Multimedia Tools and Applications (2019) 78:27957–27982
https://doi.org/10.1007/s11042-019-07859-9

Efficient representation of size functions based on moments theory

Djamila Dahmani[1] · Slimane Larabi[1] · Mehdi Cheref[1]

Received: 14 October 2018 / Revised: 4 April 2019 / Accepted: 5 June 2019 /
Published online: 1 July 2019

Abstract
Nowadays, there is a need to develop efficient and intuitive solutions such as hand gestures recognition for the Human-machine interaction. This paper presents a hand gestures recognition system based on salient geometric features extracted using size functions theory. We propose a new representation of the reduced size function based on Tchebichef moments providing more details and information for their graphs descriptions compared to existing representations. In addition, a methodical algorithm of fast Tchebichef moments computation for grey scale images is well adapted to the encoded graph of size function. Furthermore, a contour discretization based on a convexity approach is proposed for an optimal computation of the measuring functions, and new measuring functions for hand gestures classification and retrieval are proposed. The comparison with existing systems indicates that our method competes with the best ranked method for the dynamic case and surpasses the state of the art in static case; in addition it presents the advantage to be applied in both static and dynamic cases.

Keywords Hand gestures recognition · Size function · Texture · Color · Tchebichef's moments

1 Introduction

The need to conceive efficient and intuitive solutions of interaction between human and computer systems increases with the evolution of our smart environment. Touch-less interaction provides new interaction paradigms, where the hand gestures constitute an appealing alternative. Vision based gesture recognition is an easy and effective way to human machine interaction, which does not use any physical contact, since only a web camera is required. It

✉ Djamila Dahmani
ddahmani@usthb.dz

Slimane Larabi
slarabi@usthb.dz

[1] Computer Science Department, University of Science and Technology Houari Boumediene (USTHB), Algiers, Algeria

has extensive applications such as sign language interpretation and learning, teleconferencing, robotics, computer games, and virtual reality and consists generally of three main steps: hand detection, hand feature extraction, and finally hand gesture recognition and classification.

Hand shape recognition is a complex problem because such recognition must locate hand with no prior information regarding its size, rotation, background and lighting. An efficient hand gesture recognition system must be able to operate under different lighting conditions and in a variety of complex backgrounds. A common technique to deal with this issue is to apply restrictions on the user environments [13, 46, 48]. Furthermore, the hand gestures recognition task is an extremely challenging problem. This is due to the fact that the human hand is a complex articulated object with 27 degrees of freedom motion [12], which implies a great number of 2D appearances depending on the camera viewpoint.

Different hand gestures models and abstract were developed in the area. They can be sundered into two important classes:

- The first one is based on external sensors connected to the user [5, 6, 33]. Such systems capture the gestures precisely but the process is invasive and sparsely natural for real applications.
- The second class use video camera to capture and recognize the gesture. These systems are more natural than the first class and can operate in uncontrolled environments. The vision based methods can be divided into three great families the depth based [9, 22, 45], colour image based [31] and shape based features [25, 47].

The aim of this work is to develop an accurate vision-based hand gesture recognition system able to operate in uncontrolled environments under variable lighting conditions. The majority of current hand detection methods use skin colour information. However, skin colour segmentation is inadequately robust for dealing with complex backgrounds and rapid variation in illuminations. To cope with this problem, we propose in this paper to employ in addition to the skin colour clustering the skin texture cues modelled using both statistic and spatiotemporal aspects. The obtained features are used to train a neural networks classifier (ANN) to detect skin pixels. The segmentation approach was tested on hand gestures Cambridge database [28].

The rest of the paper is organized as follow; section 2 is devoted to the previous works in action recognition along with some works based on persistent homology in pattern recognition tasks. In section 3 the methodology of the proposed gestures recognition system is presented and the proposed size function representation based moments is explained. Conducted experiments and obtained results are shown and discussed in section 4. Finally, we conclude the paper and give some future works in section 5.

2 Related works

A variety of methods and approaches were proposed in the area of gesture recognition. Kobayashi and Otsu proposed in [30] a motion recognition scheme based on motion feature extraction which used co-occurrence histograms of the space-time gradients of 3D motion shape in video sequence. This method, called Space-Time Auto Correlation of Gradients (STACOG), was tested on various human action databases. Harrandi et al. proposed in [18] an action and gesture recognition method based on spatio-temporal covariance descriptors (Cov3D), and a weighted Riemannian locality preserving

projection approach that takes into account the curved space formed by the descriptors. The Cov3D descriptors are extracted from spatio-temporal window inside sample videos. A boosting approach was used to search the windows to find subset which is the best for classification. The weighted projection is then exploited to create a final multiclass classification algorithm. Lui and Beveridge [34] developed human action recognition method based on the representation of videos as a tangent bundle on Grassmann manifold. Videos are expressed as third order tensors and factorized to a set of tangent spaces. Tangent vectors are then computed between elements on a Grassmann manifold and exploited for action classification. In the same family of tensor-based methods, Su et al. [41] proposed a technique for handling classification of video sequences in unequal length of time, namely Spatial-Temporal Iterative Tensor Decomposition (S-TITD) for uniform length. In the same class of methods Hsieh and Lin developed an action recognition approach based on dual-complementary tensors [19]. The input video is normalized into dual tensors. One tensor is obtained from the raw video volume data and the other one is obtained from the histogram of oriented gradients (HOG) features. The previous cited methods have an advantage to do not require foreground detection or tracking, but they can be used in tasks when the hand posture recognition is necessary.

Outside of the tensors based approaches, Baraldi et al. [1] proposed hand gesture recognition system based on dense trajectories and hand segmentation, and Barros et al. [2] developed a dynamic gesture recognition and prediction system. The classification and prediction modules are based on Hidden Markov models and dynamic time warping. These methods have an advantage to deal with static and dynamic hand gestures recognition tasks, but they depend strongly on the efficiency of the hand region detection and tracking.

Hand gesture features extraction in our approach is inspired by shape-from-functions techniques. These techniques explore the topological shape's aspect provided by a set of real functions defined on it, and chosen to extract its salient geometric features. The size function theory is an efficient mathematical tool in this class of methods. It is derived from the persistent homology which is an algebraic concept for quantifying topological features of functions or of shapes [15]. Size functions were used for 3D object recognition [36], for 2-dimensional matching [3] and for shapes comparison under subgroups of projective groups of differential invariants [8]. The dimension of size graphs which can be arbitrarily very large, and the quantification of the most important information contained in the encoded graph of the reduced size function still presents challenges for researchers in the application of size function theory for shape comparison and classification.

Our approach differs from previous related works, the main contributions are:

- A new representation of size functions is proposed; it is based on the description of size functions using moment's theory. A methodical algorithm to compute faster the Tchebichef moments for grey scale images was adapted to the encoded graph of size functions.
- Efficient contour discretization for computation of the measuring functions using the convexity approach. The convexity approach used reduces the dimension of size graphs while preserving the shape geometry.
- Proposition of new measuring functions for hand shape description and retrieval.
- Hand region segmentation process based on texture and colour attributes which allows to the system to be operational in different lighting conditions.

3 Proposed approach

In this section we will present in Fig. 1 the general architecture of our gesture recognition system. It is composed of many important steps:

- Hand region detection and location in video frames, is based on colour and texture attributes along with an artificial neural networks classifier (ANN). More details will be presented in the section 4.
- Recognition method is based on size function theory [15]. The main idea in the theory is to compare shape proprieties described by real geometric functions, defined on topological spaces associated to the "objects" to be studied. In formal settings that mean a shape is presented by (X, φ), where X is topological space, and $\varphi : X \to R$ is continuous function called measuring function. Every pair (X, φ) is called size pair. The size function is a shape descriptor encoding of the 0th Betti Number in the sublevels sets of X induced by φ [7].

We introduce the basic concepts in the theory of size functions, in section 3.1.

3.1 Brief review of size functions theory

Given a size pair (X, φ), the reduced size function

$$\ell_{(X,\varphi)} : \{(x,y) \in R^2 : x < y\} \to N \cup \{\infty\}.$$

Can be defined by setting $\ell_{(X,\varphi)}(x,y)$ as the number of connected components of the set $S_y = \{P \in S : \varphi(p) \leq y\}$ containing at least one point of S_x.

From the computational point of view, the discrete size pair (X, φ), is a size graph(G, φ), where $G(V(G), E(G))$ is finite graph where $V(G)$ and $E(G)$ denotes the sets of vertices and edges respectively, and $\varphi : V(G) \to R$ is a measuring function labeling the nodes of the graph G

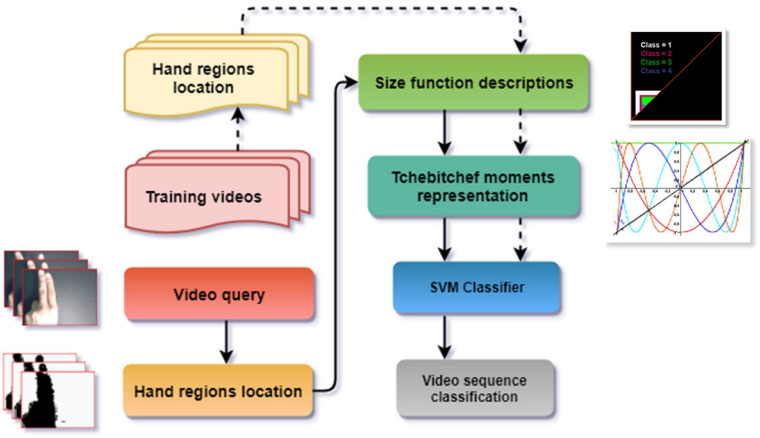

Fig. 1 Flowchart of the proposed hand gestures recognition system

. Algorithm 1 gives the required steps for calculating reduced size function $\ell_{(X,\phi)}$.

