RESEARCH ON BIOMEDICAL ENGINEERING
rbejournal.org
Volume 33, Number 1, p. 78-89, 2017

Original Article
DOI: http://dx.doi.org/10.1590/2446-4740.03816

# A fully automatic method for recognizing hand configurations of Brazilian sign language

Cicero Ferreira Fernandes Costa Filho[1]*, Robson Silva de Souza[1], Jonilson Roque dos Santos[1], Bárbara Lobato dos Santos[1], Marly Guimarães Fernandes Costa[1]

[1] *Center for Research and Development in Electronic and Information Technology, Federal University of Amazonas, Manaus, AM, Brazil.*

**Abstract** **Introduction:** Sign language is a collection of gestures, postures, movements, and facial expressions used by deaf people. The Brazilian sign language is Libras. The use of Libras has been increased among the deaf communities, but is still not disseminated outside this community. Sign language recognition is a field of research, which intends to help the deaf community communication with non-hearing-impaired people. In this context, this paper describes a new method for recognizing hand configurations of Libras - using depth maps obtained with a Kinect® sensor. **Methods:** The proposed method comprises three phases: hand segmentation, feature extraction, and classification. The segmentation phase is independent from the background and depends only on pixel value. The feature extraction process is independent from rotation and translation. The features are extracted employing two techniques: $(2D)^2$LDA and $(2D)^2$PCA. The classification employs two classifiers: a novelty classifier and a KNN classifier. A robust database is constructed for classifier evaluation, with 12,200 images of Libras and 200 gestures of each hand configuration. **Results:** The best accuracy obtained was 96.31%. **Conclusion:** The best gesture recognition accuracy obtained is much higher than the studies previously published. It must be emphasized that this recognition rate is obtained for different conditions of hand rotation and proximity of the depth camera, and with a depth camera resolution of only 640×480 pixels. This performance must be also credited to the feature extraction technique, and to the size standardization and normalization processes used previously to feature extraction step.

**Keywords** Deaf community, Sign language, Gesture recognition, Novelty classifier, kNN classifier, Libras.

## Introduction

Although the use of sign language is very popular among deaf people, other non-hearing-impaired communities do not even try to learn it, causing isolation of deaf people. Developing a system to translate sign language would be a helpful solution to this problem. Deaf communities of each country have different sign languages. Even countries with the same language may have different sign languages. For example, Brazil and Portugal have the same oral language, the Portuguese; nevertheless, the deaf communities of each country have their own sign language, Brazilian Sign Language –Libras and Portuguese Gesture Language – LGP, respectively.

In the first studies of gesture recognition 2D images, obtained with conventional cameras, were used. With 2D images, some approaches were used to facilitate hand segmentation. In some of the earliest works (Al-Jarrah and Halawani, 2001; Carneiro et al., 2009; Neris et al., 2008; Pizzolato et al., 2010), the authors employed images with a homogeneous background, of white or black color. In a second approach, the authors Bragatto et al. (2006), and Maraqa et al. (2012) employed colorful gloves. More recently, depth maps obtained with Kinect®-like depth sensors have also been used for hand gesture recognition (Dong et al., 2015; Lee et al., 2016; Rakun et al., 2013; Silva et al., 2013). The gesture segmentation with depth maps is supposed to be independent of scene illumination and background.

Dong et al. (2015) developed a recognition method for 24 American Sign Language Alphabet (excluding the dynamic signs "j" and "z"). Their approach comprised the following steps: 1) pixels classification of Kinect® depth image as belonging to hand or background, using a random forest classifier. The hand was divided

*How to cite this article:* Costa CFF Fo, Souza RS, Santos JR, Santos BL, Costa MGF. A fully automatic method for recognizing hand configurations of Brazilian sign language. Res Biomed Eng. 2017; 33(1):78-89. DOI: 10.1590/2446-4740.03816

**Corresponding author:* Center for Research and Development in Electronic and Information Technology, Federal University of Amazonas, Campus Universitário – Setor Norte, Pavilhão Ceteli, Av. General Rodrigo Otávio Jordão Ramos, 3000, Aleixo, CEP 9077-000, Manaus, AM, Brazil. E-mail: ccosta@ufam.edu.br

*Received:* 26 May 2016 / *Accepted:* 27 March 2017

in 11 regions; 2) finger joints identification using the mean-shift local mode-seeking algorithm. This algorithm estimates the mass center of probability distributions of each hand region; 3) the hand gesture is then classified using a 13-feature joint vector as input of random forest classifier. The best accuracy obtained in the recognition task was 90%.

Silva et al. (2013) also employed the Kinect® sensor for building a database of the 26 alphabet symbols of American Signal Language. For the pattern recognition task, the authors employed a template matching technique. With template matching, the authors do not present feature vectors. The main contribution of the paper is the evaluation and discussion of some comparison metrics between two templates. The best accuracy obtained in the recognition task is 99.03%.

In the method proposed by Dong et al. (2015) the main difficult is to obtain the finger joint angles. This contributes to the low accuracy presented. While the proposed method by Silva et al. (2013) requires a perfect alignment between two templates. For accomplish this task, the templates had to be obtained at the same distance from Kinect® and at the same position. The method just presented in this paper overcomes these two limitations. Additionally, both papers recognized a sequence of alphabet letters. It must be emphasized that deaf people only use finger spelling, to represent given names, acronyms or some technical or specialized vocabulary. Differently, in this paper we recognize hand gestures used in Brazilian Sign Language.

Rakun et al. (2013) extracted three features: hand shape, hand position and movement direction from Kinect® depth image. For recognition task, the Random Forest classifier and the Generalized Learned Vector Quantization (GLVQ) were employed. The authors used three recognition approaches, tested with a 10 words dataset of Indonesian sign language. The first approach uses only hand-shape data. The second one uses skeleton data. While the third one, combining the previous two approaches, obtained the best result, an accuracy of 94.37%.

The study of Lee et al. (2016) adopted a similar approach to the study of Rakun et al. (2013). The authors also determine the hand-shape, hand position and the movement direction from Kinect® sensor data. The hand position, an important parameter in Taiwanese sign language, is determined using skeleton information and a decision tree. The hand shape is determined using principal component analysis and a Support Vector Machine classifier. The movement direction is obtained using hidden Markov models. Twelve direction classes are used. For deciding the final word recognition, a confusion matrix is employed to construct a probabilistic matrix. The best accuracy obtained by the authors in recognition hand shapes is only 86.94%, and in the classification of 25 words, 85.14%.

The studies of Rakun et al. (2013) and Lee et al. (2016) adopted a different approach from the two previous related studies and from the study just presented in this paper. Instead of recognizing language symbols, these studies recognize words. Both of them used a limited casuistic to validate their methods. The accuracy obtained by Lee et al. (2016) in hand shape recognition was only 86.94%. Furthermore, these two studies do not take into account all the phonologic parameters of a hand sign language, which are described in the next paragraphs.

Another approach that does not use digital images for gesture recognition employs gloves with electrical sensors (Mehdi and Khan, 2002; Wang et al., 2006). Mehdi and Khan (2002) used seven electrical sensor signs: five from the fingers, one to measure the tilt of the hand and one to measure the rotation of the hand.

