Recognition of Static Gestures applied to Brazilian Sign Language (Libras)

Igor L. O. Bastos
Math Institute
Federal University of Bahia (UFBA),
Salvador, Brazil
igorcrexito@gmail.com

Michele F. Angelo, Angelo C. Loula
Department of Technology, Department of Exact Sciences
State University of Feira de Santana (UEFS)
Feira de Santana, Brazil
mfangelo@uefs.ecomp.br, angelocl@gmail.com

***Abstract*—This paper aims at describing an approach developed for the recognition of gestures on digital images. In this way, two shape descriptors were used: the histogram of oriented gradients (HOG) and Zernike invariant moments (ZIM). A feature vector composed by the information acquired with both descriptors was used to train and test a two stage Neural Network, which is responsible for performing the recognition. In order to evaluate the approach in a practical context, a dataset containing 9600 images representing 40 different gestures (signs) from Brazilian Sign Language (Libras) was composed. This approach showed high recognition rates (hit rates), reaching a final average of 96.77%.**

***Keywords*-Histogram of Oriented Gradients; Zernike Invariant Moments; Neural Networks; Gesture Recognition; Libras**

I. INTRODUCTION

Gesture recognition pertains to recognizing meaningful expressions of motion by a human, involving hands, arms, face, head, and/or body [1]. Different approaches and techniques have been employed to handle with gesture recognition, ranging from studies where auxiliary tools are used, to approaches exclusively based on the application of mathematical models, as Hidden Markov Models, and/or computer vision techniques, as the ones related to feature extraction and classification [2].

Gesture recognition has a wide range of applications [2], such as navigation on virtual environment, designing techniques for forensic identification, developing aids for hearing impaired, etc [1]. Among them, we highlight sign language recognition, widely used as practical field for the application of studies about gesture recognition.

Sign languages represent the principal means of communication among hearing impaired people. They are not universal. Each country has its own sign language, which is affected by aspects of their culture [3]. Some of these languages are noteworthy for the amount of work found, such as American Sign Language (ASL), British Sign Language (BSL) and Indian Sign Language (ISL).

In the present investigation, we use the Brazilian Sign Language (Libras) as a practical context. This language is used by deaf people in Brazil and is considered, since 2002, as an official Brazilian language [4].

Furthermore, the present work describes the development of an approach for recognition of static gestures in images applied to Libras. The point of this approach lies on the combination of two shape descriptors, image processing techniques and a two-stage neural network classifier to recognize the gestures. The present approach advances, when compared to related work, in terms of amount of gestures addressed, (higher than most of gesture recognition studies, especially those related to Libras); and for intending to make the dataset of images publicly available, which could auxiliate the formulation of other studies and comparison of the present approach with other proposals. In addition, we also developed an strategy for skin segmentation that can be understood as a contribution of this work, being able to be applied in contexts beyond the gesture recognition.

It is worth to mention that in this work the signs are recognized statically, representing only one Libras parameter (hand configuration). However, Libras signs also include body language, movement and orientation [4].

This paper is organized as follows: Section 2 brings papers and studies related to gesture recognition and techniques applied on it. Section 3 approaches the theory about the descriptors HOG and ZIM. On section 4, the present approach is explained and details about the methodology are described. Section 5 presents the results obtained considering the image dataset composed on this work and a public hand gesture dataset. Moreover, it brings discussions about the results found. Finally, on section 6, some concluding remarks and future work are presented.

II. RELATED WORK

In recent years, it has been noticed an increasing emergence of work related to gesture recognition [5], some unrelated to any sign languages while others, as the present work, use these languages as a practical field for the application of their approaches. The use (or not) of auxiliary tools is another variance perceived in these works.

In the work of Parvini et al. [6] and Mohandes [7], for instance, an auxiliary glove called Cyberglove is used to extract information about points and joints of the hand. These work presented hit rates of 82.32% and 99.6%, respectively. Parvini et al. [6] recognized 22 gestures from the alphabetics of the American Sign Language (ASL). On the other hand, Mohandes [7] covered 100 signs of Arabic Sign Language.

Unlike studies that use assistive devices to aid in the process of hand tracking and gesture recognition, others elect approaches that rely exclusively on the application of techniques of digital image processing and computer vision.

In the work of Panwar [8], geometrical characteristics of the hands are extracted and used as features. These characteristics are encoded in a binary word used to represent each gesture on images. That approach presented 94% of hit rate for 45 different signs.

Otherwise, works of Zhou [9] and Zhang [10] presented approaches called Finger-Earth Mover's Distance (FEMD) and Histogram of 3D Facets (H3DF), respectively. They obtained high recognition rates using public gesture recognition datasets (Zhou composed his own dataset and made it public).

