LIBRAS Sign Language Hand Configuration Recognition Based on 3D Meshes

Andres Jessé Porfirio*, Kelly Laís Wiggers†, Luiz E. S. Oliveira*, Daniel Weingaertner*
*Department of Informatic
Federal University of Parana, Curitiba, Brazil
Contact: http://web.inf.ufpr.br/vri
†Universidade Estadual do Centro Oeste, Guarapuava, Brazil.

***Abstract*—This paper presents a method for recognizing hand configurations of the Brazilian sign language (LIBRAS) using 3D meshes and 2D projections of the hand. Five actors performing 61 different hand configurations of the LIBRAS language were recorded twice, and the videos were manually segmented to extract one frame with a frontal and one with a lateral view of the hand. For each frame pair, a 3D mesh of the hand was constructed using the Shape from Silhouette method, and the rotation, translation and scale invariant Spherical Harmonics method was used to extract features for classification. A Support Vector Machine (SVM) achieved a correct classification of $Rank1 = 86.06\%$ and $Rank3 = 96.83\%$ on a database composed of 610 meshes. SVM classification was also performed on a database composed of 610 image pairs using 2D horizontal and vertical projections as features, resulting in $Rank1 = 88.69\%$ and $Rank3 = 98.36\%$. Results encourage the use of 3D meshes as opposed to videos or images, given that their direct, real time acquisition is becoming possible due to devices like LeapMotion® or high resolution depth cameras.**

***Index Terms*—sign language, hand configuration, 3D mesh, LIBRAS.**

I. Introduction

Sign languages are one of the main communication methods used by deaf people, but opposed to common thought, there is no universal sign language: every country or even regional group uses its own set of signs. The official sign language in Brazil is called LIBRAS, and given the large population that uses LIBRAS in their daily life[1], there is a great demand for tools that enable its use in digital systems.

The use of sign language in digital systems can enhance communication in both directions: animated avatars can synthesize signals based on voice or text recognition; and sign language can be translated into text or sound based on images, videos and sensors input. The latest is the ultimate goal of this research, but LIBRAS is not a simple spelling of spoken language [1], so that recognizing isolated signs or letters of the alphabet (which has been a common approach) is not sufficient for its transcription and automatic interpretation.

A LIBRAS recognition approach should be based on the language's global parameters, which, according to Guimares [2], are: hand configuration, location or point of articulation, movement, palm orientation and facial expression. These parameters are combined to compose signs in a similar manner as phonemes are used to form words in spoken (oral) language.

This work focuses on the recognition of the one of LIBRAS global parameters: hand configuration, which is the shape that the hand of the actor presents during the execution of a sign. According to Pimenta [3] LIBRAS has 61 possible hand configurations, shown in Fig.1.

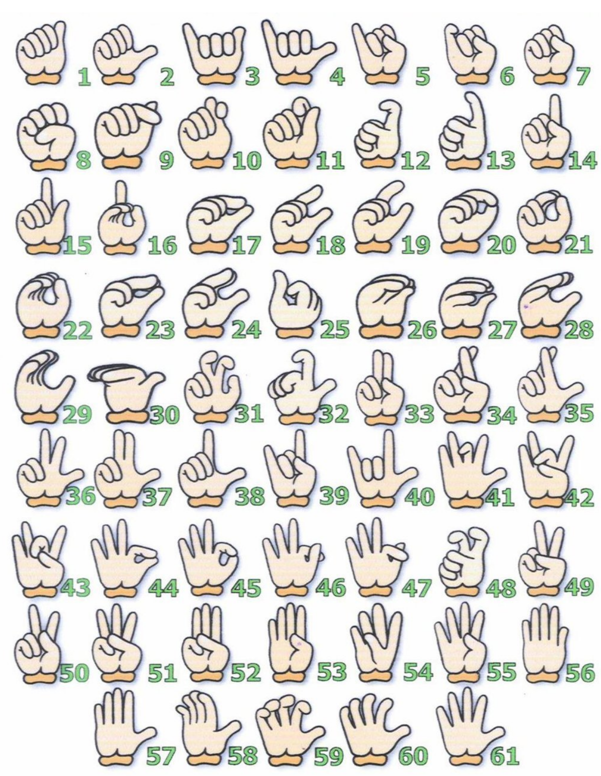

Fig. 1. LIBRAS Hand Configurations [3].