The major gesture recognition studies previously published in the literature employ different techniques for extracting characteristics: Al-Jarrah and Halawani (2001) used radial distances from gesture center to gesture border; Peres et al. (2006) employed bit signature; Neris et al. (2008) used 22 vectors with 236 coordinates corresponding to pixel intensity sums in horizontal and vertical directions; Carneiro et al. (2009) cropped the hand gesture to a region of 25x25 pixels and Pizzolato et al. (2010) extracted Hu invariant moments.

In this paper, we are concerned about the recognition of the Libras. According to the Brazilian Institute of Statistics and Geography – IBGE (Brazilian population census 2010), Brazil's population of hearing impaired people is 9.7 million, about 5% of the total population. From this total, about 1.7 million have great difficulty hearing, 344,200 are deaf and 7.5 million have some hearing impairment. Some previous studies found in the literature on Libras (Carneiro et al., 2009; Neris et al., 2008; Peres et al., 2006; Pizzolato et al., 2010) aim to translate only gestures of Portuguese alphabet letters. A sequence of alphabet letters – of finger spelling - is used by deaf people only to represent given names, acronyms or some technical or specialized vocabulary. According to Brito (2010), this is a linguistic loan and does not solve the communication problem of deaf people. Aware that the Libras language is not Portuguese letter spelling, many authors have developed studies on the phonology of sign language. According to Rossi and Rossi, cited by (Anjo, 2013), a sign language gesture is formed by combining five phonologic parameters: hand configuration, articulation point, orientation, movement and facial expression. Following this logical reasoning, this study proposes an initial step in developing a full recognition system for Libras, the recognition of one of these phonologic parameters, the Hand Configuration

(HC). HC is considered the main parameter, because it is present in almost all signs of Libras.

According to Pimenta and Quadros (2010), Libras has 61 HC. These configurations are shown in Figure 1. As noted in this figure, Libras has several similar HC. For example, some interpretation mistakes could occur between the hand configurations 12 and 13 and between the hand configurations 34 and 35.

Table 1 shows the main characteristics of studies published in the literature about Libras gestures recognition. As shown, the classifiers employed different techniques: artificial neural networks, self-organized maps, learning vector quantization and support vector machines. The databases employed varied from 26 images to 610 images.

It can be observed that only the study of Porfirio et al. (2013) translated all these 61 HC of Libras. In this study, the classification features were obtained from 3D mesh of the hands associated with features obtained from 2D images, corresponding to a frontal and a lateral view of the hand. The 2D extracted features were the following: seven Hu moments, eight Freeman directions, and horizontal and vertical histogram projections. The features obtained from 3D mesh are some 3D mesh descriptors. The authors claim that these features are scale, rotation and translation invariant. The best recognition rate obtained by the authors with rank #1 was 86.06%.

The present study aims to recognize the 61 HC of Libras. To do this we propose to:

- Develop a gesture recognition method where the segmentation step is independent of scene illumination and background. To achieve this goal, we captured depth maps with a Kinect® sensor;
- Implement a powerful method to feature extraction that is scale, rotation and translation invariant. This goal is accomplished in two steps. First, by applying geometrical transformations in the original gesture image, and second, by applying a dimensionality reduction employing one of two techniques: Bidirectional and Bi-dimensional Linear Discriminant Analysis (2D)$^2$LDA (Noushath et al., 2006) and Bidirectional and Bi-dimensional Principal Component Analysis (2D)$^2$PCA (Zhang and Zhou, 2005);
- Construct a large and robust database of Libras gestures with 12,200 images: 200 images of each 61-hand gestures, obtained from 10 different people.
- Evaluate the performance of two classifiers: the novelty classifier (Costa et al., 2013, 2014), and the k-Nearest Neighbor classifier (kNN) (Wang, 2006), using this robust database.

The HC Libras recognition task was subject of two dissertations (Santos, 2015; Silva, 2015) supervised by two of the authors. Both dissertations used the kNN classifier, however, Santos (2015) used (2D)$^2$LDA, while Silva (2015) used (2D)$^2$PCA as feature extraction technique. In this paper, we compare the results obtained in both dissertations with the results obtained with a novelty classifier.

The materials and methods section first presents the gesture image database built, the *LibrasImages*, explains the geometrical transformations applied in the original image and describes the techniques (2D)$^2$LDA and (2D)$^2$PCA employed for feature extraction. Finally,

