu_{2,1} + \mu_{0,3})] \\ \quad - (\mu_{3,0} - 3\mu_{1,2})(\mu_{2,1} + \mu_{0,3})[3(\mu_{3,0} + \mu_{1,2}^2) - (\mu_{2,1} + \mu_{0,3})^2] \end{cases} \quad (13)$$

where  $\mu_{p,q}$  is the normalized central moment.

The distance between Hu moment features is calculated as the sum of absolute differences between augmented moments  $\wedge_{H_k}(i)$  (see Eqs. (14) and (15)).

$$\wedge_{H_k}(i) = 1/(\text{sign}(H_k(i)) \times \log(H_k(i))) \quad (14)$$

$$D^{Hu}(H_k, H_l) = \sum_{i=1}^{i=7} |\wedge_{H_k}(i) - \wedge_{H_l}(i)| \quad (15)$$

#### 4.2.2. Usefulness for hand sign recognition

Hu moments have been largely used for object recognition and have a greater discrimination when classifying shapes with small contour variation [27]. We use these moments in order to distinguish between the signs meaning different letters and which have similar hand shapes such as the letters *Jiem* and *Ha* from the ArSL alphabet (see Fig. 3).

#### 4.3. The geometric features

Hand configuration and orientation are computed using three geometric features extracted based on the convex hull enclosing

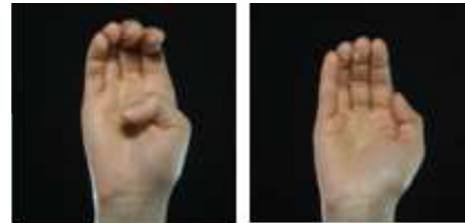


Fig. 3. Contour variation between the sign letters *Jiem* and *Ha* from ArSL alphabet.

the hand's shape: the relative area and the minimum and maximum relative distances.

##### 4.3.1. The relative area

The area of the hand has been already used as a geometric feature for sign language recognition (SLR) [6,42]. The greatest distances from the gravity center has been used to determine the hand configuration [31]. The relative area which is computed as the ratio between the area of the hand's shape and the area of the convex hull enclosing it [17] is invariant feature to scale-change.

##### 4.3.2. The minimum and maximum relative distances

The minimum and maximum relative distances are related to the farthest and closest vertex of the convex hull encompassing hand shape to the gravity center of the convex hull.

To distinguish similar hand configurations with different orientations, these distances are normalised using the height of the bounding box that encloses the hand shape. This height takes a value between the width and the length of the hand shape according to the hand orientation (see the example presented in Fig. 4).

For similar finger configurations (the Arabic sign language letters *Ayn* and *Ta* in the example), we obtain similar values for the geometric features. The height of the bounding box is not the same for the two letters which allows their distinction.

### 5. The recognition process

Assuming that:

- Hand's region in image is located using the proposed skin segmentation approach based on the skin color-texture attributes and MLP neural network,

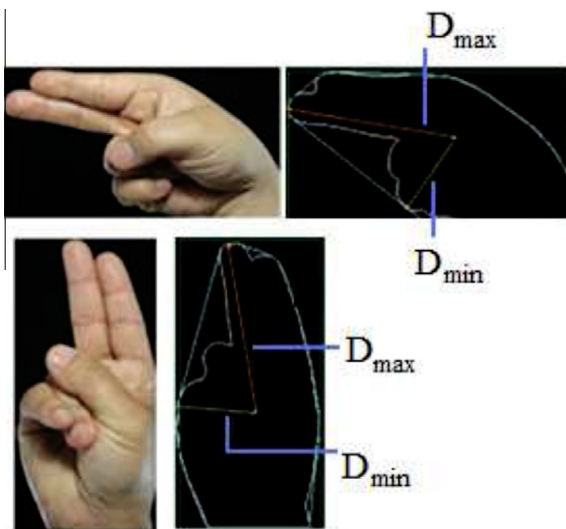


Fig. 4. Example of the Ayn and Ta hand postures from the ArSL with their convex hull in blue, their bounding box in green and the corresponding geometric features.

- Outline and internal edges of hand's shape in image are located and the three features are determined.

Hand postures classification of each feature is performed using the KNN algorithm for  $K = 3$ . The Euclidian distance is used to measure the similarity of the Tchebichef moments and geometric features, and the distance  $D^{Hu}$  previously defined is used for the Hu moments feature.

The choice of KNN classifier is motivated by the fact that each hand posture used in sign language as handwritten letter is performed naturally by different persons with different anatomies and thus must have different prototypes. Hastie et al. [20] demonstrated that, despite its simplicity, the KNN classifier is often successful when each class has possible prototypes and the decision boundary is very irregular.

To evaluate the best possible combination of the three previously defined features, several combinations are tested. We used both sequential and parallel approaches. Each pair of descriptors are combined in a sequential way using the two possible orders. The three descriptors are combined in sequential and parallel manners. Finally, SVM classifier which is successfully used for pattern and hand gesture recognition tasks [27,11,35] is also tested with the best combination scheme obtained in previous step. The comparison between the performance of both KNN and SVM classifiers is carried out.