Algorithm 1

Begin

Let G be a graph whose vertices are the points $\{p_1, p_2, ..., p_n\}$ *of topological space X and* φ *a measuring function defined on G.*

1. Find the sub graph $G_{\varphi \le v}$ *of G determined by the points* p_i *with* $\varphi(p_i) \le v$

2. Identify the connected components of $G_{\varphi \le v}$.

The size function λ_φ *at the point* $(u,v) \in R^2$ *is equal to the number of connected components of* $G_{\varphi \le v}$ *which contain at list a vertex with* $G_{\varphi \le u}$.

End

An example is presented in Fig. 2. The function $\tilde{\phi}_g$ is the normalized function of the function ϕ_g representing the distance between the center of gravity g of the hand shape and the points extracted from the hand shape contour *X*, defined by the formula:

$$\tilde{\varphi}_g = \frac{\varphi_g - \varphi_g^{\min}}{\varphi_g - \varphi_g^{max}} \tag{1}$$

Where $\phi_g{}^{\min}$ and $\phi_g{}^{\max}$ are respectively the minimum and maximum of the function ϕ_g. The reduced size function $\ell_{\tilde{\phi}_g}$ generated by the measuring $\tilde{\phi}_g$ function is invariant to scale transformation.

The domain of size function $\ell_{\tilde{\phi}_g}$ in Fig. 2d is divided into regions which size function is constant.

Given size pair (X, φ), the study and the computation of size function can be restricted to the triangular region Δ where $\Delta = \{(x,y) : \varphi^{min} \le x < y \le \varphi^{max}\}$, φ^{min} and φ^{max} are respectively the minimum and maximum of the function φ,as it was established in [44].

In order to quantify the information obtained by the reduced size function which is represented by infinite points of a triangle containing finite classes of integer values, some research efforts were deployed. The important one was the representation developed by Frosini

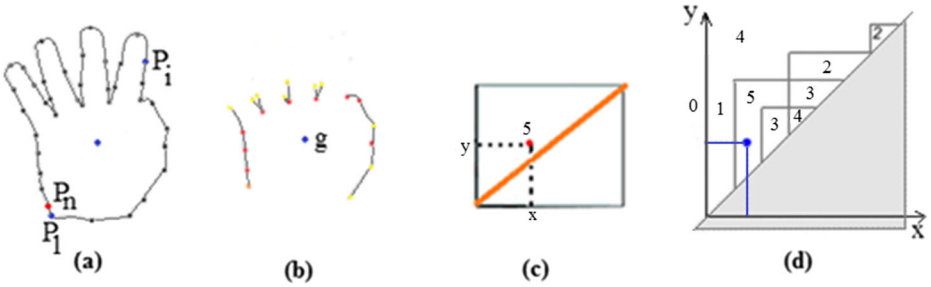

Fig. 2 **a** Points of a shape contour *X*, **b** Points p_i with $\tilde{\phi}_g(p_i) \le y$ in yellow and the points p_i with $\tilde{\phi}_g(p_i) \le x$ in red (*x*, *y* are given distances), **c** Graph of the size function $\ell_{\tilde{\phi}_g}$ with $\ell_{\tilde{\phi}_g}(x,y) = 5$, **d** the size function $\ell_{\tilde{\phi}_g}$

and Landi [16] which proposed a combinatorial representation using corner points, then considered as formal series. The size function was compared using the matching distance [7]. In [26], the authors proposed to use the eigenspace information using principal component analysis (PCA). Finding a compromise between a good quantification of information held in size functions and computation load still is an important challenge for researchers in size functions representation.

In this paper we propose to use the Tchebichef moments to collect the salient information held in size functions. In the following we explain the principle steps of our approach.

3.2 The proposed representation of the reduced size functions

The reduced size function $\ell_{(X,\phi)}$ can be entirely determined in the triangle Δ defined in the section 3.1, which contains all the relevant information about the shape under study. In addition, the reduced size function is finite and strictly positive [43], therefore we propose to describe the features of reduced size functions by computing the Tchebichef moments on its encoded –graph. Moments are scalar quantities used to characterize any function and to capture its significant features. From the mathematical point of view, moments are ″projections″ of function onto polynomial basis [14]. The main idea in this paper is to describe information contained in a 1-dimentional size function, by computing the orthogonal Tchebichef moments from its graph consisting of labelled triangles. Two size functions represented by Tchebichef moments vectors can be compared by using different metrics, including Euclidian distance. The Tchebichef moments are calculated using the method proposed in [40]. The algorithm is based on slice intensity representation of grey scale images. By analogy, we adapt this algorithm for the encoded graph of size functions. In the following we present the essential steps of our approach.

3.2.1 Convexity approach to hand contour shape discretization

Despite its beautiful formulation, in the continuous case the size function needs in depth contour discretization to reach best results. We use in this work the convexity approach based on the Douglas–Peucker polygon [10] associated to the original hand shape. The Douglas-Peucker algorithm is an efficient method to obtain a smooth contour on a finite number of vertices encoding the local geometries of the object such as the eccentricity which are the most important conditions discussed in [43] to extend the theory from the ideal continuous case of the size functions to the discrete one (see Fig. 3). Moreover by defining the measuring function on an object with well-known topology (polygon) we can cope with changes in the different topologies of original hand shapes. The Douglas-Peucker algorithm [10] reduces the number of points in curve that is approximated by a series of points. The measuring function is then defined on these selected salient points (see Fig. 3b) this also allows reducing the time costs of the size function algorithm.

3.2.2 Measuring functions

A measuring function is considered to be adequate if it produces a size function able to distinguish between different hand shapes. In this paper, we propose to use two families of measuring functions:

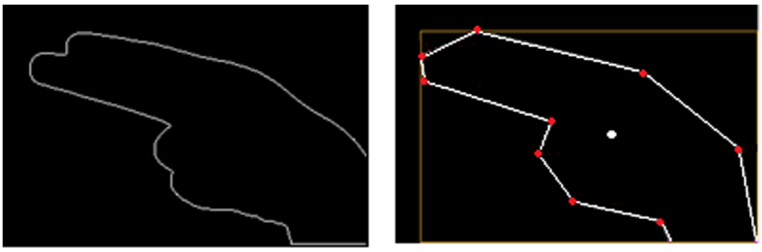

Fig. 3 **a** The original hand shape, **b** the Douglas-Peucker polygon associated to the hand shape with selected salient points

a) ***Distances from points***

The first family $\{\varphi_p\}$ of measuring functions used is the distances from points. Let $(o, e_1, e_{2)}$ be the Cartesian reference frame associated to the image.

Let $p(x_p, y_p) \in R^2$, and γ is the outline of the hand shape.

We define a measuring function $\varphi_p : \gamma \rightarrow R$ as:

$$\varphi_p(x,y) = d(p,(x,y)) = \sqrt{\left(x - x_p\right)^2 + \left(y - y_p\right)^2} \tag{2}$$

Five points were selected in this paper, the center of gravity of the hand shape and the four points of the minimum rectangle enclosing the hand shape (See Fig. 4).

b) ***Projections***

The second family $\{\varphi_\theta\}$ of measuring functions used in this paper is based on the different perpendicular projections of the outline hand shape on the lines K_θ, where K_θ is the line passing through the center of the gravity of the hand shape $G = (x_G, y_G)$, and which forms an angle θ with the horizontal axis of the reference frame associated to the image (see Fig. 5).

The coordinates of the projection point $Q = (x_Q, y_Q)$ can be obtained from the coordinates of the point $P = (x_P, y_p)$ by the eqs. (3) and (4).

$$x_Q = x_G + \frac{\left(x_p - x_G\right) + \left(y_P - y_G\right)\tan\theta}{\left(1 + \tan\theta^2\right)} \tag{3}$$

$$y_Q = y_G + \frac{\left(x_p - x_G\right) + \left(y_P - y_G\right)\tan\theta}{\left(1 + \tan\theta^2\right)}\tan\theta \tag{4}$$

Where $\tan\theta$ represent the tangent of the angle θ.

Then the measuring function $\varphi_\theta$is defined as:

$$\varphi_\theta(P) = d(Q,G) = \sqrt{\left(x_Q - x_G\right)^2 + \left(y_Q - y_G\right)^2} \tag{5}$$

where d is the Euclidian distance.