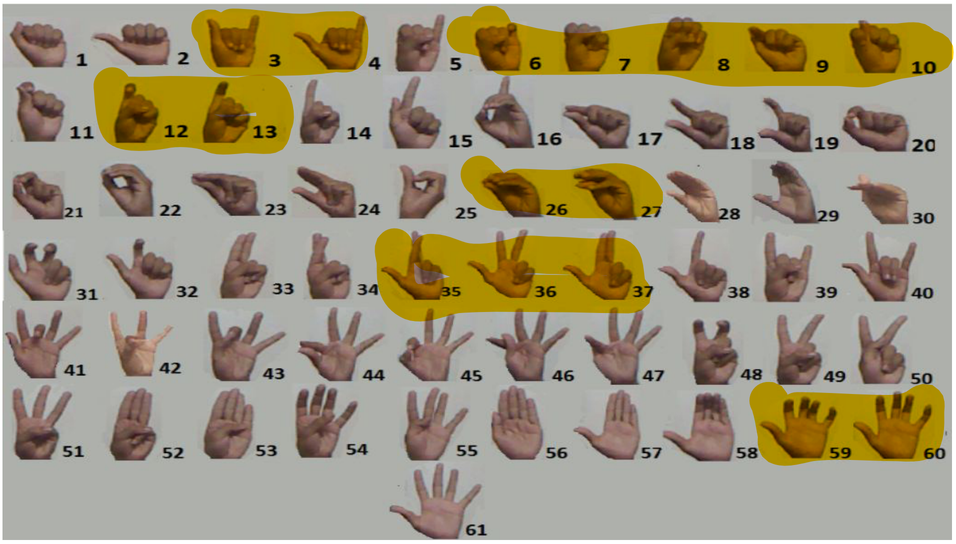


**Figure 1.** 61 Hand configurations (HC) for Brazilian Sign Language (Libras).

**Table 1.** Summary of studies concerning Libras hand gestures recognition.

| Reference | Application | Materials | Segmentation and classifier features | Classifier | Recognition rate |
|---|---|---|---|---|---|
| Peres et al. (2006) | Recognition of LIBRAS digital alphabet (26 gestures). | **Database:** 26 training binary images<br>**Background:** not applied. | **Segmentation:** not applied<br>**Classifier features:** Bit signatures | Learning Vector Quantization | 59.6% (worst result)<br>100% (best result) |
| Neris et al. (2008) | Recognition of LIBRAS digital alphabet (26 gestures). | **Database:** 26 images of gestures.<br>**Background:** black. | **Segmentation:** Threshold;<br>**Classifier features:** bits signatures (22 vectors with 236 coordinates corresponding to pixels sum in horizontal and vertical directions). | Self Organized Maps (SOM) | 98.9% |
| Carneiro et al. (2009) | Recognition of LIBRAS digital alphabet (27 gestures) and 10 words. | **Database**: 45 RGB images/gestures obtained from 45 people. They wore a black jacket;<br>**Background:** black with artificial lighting. | **Segmentation:** Threshold;<br>**Classifier features:** raw gesture image, with 625 pixels of a region of 25x25 pixels. | Artificial Neural Networks | 91.1% |
| Pizzolato et al. (2010) | Recognition of LIBRAS digital alphabet (26 gestures). | **Database:** 50 images/ gesture obtained from 3 people Images: RGB;<br>**Background:** White with artificial lighting. | **Segmentation:** Threshold in YCbCr color space; $77 \leq Cb \leq 127$ and $133 \leq Cr \leq 173$;<br>**Classifier features:** 6 Hu invariant moments. | Artificial Neural Networks | 89.67% |
| Porfirio et al. (2013) | Recognition of 61 hand configuration gestures of LIBRAS | **Database:** 610 image pairs obtained from 5 people.<br>**Background:** not cited. | **Segmentation:** not cited.<br>**Classifier features:** 2D features: seven Hu moments, eight Freeman directions, and horizontal and vertical histogram projections. 3D features: mesh descriptors. | Support Vector Machine | Rank 1: 86.06%;<br>Rank 3: 96.83% |

we present the two classifiers employed for gesture classification. The Results section presents the values of the recognition rate for both classifier methods and techniques employed in feature extraction step. The Discussion section evaluates the performance of both classifiers and feature extraction methods, and compares this performance with the performance of other methods previously published in the literature.

# Methods

This section presents the methods implemented at each stage of this pattern recognition task, namely: Image acquisition; Hand configuration (HC) segmentation; Feature extraction, and Classification.

## *Image acquisition*

The image database constructed, called *LibrasImages*, is comprised of 12,200 images. Two hundred images were captured for each one of 61 HC of Libras. These images were captured from 10 volunteers. To obtain a representative group of images, individuals belonging to different groups were selected:

- Seven individuals belong to the deaf community (deaf individuals, not only hard-of-hearing). These seven individuals are teenagers who were literate in Libras in childhood);
- Three individuals do not belong to the deaf community (they learn Libras two years ago);
- Eight individuals are men, while 2 are women;
- The individuals age ranged from 15 to 25 years.

For each HC frame, two files are generated. The first one, obtained by a RGB camera, corresponding to a true color image of 640×480 pixels, with a depth resolution of 24 bits. It is saved in *bmp* format. The second one, obtained by a depth-sensing camera, corresponding to a depth map. It has dimensions of 640×480 elements, depth resolution of 11 bits, and is saved in *txt* format. In this study, only depth map is used.

To evaluate whether the feature extraction method is scale, rotation, and translation invariant, the following set up are assumed: First, the individuals and the Kinect® are positioned as shown in Figure 2a. The depth range obtained with the Kinect® is [0.8m-3.5m]. The individuals are free to move in the Kinect® field of view. Second, to

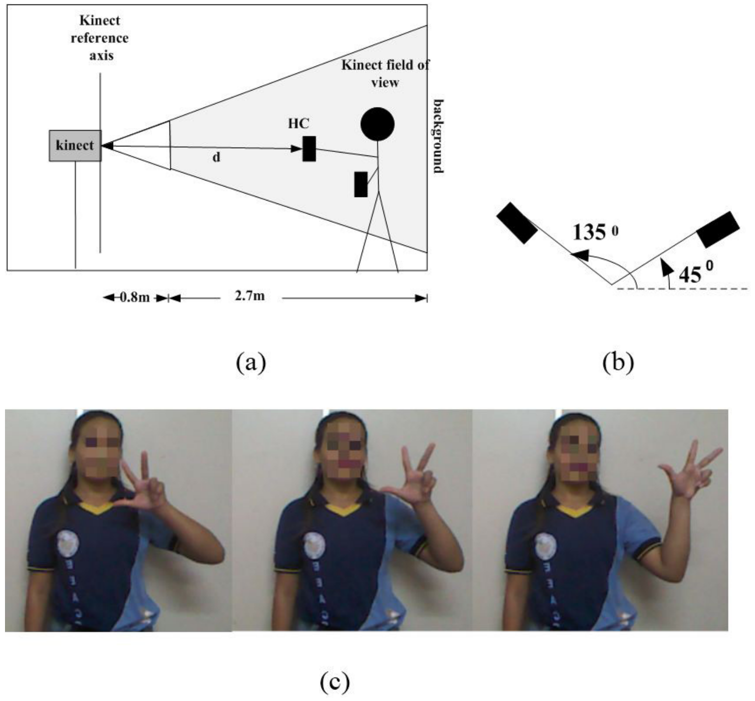


**Figure 2.** (a) Position of the Kinect® relative to the individual, *d* is the gesture distance to the Kinect®; (b) Rotation range of the HC, [45º-135º]; (c) Examples of HC images acquired in different hand orientations.

obtain the 200 images of each gesture, captured from video frames, the volunteers are free to do the gesture in different positions relative to the body (articulation point), and to rotate the gesture from 45º to 135º, as shown in Figure 2b. Figure 2c shows examples of gesture images. The scene illumination is not controlled.

### *Hand configuration segmentation*

The HC segmentation task, illustrated in Figure 3a, includes two steps of post processing: size standardization and pixels normalization. These steps aim to prepare the HC segmented to the next phases, which correspond to recognition task (feature extract and classification).

In the first step of Figure 3a, the hand+forearm are segmented using a *region growing* technique. Region growing is a procedure that groups pixels into regions based on predefined criteria for growth (Gonzalez and Woods, 2008). To implement this technique we need to set a "seed" point and from this grow region by appending to seed point those neighboring pixels that have predefined properties similar to the seed. Therefore, this technique requires two parameters: a "seed" pixel and a similarity criterion. The "seed" pixel is chosen as the pixel that is closest to the Kinect® sensor. In other words, the pixel in the depth map that has the smallest value, $d_{min}$, because the hand is always in front of the body. The similarity criteria are: pixels should be appended to seed point if they are 8-connected to a seed pixel, and the Equation 1 is satisfied:

$$d_p - d_{min} < T \qquad (1)$$

where:

$d_p$ – Distance from the pixel to the Kinect® reference axis;

$d_{min}$ – Minimum distance from the pixel to the Kinect® reference axis;

$T$ – Threshold value.

The optimal value of $T$ is obtained varying the threshold value from 50mm to 100mm in steps of 10mm and evaluating, for all gestures, which one results in the best segmentation. This procedure found that the best $T$ value is 90mm.