## 6. Experiments

### 6.1. Data collection

To assess the efficiency and capabilities of the proposed features, we conducted experiments on two different databases: Arabic sign language database and Jochen-Triesch Static hand postures Database.

#### 6.1.1. Arabic sign language database

Two datasets based on the Arabic sign language alphabet exist as benchmarks:

- The dataset of Al-Jarrah and Halawani [1] does not take into account the independence of the signers.
- The second dataset was collected from signers with colored gloves (see Fig. 5) [2,5].

Due to these restrictions, we build our database of the Arabic sign language alphabet consisting of 30 hand signs from the unified version of the Arabic sign language alphabet as illustrated by Fig. 6(a). The hand postures were performed by 24 different subjects against a uniform background and 8 other different subjects on different complex backgrounds (see Fig. 6(b)). The hand postures were taken on different scales.

The subjects did not wear gloves. Each subject performed the 30 letters an average of 4 time.

Due to the natural performance of the hand postures, a large variance was noted in the type of hand postures performed for each sign. Variations in the performance were only limited by the sign language limitations. Fig. 7 illustrates examples of the various ways that the *Sad* sign was performed by different subjects and with different scaling.

#### 6.1.2. Jochen-Triesch Static hand postures Database

The second dataset is Jochen-Triesch static hand postures database [39] which consists of 10 hand posture signs (see Fig. 8) performed by 24 subjects against a uniform light, a uniform dark and complex backgrounds. Because its simplicity (no motion or colour values of pixels), posture recognition in our system was carried out independently of complex backgrounds.

### 6.2. Extraction of outline hand shape and detection of internal edges

#### 6.2.1. ArSL database

Using the proposed method, the hand region is extracted from the colour images of ArSL postures performed against complex backgrounds. Then, the silhouette and outline of the hand shape are extracted by the Canny-Derriche algorithm (see Fig. 9(a)). To detect the internal hand edges, the segmented hand images of ArSL are converted to grey scale images and their grey-level histograms are equalized. Then, a  $5 \times 5$  Gaussian filter with  $\sigma = 0.625$  is applied and the internal edges of the hand shape are detected using an adaptive threshold filter. Finally, to reduce singular pixels, a  $5 \times 5$  median filter is applied twice.

#### 6.2.2. Triesch database

The images of Jochen-Triesch database are in grey scale and sized to  $128 \times 128$  pixels. At beginning, the histogram grey levels of the image is equalized and a  $5 \times 5$  Gaussian filter with  $\sigma = 0.5$  is applied. To extract the external contour, image segmentation is carried out using the global threshold filter. After, morphological operations (erosion and dilation with  $3 \times 3$  structuring element) are applied, and using the hand region, the contour is extracted (see Fig. 9(b)). In the same way as for the ArSL database, the internal hand edges are extracted using an adaptive threshold filter and

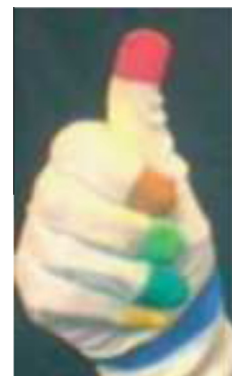


Fig. 5. Example of a signer with colored gloves.

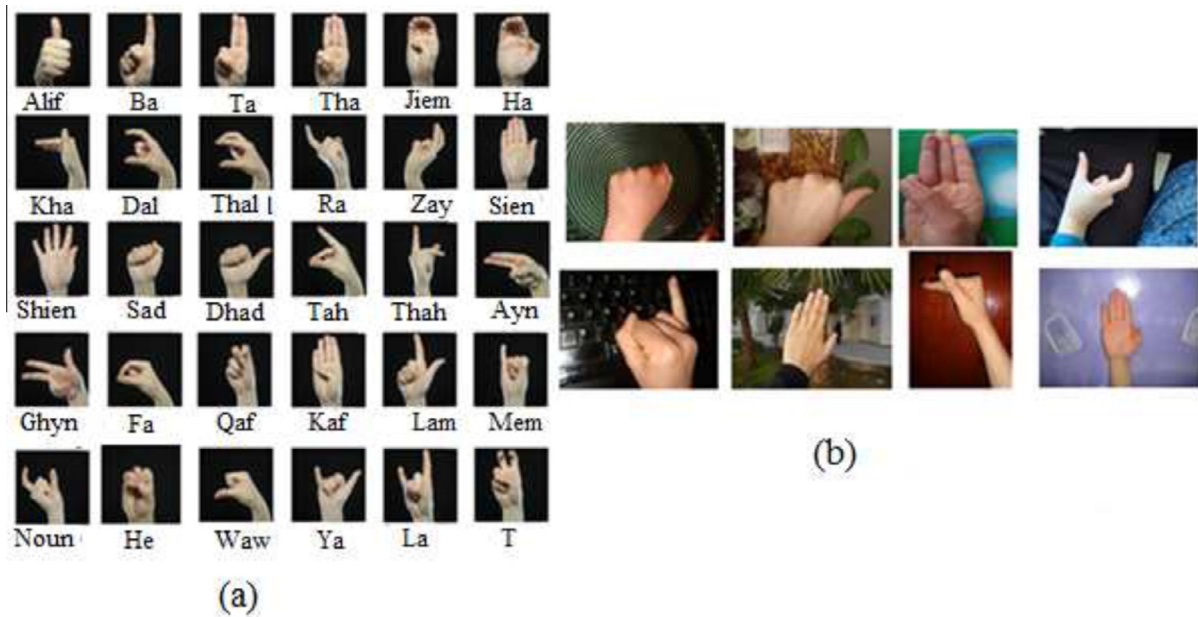


Fig. 6. (a) The Arabic sign language alphabet. (b) The different complex backgrounds used.