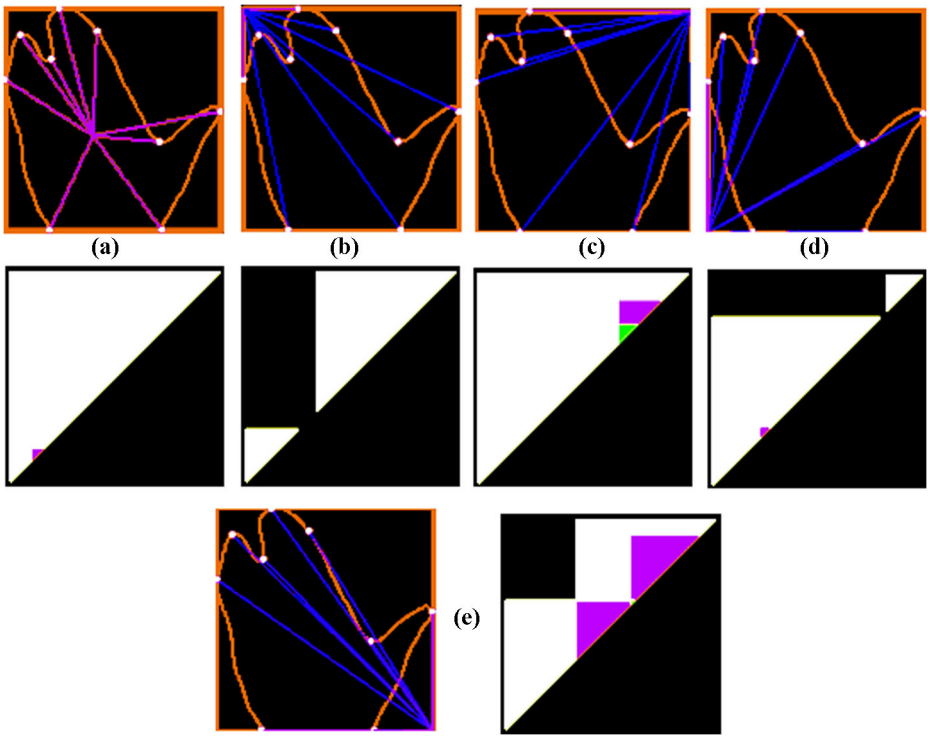

Fig. 4 The five points selected

If we consider all possible values of the angle θ, the family $\{\varphi_\theta\}$ can give a complete description of the outline's hand shape, but practically only seven angles were selected 10°, 40°, 60,130°, 160°,175° and 200°.

3.2.3 Tchebichef moments representation of size functions

The encoded graph of size function contains all information about its corresponding measuring function and the geometric properties that it translates. We propose to quantify this information by computing the Tchebichef moments of the encoded graph to faithfully represent all the information held in the size function. Figure 6 illustrates an example of two encoded graphs of

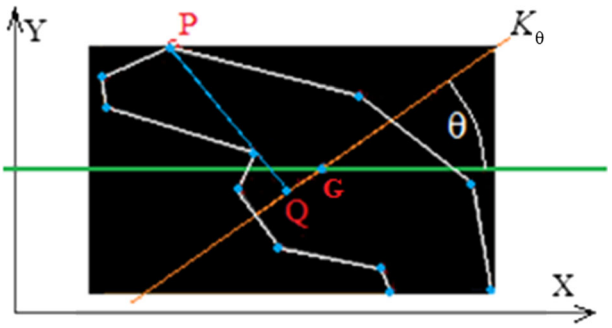

Fig. 5 The perpendicular projection

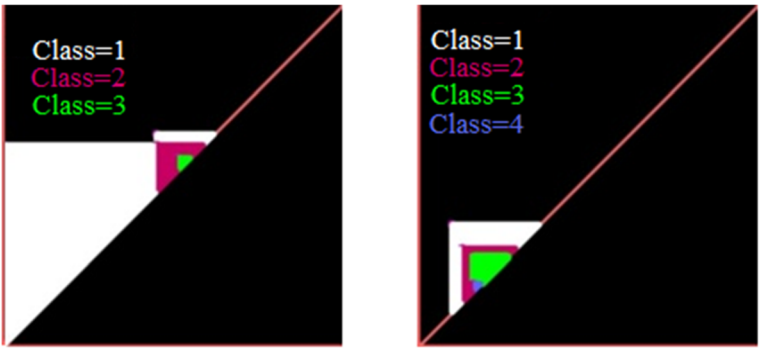

Fig. 6 The encoded graph of size function for two different measuring functions (the distance from the right top point and the distance from the center of gravity) for the same hand shape

the same hand shape of two different size functions corresponding to the measuring functions: the distance from the center of gravity and the distance from the right top point of the minimum area rectangle enclosing the hand shape. Each value of the size function is represented by a color in its encoded graph.

Differently from the original approach where the size functions are considered as formal series and determined by a finite set of 2D-corner points and lines (points and lines of discontinuities) [7], and where the values of the size function are not taken into account, our representation captures all different values that the size function can take and their emplacement in the plane.

For the computation of the Tchebichef moments we use the method proposed in [40] for grey-scale images. The authors show that their proposed method can speed up the computational efficiency as far as the number of blocks is smaller than the image size. However the region of interest (Δ) in the encoded graph of size functions (see section 3.1) is constituted from small number of regions with the same integer value j (see Fig. 6). As a result, the algorithm in [40] is well appropriate in our case.

The algorithm is based on a number of steps that we have modified as follows, for mathematical proofs of results refer to the article [40].

- **Step 1.**

The decomposition of the region of interest namely the reduced size function $\ell_\varphi(x,y)$ defined on $\Delta = \{(x,y) : \varphi^{min} \le x < y \le \varphi^{max}\}$ on into series of two level images $\ell_\varphi{}^j(x,y)$(Blocks inside the triangle Δ which have the same value) is given by eq. (6).

$$\ell_\varphi(x,y) = \sum_{j=1}^{L} \ell_\varphi{}^j(x,y) \tag{6}$$

Where L is the number of slices (equal to the number of the different values of the given size function). The method used is based on the algorithm PIBR proposed in [39].

- **Step 2.**

Computation of the Tchebichef polynomials values at corners at each block using the recurrence formula (7) [37]:

$$t_n(x) = \alpha_1 t_n(x-1) + \alpha_2 t_n(x-1) \text{ for } n = 1, 2, \ldots, N, x = 2, 3, \ldots, \frac{N}{2} \tag{7}$$

and

$$\begin{aligned} t_n(0) &= -\sqrt{\frac{(N-n)(2n+1)}{(N+n)(2n-1)}} t_{n-1}(0); \quad t_n(1) = \left(1 + \frac{n(n+1)}{1-N}\right) t_n(0); \\ t_0(0) &= \sqrt{\frac{1}{N}} \end{aligned} \tag{8}$$

Where $N = \varphi^{max}$ the maximum of the measuring function φ and so the dimension of $\ell_\varphi(x, y)$, and:

$$\begin{aligned} \alpha_1 &= \frac{-n(n+1) - (2x-1)(x-N-1) - x}{x(N-x)}; \\ \alpha_2 &= \frac{(x-1)(x-N-1)}{x(N-x)} \end{aligned} \tag{9}$$

Finally the computation of the vector $R_n(\delta_{b_i})$ is performed using the formula proved in [40].

$$R_n(\delta_{b_i}) = t_{n+1}\left(x_{1,b_i} + \delta_{b_i}\right) - t_{n+1}\left(x_{1,b_i}\right) \tag{10}$$

δ_{b_i} is the number of pixels in each block b_i, $\delta_{b_i} = \left(x_{2,b_i} - x_{1,b_i} + 1\right)$, where $\left(x_{1,b_i}, y_{1,b_i}\right)$ and $\left(x_{2,b_i}, y_{2,b_i}\right)$ are the left-up and the right-bottom coordinates of the block b_i.

- **Step 3.**

It was proved in [40], that $T_{nm}^{b_i}$ the Tchebichef moment of block b_i can be given by the formula:

$$T_{nm}^{b_i} = S_n\left(x_{1,b_i}, x_{2,b_i}\right) S_m\left(y_{1,b_i}, y_{2,b_i}\right) \tag{11}$$

where

$$S_n\left(x_{1,b_i}, x_{2,b_i}\right) = \sum_{x=x_{1,b_i}}^{x_{2,b_i}} t_n(x); S_m\left(y_{1,b_i}, y_{2,b_i}\right) = \sum_{y=y_{1,b_i}}^{y_{2,b_i}} t_m(y) \tag{12}$$

$S_n\left(x_{1,b_i}, x_{2,b_i}\right)$ and $S_m\left(y_{1,b_i}, y_{2,b_i}\right)$ correspond to the sum of the discrete Tchebichef polynomials values according to the axe (Ox) in the segment $\left[x_{1,b_i}, x_{2,b_i}\right]$, and according to the axe (Oy) in the segment $\left[y_{1,b_i}, y_{2,b_i}\right]$ respectively. Underline that $S_n\left(x_{1,b_i}, x_{2,b_i}\right)$ and $S_m\left(y_{1,b_i}, y_{2,b_i}\right)$ could be calculated in a similar manner.