Triesch and Malsburg [11] presented on his study a technique called Elastic Graph to recognize 10 different hand postures. That technique shows to be robust to the presence of complex backgrounds, achieving a recognition rate of 86.2%.

As well as the work of Parvini et al. [6] and Mohandes [7], other approaches applied the gesture recognition to the field of sign languages. However, these differ from the studies of Parvini et al. [6] and Mohandes [7] for not using auxiliary tools. Among them, we could highlight those proposed by Zahedi et al. [12], Rahman and Afrin [13], and Uebersax et al. [14], which recognize signs of the American Sign Language (ASL); and Bowden et al. [15] and Liwicki and Everingham [16], which operate with the recognition of signs of British Sign Language (BSL).

The recognition of Libras signs is also approached in some works, as the ones proposed by Pizzolato, Anjo e Pedroso [17] and Anjo, Pizzolato and Feuerstack [18]. In their study, Pizzolato, Anjo e Pedroso [17] obtained 90.7% for the recognition of 27 different Libras signs (8 dynamic gestures). Anjo, Pizzolato and Feuerstack [18] used the Microsoft Kinect as an aid tool and their approach obtained 100% hit rate for the recognition of 10 different static Libras signs.

III. THE DESCRIPTORS

Feature descriptors correspond to methods and/or techniques applied to obtain information in digital images. This information allows their representation in a domain according to some characteristic, such as shapes and edges in images, colors, textures, etc. [19]. These descriptors may present characteristics that make their use desirable in some contexts, as invariance to rotation, scale, and translation; insensitivity to lighting and orientation of objects in the images [19].

In our approach, two descriptors were used: the Histogram of Oriented Gradients (HOG) and Zernike Invariant Moments (ZIM). Both techniques are able to provide information related to the shapes and contours on the images. Furthermore, these descriptors were and have been successfully used in several studies aimed at recognizing objects in images, as in the researches of Tian et al. [20], Tsolakidis, Kosmopoulos and Papadourakis [21] and Qader, Ramli and Al-haddad [22].

A. Histogram of Oriented Gradients

The Histogram of Oriented Gradients (HOG) is a descriptor used for recognizing objects in digital images. The idea behind this descriptor is that the local appearance an object as well as its form, can be described by the distribution of the intensity of gradients and direction of edges [23].

The application of HOG is based on the division of the image into smaller regions, called cells [23], [24]. For each of these cells, a 1-D histogram is computed taking into account directions and intensities of gradients. At the end of the process, these 1-D histograms are combined in order to provide a representation of the entire image [23].

Due to the ability to provide information regarding gradients, HOG has been widely used for recognition of objects and/or people, as on work of Tian et al. [20], Tsolakidis, Kosmopoulos and Papadourakis [21], Dalal and Triggs [24] and Misra, Abe and Deguchi [25].

B. Zernike Invariant Moments

The Zernike Invariant Moments (ZIM) are a class of orthogonal moments used in terms of representation of images [26]. These moments are ajusted by 2 parameters: repetition and order, which are related to the capacity of representing details in the images. For instance, moments of high order can better represent thin and small details on images. Otherwise, coarse shapes are better represented by low order moments.

Khotazand and Hong [27] demonstrated that the magnitude of the ZIM, calculated by equation 1, are not affected by any rotation on the objects on images. Because of this property, the magnitude of ZIM is used as a rotation invariant feature. Another important feature of ZIM is their orthogonality property. It ensures that there is no duplication or overlap of information among moments that have different order and/or repetitions [28].

$$A_{nm} = (n+1)/\pi \sum_{x} \sum_{y} f(x,y) V_{nm}(x,y), x^2 + y^2 \leq 1 \tag{1}$$

where 'n' is a non-negative integer and 'm' is an integer submitted to: (n -m is pair) and (m $\leq n$).

As HOG, ZIM has been widely used on recognition works, as on the ones proposed by Qader, Ramli and Al-haddad [22], Hse and Newton [26] and Oujaoura et al. [29].

IV. Materials and Methods

The present approach consists in the use of shape descriptors (HOG and ZIM) for extracting information related to edges and shapes of hands in digital images. This information is combined in a feature vector, which is associated with a Multilayer Perceptron classifier for the recognition of gestures. In addition, the present approach adopts a two-stage classification process and uses image processing techniques such as skin detection to improve the significance of the data extracted with the descriptors.

As the case study for this approach, the gesture recognition was applied on the Brazilian Sign Language. Since there are no public Libras datasets, the approach encompasses the creation of a dataset of images. Other steps of the approach are also presented, such as image acquisition and results validation. These results also include tests obtained through a public image dataset (NTU Hand Digit Dataset [9]).