Fig. 7. Examples of the variation in the performance of the letter *Sad* from the ArSL.

singular pixels are reduced by means of  $5 \times 5$  median filter (see Fig. 10).

### 6.3. Wrist cropping

Images of ArSL database require that the hand must be separated from the forearm. The wrist-cropping procedure used for the binary images of the ArSL database, is based on the width of the arm as proposed in [40,30] (see Fig. 11).

This width-based cropping algorithm is based on observation that there is sudden increase in hand width as one moves from the lower arm to the hand. Before the wrist, the width of the lower arm is approximately constant. At the wrist, the width increases suddenly. In our work, at first the hand orientation is determined using the algorithm of Fitzgibbon et al. [14]. It is an efficient method for fitting ellipses using least-square approximation to scattered data. The orientation of the hand is then determined using the ellipse orientation given by an angle taking a value between 0 and  $2\pi$ . After, the width values of the arm are measured along the main axis of the ellipse as shown in Fig. 11.

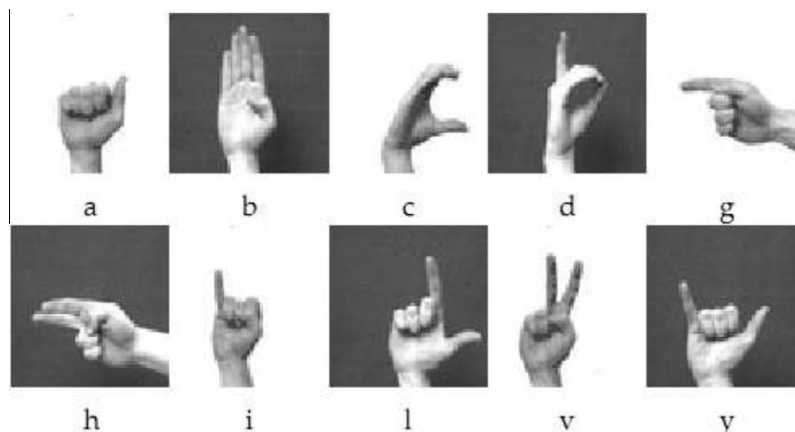


Fig. 8. The 10 hand posture signs from the Triesch database performed against light and dark backgrounds.

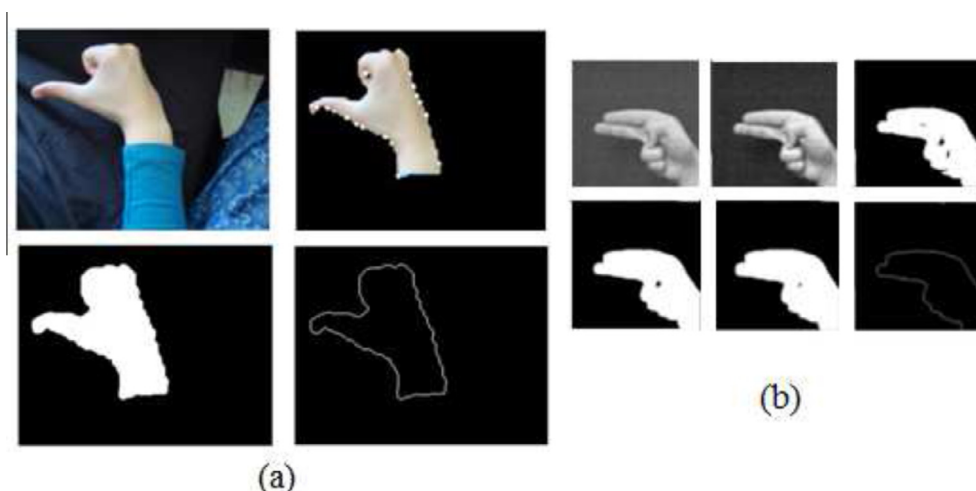


Fig. 9. (a) Steps of contour extraction: (a) Waw hand posture of ArSL database, (b) h hand posture from Triesch database.

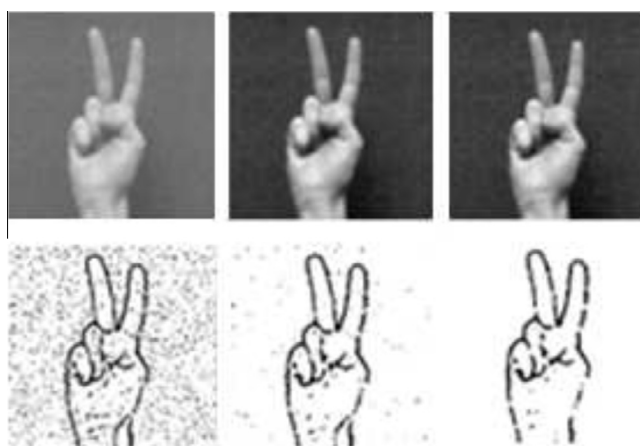


Fig. 10. Example of the internal and external edges extracted for the V hand posture from the Triesch dataset.