The 1-dimensional Tchebichef moments vector of each block can be written as

$$V_m\left(x_{1,b_i}, x_{2,b_i}\right) = \left(s_0\left(x_{1,b_i}, x_{2,b_i}\right), s\left(x_{1,b_i}, x_{2,b_i}\right)_1, \ldots, s_{m-1}\left(x_{1,b_i}, x_{2,b_i}\right)\right)^t \tag{13}$$

It was showed in [40] that V_m is the solution vector of the linear system eq. (14).

$$U_m(\delta_{b_i}) = -A_m V_m\left(x_{1,b_i}, x_{2,b_i}\right) \tag{14}$$

where the vector:

$$U_m(\delta_{b_i}) = \left((R_0(\delta_{b_i}), R_1(\delta_{b_i}), \ldots, R_{m-1}(\delta_{b_i})\right)^t \tag{15}$$

whose coordinates are obtained in step 2, and.

A_m is m × m lower triangular matrix given by:

$$A_m = \begin{pmatrix} g_1(1,0) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ g_1(m,0) & \cdots & g_1(m,m-1) \end{pmatrix} \tag{16}$$

$$\text{The elements } g_1(n,n-1) = 2\sqrt{\frac{(2n-1)(2n+1)}{(N-n)(N+n)}} \tag{17}$$

The matrix A_m is invertible so we can obtain easily the solution of the system (14).

- **Step 4.**

Computation of Tchebichef moments valuesT_{nm} of the region Δ of the size function using the equations:

$$T_{nm} = \sum_{j=0}^{L} T_{nm}(j)\ell_\varphi{}^j(x,y) \tag{18}$$

where $T_{nm}(j)$ is (n + m)th order of Tchebichef moments of the encoded graph region of value j. Then we can compute $T_{nm}(j)$ using the formula

$$T_{nm}(j) = \sum_{i=0}^{K-1} T_{nm}^{b_i}(j) \tag{19}$$

Where K is the number of blocks of slices j in the region Δ and the $T_{nm}^{b_i}(j)$ are determined in the step 3.

3.2.4 Recognition framework

At first reduced size $\ell_{\tilde{\phi}_g}$ function is represented by the Tchebichef moments vector computed from its associated encoded graph. The Tchebichef moments order is determined experimentally.

The classification task is performed using SVM classifier. The SVM classifier is extended to obtain class probabilities [32], in this paper we use the three first classes with the strongest probabilities.

The SVM classifier is trained as follows:

For each hand posture or gesture class c we note ψ_C the matrices computed from the Tchebichef moments of the encoded graph of size functions obtained from the different measuring functions presented previously (section 3.2.2). These measuring functions are

computed from the training images corresponding to the hand posture or gesture class c. To train the SVM the matrix ψ_C is used as the 1 labelled training data and $\psi_C^{\check{}}$ the matrix computed from the Tchebichef moments of the encoded graph of size functions on the training images not corresponding to the hand posture class c is used as the 0 labelled training data.

4 Experimental results

4.1 Experimental setup

We evaluate our techniques using two separate data sets:

4.1.1 Triesch static hand postures database

This database consists of 10 hand posture signs (see Fig. 7) performed by 24 subjects against uniform light uniform dark and complex backgrounds [42]. Since there are no color values of pixels available in Triesch static hand postures database. In our system, posture recognition is carried out independent of complex backgrounds.

Images of Triesch static hand posture dataset are in grey scale sized to 128 by 128 pixels. To extract their external contour we equalize the histogram grey levels of the image, we apply after a 5 × 5 Gaussian filter. Image segmentation is carried out using the global threshold filter. Finally the morphological operation (erosion and dilation) are applied and using the region of the hand gesture we extract the outline hand shape (see Fig. 8).

4.1.2 Cambridge hand gesture data set

The Cambridge hand gesture dataset [28] is composed from 900 image sequences separated into nine hand gestures classes. The dataset is divided in five sets of illumination, that it is illustrated in Fig. 9.

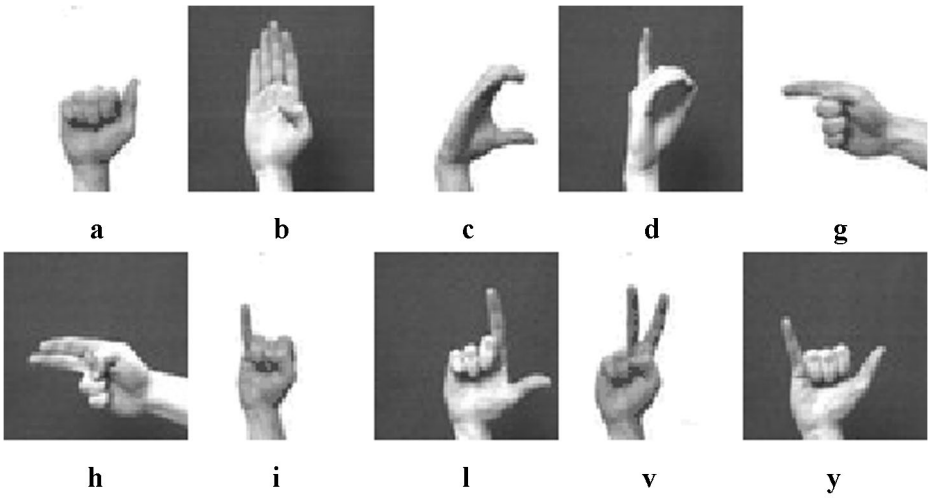

Fig. 7 The 10 hand posture signs from the Triesch database performed against light and dark background

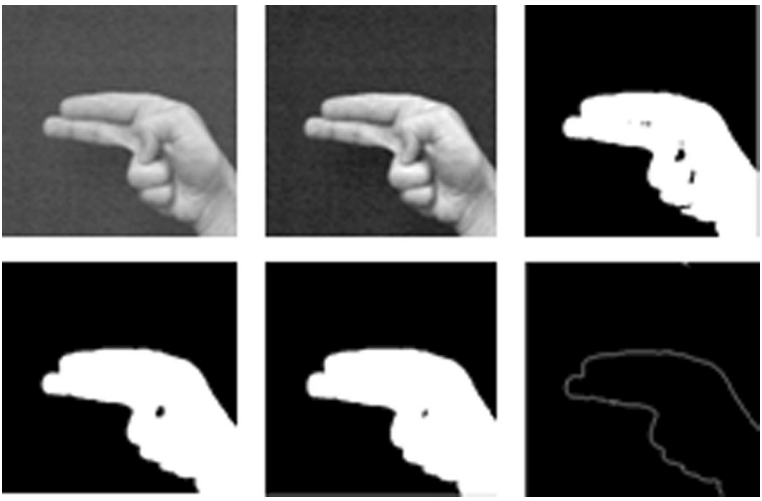

Fig. 8 Example of contour extraction for Triesch dataset. From left to right and top to bottom: Original image, Histogram equalization, Smoothing with Gaussian filter, Erosion, Dilatation, Contour extraction

4.1.3 Cambridge hand gesture database segmentation

Hand segmentation from image sequences is an important stage for gesture recognition using the shape based methods. It has a direct impact on the accuracy of the hand shape description and the classification process. The majority of approaches proposed for the recognition of gestures in Cambridge hand gestures dataset use the tensor-based methods, [28–30, 38, 41]. Only few works in the literature have attempted to segment this dataset, among the most recent we can cite [1, 2].

In this work we use to complement the skin color characteristics, the skin texture cues. This allows distinguishing between the skin regions and the light regions in video sequences due to the five different types of illumination contained in this data base.

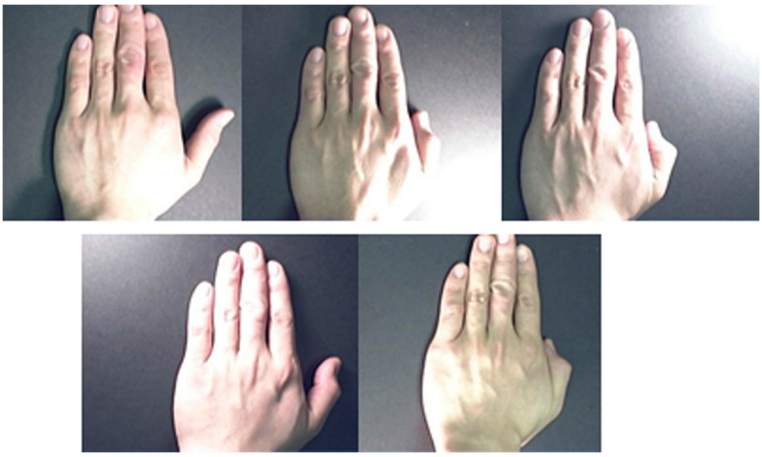

Fig. 9 The Five sets of illumination of Cambridge hand gestures dataset [28]

The use of color and texture information collectively has strong links with the human perception and in many practical scenarios the colour alone or texture alone is not sufficiently robust to accurately describe the image content [20]. To characteristic the colour, the chrominance components blue and red Cb and Cr are used, along with the hue H and saturation S components from the HSV colour space. The representation of the texture is based on two aspects: The spatiotemporal aspect using the Gabor filters [21], and the statistic one based on the Grey-Level Co-occurrences Matrices (GLCM) with the first 13 Haralick indices [17].

