Static Gestures Recognition for Brazilian Sign Language with Kinect and Eigenhands

Sérgio Bessa Carneiro
Department of Computing
Pontifical Catholic University of
Goiás, PUC-GO
Goiânia, Brazil

Edson D. F. de M. Santos
Department of Computing
Pontifical Catholic University of
Goiás, PUC-GO
Goiânia, Brazil

Talles M. de A. Barbosa
Department of Computing
Pontifical Catholic University of
Goiás, PUC-GO
Goiânia, Brazil

José O. Ferreira
Department of Computing
Pontifical Catholic
University of Goiás,
PUC-GO
Goiânia, Brazil

Adson F. Da Rocha
Department of Computing
UnB Gama Faculty, FGA
Brasília, Brazil

ABSTRACT

The automatic recognition of sign languages to help hearing impaired people is an area that has been explored for quite some time. However, this is still a practical problem due to the complexity involved making it a big challenge. Environments with complex backgrounds and different lighting conditions, hinders the recognition of hand gestures and other expressions when performed. The use of devices that are developed to increase the man machine interaction for entertainment, such as the Microsoft Kinect, has been shown to be promising in gesture recognition due to the amount of resources that it provides in the development of applications involving movements. The ease in detecting users and regions of interest such as hands, greatly reduces the complexity in the process of capturing gestures made by a user in complex environments. Therefore, this paper proposes an application for users of the Brazilian Sign Language (Libras), which brings a main feature of voice generation based on static hand gestures. The application utilizes the Kinect technology concomitantly with the *eigenhands* technique linked with the process of lighting normalization in the gesture images. The preliminary results obtained through testing had an accuracy reaching up to 89% in the static gesture recognition of the language LIBRAS.

General Terms

Static Gesture Recognition.

Keywords

Microsoft Kinect, Brazilian Sign Language.

1. INTRODUCTION

The Brazilian Sign Language (LIBRAS) is used in Brazil by individuals especially those that have hearing problems, to aid in the communication process. It is a visual language since it requires vision in the process of capturing the messages and movements especially of the hands to convey meaning [1]. In addition, since the enactment of law No. 10,436 in 2002 it is an official language. A large part of the Brazilian population uses primarily the Portuguese language as a mean in the process of communication. This often builds a barrier during the communication process for individuals with special needs. Thus, computer systems that are able to do automatic translations of the sign language can be used as an intermediate between the communication process of individuals who use oral language and sign language. There has been several published papers in the field of translation and automatic recognition of Sign Languages. In [2], [3], [4], and [5] there are systems that utilizes gloves for gesture recognition. This type of approach facilitates the tracking of hands and recognition of similar gestures, presenting good results. However, these devices provide its users with inconveniences for being little ergonomic. Other methods widely used in gesture recognition are those based on computer vision. In [6] a system was developed for the recognition of the Marata sign language used in India which had results of 90% accuracy in recognizing gestures. The system used hand segments and a webcam. In [7] a Libras image gesture recognition system is presented that uses hand shaped descriptors and neural networks as its classifiers producing results of 96.77% accuracy. In [8] the authors propose a gesture recognition system for the ASL utilizing the Wavelet transformation showing results of 97.4% accuracy. Aiming to improve the recognition time in [9] the authors propose an architecture of an embedded platform to recognize the Brazilian Sign Language (LIBRAS) based on FPGA and images obtained from a camera. This paper proposes a system for static gesture recognition of LIBRAS using Microsoft Kinect concomitantly with the eigenhands technique linked with the process of lighting normalization in the gesture images obtained through the Kinect. Differently from the aforementioned studies where images were obtained in controlled environments, often with simple backgrounds, and not considering different lighting conditions.