#### 6.4. Conducted tests

We evaluated the proposed recognition system using datasets described in the previous subsection. The hand postures of ArSL dataset were classified using the hand postures of 4 from the 24 subjects, for the training; the hand postures of the remaining 20 subjects were used for the test.

Hand postures of Triesch database were classified using two evaluation protocols  $P_1$  and  $P_2$ .  $P_1$  is the same protocol used by [39]: 3 subjects for the training and 21 subjects for the test dataset.

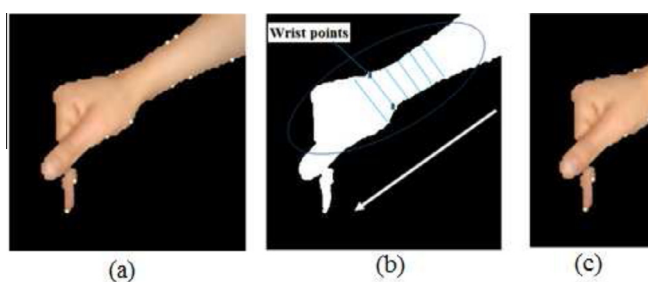


Fig. 11. The wrist-cropping process: (a) initial segmented hand, (b) located wrist points and the ellipse's main axis, (c) cropped hand.

The protocol  $P_2$ , 8 subjects were used for the training and the remaining 16 subjects were used for the test [24].

To evaluate the discriminatory properties of the proposed features, experiments are conducted in order to:

- Obtain the optimal order of Tchebichef moments for discriminating hand postures.
- Evaluate the performance of the Tchebichef moments using both the internal and external edges of hand shape compared to those using only external contour.
- Determine whether or not combining Tchebichef moments, Hu moments and geometric features offer complementary information regarding the hand posture patterns.
- Show if the results can be improved using SVM classifier.
- Test the robustness of the method in more realistic conditions.

##### 6.4.1. The optimal order of the Tchebichef moments

We first conducted experiments to find the optimal order of the Tchebichef moments. The order  $\eta$  of the calculated moments potentially affects the quality of the representation. However, the number of the calculated moments increases quadratically with the order as shown by Eq. (16) [9]. Consequently, we must choose the lowest order for reducing computation time, while having a correct and sufficient description, for the recognition of the object

$$\rho = (\eta + 1)(\eta + 2)/2 \quad (16)$$

We tested the effect of changing  $\rho$  on the recognition rate using only the Tchebichef moments. We took 60 hand postures from the Triesch dataset performed by 6 subjects against a uniform light background. 10 hand postures were chosen for the training set and the remaining 50 for the test. The distance between the Tchebichef moments was computed using the Euclidian distance.

Theoretically, the image function  $f(x, y)$  can be expressed completely as linear combinations of Tchebichef polynomials and so reconstructed for  $\eta$  sufficiently large. But in the recognition task the computation of the Tchebichef moments up to any order means the projection of the image function  $f(x, y)$  in any subspace. Each subspace represents a set of features used to describe the image function  $f(x, y)$ . Without experiments we cannot predict what are the relevant features i.e. the best projection of our image function  $f(x, y)$  in the different subspaces. We calculated the Tchebichef moments up to 13<sup>th</sup> order and found that the best recognition rate was obtained for  $\eta = 9$  and the worst one for  $\eta = 6$ . We can deduce that for  $\eta = 6$  the projection of the hand image function is not relevant to distinguish the different hand shapes. Thus, we used these

moments up to the 9th order (55 moments) in our recognition process (see Fig. 12).

#### 6.4.2. Performance of internal and external hand edges

The second experiments were conducted to examine how the use of the internal and external hand edges for computing the Tchebichef moments led to better recognition than the use of the outline of the hand shape. Firstly, Tchebichef moments descriptor was tested on five signs (A, B, C, D and G) from the Triesch database performed against uniform light background using the  $P_1$  protocol.

Tchebichef moments up to the 9th order were used as previously shown. Fig. 13(a) indicates that better recognition rates are obtained with the Tchebichef moments descriptor computed from the internal and external edges of the hand shape for the letters B, C and D and there is no difference in the recognition rates for the letters A and G. For the letters signs B, C and D, the use of internal edges additionally to the external one allowed to distinguish them from the other letters in the reference base. For example we can see that the letters D and I have very similar outline shapes but very different internal edges.

We compared the letters of the Arabic sign language that have similar outlines (He and Sad) and (Dal and Dhal) (see Fig. 2). The Tchebichef moment descriptor was tested on the ArSL database described earlier using 4 subjects for the training dataset and the remaining 20 subjects for the test dataset (see Fig. 13(b)). The obtained results indicate that the use of internal and external edges of the hand's shape enhanced the recognition rates of these letters and reduced the confusion between them. However with both representations of the hand's shape in the Tchebichef moments descriptor classification, the letter Dhal can be misclassified as Dal and the letter sign Sad can be misclassified as He.