Skin colour modelling The orthogonal color spaces (YC_bC_r, YIQ, YUV, YES) reduce the redundancy present in RGB color channels and represent the color with statistically independent components [24]. In this work the first color space used is YC_bC_r.

The YC_bC_r represents the color as luminance Y and C_b , C_r the blue and red chrominance. Where the values Y, C_b and C_r are given by the relations:

$$\begin{cases} Y = 0.299\,R + 0.587G + 0.114B \\ C_b = 128 - 0.168736R - 0.331264G + 0.5B \\ C_r = 128 + 0.5R - 0.418688G - 0.081312B \end{cases} \tag{20}$$

The second color space used is HSV color space. The hue and saturation components are deeply inspired by human perception. The HSV color space is obtained by a nonlinear transformation of the RGB color space, given by equations below:

$$if \max(R,G,B) \neq \min(R,G,B) then:$$
$$H = \begin{cases} 60 \times \dfrac{\mathrm{G-B}}{\max(\mathrm{R,G,B}) - \min(\mathrm{R,G,B})} & \text{if } \mathrm{R} = \max(\mathrm{R,G,B}) \\ 60 \times \dfrac{\mathrm{g-b}}{\max(\mathrm{R,G,B}) - \min(\mathrm{R,G,B})} + 120 & \text{if } \mathrm{G} = \max(\mathrm{R,G,B}) \\ 60 \times \dfrac{\mathrm{g-b}}{\max(\mathrm{R,G,B}) - \min(\mathrm{R,G,B})} + 240 & \text{if } \mathrm{B} = \max(\mathrm{R,G,B}) \end{cases} \tag{21}$$

$$\mathrm{S} = \frac{max(R,G,B) - min(R,G,B)}{max(R,G,B)} \tag{22}$$

$$else\ H = S = 0$$

Skin texture modelling Texture is an important characteristic used in identifying regions of interest in image. In this work we combine spatiotemporal aspect with the statistical one.

At first an RGB patch of an image frame in video sequence is converted to HSV color space (using the formulas (21) and (22) cited upper). After, the Gabor filter transform is applied to the obtained HSV patch. Wavelet transform could extract both the time (Spatial) and frequency information from a given signal. Among kinds of wavelet transforms, the Gabor wavelet transform has both the multi-resolution and multi-orientation proprieties. Moreover the simple cells of the visual cortex of mammalian brains are best modelled as a family of self-similar 2D Gabor filters. The 2-D Gabor filter is defined as [21].

$$\varphi(x,y) = e^{-\frac{1}{2}\left(\frac{x_\theta^2}{\sigma_x^2}+\frac{y_\theta^2}{\sigma_y^2}\right)}\cos(2\pi f x_\theta) \tag{23}$$

with

$$x_\theta = x\cos\theta + y\sin\theta \; et \; y_\theta = y\cos\theta - x\sin\theta \tag{24}$$

where θ is the orientation of the major axis of the elliptical Gaussian, f the frequency, and σ_x, respectively σ_y are the standard deviation of the Gaussian envelope along x, y axes respectively. The Gabor filters used are from different orientations from 0 to π, with a step of $\frac{\pi}{6}$.

Then, the obtained patches are converted to grey scale images for computing the co-occurrences matrices (GLCM) [17]. Grey-Level Co-occurrences Matrices (GLCM) [17] is one of the most widely used approaches for the texture feature extraction and analysis.

Each value x at 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 θ. Formally for an $N \times M$ image f the normalized GLCM $P_{d,\theta}$ is given as:

$$P_{d,\theta}(i,j) = \frac{|\{(n,m) : f(n,m) = i, f(n + d\cos\theta, m + d\sin\theta) = j\}|}{N \times M} \tag{25}$$

Fourteen features were extracted by Haralick [17] from the (GLCM) to characterize texture. In this paper the first 13 indices are used namely(Angular Second Moment, Contrast, Correlation, Variance, Inverse Difference Moment, Sum Average, Sum Variance, Sum Entropy, Entropy, Difference Variance, Difference Entropy,Info. Measure of correlation 1, Info. Measure of correlation 2). Since rotation invariance is a primary criterion for all these statistics (indices), a kind of invariance can be obtained for each of these statistics by assigning their mean to the four directional co-occurrence matrices.

So the 13 indices are extracted from average matrix of the co-occurrences matrices (GLCM) computed with distance d = 1 and four different orientations 0, $\frac{\pi}{2}$, π, and $\frac{2\pi}{3}$. The maximum correlation coefficient was not calculated due to computational instability.

Finally the characteristics vector composed by the 13 Haralick indices, Cb, Cr, H, and S for the color representation is used to train Artificial Neural Network classifier(ANN) for skin pixel detection.

The neural network is deployed with patch size 8 × 8 (pixel). However, during the detection process, since the images in the video sequences of the Cambridge hand gestures dataset [28] are small, the analysis patch is 4 × 4 (pixel) in size.

After running many simulations, the architecture of the best ANN obtained is with 17 input nodes, 4 Hidden Layers and 6 neurons. The correct prediction rate on the test data is about 98.89%. The Fig. 10 shows an example of obtained results.

Since the segmented hand has the largest area, the widest contour has been drawn and, to smooth out its shape, we used a median blur filter.

4.2 Performance of our proposed approach based Tchebichef moments

In order to compare the performance of our proposed representation based moments with the original representation of corner points, we test the recognition performance on hand postures

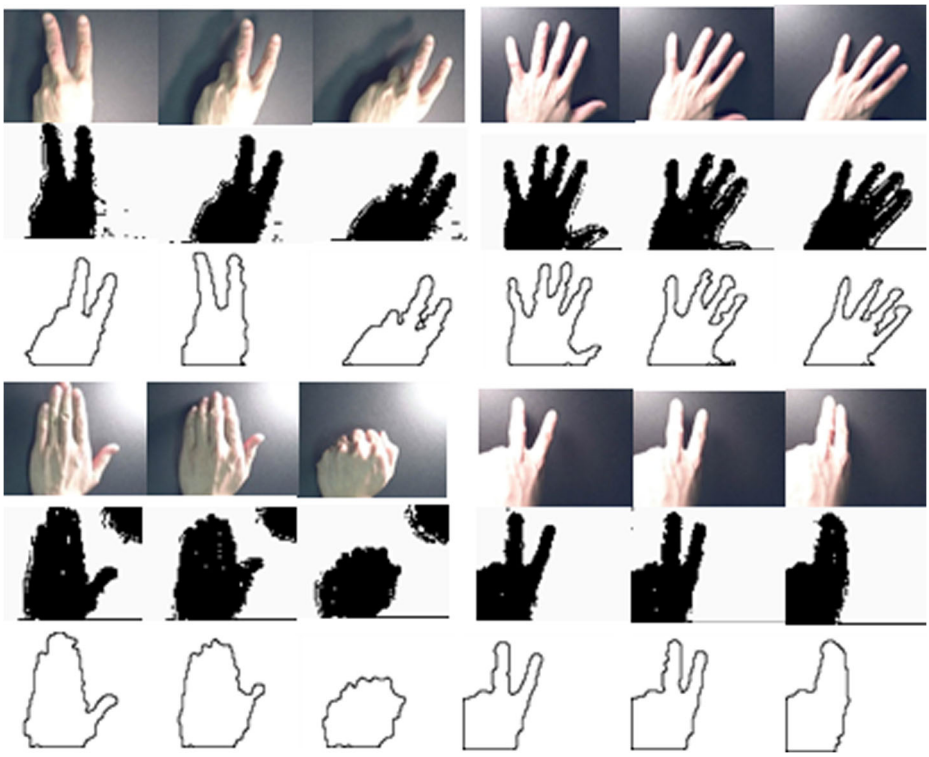

Fig. 10 Example of segmented images from Cambridge dataset

of Triesch database of two size functions inducted from two measuring functions namely the distance from the center of gravity and the distance from the right bottom point of the minimum rectangle of hand shape (see Fig. 11). The Tchebichef moments were calculated up to 7th order. The hand postures of Triesch database were classified using 1-nn classifier and the distance used is the Euclidian distance, with the same protocol used in [23]: 8 subjects in the reference dataset and 12 remaining subjects in the test data-set.

The obtained results show that for both measuring functions and for most letters, the best recognition rate was achieved with the proposed representation based moments, despite the fact that the order used is relatively small namely up the 7th order, except for the letter B in the size function corresponding to the distance from the center of gravity measuring function and letter Y in the size function corresponding to the distance from the right bottom point measuring function. We can see also that the letters B, C, I, L and V were not recognized at all using the corner point representation of the right bottom point size function, while we obtained a recognition rate more than 10% using the Tchebichef moments and who has even reached 100% for some letters. In both cases the global recognition rates are low; this is due the 1nn classifier used.