We next describe the steps of the approach, in which we detail the creation of the image dataset and the selection of descriptors parameters. We also describe the way we performed the classification, presenting the arrangement of images into groups and the architecture of the classifier.

A. Image Dataset

The first step of our approach was the creation of an image dataset. As this work intended to use Libras as a practical field, images from different Libras signs were used to compose it. Three Libras experts and two deaf students volunteered to be models to compose the dataset, which contains 9600 images. The set of signs is composed by: (i) letters from the Libras alphabet: A, B, C, D, E, F, G, I, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, (ii) numbers: 1, 2, 4, 5, 7 and 9; and (iii) words: Plane, Word, Adult, America, House, Gas, Law, Identity, Together, Stone, Little and Verb, totaling 40 different signs.

To select the signs that would be recognized in this work, Libras experts were asked to select those signs recognizable by only the hand configuration parameter. Signs that have other parameters, like distance to certain parts of body or hand movements, were not considered except those in which the hand configuration is sufficient to allow the recognition. For each one of the aforementioned signs, 240 images (with resolution of 50x50 pixels) were acquired, totaling 9600 images. Half of them are grayscale images representing the gestures. The other half corresponds to binary masks obtained with a skin detection approach (applied before converting images to grayscale) and represents the skin zones of the 4800 images obtained before, as shown on Fig. 1.

The acquisition process was performed considering a standard distance from the camera to the individuals used as models. In addition, the images were acquired considering some small variations on lightning and a simple (white) background. The image dataset also considers some variations in terms of hand postures and hand sizes, which are particular to each individual, as shown on Fig. 2.

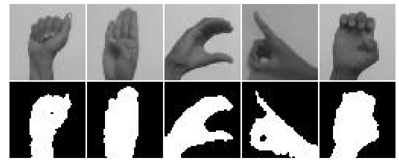

Figure 1. Grayscale images and corresponding skin binary masks.

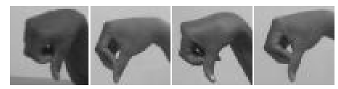

Figure 2. Different hand postures and sizes for sign ’9’.

B. Skin detection approach

In order to generate the binary masks used on the present work, a skin detection approach was applied. It combined 3 of the most applied algorithms for skin recognition: the ones proposed by Kovac et al. [30], Gomez et al. [31] and Bhuiyan et al. [32].

The algorithms presented on [30], [31] and [32] use the components of RGB (red, green and blue), HSV (hue, saturation and value) and YIQ (luminance, in-phase, quadrature) color spaces to label pixels as skin or not using threshold values (calculated from experiments) for this.

In the present study, the components of these 3 color spaces were combined to generate a feature vector, shown on Fig. 3.

R	G	B	H	S	V	Y	I	Q

Figure 3. Feature vector used on the skin approach.

The feature vector shown on Fig. 3 was associatated to a Multilayer Perceptron classifier. The adjustment of parameters, training and testing of this neural network was performed using images randomly picked from 2 public skin image datasets and a own image dataset. They were: (i) 90 images from the Annotated Skin Database [33], (ii) 50 images from the Labelled Faces in the Wild Dataset [34] and (iii) 103 images acquired with a camera (Microsoft HD Lifecam 3000), totalling 243 images with different resolutions.

Images from the Annotated Skin Database [33] are provided along with binary masks that highlight skin zones. For the other images, it was necessary to create these masks using manual markings. The binary masks were used to inform to the classifier which pixels corresponded to skin zones on the training phase and were also used to evaluate the tests and calculate the accuracy of the classifier.

In order to adjust the parameters of the neural network applied on this skin approach, as the number of hidden neurons, the weight values and the activation function, tests were executed considering part of the 243 selected images (10 images from each dataset). From this adjustment, the architecture of this classifier was designed, being shown on Fig. 4. It is noticeable that 9 neurons were used on the input layer. 45 neurons were used on the hidden layer (tests shown that for this value, the approach obtained the highest accuracies) and 1 single neuron on the output layer (the network has to 'answer' if a pixel is skin or not). The activation function choosen was the sigmoid function.

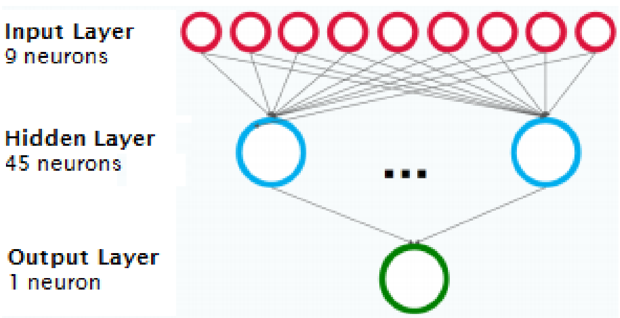

Figure 4. Architecture of the skin classifier.