#### 6.4.3. Combining the sets of features

To show that no redundancy exists between the three feature sets used, several combinations were examined. Firstly, each pair



Fig. 12. The effect of changing  $n$  on the recognition rate.

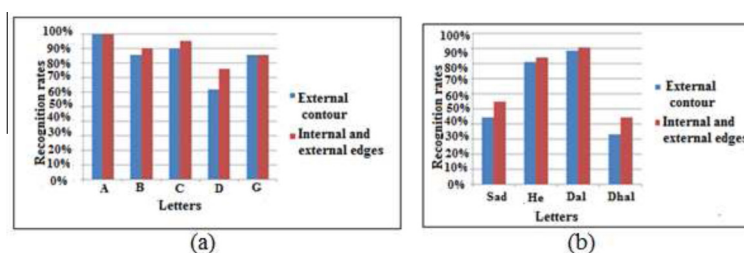


Fig. 13. (a) Comparison between the recognition rate performance on the letters A, B, C, D and G from the Triesch database. (b) Comparison between the recognition rate performance on letters Sad, He, Dal and Dhal from the ArSL database.

of features were combined sequentially, and in different orders. Then, for the best pair obtained from the three features, we added the third feature and the sequential combination of the three features in this order was tested. Finally, the three feature sets were combined in a parallel manner.

We present the recognition rates for the different combinations tested using the ArSL database and the Triesch dataset with the  $P_2$  protocol against light and dark backgrounds (see Fig. 14).

We use the following abbreviations:

- T, H, and D represent the Tchebichef moments, the Hu moments and the geometric features, respectively.
- The sequential combination of the features is noted using the hyphen; for example, T-H represents the sequential combination of the Tchebichef and the Hu moments.
- The parallel combination of the three features is noted using commas T, H, D.

The sequential combination of two feature sets was performed as follows: First, we obtained the three nearest neighbours of the query image according to the first feature using the KNN classifier,  $K = 3$ . Then, from the obtained classes, the class with the minimum distance computed from the second feature was recognized as the query image. The sequential combination of the three features was performed as follows: First we obtained the three nearest neighbours of the query image according to the first feature, then from the obtained classes two classes with the minimum distances computed from the second feature were retained and finally from the obtained two classes, the class with the minimum distance computed from the third feature was recognized as the query image.

The parallel combination of the three classifiers is performed as follows: the KNN classifier is used for each descriptor with  $K = 3$ . So, for each descriptor we obtain the three nearest classes of the query image. The class with the largest number of occurrences where the three descriptors are taken at once is recognised as the selected image.

Fig. 14 indicates that the best recognition rates were obtained from the parallel combination scheme of the three feature sets tested on the ArSL database, and the Triesch database against light and dark backgrounds. The recognition rates of the parallel combination were 94.67% for the ArSL database, 97.5% for the Triesch database against a light background, and 88.70% for the Triesch database against a dark background.

In addition, when only one feature was used, the Tchebichef moments was found to be the best feature with recognition rates of 82.33% for the ArSL database, 90.62% for the Triesch database against a light background and 78.27% against a dark background. In each case, if two features were used, the sequential combination of the two features improved the recognition rate over that of the first one, which indicates that the feature sets complemented one another.

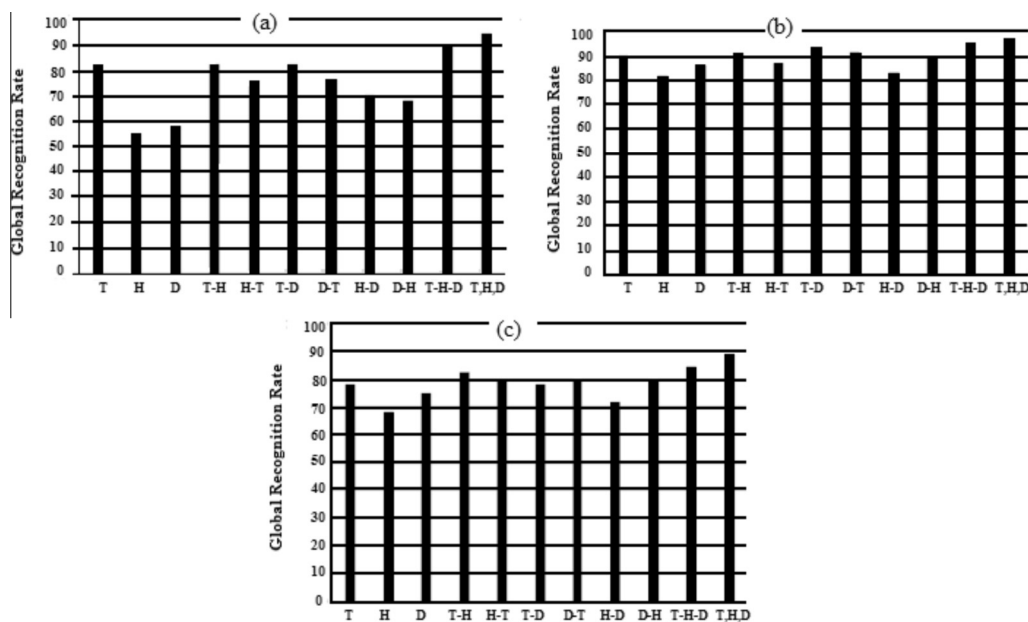


Fig. 14. (a) Global recognition rates obtained for the ArSL sign language Database. (b) Global recognition rates obtained for the Triesch database using the  $P_2$  protocol against a light background. (c) Results for the Triesch database using the  $P_2$  protocol on a dark background.