2. MATERIALS AND METHODS

2.1 Architecture of Proposal Application

The Block diagram of the proposed application is shown in Figure 1. The gesture recognition process performed by the application is done in the following steps: user detection, capture of gestures, normalization of lighting, recognition, and classification.
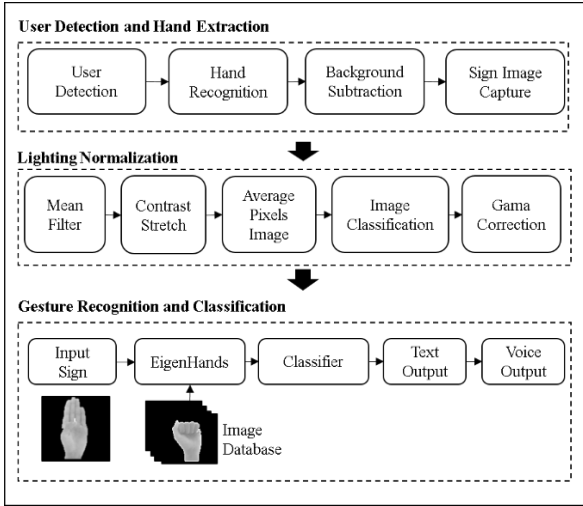
Figure 1: Block Diagram proposal application

2.2 User Detection and Hand Extraction

Figure 2 shows the application's interface in the process of detection and capturing of static gestures made by a user.

Figure 2: Application's interface

Initially, the aplication detects the user using the Kinect. It then recognizes the hand and performs a background replacement operation in real time. Resulting in an image with the dimensions of 100x100 pixels in grayscale with a dark background. This process allows the aplication to run in environments that have complex backgrounds, helping to greatly increase its accuracy.

2.3 Lighting Normalization

Lighting conditions greatly affect the applications recognition rate of gestures. Therefore, a process of lighting normalization is used in the captured images of gestures. Seen in [10] a similar approach is used, where an adaptation is made to fit the context of this work. First, an average filter with a 3x3 mask is applied in order to reduce the influence of stray pixels in the gesture images. Next, to improve the images contrast a method known as contrast stretch [11] is applied. The average value for every image is calculated using all its pixels, with the objective of classifying the image in a group according to their lighting condition. Based on the group each image was classified, a gamma correction is then applied. The values used by the application is shown in Table 1**.**

Table 1. Gamma Values Based on Classification

Average Image (interval)	Image Group	Gama Factor (γ)
[0, 62]	Very Dark	2.0
[63, 123]	Dark	1.5
[124, 190]	Light	0.5
[191, 255]	Very Light	0.2

The gamma correction is seen in (1) where *c* is a constant, *r* is the input image, and *s* is the normalized image.

$$s = cr^{\gamma} \tag{1}$$

The values for gamma were selected after tests with the application. Thus, the calibration is performed automatically by the application according to the light intensity from where the gesture image was collected.

2.4 Gesture Recognition and Classification

After the image normalization, the recognition process is initiated using the technique *eigenfaces*[12], named *eigenhands* when applied to hand images. This extracts all the relevant information from the analyzed image. Every image is broken down into a series of main components, thus representing images efficiently through the technique PCA [13].

A hand image L(x, y) in a two-dimensional space NxN can be represented as a vector of dimension N², so an image of 100x100 pixels becomes a vector of dimension 10,000. The objective of the algorithm is to find vectors that best represents the distribution of the hand images within the entire image space [12]. Each of these vectors represents an NxN image and constitutes the hands linear combination of original images.It is necessary to mount a base images to make the implementation of training. All images must be the same size in pixels and be in gray gradient with values between 0 and 255. Each image of this training set with M elements represents a vector $\Gamma_i(\Gamma_1, \Gamma_2, ..., \Gamma_k)$. After being defined the training set is necessary to calculate the average hand image (μ) according to (2).

$$\mu = \frac{1}{M}\sum_{n=1}^{M}\Gamma_n \tag{2}$$

Every image in the training set differs from the average by the vector (3). This set of vectors is then subject to PCA [13].

$$\emptyset_i = \Gamma_i - \mu \tag{3}$$

This helps to determine a set of M orthonormal vectors (a_n) which best represent the distribution of the data. The vector of order k, a_k are chosen such that (4) is a maximum.

$$\Delta_k = \frac{1}{M}\sum_{n=1}^{M}(a_k^T\emptyset_n)^2 \tag{4}$$

Subject to:

$$a_l^T a_k = \delta_{lk} = \begin{cases}1, & se\ l = k \\ 0, & otherwise\end{cases} \tag{5}$$

The vectors a_k and values Δ_k are respectively the vectors and values of the covariance matrix (6).