The skin segmentation approach developed presented high accuracy rates, mainly for images from the dataset acquired with a camera. The images of this dataset, in the same way the gesture images we intended to recognize, present simple background and few variations in terms of lightning. It was found a final average accuracy of 86,8% for the approach. It is important to mention that for some images this skin approach presented accuracy rates superior to 99%. Fig. 5 shows the result of the skin recognition for an image. Note that the highlight of skin areas was faithfully, with little presence of false positives and false negatives.

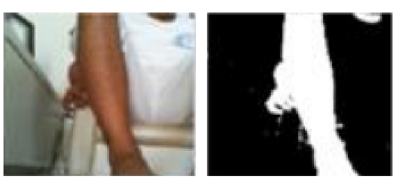

Figure 5. Application of the skin recognition approach.

C. Descriptors and parameters

The two feature descriptors used in this work were the Histogram of Oriented Gradients (HOG) and Zernike invariant moments. They were applied on the images of the dataset in order to generate a feature vector. The HOG descriptor, which is associated to the edges (gradients) of the image, was directly applied on the grayscale images. On the other hand, for the Zernike moments, related to the shape of objects, it was necessary to use the binary masks over the grayscale images, applying the Zernike descriptor on the resulting images.

Both descriptors have parameters that had to be adjusted. In order to select the parameters of HOG, as number of bins and size of blocks and cells; and the order and repetition of Zernike moments that would be used, tests were performed and the hit rate of the classifier was measured, considering a variation of these parameters. The selected parameters of both descriptors were those for which the classifier had the highest recognition rates (Table I). Notice that 9 different Zernike moments were computed using different pairs of parameters. These different parameters were used in order to obtain information about the details and the gross shape present on the images. The orthogonal characteristic of the moments was also considered to select of them. We performed these tests in a portion of each fold of the dataset called 'Adjust'. For each fold, 180 out of 240 images of each gesture were used for 'Training', while 40 out of 240 were used for 'Testing' and 20 out of 240 were used for 'Adjusting' the parameters of classifiers and descriptors.

Table I
SELECTED FEATURES

Zernike Moments (order and repetition)	**HOG (bins, cell dimensions, block dimensions, signed)**
(10,4), (9,3), (8,4), (7,3), (6,2), (5,3), (4,2), (3,3), (2,2))	9, 8x8, 16x16, not signed

D. Assembling the groups

In order to minimize difficulties in the classification of a wide range of patterns (40 signs) by a single classifier, the signs were arranged in 12 small groups and the classification task was divided in 2 stages: (1) recognition of the group which the input image belongs to; (2) recognition of which sign the input image corresponds to.

This two-stage classification approach reduced problems related to excessive growth of the neural network. With the use of a single classifier, a large number of hidden neurons and/or layers had to be employed to ensure the ability to classify so many patterns. This growth would also caused other problems such as an increase in computational cost and time for training and classificating, as well as difficulties regarding to excessive adaptation to the trained patterns (overfitting).

As the HOG and Zernike are shape descriptors, we decided to group signs that had conformable shape in the same groups, using visual inspection as the only criterion. Signs that have very similar shape, as F and T, were grouped together with no other sign, in order to ease the separation by the classifier. Fig. 6 shows all the signs and their respective groups.

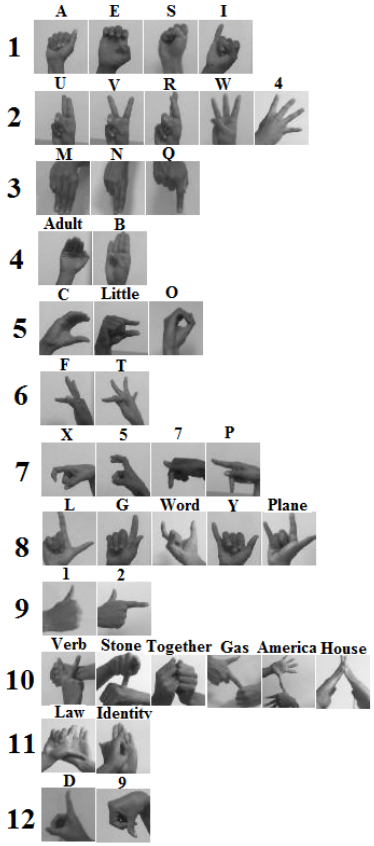

Figure 6. Signs and respective groups.