In the ArSL database, the best results of the sequential combinations were obtained from the combination of the Tchebichef moments and Hu moments features (T–H) with an 82.67% recognition rate. For Triesch database, the best results of the sequential combinations were obtained:

- against a light background, from the combination of the Tchebichef moments and geometric features (T–D) with a recognition rate of 93.75%,
- against a dark background, from the combination of the Tchebichef moments and Hu moments features, with a recognition rate of 82.37%.

The recognition rates of the sequential combination of the three features were 90.33% for the ArSL database, and 84.63% for the Triesch database against dark background with the scheme T–H–D, and 95.55% for the Triesch database against light background with the scheme T–D–H.

#### 6.4.4. The use of the SVM classifier

SVM is a set of supervised learning methods used in classification and regression. We used SVM classifier with the Radial Basis function (RBF) Gaussian kernel. The hand posture is represented by the concatenate vector of the three descriptors which has been found as the best combination scheme.

To classify an image  $x$  containing an unknown hand posture, it must be assigned to one of the  $n$  possible hand posture classes  $\alpha_1, \alpha_2, \dots, \alpha_n$ . The SVM type used in this work is based on one-versus-rest classes decision. For each hand posture  $\alpha_i$ , a support vector machine  $SVM^i$  is trained and used to calculate the probability that image hand posture  $x$  belong to the class  $i$ .

The classifier  $SVM^i$  is trained as follows: For each hand posture class  $i$  we note  $\alpha_i$  the matrix computed from the concatenate vector of the three descriptors (Discrete orthogonal Tchebichef moments, Hu moments and geometric features) on the training images corresponding to the hand posture class  $i$ . To train  $SVM^i$  the matrix  $\phi_i$  is used as labelled training data by  $y = 1$  and the matrix  $\phi_{\bar{i}}$  computed from the three descriptors on the training images not corresponding to the hand posture class  $i$  is used as labelled training data by  $y = -1$ .  $SVM^i$  is then trained to maximize the hyper plane margin

between  $\phi_i$  and  $\phi_{\bar{i}}$ . For ArSL database 30 SVM were trained and 10 SVM were trained for the Triesch database.

Then each  $SVM^i$  is extended to calculate the probability that  $x$  belong to the class  $c$  using Platts method [33]. Platts proposed approximating the posterior class probability  $pr(y = 1/x)$  by a sigmoid function:

$pr(y = 1/x) = 1/(1 + \exp(Af + B))$ , where  $A$  and  $B$  are estimated by minimizing the negative Log likelihood function using known training data and their decision value  $f$ .

Finally the class  $c$  with the strongest probability among all possible hand postures classes is recognized as the unknown image  $x$ . The obtained results are presented in Table 1. The protocol represents the number of persons used in the training dataset and in test dataset.

We can see that KNN classifier outperforms the SVM when the number of training samples is very small and in the user-independent mode. However the SVM classifier obtains best results compared to the KNN classifier tested on the Triesch database with the protocol  $P_2$  when the number of training samples increases.

#### 6.4.5. The robustness of the method

To illustrate the robustness of the proposed method in more realistic and harsh conditions, the system was tested with the best combination scheme obtained on images of ArSL alphabet performed on complex backgrounds and under different lighting conditions. Moreover we maintain the same reference consisted on hand postures performed against a simple background from only 4 subjects. The hand postures performed on complex backgrounds were used only for the test. Both the classification techniques KNN and SVM were explored. The obtained results are summarized in the Table 2. The proposed system can obtain good results in realistic conditions, with small number of training samples and in the user-independent mode. We can show also that the KNN classifier outperforms the SVM classifier when the number of samples used in the reference database is limited.

#### 6.4.6. Recognition performance

Table 3 gives a comparison of the results of our system with those obtained in [39,24,27]. The results obtained by our system

**Table 1**  
Recognition performance using the SVM classifier.

Dataset	Protocol	Recognition rate
ArSL data base	4 – 20	88.87
Triesch database	$P_1$ : 3 – 21	85.33
Triesch database	$P_2$ : 8 – 16	96.88

**Table 2**  
Recognition performance for the Arsl database performed against complex backgrounds Classification

Technique	Recognition rate
KNN	89.35
SVM	86.9

**Table 3**  
Recognition performance

Database, Protocol	Recog. rate	Alphabet	Instruments used	Recogn. rate
Our method, $P_1$	86.66	Our method	free	94.67
[39], $P_1$	93.8	[1]	free	93.55
[27], $P_1$	85.1	[2]	Coloured gloves	88
Our method, $P_2$	93.1	[5]	Coloured gloves	93.4
[24], $P_2$	89.9	[38]	free	90.4
[27], $P_2$	91.8	[13]	free	92

for the ArSL alphabet are compared with those obtained by other state of the art ArSL systems.