$$C = \frac{1}{M}\sum_{n=1}^{M}\emptyset_n\emptyset_n^T = AA^T \tag{6}$$

Where the matrix $A = \emptyset_1 \emptyset_2 ..., \emptyset_M$. Since matrix C has the dimensions of N²xN², it is extremely complex to calculate the values of N² eigenvectors. Based on [12] it is possible to reduce the number of N² eigenvectors to M (number of images in the training set), since M<<N² simplifying calculations. To solve for Matrix C the first step is solving a matrix MxM and then applying linear combinations of the hand images $\emptyset_i$. Consider eigenvectors v_i of $A^T A$ such that it satisfies (7).

$$A^T A v_i = \beta_i v_i \qquad (7)$$

Multiplying both sides of (7) by A, note that $A v_i$ represents the eigenvectors of C. With this it is possible to construct a matrix $H = A^T A$, where:

$$H_{mn} = \emptyset_m^T \emptyset_n \qquad (8)$$

Determining the M eigenvectors (v_i) of H. These vectors determine the linear combinations of the M training set hand images, yielding the eigenhand (ρ_i).

$$\rho_l = \sum_{k=1}^{N} v_{lk} \emptyset_k \qquad (9)$$

The eigenhand represents the greatest variance in the hands space, making it possible to reconstruct any image from the training set with just M eigenvectors [12].

2.5 Gesture Classification
The eigenhand determined from H (8) that exhibits the highest eigenvalues will be used in the classification process. Let an input gesture image (Γ) this will be projected into the hands space by (10). Where, k=1, 2...M.

$$\delta_k = \rho_k^T (\Gamma - \mu) \qquad (10)$$

Each weight (δ_k) describes the contribution of each eigenhand in representing the input hand image. The input image (Γ) is classified based on the Euclidian distance (11).Where $\tau^T = \delta_1, \delta_2, \delta_M$.

$$\epsilon_k^2 = \|(\tau - \tau_k)\|^2 \qquad (11)$$

3. RESULTS
3.1 Data Acquisition
To carry out the training of the developed application, static gesture images were collected from four volunteers. Each volunteer performed the same set of gestures in three different scenarios: well-lit indoor environment, indoor low lighting, and outdoors with natural light, shown in Figures 3, 4 and 5. These settings were chosen based on the lighting thereof.

Figure 3: Well-lit indoor environment

Figure 4: Indoor low lighting

Figure 5: Outdoors with natural light

At the end of the collection the training set consisted of 480 images that corresponded to the letters in the LIBRAS alphabet, excluding H, J, K, X and Z since they are not static gestures. Figure 6 shows the type of gestures that were used for the collection and testing.
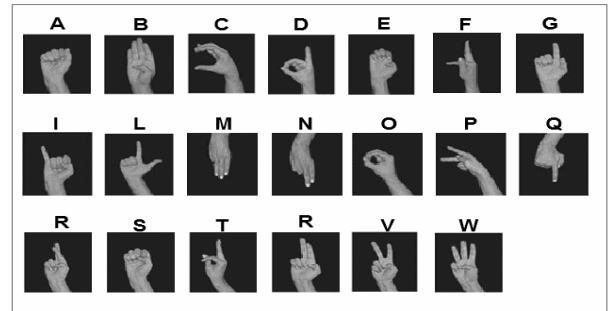

Figure 6: Static gestures Brazilian sign language [14]

3.2 Tests
To validate the application, tests were conducted with two new volunteers of the LIBRAS language, unknown to the application. Each volunteer performed five gestures of different letters corresponding to the alphabet in a sequential order. Within an indoor low lighting environment, and at a distance of 2 meters from the Kinect. Figure 7 shows the confusion matrix of the tests.