E. Architecture of the classifier

Fig. 7 shows the architecture of the classifier used in the present approach. Each circle corresponds to a Multilayer Perceptron network, trained with the Backpropagation algorithm and using sigmoid transfer function. The training stopping criterion is associatated to the mean squared error (MSE) variation. The process stops if this variation is lower than 1%.

The recognition of the input gesture was performed in two stages. The first one corresponds to a general classification, performed by a neural network trained with images from the 40 signs. The role of this network is to recognize the group which the input image belongs and to direct the recognition process to a next neural network. The result of this first stage determines which subsequent network is activated. So, the classification process goes to stage 2, where these subsequent networks perform the effective classification of the sign. If the general network misses the group which a sign belongs, a wrong subsequent network is activated and the output is considered an error. Each subsequent network is trained with a small set of gestures, as shown on Fig. 7.

The number of hidden neurons of each network, as the same way the architecture presented on Fig. 7, were determined through tests performed with the 'Adjust' images

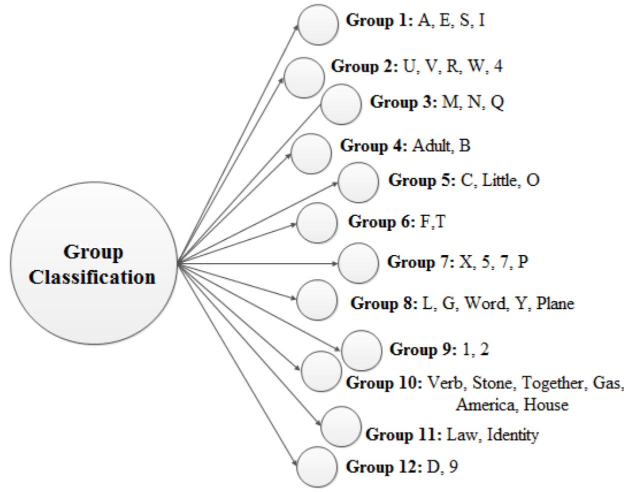

Figure 7. Architecture of the classifier.

of the dataset. The selected values are shown on Table II. Using them, the approach presented the highest recognition rates.

Table II
NUMBER OF HIDDEN NEURONS OF EACH NETWORK.

Network (Group)	Number of hidden neurons
1	20
2	23
3	16
4	15
5	16
6	16
7	22
8	23
9	16
10	22
11	16
12	13
General	44

V. RESULTS AND DISCUSSION

To obtain results of the present approach, tests were performed using the image dataset. In order to do this, the dataset was split into 6 folds, in which a cross validation was applied. After adjusting the parameters of the classifiers and descriptors, we performed the tests utilizing the testing images of the dataset. Considering all 40 signs, we found a recognition rate of 96,77% for the two stage classifier. The average recognition rate and standard deviation for each sign, considering all folds, can be seen on Table III. For most of the signs, the approach obtained high rates and even the worst results were higher than 86%.

The major loss of hit rate resulted of the group recognition stage. Fig. 8 shows the rate (%) obtained for each

Table III
RECOGNITION RATE FOR ALL SIGNS.

Signs	Mean recognition rate (%) +/- Std. Deviation
1	100.00 +/- 0.00
2	95.83 +/- 3.76
4	99.16 +/- 2.04
5	100.00 +/- 0.00
7	96.67 +/- 6.05
9	95.83 +/- 4.91
A	95.00 +/- 6.32
B	86.67 +/- 10.80
C	90.00 +/- 7.07
D	96.67 +/- 4.08
E	98.33 +/- 2.58
F	97.5 +/- 4.18
G	99.16 +/- 2.04
I	89.16 +/- 8.01
L	95.00 +/- 8.36
M	96.67 +/- 6.05
N	99.16 +/- 2.04
O	96.67 +/- 2.58
P	100.00 +/- 0.00
Q	98.83 +/- 2.58
R	96.67 +/- 2.58
S	96.67 +/- 5.16
T	100.00 +/- 0.00
U	90.83 +/- 5.84
V	95.83 +/- 5.84
W	95.00 +/- 4.47
X	98.33 +/- 2.58
Y	95.00 +/- 8.36
Adult	95.83 +/- 4.91
America	100.00 +/- 0.00
Gas	100.00 +/- 0.00
House	100.00 +/- 0.00
Identity	98.33 +/- 4.08
Law	99.16 +/- 2.04
Little	90.83 +/- 9.17
Plane	100.00 +/- 0.00
Together	98.33 +/- 4.08
Stone	99.16 +/- 2.04
Word	100.00 +/- 0.00
Verb	95.00 +/- 4.47

sign considering stages 1 (group recognition), 2 (specific recognition) and final recognition.