The second step shown in Figure 3a is vertical alignment of the hand + forearm. This step is accomplished by a rotation of angle β, calculated by Equation 2, where the angle orientation θ, shown in this figure, is defined as the angle that a line passing through the centroid of the hand+forearm, in the direction of the forearm, forms with the vertical line. θ value is calculated by Equation 3. This alignment operation uses bilinear interpolation (Gonzalez and Woods, 2008).

$$\beta = 90 - \theta \qquad (2)$$

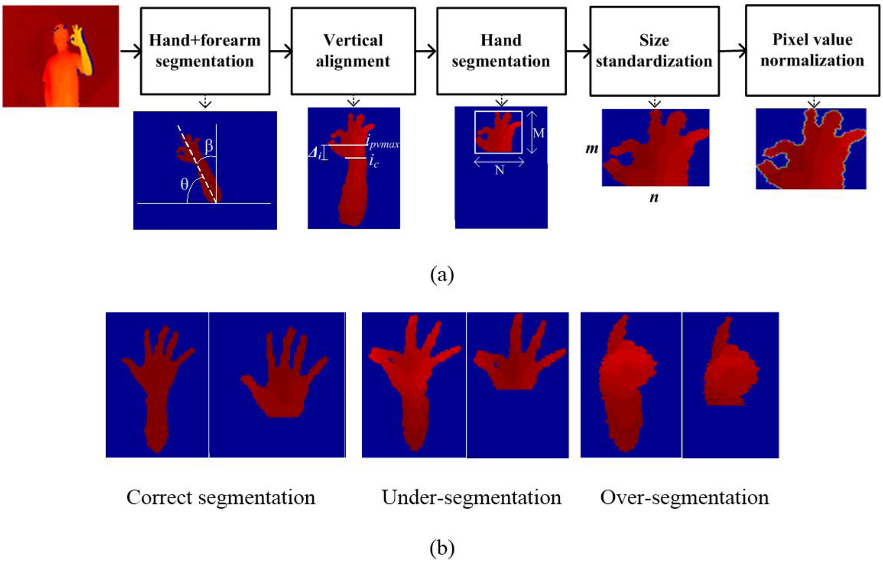


**Figure 3.** Segmentation: (a) Block diagram of segmentation and geometrical operations illustrated with example of resultant images of each step, where: θ is the orientation angle of the gesture; β is the rotation angle; the forearm cut, $i_{Pv_{max}}$, is the line number corresponding to the higher value of vertical projection and $i_c$ corresponds to the forearm cut line. The dimension of HC segmented, MxN, is resized to a standard size *m x m*; (b) HC segmentation examples. The left images show the HC segmented from the forearm. The central images show the HC under segmented from the forearm and the right images show the hand over segmented from the forearm.

$$\theta = \frac{1}{2} arctg\left(\frac{2\mu_{1,1}}{\mu_{2,0}-\mu_{0,2}}\right) \qquad (3)$$

where:

$\mu_{1,1}$, $\mu_{2,0}$ and $\mu_{0,2}$ – Second order Hu moments.

After the hand+forearm is aligned with the vertical direction, the hand segmentation is accomplished in the following steps:

1. Obtaining the vertical projection $P_v$ of the hand + forearm aligned image. This projection is obtained counting the number of pixels in each line. $P_v$ is a column vector with Mx1 dimensions, where M is the number of image lines. $P_{v_i}$ (with $1 < i < M$) is given by Equation 4.

$$Pv_i = \sum_i I_P(i,j) \qquad (4)$$

where:

$I_P(i,j)$ – intensity of pixel *j* in line *i* (1 or 0)

2. Identifying the line $i_{Pv_{max}}$, corresponding to the maximum value of $Pv_i$, as shown in Figure 3a.
3. Obtaining the forearm cut line, $i_c$ using the Equation 5. This cut line is shown in Figure 3a.

$$i_c = i_{Pv_{max}} + \overline{\Delta_i} \qquad (5)$$

where:

$$\overline{\Delta_i} = \frac{1}{N}\sum_{i=1}^{N}\Delta_i \qquad (6)$$

$\Delta_i$ – Experimentally determined for a set of images (N=610).

The value of $\overline{\Delta_i}$, calculated by Equation 6, is equal to 26.

According to the block diagram of Figure 3a, hand segmentation is followed by hand-size standardization and pixel normalization. These steps are required as a pre-processing to the feature extraction techniques used in this study, (2D)$^2$LDA and (2D)$^2$PCA. Hand-size standardization is accomplished in two ulterior steps. In the first one, the hand is cropped according to the minimum rectangle that encases it, as shown in Figure 3a. After, the hand is resized to a standard size, *m x n*, using bilinear interpolation (Gonzalez and Woods, 2008). The values of *m* and *n* are 135, and 139, respectively. These dimensions correspond to the maximum gesture sizes.

As the depth maps are not obtained at the same distance of the Kinect®, the depth maps of the same HC could have different value ranges. To obtain the same value range for a given HC depth map, the following normalization procedure is adopted: subtract the maximum value from the minimum value of a depth map and scale the resulting values to the range 0-2047 (11bits).

Figure 3b shows the most frequently cases found in the segmentation process. The left images show the hand corrected segmented from the forearm. The central images show the hand under segmented from the forearm, and the left images show the hand over segmented from the forearm.

### *Feature extraction*

In feature extraction phase, the characteristic matrices, $C$, are obtained from the processed depth maps, $A(mxn)$, using one of two dimensionality reduction techniques: (2D)²LDA and (2D)²PCA. For each technique, four dimensions of characteristics matrices are obtained: 5×5, 10×10, 15×15 or 20×20. Next, these matrices are converted to column vectors ([25,1], [100, 1], [225, 1] and [400, 1]). Each one of these vectors will be the classifier's input.

In the sequence, we will describe the two aforementioned techniques used for dimensionality reduction of processed Libras depth maps.

Technique (2D)²LDA is intended to reduce the dimensions of the processed depth map, optimizing the separation of the classes (hand configurations). Its origin is the method known as Image Matrix-based Linear Discriminant Analysis (IMLDA) (Yang et al., 2005) which is, in turn, based on the Fisher criteria, applied to the matrix that describes the processed depth map, $A$. Let $c$ be the number of standard classes, $N$ the total samples for training, $N_i$ the number of samples of class $i$, $A_j^{(i)}$ the $j_{th}$ processed depth map of class $i$ with dimension $m x n$, $\overline{A}_j^{(i)}$ the average of the depth maps of class $i$, and $\overline{A}$ the total average of the training images. Based on the matrices of the images used for training, the scattering matrix between classes and the scattering matrix inside the class are given respectively by:

$$S_B = \frac{1}{N}\sum_{i=1}^{c} N_i \left(\overline{A}_i - \overline{A}\right)^T \left(\overline{A}_i - \overline{A}\right) \tag{7}$$

and

$$S_W = \frac{1}{N}\cdot\sum_{i=1}^{c}\sum_{j=1}^{N_i}\left(A_j^{(i)} - \overline{A}^i\right)^T\left(A_j^{(i)} - \overline{A}^{(i)}\right) \tag{8}$$

where $S_B$ and $S_W$ are positive definite.

The generalized Fisher criterion aims to obtain a projection $H$ matrix that maximizes the following quotient:

$$\varnothing(H) = \frac{H^T S_B H}{H^T S_W H} \tag{9}$$

The solution of (9) is the matrix $H = \left[h_1, h_2, \dots h_q\right]$ formed by the eigenvectors of $S_w^{-1} S_B$ corresponding to $q$ largest eigenvalues. Matrix $H$ is a linear transformer, conventionally called projection matrix. On IMLDA, making $B = A_j H$, we obtain $B$ with dimension $mxq$, with $q < n$, which is used to describe image $A_j$ in the classification step. IMLDA performs a dimension reduction on the horizontal direction of the processed depth map's matrix.