4.3 Results on the Triesch database

The classification was performed using support vector machine classifier (SVM) as it was explained in the previous section 3.2.4. We trained the SVM's on each of the 10 hand signs of

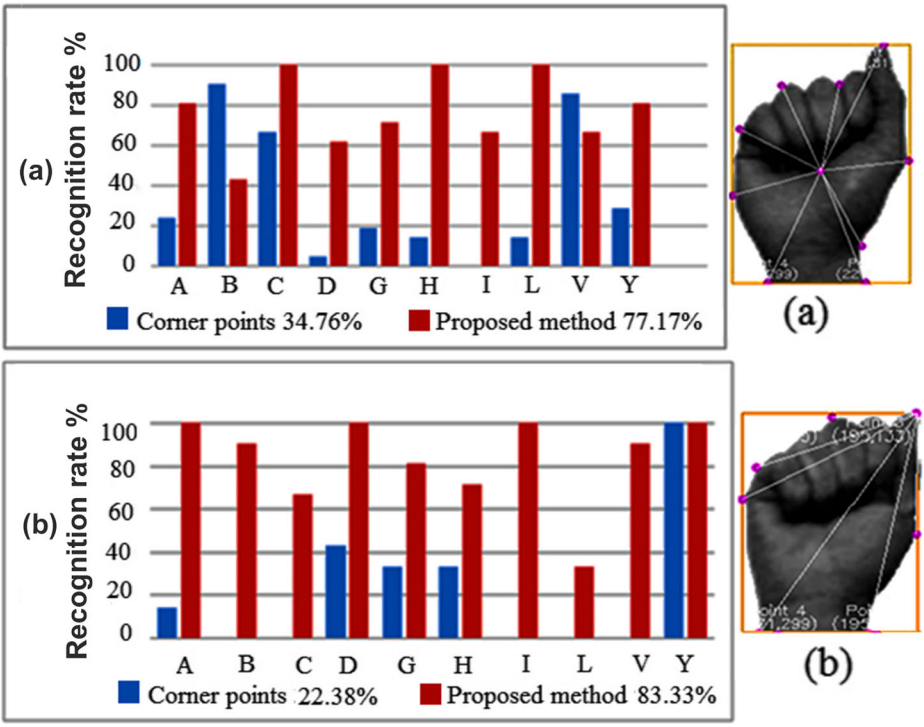

Fig. 11 **a** The centre of gravity size function, **b** Right bottom point of the minimum rectangle size function

the Triesch database. The SVM's received as input a vector of the 12 size functions corresponding to the distance measuring functions and projections (see section 3.2.2). Each size function is represented by a Tchebichef moments vector up to 7th order with 36 coordinates. Different types of SVM kernels were tested (polynomial, Gaussian and linear). The linear support vector machine gave the best results. The Fig. 12

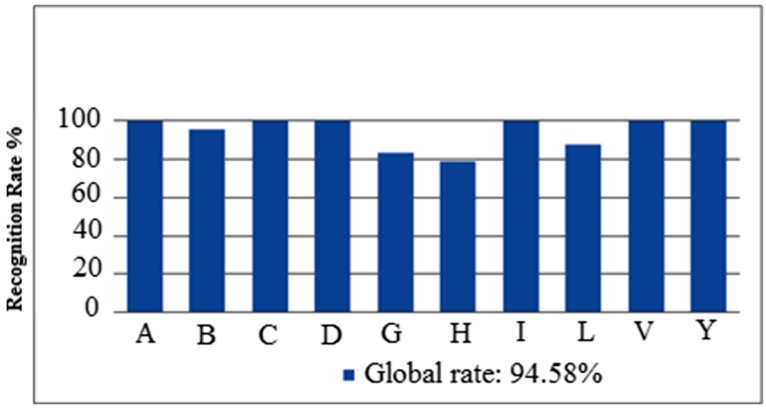

Fig. 12 Recognition rates in Treisch database

presents the detailed results for each letter. We noticed that for 6 letters out of 10, the recognition rate was 100%. Low rates are obtained for the letters G and H. Mostly our approach obtained a recognition rate of 94.58% in user independent scheme; the persons in the training database are different from those in the test database. The convexity approach in the outline's discretization for the computation of the measuring functions, allowed the system to cope with the different anatomies of hands of different persons.

Another manner to illustrate the results obtained by the proposed hand gestures recognition system is the computation of the confusion matrix. Table 1 presents the confusion matrix obtained when testing the proposed system on Triesch database [42].

We remark that the letters G and H are the most confused letters in the proposed hand gestures recognition system because of the similarity between the two letters. Six letters among 10 letters were recognized entirely.

Additionally, some metrics are computed to measure the classification performance of the proposed system. The metrics used are: Recall, Precision, F-measure and Specificity. The computed metrics are presented in the Table 2.

It is interesting to underline that the selected metrics highlight different aspects of the pattern classification methods. The obtained results are generally satisfactory.

Finally, we have compared our results with the pertinent state of art works for hand postures recognition in Treisch database: Modified Census Transform [23], Eigenspace representation of size function [26], Fourier Descriptor with K-NN classifier [4].

In our system we respect the same protocol used in [23], 8 subjects in the training dataset and 12 remaining subjects in the test data-set. The conditions of the system in [23] were not clearly specified; otherwise the method in [26] was tested on two datasets: the Triesch database, and an Irish sign language dataset, where the user should wear a colored glove. In the technique in [4], some restrictions were imposed among them: the gestures should be realized over a desk and the users wear long sleeve clothes. Table 3 shows the results obtained by each technique.

As the Table 3 shown, our method obtained the best result comparing to the other methods. Compared to the eigenspace representation of size function proposed by Kelly et al. (2010) in [26] where the tests was performed on the same database with the same Protocol in user independent mode, our representation of size functions based Tchebichef moments achieved a better result. This is due to the robustess of our representation wich by the computation of Tchebichef moments

Table 1 Confusuion Matrix obtained on Triesch database

	a	b	c	d	g	h	i	l	v	y
a	1	0	0	0	0	0	0	0	0	0
b	0.01	0.95	0	0.02	0	0	0.01	0.01	0	0
c	0	0	1	0	0	0	0	0	0	0
d	0	0	0	1	0	0	0	0	0	0
g	0	0	0	0	0.83	0.17	0	0	0	0
h	0	0	0	0	0.21	0.79	0	0	0	0
i	0	0	0	0	0	0	1	0	0	0
l	0.05	0	0	0	0	0	0	0.87	0	0.08
v	0	0	0	0	0	0	0	0	1	0
y	0	0	0	0	0	0	0	0	0	1

Table 2 Quantitative evaluation of the proposed system on Triesch database

Measures	Recall	Precision	F-measure	Specificity
Obtained Results	94.4	93.81	94.1	93.13

on the encoded graph of size functions take into account all the different values of the size function and their coordinates in the plane.

4.4 Results on the hand gestures Cambridge database

For the gestures recognition system, each gesture in the Cambridge database is considered as a set of pertinent postures (partial gesture), We train SVM's for each pertinent posture using a feature vector of the Tchebichef moments up to 7th order applied to the 12 size functions on pertinent posture frames. The gesture is then recognized through the SVM classification of the pertinent postures that compose it.

Table 4 presents the confusion matrix obtained, when testing the proposed gesture recognition system on Cambridge database. The last column of the matrix indicates the 9 different classes of gestures in the database.

We remark that 5 gestures among 9 were correctly recognized. We remark also that generally the confusion is the most important for the gestures with strong correlation in the hand postures frames, as for example for the gestures spread- Contract and flat -Contract, or gestures V-shape-Rightward and Flat-Rightward.

We note however, some exceptions, as for example for the gestures flat-Contract and V-shape-Rightward this is principally due to the bad segmentation of the hand region in a few videos images.

We also present in the Table 5 some quantitative results using the measures selected in Table 2, in order to evaluate the proposed hand gestures recognition system when testing on the Cambridge database.

The different metrics give satisfactory results which reflect a good performance of the proposed system.

The obtained results are then compared with some techniques proposed in the litterature which respect the experimental procedure described in [28], the set 5 for training and the other sets for testing, namely: Tensor Canonical Correlation Analysis (TCCA) [27], Product Manifolds (PM) [35], Tangent bundles (TB) [34], space–time auto-correlation of gradients (STACOG) [30]; spatio-temporal covariance descriptors (Cov3D) [18]; DTW + CSURF [2], Dense Trajectories (DT) [1], Spatial-Temporal Iterative Tensor (STIT) [41],(D-C T) Dual complementary tensors [19]. Table 6 shows the results comparison.