On Fig. 8, it can be seen that the first stage limited the final recognition rate of the approach, being lower than the group stage for most signs. In terms of hit rate, it was obtained an average hit rate of 97,00% on the first stage. On the second one, it was obtained 98,43%. Signs 'B' and 'I' presented the lowest rates. Both cases can be associatated to mistakes committed by the classifier on group stage.

It can be noticed that high recognition rates were obtained for the two-handed signs (group 10). This can be associated to their hand configurations that are quite different from the others present on the database.

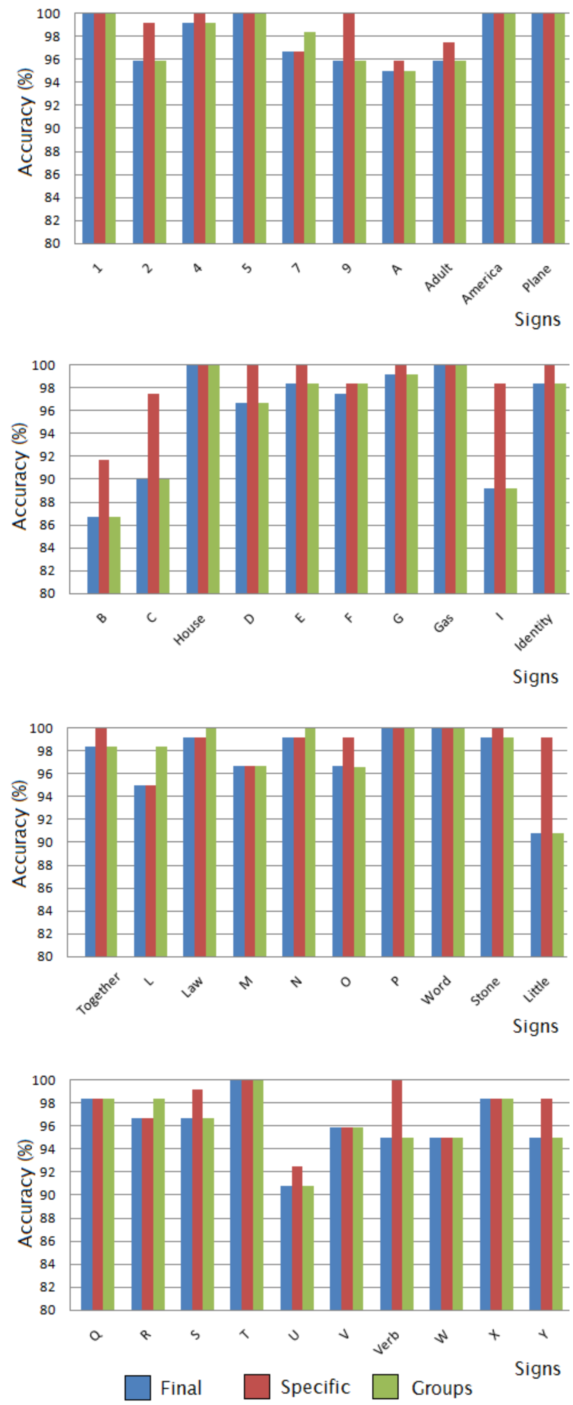

Figure 8. Comparison between stages

A. Validating the approach and the results

In order to ensure the validity of the present approach and the results obtained, variations were performed to justify the architecture adopted and understand the impact of each descriptor (tests 1 and 2) on the classification. Furthermore, another variation was carryed out to evaluate the robustness of the results found (test 3).

These were: (test 1) use of a single classifier and classification in a single stage; (test 2) use of each descriptor isolated; (test 3) tests considering an individual not present in the training set (20 new test images were acquired for each gesture). The Wilcoxon rank test (5% significance level) was performed to compare data shown on Table III and the results of tests 1 and 2. It was proven that the 2-stage approach using both descriptors showed outcomes statistically superior than the ones obtained with a single one. It was also noticed that the test 3 showed high recognition rates, evidencing the robustness of the approach, since the individual used as a model for this test showed different hand characteristics (hand size, distance among fingers, skin color, etc.) and different hand postures when compared to any other individual present in the training dataset. Not changing the weights of the neural networks or any aspect of the configuration of their layers also demonstrate this robustness. Table IV shows the average hit rate obtained on each test.

Table IV
AVERAGE RECOGNITION OF TESTS 1,2 AND 3

Test	**Recognition rates (%)**
Result of the approach	96.77
Using single classifier (1)	91.02
Using only HOG (2)	94.33
Using only Zernike (2)	86.62
Individual not present (3)	93.50

As a final test, we evaluated our approach using a public hand gesture dataset, the NTU Hand Digit Dataset [9]. This test was performed in order to test the robustness and compare it to other methods applied to gesture recognition.