All these systems were working in simple backgrounds. A direct comparison with the ArSL alphabet systems cannot be performed because different methods, and databases were used, and all of which were in a signer-dependent mode.

Table 3 indicates that, for the Triesch database, our system performed well with the protocol used in [24]. The results also indicate that the elastic graph matching algorithm [39] achieved a higher recognition rate than our method.

The main reason of the superiority of the method from [39] compared to our method is that the elastic graph matching algorithm is designed to be very efficient when the number of hand postures classes is limited (10 hand postures classes in this special case). The method is sensitive to hand shape distortions when the number of hand postures classes is more large. Although this method requires several seconds to analyse a single image unless special hardware is utilized.

The computation time for our algorithm was obtained as the average of the computation times for the different images of the hand postures with different scaling and included time required for the image processing, feature extraction, and classification. The system implemented using C++ with a 2.10 GHz Intel core2 CPU and the computation time is 1.36 s.

Although our method had a better computation time than the elastic graph matching algorithm, it still requires improvement for its use in real time applications. The second section of the Table 3 indicates that with the user-independent mode and free hands, our system outperforms the other recognition systems of Arabic sign language alphabet.

## 7. Conclusion

We presented a new signer-independent system for hand posture recognition of Arabic sign language alphabets without using

gloves or any other devices. The system is working in different backgrounds and under various lighting conditions.

Our system efficiently combines three hand shape descriptors which allow representing finger configuration and hand orientation. In addition, internal edges of the hand shape are used for describing and discriminating hand postures. Moreover, a set of geometric features was efficiently adapted to the sign language recognition specificities. The proposed system performed well the classification of the hand postures from the ArSL alphabet.

Recognition rates of 94.67% in simple background and 89.35% in complex backgrounds were achieved in a signer-independent mode. We also obtained a good performance of 96.88% with the SVM classifier using the  $P_2$  protocol on the benchmark Triesch hand postures database.

The computation time of the proposed system could be improved by integrating the computation of the Tchebichef moments in a parallel processes in order to integrate the proposed approach in a real-time video application.

## References

- [1] O. Al-Jarrah, A. Halawani, Recognition of gestures in Arabic sign language using neuro-fuzzy systems, *Artif. Intell.* 133 (1-2) (2001) 117–138.
- [2] M. Al-Roussan, M. Hussain, Automatic recognition of Arabic sign language finger spelling, *Int. J. Comput. Appl. (IJCA)* 8 (2) (2001) 80–88 (Special issue on Fuzzy Systems).
- [3] M. Al-Roussan, K. Assaleh, A. Talaa, 2009: Videobased Signer independent Arabic sign language recognition using hidden-Markov models, *Appl. Softw. Comput.* 9 (2009) 990–999.
- [4] O. Aran, T. Burger, A. Caplier, L. Akarun, A belief-based sequential fusion approach for fusing manual signs and non-manual signals, *Pattern Recognit.* 42 (2009) 812–822.
- [5] K. Assalaeh, M. Al-Roussan, Recognition of Arabic sign language alphabet using polynomial classifier, *EURASIP J. Appl. Sign. Proces.* 13 (2005) 2136–2145.
- [6] B. Bauer, H. Hienz, Relevant features for video-based continuous sign language recognition, in: *FG00 Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000, pp. 440–445.
- [7] B. Bauer, K.F. Kraiss, Video-based sign recognition using self-organizing subunits, in: *Proceedings of the 16th International Conference on Pattern Recognition*, 2002, pp. 434–437.
- [8] R. Bowden, D. Windridge, T. Kadir, A. Zisserman, M. Brady, A linguistic feature vector for the visual interpretation of sign language, in: *Proceedings of the Eighth European Conference on Computer Vision*, 2004, pp. 391–401.
- [9] C.-C. Chang, J.J. Chen, W.K. Tai, C.C. Han, New approach for static gesture recognition, *J. Inf. Sci. Eng.* 22 (2006) 1047–1057.
- [10] H. Cooper, B. Holt, R. Bowden, Sign language recognition, Chapter in *Visual Analysis of Humans: Looking at People*, 2011, pp. 539–562.
- [11] N.H. Dardas, N.D. Georganas, Real time hand gesture detection and recognition using bag of features and support vector machine techniques, *IEEE Trans. Instrum. Meas.* 60 (11) (2011) 3592–3607.
- [12] P. Dreuw, D. Stein, T. Desealers, D. Rybach, M. Zahedi, H. Ney, Spoken language processing techniques for sign language recognition and translation, *Technol. Disability* 20 (2008).
- [13] S.A. Elons, M. Abul-ela, M.F. Tolba, Neutralizing lighting non-homogeneity and background size in PCNN image signature for Arabic sign language recognition, *Neural Comput. Appl.* (2012).
- [14] A. Fitzgibbon, M. Pilu, R.B. Fisher, Direct least square fitting of ellipses, *Pattern Anal. Mach. Intell.* 21 (5) (1999) 476–480.
- [15] J. Flusser, B. Zitova, T. Suk, *Moments and Moment Invariants in Pattern Recognition*, John Wiley & Sons Ltd, 2009.
- [16] W. Gao, G.L. Fang, D.B. Zhao, Y.Q.A. Chen, A chinese sign language recognition system based on SOFM/SRN/HMM, *Pattern Recognit.* 37 (2004) 2389–2402.
- [17] R.L. Graham, An efficient algorithm for determining the convex hull of a finite planar set, *Inf. Process. Lett.* 1 (1972) 132–133.
- [18] L. Gu, J. Su, Natural hand posture classification based on Zernike moments and hierarchical classifier, *IEEE Int. Conf. Robotics Autom.* (2008) 3088–3093.
- [19] R. Haralick, K. Shanmugan, I. Dinstein, Textural features for image classification, *IEEE Trans. Syst. Man Cybern.* SMC-3 (1973) 610621.
- [20] T. Hastie, R. Tibshirani, J. Freidman, *The elements of statistical learning*, Data Mining, Inference and Prediction, Springer Series in Statistics, second ed., 2008.
- [21] M.-K. Hu, Visual pattern recognition by moment invariants, *IEEE IRE Trans. Inf. Theory* 8 (2) (1962) 179187.
- [22] C.L. Hwang, H.W. Lee, The command control by hand gesture with Hu and contour sequence moments and probability neural network, in: *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 2056–2061, 2011.
- [23] D.E. Ilea, P.F. Whelan, Image segmentation based on the integration of colour-texture descriptors – a review, *Pattern Recognit.* 44 (10-11) (2011) 2479–2501.