Gestures Predicted

	A	B	C	D	E	F	G	I	L	M	N	O	P	Q	R	S	T	U	V	W
A	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
B	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	8	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	1	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	2	0	0	0
G	0	0	0	0	1	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	1	8	1	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	2	8	0	0	0	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	2	0	0	0	0	8	0	0	0	0	0	0
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0
S	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0
T	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	8	0	0	0
U	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	9	0	0
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	8	1
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	9

Figure 7: Confusion matrix

4. DISCUSSIONS

The application presented results that were most satisfactory for the evaluated samples, reaching 89% accuracy. It is important to highlight that for some gestures the application obtained 100% accuracy. The use of the Kinect technology concomitantly with the *eignhands* technique linked with the process of lighting normalization, allowed the use of the application in environments with complex backgrounds and non-uniform lighting conditions. It was possible to observe some cases where the application classified some gestures erroneously, as seen in Figure 7 in the areas close to the diagonal fields that have non-zero values. This is often due to the similarity between the gestures, often changing the position of the thumb as shown in Figure 6.

5. CONCLUSIONS

This paper proposes an application for static gesture recognitions of Libras. The use of the Kinect in the application allowed for the detection of users and capturing of gestures with the lowest complexity. Due to the use of the programming libraries previously developed for this purpose. The technique *eignhands* concomitantly with the process of lighting normalization allowed for higher accuracy in the application for environments with non-uniform lighting. In future works the application should be improved in order to increase its accuracy, while maintaining the same operating conditions. Furthermore, the use of advanced technologies for better image acquisition for instance the Kinect version 2 and newer techniques for automatic calibration such as neural networks is being researched. In addition, the recognition samples will be expanded to include dynamic gestures of Libras, gestures that requires movements of the arms, body or facial expressions.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Quadros, Ronice Müller. "Estudos de línguas de sinais: uma entrevista com Ronice Müller de Quadros". ReVEL, vol. 10, n. 19, 2012.

[2] P. Lokhande, R. Prajapati e S. Pansare, "Data Gloves for Sign Language Recognition System", In: International Journal of Computer Applications, National Conference on Emerging Trends in Advanced Communication Technologies, 2015.

[3] S. Garg, S. Singhal e S. Kumari. "A review paper on Hand Gesture Recognition and Voice conversion system". In: International Journal of Scientific &Engineering Research, Volume 6, Issue 10, October-2015.

[4] A.Z Shukor, M.F Miskon, et al. "A New Data Glove Approach for Malaysian Sign Language Detection". IEEE International Symposium on Robotics and Intelligent Sensors, Volume 76, 2015, p. 60–67.

[5] Hernandez-Rebollar, J.L, Lindeman, R.W Kyriakopoulos. "A Multi-Class Pattern Recognition System for Practical Finger Spelling Translation", Fourth IEEE Int'l Conference on Multimodal Interfaces, Pittsburgh, USA, October 14-16, 2002, pp. 185-190.

[6] S. Amitkumar, R.Kagalkar. "Advanced Marathi Sign Language Recognition using Computer Vision". In: International Journal of Computer Applications, Volume 118 – No. 13, May 2015.p. 2-7.

[7] I.L.O. Bastos, M.F. Angelo, A.C. Loula. "Recognition of Static Gestures applied to Brazilian Sign Language (Libras)" (2016, 11, April). [Online]. In: http://sigrapi.sidinpe.br/col/sid.inpe.br/sigrapi/2015/06.1 8.19.06/doc/PID3766353.pdf.

[8] A. Thalange and S. Dixit. "Sign Language Alphabets Recognition Using Wavelet Transform". In: Conference on Intelligent Computing, Electronics Systems and Information Technology. Kuala Lumpur (Malaysia) Aug 25-26, 2015.

[9] F. M.P. Neto, et. al."Extreme Learning Machine for Real Time Recognition of Brazilian Sign Language".In: IEEE International Conference on Systems, Man, and Cybernetics, 2015.

[10] M.M. Mendonça, J. G. Denipote, R.A. S. Fernandes, M.S.V. Paiva. "Novo Método para Normalização da Iluminação em Imagens de Face". Biblioteca Brasileira de Computação, 2008.

[11] R. C. Gonzalez and R E. Woods. Digital Image Processing. 3th ed. Pearson International Edition, 2009.

[12] M. Turk and A. Pentland, "Eigenfaces for Recognition", Journal of Cognitive Neuroscience, vol.3, no. 1, pp. 71-86, 1991

[13] Jolliffe IT. Principal Component Analysis. 2th ed. New York: Springer, 2002.

[14] Alfabeto Libras (11 de Abril 2016). [Online]. In: http://www.surdo.com.br/alfabeto-libras.html.