The NTU is not a Libras dataset and contains 1000 images from 10 diferent gestures and 10 different subjects. It contains images and a distance map related to them.

In order to apply this approach on the NTU Dataset, it was necessary to use the distance map to section the hands on the images through bounding boxes (hands are closer than any other object) and the skin approach detection was also applied. On Fig. 9, images of NTU Dataset can be seen after sectioning the hand regions. It is also noticiable that even after the sectioning, elements of background are still present.

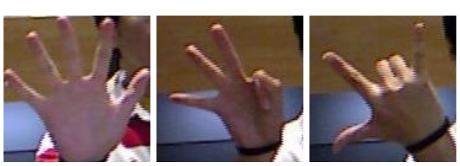

Figure 9. Images from NTU Dataset after sectioning.

It is noticed from Table V that our approach obtained similar results compared to previous work applied to this dataset.

Table V
RECOGNITION RATE OF METHODS USING NTU DATASET

Algorithm (method)	**Recognition rates (%)**
Our approach	95,8
HOG [10]	93,1
H3DF [10]	95,5
FEMD [9] (Near Convex)	93,9

VI. CONCLUSION

The present approach showed high hit rates, with an average recognition rate of 96,77% and low standard deviation values for most signs.

The skin recognition strategy, the Libras image dataset and the gesture recognition approach itself represent contributions of the present study. They showed good results and could be applied in contexts beyond the present work. In addition, the results found proved to be robust and comparable to works used as a reference regarding the gesture recognition, such as those proposed by [9] and [10].

It was realized that the signs with the lowest recognition rates were those where the first stage did not show the best results, it evidences that a better group distribution could provide better results.

As future work, we intend to perform tests varying the group distribution (signs of each group) and evaluate if it is necessary to apply a strategy or technique to help on this task or to handle with the signs wrongly recognized on the first stage. We also intend to improve our approach to cover Libras signs that present motion.

REFERENCES

[1] S. Mitra and T. Acharya, *Gesture Recognition: A Survey*, Transactions of System, Man Cyber Part C, pp. 311-324, v.37, n.3, 2007.

[2] V. I. Pavlovic, R. Sharma and T. S. Huang, *Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review*, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 677-695, v. 19, n. 7, 1997.

[3] B. Teodoro and L. Digiampetri, *A local alignment based on sign language recognition system*, Proceedings of SibGRAPI 2013, Works in Progress (WIP), Brazil, 2013.

[4] T. A. Felipe, *A Libras em Contexto*, Walprint Grafica e Editora, Brasilia, 2007.

[5] R. Z. Khan and N. A. Ibraheem, *Survey on Gesture Recognition for Hand Image Postures*, Computer and Information Science Journal, pp. 110-121, v. 5., 2012.

[6] F. Parvini, D. Mcleod, C. Shahabi, B. Navai, B. Zali and S. Ghandeharizadeh, *An Approach to Glove-Based Gesture Recognition*, Proceedings of the 13th International Conference on Human-Computer Interaction. Part II: Novel Interaction Methods and Techniques, pp. 236-245, 2009.

[7] M. A. Mohandes, *Recognition of Two-handed Arabic Signs using the CyberGlove*, Proceedings of the 4th International Conference on Advanced Engineering Computing and Applications in Sciences, Italy, 2010.

[8] M. Panwar, *Hand Gesture Recognition based on Shape Parameters*, International Conference on Computing, Communication and Applications (ICCCA' 12), 2012.

[9] R. Zhou, Y. Junsong, Z. Zhengyou, *Robust Hand Gesture Recognition Based on Finger-earth Mover's Distance with a Commodity Depth Camera*, Proceedings of the 19th ACM International Conference on Multimedia, pp. 1093-1096, 2011.

[10] C. Zhang, X. Yang and Y. Tian, *Histogram of 3D Facets: A characteristic descriptor for hand gesture recognition*, International Conference on Automatic Face and Gesture Recognition, pp. 1-8, 2013.

[11] J. Triesch and C. Malsburg, *Robust classification of hand postures against complex backgrounds*, International Conference on Automatic Face and Gesture Recognition , pp. 170-175, 1996.

[12] M. Zahedi, D. Rybach, T. Deselaers and H. Ney, *Using Geometric Features to Improve Continuous Appearance-based Sign Language Recognition*, Proceedings of BMVC 06, 17th British Maschine Vision Conference, pp. 1019-1028, 2006.