On (2D)²LDA, IMLDA is applied a second time, aiming now to reduce the vertical direction of matrix $B$. Applying IMLDA in the vertical direction consists of designing the dispersion matrix between class $G_B$ and the dispersion matrix inside class $G_W$, having as input matrices $B$:

$$G_B = \frac{1}{N}\sum_{i=1}^{c} N_i \left(\overline{B}_i - \overline{B}\right)\left(\overline{B}_i - \overline{B}\right)^T \tag{10}$$

$$G_W = \frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{N_i}\left(\overline{B}_j^{(i)} - \overline{B}^{(i)}\right)\left(\overline{B}_j^{(i)} - \overline{B}^{(i)}\right)^T \tag{11}$$

where:

$$B_j^{(i)} = A_j^{(i)} H \tag{12}$$

$$\overline{B}^{(i)} = \overline{A}^{(i)} H \tag{13}$$

$$\overline{B} = \overline{A} H \tag{14}$$

Afterwards, Fisher's criterion is applied to optimize and obtain the projection matrix $V = \left[v_1, v_2, \dots v_p\right]$, which is formed by the eigenvectors of $G_w^{-1} G_B$ corresponding to $p$ greatest eigenvalues. Thus, the characteristic matrix $C$, which represents image $A$ in the classification step, is obtained for the following transformation:

$$C = V^T B = V^T A H B \tag{15}$$

where $C$ has dimensions $p \times q$, being much lower than the matrix of the processed image $A$ with dimensions $m \times n$.

The technique (2D)²PCA aims at reducing the dimensions of the depth map space, optimizing the variance of projections in horizontal and vertical directions. Based on the matrices of the depth maps used for training the classifiers, the dispersion matrix is given by:

$$G_H = \frac{1}{N}\sum_{i=1}^{N}\left(A_i - \overline{A}\right)^T . \left(A_i - \overline{A}\right) \tag{16}$$

To maximize the projections in horizontal, the projection matrix $U = [u_1, u_2, \dots u_d]$ is employd, being formed by the eigenvectors of $G_H$ corresponding to $d$ largest eigenvalues. Matrix $U$ is a linear transformer, conventionally called projection matrix. For a depth map $A$, the projected matrix is $B = AU$ with dimension $mxd$, with $d < n$.

Reducing dimensions in vertical direction of matrix $B$ consists of building the dispersion matrix $G_V$, given by:

$$G_V = \frac{1}{N}\sum_{i=1}^{N}\left(A_i - \overline{A}\right)\left(A_i - \overline{A}\right)^T \tag{17}$$

To maximize the projections in the horizontal direction, projection matrix $V = [v_1, v_2, \dots v_r]$ is employd, being formed by the by the eigenvectors of $G_V$ corresponding to $d$ largest eigenvalues. Matrix $V$ is a linear transformer, conventionally called a projection matrix. For image $B$, the projected matrix is:

$$C = V^T B = V^T AU \tag{18}$$

where $C$ has dimensions $r \times d$, being much lower than the matrix of the original image $A$ with dimensions $m \times n$.

## *Classification*

As classifiers, are used the Novelty classifier and the k-Nearest Neighbors (kNN) classifier. In the following, we present both of them. The novelty classifier was previously proposed in Costa et al. (2013, 2014). In these previous studies, its mathematical formulation was based on the Gram-Schmidt orthogonalization process. In the present study, the mathematical formulation of the novelty classifier is based on the pseudo-inverse matrix.

The motivations to use the novelty classifier in this study are the higher recognition rates obtained in the previous studies and the excellent generalization capability, even with a low number of samples in the training set (Costa et al., 2013, 2014).

For explaining the novelty classifier, we will first explain the novelty filter concept.

Consider a group of vectors $\{x_1, x_2, \ldots, x_\mathrm{m}\} \subset R^n$ forming a base that generates a subspace $L \subset R^n$, with $m<n$. An arbitrary vector $x \in \mathrm{R}^n$ can be decomposed in two components, $\hat{x}$ and $\tilde{x}$, where $\hat{x}$ is a linear combination of vectors $x_\mathrm{k}$. In other words, $\hat{x}$ is the orthogonal projection of $x$ on subspace L and $\tilde{x}$ is the orthogonal projection of $x$ on a subspace $L$ ⊥(orthogonal complement of $L$). Figure 4a illustrates the orthogonal projections of $x$ in a tridimensional space. It can be shown, through the projection theorem, that $\tilde{x}$ is single and has a minimum norm. So, $\hat{x}$ is the best representation of $x$ on subspace $L$.

The $\tilde{x}$ component of the vector can be thought of as the result of an operation of information processing, with very interesting properties. It can be assumed that $\tilde{x}$ is the residue remaining when the best linear combination of the old patterns (base vectors $x_k$) is adjusted to express vector $x$. So, it is possible to say that $\tilde{x}$ is the new part of $x$ that cannot be explained by the "old" patterns. This component is named "novelty" and the system that extracts this component from $x$ is named the "novelty filter". Vectors base, $x_k$, can be understood as the memory of the system, while $x$ is a key through which information is associatively searched in the memory. It can be shown that the decomposition of an arbitrary vector $x \in R^n$ in its orthogonal projections $\hat{x} \in \mathrm{L} \subset R^n$ and $\tilde{x} \in \mathrm{L}\perp$ can be obtained from a linear transformation, using a symmetric matrix $P$, so:

$$\hat{x} = P.x \tag{19}$$

$$\tilde{x} = (I - P).x \tag{20}$$

The matrix $(I - P)$ is named orthogonal projector operator in $L$ and is named novelty filter, as described by Kohonen (1989).

Consider a matrix $X = [x_1, x_2, \ldots x_k]$, with $k < n$, as $x_i$ its columns. Suppose that the vectors $x_\mathrm{i} \in R^\mathrm{n}$, $i = 1, 2 \ldots k$, span the subspace $L$. As cited above, the decomposition of $x = \hat{x} + \tilde{x}$ is unique and $\tilde{x}$ can be determined through the condition that it is orthogonal to all columns of $X$. In other words:

$$\tilde{x}^T.X = 0 \tag{21}$$

The Penrose solution (Penrose and Todd, 1955) to Equation 3 is given by:

$$\tilde{x}^T = y^T \left(I - X.X^+\right) \tag{22}$$

where:

$y$ is an arbitrary vector with the same dimension of $\tilde{x}$;

$X^+$ is the pseudo-inverse matrix of $X$.

Using the properties of symmetry and idempotence of the pseudo-inverse matrix, it follows that:

$$x^T.\tilde{x} = x^T.\left(I - X.X^+\right).y \tag{23}$$

$$x^T.\tilde{x} = \tilde{x}^T.x = y^T.\left(I - X.X^+\right)^T.x \tag{24}$$

Comparing Equations 23 and 24, it follows that $y = x$. So $\tilde{x}$ can be written as:

$$\tilde{x} = \left(I - X.X^+\right).x \tag{25}$$

As $\tilde{x}$ is unique, it follows that: $\mathrm{I} - \mathrm{P} = \mathrm{I} - X.X^+$ and

$$\mathrm{P} = X.X^+ \tag{26}$$

The novelty classifier training consists of determining the novelty filter of each hand configuration of Libras. For each HC training set, a novelty filter is designed. For a given HC, consider that $X = [x_1, x_2, \ldots, x_{100}]$ is the set of 100 vectors. The P matrix for this HC is calculated using Equation 26. Given an HC depth map sample, the novelty is calculated using Equation 20. Figure 4b illustrates the novelty vector calculation for a $x$ sample of a depth map HC with a training matrix $X$.