- [24] A. Just, Y. Rodriguez, S. Marcel, Hand posture classification and recognition using the modified census transform, in: 7th Internat. Conf. on Automatic Face and Gesture Recognition, FGR, pp. 351–356, 2006.
- [25] M.W. Kadous, Machine recognition of Australian signs using Powergloves: Towards large-lexicon recognition of sign languages, in: Workshop on the Integration of Gestures in Language and Speech, Wilmington Delaware, 1996.
- [26] P. Kakumanu, S. Makrogiannis, N. Bourbakis, A survey of skin-color modeling and detection methods, *Pattern Recognit.* 40 (2007) 1106–1122.
- [27] D. Kelly, J. McDonald, C. Markham, A person independent system for recognition of hand postures used in sign language, *Pattern Recognit. Lett.* 31 (2010) 1359–1368.
- [28] J.S. Kim, W. Jang, Z. Bien, A dynamic gesture recognition system for the Korean sign language (KSL), *IEEE Trans. Syst. Man Cybern.* 26 (2) (1996) 354–359.
- [29] M. Last, H. Bunke, A. Kandel, A featurebased serial approach to classifier combination, *Pattern Anal. Appl.* 5 (4) (2002) 385–389.
- [30] A. Licsar, T. Szirnyi, User-adaptive hand gesture recognition system with interactive training, *Image Vision Comput.* 23 (12) (2005) 1102–1114.
- [31] A. Malima, E. Ozgur, M. Cetin, A fast algorithm for vision-based gesture recognition for robot control, in: 14th IEEE conference on Signal processing and Communication Applications, 2006, pp. 1–4.
- [32] R. Mukundan, S.H. Ong, P.A. Lee, Image analysis by Tchebichef moments, *IEEE Trans. Image process.* 10 (9) (2001) 1357–1364.
- [33] J.C. Platt, Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, in: *Advances in Large Margin Classifiers*, MIT Press, 1999.
- [34] S.P. Priyal, P.K. Bora, A study on static hand gesture recognition using moments, in: IEEE International Conference on Signal Processing and Communications (SPCOM), 2010, pp. 1–5.
- [35] P. Premartne, S. Ajaz, M. Premartne, Hand gesture tracking and recognition system using Lucas–Kanade algorithms for control of consumer electronics, *Neurocomputing* 116 (2012) 242–249.
- [36] K.W. See, K.S. Loke, P.A. Lee, K.F. Loe, Image reconstruction using various discrete orthogonal polynomials in comparison with DCT, *Appl. Math. Comput.* 93 (2) (2007) 346–359.
- [37] T. Starner, A. Pentland, Real-time american sign language recognition from video using hidden Markov models, in: AAAI Fall Symposium on Disabilities, Cambridge, MA, 1996.
- [38] MF. Tolba, MS. Abdellwahab, M. Aboul-Ela, A. Samir, Image signature improving by PCNN for Arabic sign language recognition, *Can. J. Artif. Intell. Mach. Learn. Pattern Recognit.* 1 (1) (2010) 1–6.
- [39] J. Triesch, C. Von der malsuburg, Classification of hand postures against complex backgrounds using elastic graph matching, *Image Vision Comput.* 20 (13–14) (2002) 937–943.
- [40] C. Wah Ng, S. Ranganath, Real-time gesture recognition system and application, *Image Vision Comput.* 20 (1314) (2002) 993–1007.
- [41] C.Y. Wee, R. Paramesran, R. Mukundan, X. Jiang, Image quality assessment by discrete orthogonal moments, *Pattern Recognit.* 43 (12) (2010) 4055–4068.
- [42] M. Zahedi, P. Dreuw, D. Rybach, T. Desealers, H. Ney, Using geometric features to improve continuous appearance-based sign language recognition, in: *British Machine Vision Conference (BMVC)*, vol 3, Edinburgh, UK, 2006.