Our method obtained a best result than 7 methods among the 9 methods cited, and the same result as the Spatial –Temporal Iterative technique (S-TITD), only the Dual Complementary

Table 3 Comparison between different techniques applied to the Triesch database

Techniques	Rate (%)
MCT [23]	92.79
Eigenspace Size functions [26]	91.8
Fourier Descriptor K-NN [4]	93.3
Size functions based moments	94.58

Table 4 Confusion Matrix obtained on Cambridge database

	FL	FR	FC	SL	SR	SC	VL	VR	VC	Hand gestures classes
FL	1	0	0	0	0	0	0	0	0	FL: Flat-Leftward
FR	0	0.9	0	0	0.05	0.05	0	0	0	FR: Flat-Rightward
FC	0	0	0.85	0	0	0.05	0.02	0.05	0.03	FC: Flat-Contract
SL	0	0	0	1	0	0	0	0	0	SL: Spread-Leftward
SR	0	0	0	0	1	0	0	0	0	SR: Spread-Rightward
SC	0.01	0	0.08	0.02	0	0.85	0	0	0.04	SC: Spread-Contract
VL	0	0	0	0	0	0	1	0	0	VL: V-shape-Leftward
VR	0	0.07	0	0	0.03	0	0	0.9	0	VR: V-shape-Rightward
VC	0	0	0	0	0	0	0	0	1	VC: V-shape-Contract

tensors technique (DC-T) achieved a better result than our approach. Differently from all the techniques cited, in our method a hand segmentation process was developed. The segmentation approach proposed can be applied for different databases performed against complex backgrounds. All the techniques cited in the Table 3 apart from DTW + CSURF and (DT) techniques, are based on the pixel intensity change in video, so they do not take into account the shape deformation.. Our model is based on the topological aspect of the hand posture shape and can be useful in different hand gestures recognition systems, where a huge of hand shapes composed the gesture. This capability is not be possible with the pixel intensity change methods.

4.5 Time computation performance

The computation time in the proposed hand gestures recognition system was optimized using two approaches. The first one was in the contour discretization using the convexity approach based on Douglas-Peucker polygon [10]. This approach allows to the system to calculate the measuring function values only in some selected points, and to reduce the size graph algorithm of the size function computation. The second approach was in the Tchebichef moments computation algorithm proposed in [40] for grey-scale images and which we well adapted to the encoded graph of size function. The algorithmic complexity of the Douglas-Peucker polygon is polynomial of degree 2. The algorithmic complexity of the method proposed in [40] has a polynomial complexity of degree1 according to the number of blocks. In the case of the encoded graph of size functions, the number of blocks is generally small, which allows to faster the computation of the Tchebichef moments. We use Python to implement our proposed method, the computation time is measured on Intel Core 3 2.27 GHZ CPU.

The Table 7 presents the average computational cost performance against some related works in Triesch database [42]. The computation time is presented per single image.

We remark that the proposed method is faster than the related works when testing in the hand postures recognition task, in Triesch database.

In Table 8, we present the computation time performance of the proposed hand gestures system in Cambridge database compared to some related works.

Table 5 Quantitative evaluation of the proposed system on Cambridge database

Measures	Recall	Precision	F-measure	Specificity
Obtained Results	95.93	94.44	95.17	95.43

Table 6 Comparison between different techniques applied to the Cambridge database

Techniques	Rate
(TCCA) [27]	0.82
(PM) [35]	0.88
(TB) [34]	0.91
(STACOG+Sobel) [30]	0.92
(Cov3D) [18]	0.93
DTW + CSURF [2]	0.93
(DT) [1]	0.94
(S-TITD) [41]	0.95
Size Functions based moments	0.95
(D-C T) [19]	0.97

The computation time of the proposed approach is interesting compared to the low material performance used.

4.6 Unknown hand gesture classification

It is interesting to test the performance of a hand gestures recognition system face to unknown hand gesture. In the proposed hand gestures recognition system, the classification is performed using an extended SVM to obtain class probabilities (see section 3.2.4). The algorithm tends to classify the unknown gesture to the closest existent class, but this classification is carried with very small probability. In this case, it will be judicious to reject the gesture rather to classify it with an important probability of error. In order to test the proposed system face to these situations, we have varied the threshold of acceptance and reject, according to the probability that the gesture is in the assigned class. If this probability is inferior to the threshold the gesture is rejected else it is classified. Table 9 presents the results obtained for different thresholds, when testing on some unknown gestures in Cambridge database [28].

We remark that if we impose to the unknown gesture to be assigned to a gesture class only if the corresponding probability is superior to a given threshold, the system can be able to distinguish the inappropriate or erroneous gestures from the true gestures of the database.

4.7 Discussion

In this section we present the advantages and disadvantages of the proposed method against some related works. This comparison is carried out firstly according to the new size function representation based on Tchebichef moments and secondly according to the architecture of the proposed hand gestures recognition system. The size function representation based on set of corner points is inspired by the algebraic representation of size functions as formal series [16]. This representation has an advantage to reduce all the

Table 7 Computation time comparison in Triesch database

Techniques	Hard used	Computation time
Eigenspace Size functions [26]	2.16 GHz Intel Core 2 CPU.	60 ms
Fourrier Descriptor (knn) [4]	2GHz PC	10 ms
Proposed method	2.27GHZ Intel core 3CPU	5 ms

Table 8 Computation time comparison in Cambridge database

Techniques	Hard used	Computation time
DTW + Csurf [2]	unknown	79.75 ms
DT [1]	Work station i7–2600 CPU that runs at 3.40 GHz	1091,71 ms
DT [1]	Embeded system ARM processor (Cortex A15 and A7)	787,68 ms
Proposed method	2.27GHZ Intel core 3 CPU	4576 ms

information held in size functions in the computation of finite set of points and to do the shape comparison using the matching distance developed in [7]. A corner point (u, v) is encoding the level u at which a new connected component is born and the level v at which it get merged to another connected component. Consequently the combinatorial representation based on corner points do not take into account the value of size function at the point (u, v), and so do not translate all the information held in the encoding graph of the size function. On the other hand, a given size function graph has not always proper corner points. A further critical insight is the existence of a pairing in which positive simplices mark the appearance (birth) of topological features while negative simplices mark their disappearance (death) as indicated in the survey [11]. D. Kelly et al. proposed to use the eigenspace information using principal component analysis (PCA) [26]. This representation has an advantage to be very efficient for reducing the dimensionality of data. However, PCA assumes that the principal components are a linear combination of the original features. If this is not true, PCA will not give sensible results. In this paper we propose to use the Tchebichef moments to collect the salient information held in size functions. Moments are scalar quantities used to characterize a function and to capture its significant features [14]. So it will be appear judicious to describe the size functions using moment's theory The main advantage of the proposed representation is that it takes into account all the information held in size function. The most important disadvantage is that the computation of the Tchebichef moments can be very expensive in time computation. This disadvantage was be seriously reduced when adapting the algorithm proposed in [40] to faster the computation of Tchebichef moments.

The architecture of the proposed hand gestures recognition system contains three important steps (hand region location, features extraction and classification). The hand segmentation and location in the image video is based on the color and texture attributes along with an artificial neural networks ANN. Differently from the segmentation methods proposed in [1, 2] which depend to the Cambridge database, our proposed segmentation method has an advantage to be applicable for any another hand gestures database. The important inconvenient of the methods based on hand region segmentation according to the methods cited in [18, 19, 27, 30, 34, 35, 41], is that the recognition accuracy depends strongly on the hand region segmentation efficiency. However the intensity change pixel methods have a disadvantage to don't take into account the hand postures recognition task, the system proposed can be applied to hand gestures and postures recognition.

Table 9 Rejection rate with unknown gestures on Cambridge database

Threshold	0.1	0.2	0.3	0.4	0.5
Rejection rate	45%	52.5%	67.5%	75%	82.5%

5 Conclusion

This study proposes a new representation of reduced size functions which is based on the computation of the Tchebichef moments from the size function's encoded graph. The encoded graph of size function contains all information about its corresponding measuring function and the geometric properties that it translates. The computation of the Tchebichef moments from the encoded graph captures all different values that the size function can take and their emplacement in the plane. A proposed technique for the computation of the Tchebichef moments from grey scale images is adapted in our system for the region of interest in the encoded graph of size functions. This allows speeding up the computational efficiency of Tchebichef moments.

The shape contour discretization is one of the problems posed in the size function theory. To deal with this problem our model uses the convexity approach technique to select the salient points that can represent the hand shapes.

In addition we propose a set of measuring functions, which proved to be interesting in the hand shapes classification. We also introduce an accurate hand region segmentation process based on color and texture attributes of skin pixels to detect hand in video frames.

Our segmentation technique can be applied in different databases with complex backgrounds and under variable lighting conditions.

Compared of the other state-of –the- art methods in both the Cambridge Hand Data and the Triesch database, our method competes the best ranked methods, which proves that it can deal with static and dynamic gestures.

In future research, we plan to adapt our approach to real time. This is motivated by the increasing interest in ego-centric human-machine interfaces.