[13] M. H. Rahman and J. Afrin, *Hand Gesture Recognition using Multiclass Support Vector Machine*, International Journal of Computer Applications, pp. 39-43, v. 74, 2013.

[14] D. Uebersax, J. Gall, M. Van den Bergh and L. V. Gool, *Real-time sign language letter and word recognition from depth data*, 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp. 383-390, 2011.

[15] R. Bowden, D. Windridge, T. Kadir, A. Zisserman and M. Brady, *A Linguistic Feature Vector for the Visual Interpretation of Sign Language*, Proceeding of European Conference on Computer Vision (ECCV '2004), pp. 390-401, v. 3021, 2004.

[16] S. Liwicki and M. Everingham, *Automatic recognition of fingerspelled words in British Sign Language*, Proceedings of the 2nd IEEE Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB' 09), pp. 50-57, 2009.

[17] E. B. Pizzolato, M. S. Anjo and G. C. Pedroso, *Automatic Recognition of Finger Spelling for LIBRAS Based on a Two-layer Architecture*, Proceedings of the 2010 ACM Symposium on Applied Computing SAC '10, pp. 969-973, 2010.

[18] M. S. Anjo, E. B. Pizzolato and S. Feuerstack, *A Real-time System to Recognize Static Gestures of Brazilian Sign Language (Libras) Alphabet Using Kinect*, Proceedings of the 11th Brazilian Symposium on Human Factors in Computing Systems, IHC '12, pp. 250-268, Brazil, 2012.

[19] S. Maji, *A Comparison of Feature Descriptors*, Scientific Literature Digital Library, 2005.

[20] S. Tian, S. Lu, B. Su and C. L. Tan, *Scene Text Recognition Using Co-occurrence of Histogram of Oriented Gradients*, 12th International Conference on Document Analysis and Recognition (ICDAR), 2013.

[21] D. G. Tsolakidis, D. I. Kosmopoulos and G. Papadourakis, *Plant Leaf Recognition Using Zernike Moments and Histogram of Oriented Gradients*, 8th Hellenic Conference on Artificial Intelligence: Methods and Applications, 2014.

[22] H. A. Qader, A. R. Ramli and S. Al-Haddad, *Fingerprint Recognition Using Zernike Moments*, The International Arab Journal of Information Technology, v.4, 2006.

[23] T. Gritti, S. Caifeng, V. Jeanne and R. Braspenning, *Local features based facial expression recognition with face registration errors*, 8th International Conference on Automatic Face and Gesture Recognition (FG '08), pp. 1-8, 2008.

[24] N. Dalal and B. Triggs, *Histograms of Oriented Gradients for Human Detection*, Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR '05), pp. 886-893, v.1, USA, 2005.

[25] A. Misra, T. Abe and K. Deguchi, *Hand Gesture Recognition Using Histogram of Oriented Gradients and Partial Least Squares Regression*, Proceedings of the MVA 2011, pp. 479-482, 2011.

[26] H. Hse, A. R. Newton, *Sketched Symbol recognition using Zernike Moments*, Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004), pp. 367-370, United Kingdom, 2004.

[27] A. Khotazand and H. H. Yaw, *Invariant image recognition by Zernike moments*, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 489-497, v. 12, 1990.

[28] S. K. Hwang and K. Whoi-Yul, *A novel approach to the fast computation of Zernike moments*, Pattern Recognition Journal (PR), pp. 2065-2076, v. 39, 2006.

[29] M. Oujaoura, R. Ayachi, M. Fakir, B. Bouikhalene and B. Minaoui, *Zernike moments and neural networks for recognition of isolated Arabic characters*, International Journal of Computer Engineering Science (IJCES), v. 2, Korea, 2012.

[30] K. Kovac, P. Peer and F. Solina, *Human Skin color Clustering for face detection*, Proceedings of EUROCON 2003. Computer as a Tool, pp. 144-148, 2003.

[31] G. Gomez, M. Sanchez and L. E. Sucar, *On selecting an appropriate colour space for skin detection*, Proceedings of Mexican Int. Conference on Artificial Intelligence, pp. 70-79, Mexico, 2002.

[32] A. Bhuiyan, V. Ampornaramveth, S. Muto and H. Ueno, *Face Detection and Facial Feature Localization for Human-machine Interface*, NII Journal 2003, v. 5, 2003.

[33] J. Ruiz and R. Verschae, *Skin Detection using Neighborhood Information*, Proceedings of the 6th International Conference on Automatic Face and Gesture Recognition (FG2004), pp. 463-468, Korea, 2004.

[34] G. B. Huang, M. Ramesh, T. Berg and E. Learned-Miller, *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*, University of Massachusetts technical report, pp. 07-49, 2007.