The novelty classifier is constructed using the block diagram of Figure 4c. In this figure, there are 61 novelty filters, one for each Libras hand configuration. For a sample depth map presented at classifier input, are calculated 61 novelty vectors $\tilde{x}_i, 1 < i < 61$. After the calculation of each novelty vector, $\tilde{x}_i$, the vector norm is extracted. The 61 novelty filter norms are the inputs to a comparator block and the lowest value of vector norm is selected. The HC corresponding to the novelty filter that presents the lowest value is the one to which sample $x$ belongs.

In this study, the training matrix $X$ is formed only with vectors that are Linearly Independent (LI). In the results section, we show the novelty classifier performance with different sizes of training matrix, $X$.

The other classifier used in this study is the kNN classifier. For this classifier, which is well known in the literature, the value of k is varied from 1 to 15.

For pattern classification, two metrics are employed, the Manhattan distance and the Euclidian distance.

### *Implementation*

The simulations were made using a computer with Intel(R) Core (TM) i3, 2.0GHz Processor, with 3.0GB of RAM, running Matlab® 2014.

## Results

All the 12,200 depth maps were successfully segmented with the method proposed in this paper. Figure 3b shows examples of hand segmentation. The cases of under and over hand segmentation do not affect the HC recognition process, because the main details of a gesture are located in the upper part thereof.

The accuracies obtained for gesture classification with the novelty classifier and with the kNN classifier are shown in Tables 2 and 3, respectively. For the novelty classifier, the number of vectors of the 61 training matrices, X, is shown in the top line of Table 2. As shown in this Table, the maximum number of training vectors of training matrix, X, is 86.

For the kNN classifier, the number of neighbors was varied from 1 to 15, in steps of 5.

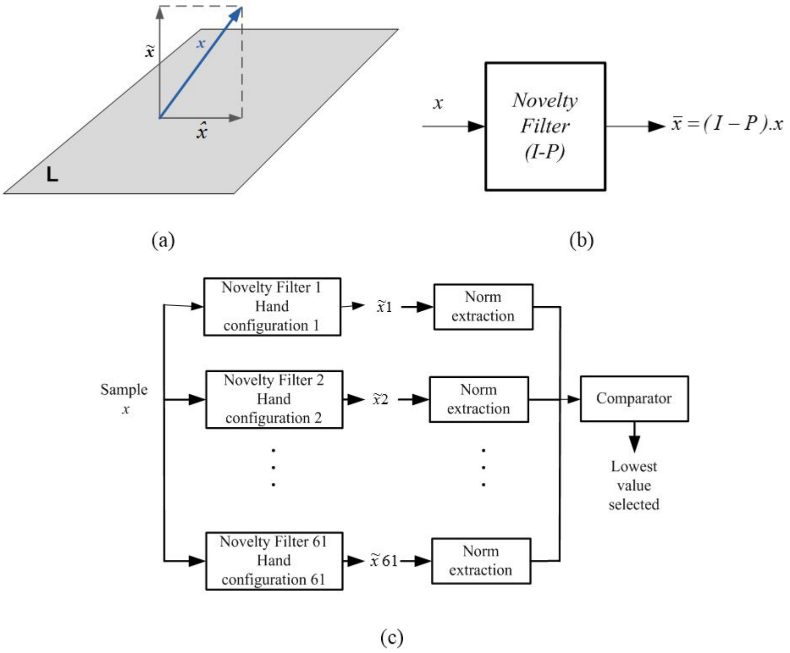


**Figure 4.** (a) Orthogonal projections of a vector in a subspace L (b) novelty filter concept (c) novelty classifier for the classification of the 61 HC.

**Table 2.** Accuracy of Novelty Classifier for Libras hand configuration.

| Feature extraction | Training vector size | Accuracy | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean number of vectors LI* in the training matrix X | | | | | | | | | |
| | | 10 | 20 | 28 | 36 | 45 | 54 | 63 | 72 | 81 | 86 |
| $(2D)^2$PCA | [100, 1] | 72.39 | 83.81 | 88.54 | 90.91 | 91.45 | 92.13 | 90.29 | 86.62 | 66.77 | 40.47 |
| | [225, 1] | 71.54 | 83.55 | 88.07 | 90.57 | 92.39 | 93.95 | 94.37 | 94.51 | 94.75 | 94.56 |
| | [400, 1] | 70.15 | 83.49 | 88.98 | 89.09 | 93.00 | 93.70 | 94.74 | 95.03 | 95.19 | 95.41 |
| $(2D)^2$LDA | [100, 1] | 62.59 | 75.62 | 83.01 | 85.21 | 86.39 | 87.24 | 84.83 | 80.56 | 67.67 | 46.57 |
| | [225, 1] | 63.85 | 77.04 | 82.57 | 86.90 | 87.34 | 89.69 | 89.78 | 90.98 | 91.08 | 91.61 |
| | [400, 1] | 62.59 | 75.70 | 81.54 | 85.68 | 87.64 | 89.34 | 90.18 | 91.23 | 92.13 | 91.96 |

*LI: linearly independent.

The main reason for the errors observed in both classifiers is the similarity between hand configurations of Libras. Table 4 shows mean values of errors observed in both classifiers when classifying some hand configurations of Libras.

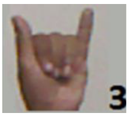
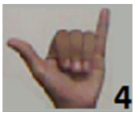
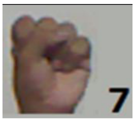
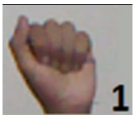
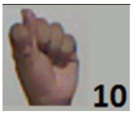
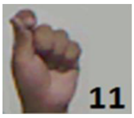
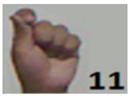
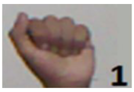
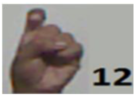
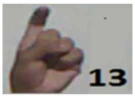
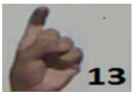
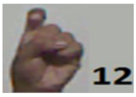
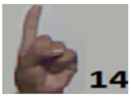
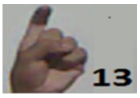
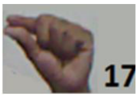
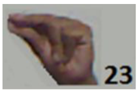
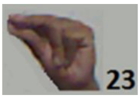
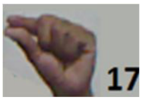
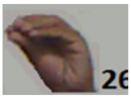
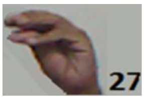
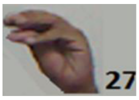
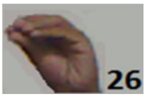
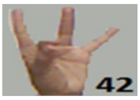
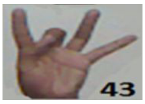
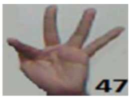
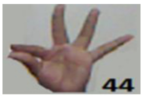
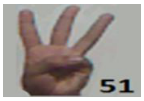
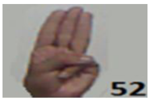
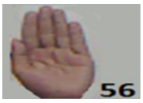
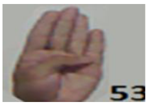
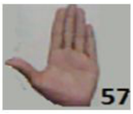
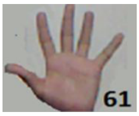
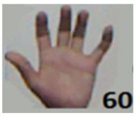
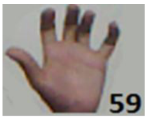
## Discussion