References

1. Baraldi L, Paci F, Serra G, Benini L, Cucchiara R (2014) Gesture recognition in ego-centric videos using dense trajectories and hand segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops. IEEE, pp. 702–707
2. Barros P, Maciel-Junior NT, Fernandes BJT, Bezerra BLD (2017) A dynamic gesture recognition and prediction system using the convexity approach. Comput Vis Image Underst 155:139–149
3. Biasoti S, Cerri A, Frosini P, Giorgi D (2011) A new-algorithm for computing the 2-dimentional matching distance between size functions. Pattern Recogn Lett 32(14):1735–1746
4. Bourennane S, Fossati C (2012) Comparison of shape descriptors for hand posture recognition in video. SIViP 6:147–157
5. Chan L, Liang R-H, Tsai M-C, Cheng K-Y, Su C-H, Chen MY, Cheng W-H, Chen B-Y (2013) Finger Pad private and subtle interaction using fingertips. In: Symposium on User Interface Software and Technology, ACM, pp. 255–260
6. Cheng J, Chen X, Liu A, Peng H (2015) A novel phonology-and radical-coded chinese sign language recognition framework using accelerometer and surface electromyography sensors. Sensors 15(9):23303–23324
7. d'Amico M, Frosini P, Landi C (2006) Using matching distance in Size Theory: A survey. Int J Imaging Syst Technol 16(5):154–161
8. Dibos F, Frosini P, Pasquignon D (2004) The use of size functions for comparison of shapes through differential invariants. J Math Imaging Vis 21(2):107–118
9. Dominio F, Donadeo M, Zanuttigh P (2014) Combining multiple depth-based descriptors for hand gesture recognition. Pattern Recogn Lett 50(C):101–111
10. Douglas DH, Peucker TK (1973) Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. Cartographica 10(2):112–122

11. Edelsbrunner H, Harer J (2008) Persistent homology—a survey. In: Surveys on discrete and computational geometry, vol. 453 of Contemp. Math. Amer. Math.Soc., Providence, pp. 257–282
12. Erol A, Bebis G, Nicolescu M, Boyle RD, Twombly X (2007) Vision-Based Hand Pose Estimation : A review. Comput Vis Image Underst 108(1–2):52–73
13. Flasinski M, Myslinski S (2010) On the use of graph parsing for recognition of isolated hand postures of polish sign language. Pattern Recogn 43(6):2249–2264
14. Flusser J, Zitova B, Suk T (2009) Moments and moment invariants in pattern recognition. John Wiley & Sons, Ltd, Hoboken
15. Frosini P (1992) Discrete computation of size functions. J Combin Inform System Sci 17(3–4):232–250
16. Frosini P, Landi C (2001) Size functions and formal series. AAECC 12(4):327–349
17. Haralick R, Shanmugan K, Dinstein I (1973) Textural features for image classification. IEEE Transactions On Systems, Man, and Cybernetics SMC 3:610–621
18. Harandi MT, Sanderson C, Sanin A, Lovell BC (2013) Spatio-temporal covariance descriptors for action and gesture recognition. In: Proceedings of the 2013 IEEE Workshop on Applications of Computer Vision (WACV). IEEE Computer Society, Washington, DC, pp. 103–110
19. Hsieh CY, Lin WY (2017) Video-based human action and hand gesture recognition by fusing factored matrices of dual tensors. Multimed Tools Appl 76(6):7575–7594
20. Ilea DE, Whelan PF (2011) Image segmentation based on the integration of colour-texture descriptors- A review. Pattern Recogn 44(10–11):2479–2501
21. Jones JP, Palmer LA (1987) An evaluation of the two dimensional Gabor filter model of simple receptive fields in cat striate cortex. J Neurophysiol 58(6):1233–1258
22. Ju Z, Gao D, Cao J, Liu H (2016) A novel approach to extract hand gesture feature in depth images. Multimed Tools Appl 75(19):11929–11943
23. Just A, Rodriguez Y, Marcel S (2006) Hand posture classification and recognition using the modified census transform. In: 7th Internat. Conf. on Automatic Face and Gesture Recognition, FGR, pp. 351–356. doi: https://doi.org/10.1109/FGR.2006.62
24. Kakumanu P, Makrogiannis S, Bourbakis N (2007) A survey of skin-color modeling and detection methods. Pattern Recogn 40:1106–1122
25. Kane L, Khanna P (2015) A framework for live and cross platform fingerspelling recognition using modified shape matrix variants on depth silhouettes. Comput Vis Image Underst 141:138–151
26. Kelly D, McDonald J, Markham C (2010) A person independent system for recognition of hand postures used in sign language. Pattern Recogn Lett 31(11):1359–1368
27. Kim T-K, Cipolla R (2009) Canonical correlation analysis of video volume tensors for action categorization and detection. IEEE Trans Pattern Anal Mach Intell 31(8):1415–1428
28. Kim T-K, Wong S-F, Cipolla R (2007) Tensor canonical correlation analysis for action classification. IEEE Conference on Computer Vision and Pattern Recognition:1–8
29. Kobayashi T, Otsu N (2009) A three-way auto-correlation based approach to motion recognition. Pattern Recogn Lett 30(3):185–192
30. Kobayashi T, Otsu N (2012) Motion recognition using local auto-correlation of space–time gradients. Pattern Recogn Lett 33:1188–1195
31. Li C, Xie C, Zhang B, Chen C, Han J (2018) Deep Fisher discriminant learning for mobile hand gesture recognition. Pattern Recogn 77:276–288
32. Lin H-T, Lin C-J, Weng RC (2007) A note on Platt's probabilistic outputs for support vector machines. Mach Learn 68(3):267–276 ISSN 0885-6125
33. Liu S, Xiao Q (2015) A signer-independent sign language recognition system based on the weighted KNN/KMM. In: Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2015 7th International Conference on, vol. 2, pp. 186–189
34. Lui YM, Beveridge JR (2011) Tangent bundle for human action recognition. In proc. of Automatic Face & Gesture Recognition and Workshops, pp. 97–102
35. Lui YM, Beveridge JR, Kirby M (2010) Action classification on product manifolds. In Proc. of CVPR, pp. 833–839
36. Mhamedi MAA, Ziou D (2014) A local approach for 3D object recognition through a set of size functions. Image Vis Comput 32(12):1030–1044
37. Mukundan R (2004) Some computational aspects of discrete orthonormal moments. IEEE Trans Image Process 13(8):1055–1059
38. Niebles J, Wang H, Fei-Fei L (2006) Unsupervised learning of human action categories using spatial-temporal words. In: British Machine Vision Conference, pp 1249–1258
39. Papakostas GA, Karakasis EG, Koulourisotis DE (2008) Efficient and accurate computation of geometric moments on grey-scale images. Pattern Recogn 41(6):1895–1904

40. Shu H, Zhang H, Beijing C, Haigron P, Luo L (2010) Fast computation of Tchebichef moments for binary and greyscale images. IEEE Transactions on Image Processing, Institute of Electrical and Electronics Engineers 19(12):3171–3180
41. Su Y, Wang H, Jing P, Xu C (2017) (2017) A spatial-temporal iterative tensor decomposition technique for action and gesture recognition. Multimed Tools Appl 76:10635–10652
42. Triesch J, Von der malsuburg C (2002) Classification of Hand postures Against Complex Backgrounds Using Elastic Graph Matching. Image Vis Comput 20(13–14):937–943
43. Verri A, Uras C (1996) Metric-topological approach to shape representation and recognition. Image Vis Comput 14:189–207
44. Verri A, Uras C, Frosini P, Ferri M (1993) On the use of size functions for shape analysis. Biol Cybern 70: 99–107
45. Wang C, Liu Z, Zhu M, Zhao J, Chan S-C (2017) A hand gesture recognition system based on canonical superpixel-graph. Signal Process Image Commun 58:87–98
46. Wang JW, Wang CC, Lee JS (2013) Genetic Eigen Hand Selection for hand shape classification based on compact hand extraction. Eng Appl Artif Intell 26:2215–2226
47. Yang C, Han D-K, Ko H (2017) Continuous hand gesture recognition based on trajectory shape information. Pattern Recogn Lett 99:39–47
48. Yang H-D, Sclaroff S, Lee S-W (2009) Sign language spotting with a threshold model based on conditional random field. IEEE Trans Pattern Anal Match Intell 31(7):1264–1277

Djamila Dahmani received the DES degree in 1994 and Magister degree in Algebra and Number Theory in 1997 from the mathematical faculty of University of Science and Technology Houari Boumediene (USTHB), and PhD degree in computer vision from the faculty of Electronics and Computer science in USTHB in 2014. His recent works concentrate on hand gesture and posture recognition, human machine interaction and image processing. She is currently Assistant Professor in the department of computer science of USTHB.

Slimane Larabi received the PhD degree in computer vision from the National Polytechnic School of Toulouse, France, in 1991.He is currently Professor in the department of Computer science of the University of Science and Technology Houari Boumediene (USTHB). His current researches include Shape description and Recognition Contour detection and Image segmentation Human gesture and action recognition Image and video database indexing and retrieval Image Analysis and Pattern recognition.

Mehdi Cheref received the B.S. degree and the M.S. degree from University of Science and Technology – Houari Boumediene Algiers (**USTHB**), Algeria, in 2016 and 2018 respectively, the B.S in Information Systems and Software Engineering and the M.S. in Visual Computing, He is currently a PHD candidate in USTHB. His research interests include artificial intelligence, multimedia content and computer vision.