The first inference that can be drawn from Table 2 about the novelty classifier is that the best performance is obtained with the $(2D)^2$PCA feature extraction technique,

**Table 3.** Accuracy of kNN classifier for Libras hand configuration.

| Training vector size | k-Neighboors | Accuracy | | | |
|---|---|---|---|---|---|
| | | $(2D)^2$LDA | | $(2D)^2$PCA | |
| | | Euclidian distance | Manhattan distance | Euclidian distance | Manhattan distance |
| [25, 1] | 1 | 89.04 | 92.27 | 94.98 | 95.26 |
| | 5 | 85.16 | 88.96 | 91.95 | 92.91 |
| | 10 | 80.90 | 85.51 | 87.44 | 89.43 |
| | 15 | 77.24 | 82.28 | 83.87 | 85.54 |
| [100, 1] | 1 | 92.92 | 94.47 | 96.31 | 96.13 |
| | 5 | 89.11 | 90.44 | 93.20 | 93.11 |
| | 10 | 84.55 | 85.90 | 89.13 | 88.49 |
| | 15 | 80.55 | 82.14 | 84.72 | 84.15 |
| [225, 1] | 1 | 92.14 | 92.16 | 95.93 | 94.57 |
| | 5 | 85.16 | 86.67 | 92.56 | 89.34 |
| | 10 | 82.96 | 80.95 | 88.57 | 84.07 |
| | 15 | 78.62 | 75.65 | 84.00 | 78.97 |
| [400, 1] | 1 | 90.75 | 87.08 | 95.82 | 92.16 |
| | 5 | 85.72 | 78.80 | 92.41 | 86.03 |
| | 10 | 80.28 | 71.61 | 88.16 | 78.44 |
| | 15 | 75.61 | 64.95 | 83.59 | 71.57 |

**Table 4.** Mean error values of both classifiers when classifying HC of Libras.

| Submitted hand configuration | Recognized hand configuration | Mean Error (%) | Submitted hand configuration | Recognized hand configuration | Mean Error (%) | Submitted hand configuration | Recognized hand configuration | Mean Error (%) |
|---|---|---|---|---|---|---|---|---|
|  |  | 8.0 |  |  | 3.0 |  |  | 8.0 |
|  |  | 3.0 |  |  | 9.0 |  |  | 7.0 |
|  |  | 5.0 |  |  | 3.0 |  |  | 6.0 |
|  |  | 4.0 |  |  | 9.0 |  |  | 4.0 |
|  |  | 3.0 |  |  | 6.0 |  |  | 7.0 |
|  |  | 3.0 |  |  | 9.0 | | | |

with input vectors of size [225, 1] or [400, 1] and with training matrices $X$ formed by LI vectors in the range of 63-86. The best result of the novelty classifier is an accuracy of 95.41% and is obtained in the following condition: $(2D)^2$PCA feature extraction technique, with input vector of size [400, 1] and with training matrices, $X$, formed by 86 LI vectors.

Table 2 also shows that, for the novelty classifier, the classification performance obtained varies with the mean number of vectors LI in the training matrix. For vector size [100,1], performance increases and reaches a maximum value with 54 vectors LI, and decreases thereafter. With vector sizes [225,1] and [400, 1] the classification performance increases continuously from a mean number of vectors LI equal to 10 up to 86.

From Table 3, that shows the accuracies of the kNN classifier, we can observe that the best performance is also obtained with the $(2D)^2$PCA feature extraction technique and with input vectors of size [100, 1]. Except for *k=10*, the accuracy reaches maximum values for an input vector of size [100, 1], for the Euclidian Distance and for *k=1*. The best result of the kNN classifier is an accuracy of 96.31%, which is obtained in the following condition: $(2D)^2$PCA feature extraction technique, with input vector of size [100, 1], using Euclidian Distance and *k=1*.

We can observe in Table 3 that, for a fixed vector size, best values are obtained with a fewer k-neighbors.

The vector that generates the best results with novelty classifier has size of [400,1], while with the kNN classifier, it has a size of [100,1]. Seemingly, there is no apparent reason for this behavior.

From the aforementioned we can say that: a) the performance of the novelty classifier and the performance of the kNN classifier are similar; b) with both classifiers the performance of the $(2D)^2$PCA feature extraction technique is better than the performance of the $(2D)^2$LDA feature extraction technique.

Although the novelty concept is an old one, the novelty classifier concept is new. This classifier has recently been used for iris recognition (Costa et al., 2013) and for face recognition (Costa et al., 2014). In this study, we show the suitability of the novelty classifier for gesture recognition, more specifically, for Hand Configurations used by the Brazilian-deaf community.

The literature review of Libras hand configuration recognition shows that, except for the study of Porfirio et al. (2013) the other studies used a controlled background (Carneiro et al., 2009; Neris et al., 2008; Pizzolato et al., 2010). Other studies not on Libras used colorful gloves (Bragatto et al., 2006; Maraqa et al., 2012) or gloves with sensors (Mehdi and Khan, 2002; Wang et al., 2006), which require complex interfaces with the computer system. Methods that use depth maps, as the one proposed in this study, make a major contribution to this research field, since they do not depend on controlled environments and do not require complex interfaces with computers, neither wearable sensors such as gloves with sensors.

The best gesture recognition accuracy of Libras obtained in this study, 96.31%, is much higher than the one obtained in (Porfirio et al., 2013), 86.06%. It must be emphasized that this recognition rate is obtained for different conditions of hand rotation and proximity of the depth camera, and with a depth camera resolution of only 640×480 pixels. This performance must be also credited to the feature extraction technique and to the standardization and normalization processes used.

The mean times spent in each phase of the method for recognizing a HC are the following: segmentation: 0.32s; feature extraction: 0.086s and classification: 0.022s. As observed, the segmentation consumes more time than the other phases. It is important remember that these mean times were obtained using a computer with Intel(R) Core $^{(TM)}$ i3, 2.0GHz Processor, with 3.0GB of RAM, running Matlab 2014.

We intend to continue this study in different ways. The first one would be to develop tools to recognize other Libras phonological parameters, such as face expression. The second one would be to improve gesture recognition phase accuracy. Although hand configurations were captured from a video stream, the volunteers were not actually communicating in Libras. They were simply performing individual hand configurations. Based on that, the third area we would further develop is integrating the different phonological parameters and building a full Libras translator, probably using convolutional neural networks as pattern recognition tool. This research is now under way.

## Acknowledgements

Part of the results presented in this paper were obtained from a project for research and human resources qualification, at the undergraduate and graduate levels, in the areas of industrial automation, mobile software and Digital TV devices, sponsored by Samsung Electronica da Amazonia Ltda, under the terms of the Brazilian Federal Law number 8.387/91. We would like to thank to Marcos Roberto dos Santos, professor of the State University of Amazonas, for his support in Libras.

## References

Al-Jarrah O, Halawani A. Recognition of gestures in Arabic sign language using neuro-fuzzy systems. Artificial Intelligence. 2001; 133(1-2):117-38. http://dx.doi.org/10.1016/S0004-3702(01)00141-2.

Anjo MS. Avaliação das técnicas de segmentação, modelagem e classificação para o reconhecimento automático de gestos e proposta de uma solução para classificar gestos da libras em tempo real [dissertation]. São Carlos: UFSCar; 2013.

Bragatto TA, Ruas GI, Lamar MV. Real-time video based finger spelling recognition system using low computational complexity Artificial Neural Networks. In: Proceedings of the IEEE International Telecommunications Symposium; 2006 Sept 3-6; Fortaleza, BR. New York: IEEE; 2006. p. 393-7.

Brito LF. Por uma gramática de Língua de Sinais. 1st ed. Rio de Janeiro: Tempo Brasileiro; 2010.

Carneiro TS, Cortez PC, Costa RCS. Recognizing Libras gestures with neural classifiers, using Hu invariant moments. In: Proceedings of Interaction 09 - South America; 2009 Nov 26-28; São Paulo, BR. São Paulo: IXDA; 2009. p. 190-5.

Costa CFF Fo, Falcão AT, Costa MGF, Gomes JR. Proposing the novelty classifier for face recognition. Revista Brasileira de Engenharia Biomédica. 2014; 30(4):301-11. http://dx.doi.org/10.1590/1517-3151.0543.

Costa CFF Fo, Pinheiro CFM, Costa MGF, Pereira WCA. Applying a novelty filter as a matching criterion to iris recognition for binary and real-valued feature vectors. Signal Image Video Process. 2013; 7(2):287-96. http://dx.doi.org/10.1007/s11760-011-0237-5.

Dong C, Leu MC, Yin Z. American sign language alphabet recognition using microsoft kinect. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops; 2015 June 7-12; Boston, MA. New York: IEEE; 2015. p. 4452.

Gonzalez RC, Woods RE. Digital image processing, 3rd ed. New Jersey: Pearson Prentice Hall; 2008.

Kohonen T. Self-organization and associative memory. 3rd ed. New York: Springer-Verlag; 1989. http://dx.doi.org/10.1007/978-3-642-88163-3.

Lee GC, Yeh FH, Hsiao YH. Kinect based Taiwanese sign language recognition system. Multimedia Tools and Applications. 2016; 75(1):261-79. http://dx.doi.org/10.1007/s11042-014-2290-x.

Maraqa M, Al-Zboun F, Dhyabat M, Zitar RA. Recognition of Arabic sign language (ArSL) using recurrent neural networks. J Intell Learn Sys Appl. 2012; 4(1):41-52. http://dx.doi.org/10.4236/jilsa.2012.41004.

Mehdi SA, Khan YN. Sign language recognition using sensor gloves. In Wang L, Rajapakse JC, Fukushima K, Lee SY, Yao X, editors. ICONIP'02: Proceedings of the 9th International Conference on Neural Information Processing: Computational Intelligence for the e-age. 2002 Nov 18-22; Singapore, SG. New York: IEEE; 2002. p. 2204-6.

Neris MN, Silva AJ, Peres SM, Flores FC. Self organizing maps and bit signature: a study applied on signal language recognition. 2008 IEEE International Joint Conference on Neural Networks. 2008 June 1-8; Hong Kong. New York: IEEE; 2008. p. 2934-41.

Noushath S, Kumar GH, Shivakumara P. (2D)(2)LDA: an efficient approach for face recognition. Pattern Recognition. 2006; 39(5):1396-400. http://dx.doi.org/10.1016/j.patcog.2006.01.018.

Penrose R, Todd JA. A generalized inverse for matrices. Proceedings of Mathematical Proceedings of the Cambridge Philosophical Society. 1955; 51(3):406-13. http://dx.doi.org/10.1017/S0305004100030401.

Peres SM, Flores FC, Veronez D, Olguin CJ. Libras signals recognition: a study with learning vector quantization and bit signature. In: Canuta AMP, Souto MCP, Silva ACR, editors. SBRN'06 Ninth Brazilian Symposium on Neural Networks. 2006 Oct 23-27; Ribeirão Preto, SP. New York: IEEE; 2006. p. 119-24.

Pimenta N, Quadros RM. Curso de Libras 1. 4th ed. Rio de Janeiro: Vozes; 2010.

Pizzolato EB, Anjo MS, Pedroso GC. Automatic recognition of finger spelling for libras based on a two layer architecture. In: Proceedings of the 25th Symposium on Applied Computing - SAC'10; 2010 Mar 22-26; Sierre, CH. New York: ACM; 2010. p. 970-4.

Porfirio AJ, Wiggers KL, Oliveira LES, Weingaertner D. Libras sign language hand configuration recognition based on 3D meshes. In: Proceedings of SMC2013 IEEE International Conference on Systems, Man, and Cybernetics. 2013 Oct 13-16; Manchester, UK. Los Alamitos: CPS; 2013. p. 1588-93.

Rakun EM, Andriani I, Wiprayoga W, Danniswara K, Tjandra A. Combining depth image and skeleton data from Kinect for recognizing words in the sign system for Indonesian language (SIBI [Sistem Isyarat Bahasa Indonesia]). In: Proceedings of International Conference on Advanced Computer Science and Information Systems (ICACSIS); 2013 Sept 28-29; Bali, ID. Indonesia: Faculty of Computer Science Universitas Indonesia; 2013. p. 387-92.

Santos JRD. Reconhecimento das configurações de mãos Libras baseado na análise discriminante de Fisher bidimensional utilizando imagens de profundidade [dissertation]. Manaus: Universidade Federal do Amazonas; 2015.

Silva JP, Lamar MV, Bordim JL. A study of the ICP algorithm for recognition of the hand alphabet. In: Aguilar J, Cerqueira E, editors. In: Proceedings of XXXIX Latin American Computing Conference (CLEI 2013); 2013 Oct 7-11; Naiguata, VE. New York: IEEE; 2013 p. 1-9.

Silva RS. Reconhecimento das configurações de mão da língua brasileira de sinais, Libras, em imagens de profundidade através da análise de componentes principais e do classificador k-vizinhos mais próximos [dissertation]. Manaus: Universidade Federal do Amazonas; 2015.

Wang H, Leu MC, Oz C. American Sign Language recognition using multi-dimensional hidden Markov models. Journal of Information Science and Engineering. 2006; 22(5):1109-23.

Wang H. Nearest neighbors by neighborhood counting. IEEE T Pattern Anal. 2006; 28(6):942-53. PMid:16724588. http://dx.doi.org/10.1109/TPAMI.2006.126.

Yang J, Zhang D, Yong X, Yang JY. Two-dimensional discriminant transform for face recognition. Pattern Recognition. 2005; 38(7):1125-9. http://dx.doi.org/10.1016/j.patcog.2004.11.019.

Zhang D, Zhou Z 2nd. PCA:Two-directional two-dimensional PCA for efficient face representation and recognition. Neurocomputing. 2005; 69(1-3):224-31. http://dx.doi.org/10.1016/j.neucom.2005.06.004.