The contribution of this work is two-fold. Firstly, we introduce a database composed of 610 videos acquired with a Kinect® sensor, where 5 actors perform all 61 hand configurations of the LIBRAS twice, including rotations and translations. A derived database containing a frontal and a

[1]There is no official estimate of LIBRAS speakers, but according to the 2010 census, Brazil has more than two million persons declared deaf or with severely impaired hearing.

lateral view frame image is extracted and also available[2].

Secondly, we propose a classification method based on 3D meshes of the hand. The main reason for that choice, as opposed to more commonly used video or sensor data, is the fact that 3D meshes suffer no influence from illumination or occlusion issues, and can be used to extract scale, rotation and translation invariant features for classification. Since the direct capture of mesh data was not available at the time of this research, artificially generated meshes from video input was used. But recently developed technologies such as LeapMotion[3] and Kinect based software [4] promise to provide real time 3D meshes of the hand in a near future.

Through a comprehensive set of experiments we show that the proposed method is a good alternative to recognize hand configurations. The results reported in this study show that SVM trained with spherical harmonic features achieve a performance of $Rank1 = 86.06\%$ and $Rank3 = 96.83\%$ on a database composed of 61 classes (10 meshes per class). These results compare favorably to the literature, and a "scaled down" experiment shows equivalent results to a ten classes problem reported at [5].

The next sections are organized as follows: Section II reviews related works. Section III describes the hand configuration data acquisition and preprocessing. Section IV presents the proposed method for hand configuration classification and Section V presents the experimental results. In the last section the conclusions are stated, as well as the perspectives for future work.

II. Related Work

Recognition of sign languages is a widely studied topic in literature, beginning with the work of Shantz [6] in 1982. There are two main approaches for sign language recognition: an approach based on sensory gloves or wearable sensors, and a vision based approach that uses videos and static pictures [7], [8]. The latest is the one considered in this paper, although both approaches are very similar in the classification step. The main difference is on data acquisition and features extraction, with vision based approaches having a clear usability advantage over sensory approaches.

Marcotti [9] built a database with four sets of images, each set containing 21 photos of different LIBRAS signs. The relative position of the hand in the image plane, the relative area of the object and the moment of inertia around the horizontal and vertical axes were extracted and classified using Weka's J48 software. Unfortunately the hit rates of the classifier were not reported.

Wang [10] used two cameras capturing different angles of the scene and reconstructed a standardized frontal view of the actor. The vocabulary was composed of 100 signs of Chinese sign language, and colored gloves were used to facilitate segmentation of hand and fingers. Hidden Markov Models (HMM) fed with the fundamental matrices of stereo vision system as features achieved a recognition rate of 94%.

Quan [8] described experiments classifying hand signals based on spatial and temporal information extracted from video sequences. The database captured by the author consisted of 30 letters of the Chinese alphabet, with 195 images representing each letter, totaling 5850 images. Color histogram, Hu moments, Gabor filters and Fourier descriptors were used as input features for a Support Vector Machine (SVM), and hit rates were 95.55%.

Starner [11] worked in recognition of sentences of American Sign Language (ASL) using two cameras: one positioned in front of the signer and the other fixed on the signer's cap, pointing downwards (focus on hands). The recorded vocabulary consisted of 40 words that were combined to form sentences. Among the features used for the classification are: the position of the hand in the image, the hand translation during the frames and the size of the area corresponding to the hand in the image. The HMM classifier achieved a recognition rate of 92% for the frontal camera, and 97% for the cap camera.

Li [12] also worked with sentence recognition of ASL and used the Kinect sensor to acquire data. The feature vector was based on the joints of the body calculated by the Kinect and the author used template matching technique for comparing the signals. There were no reported hit rates, and the author states that the prototype is able to accurately recognize a list of 11 phrases.

Keskin [5] performed the recognition of ASL hand configurations of the 10 digits with videos acquired using Kinect. The author used a synthetic database (rendered from a model) for training and testing, and later applied the system in a real environment (using images from Kinect). The method is based on obtaining a 3D skeleton of the hand which, combined with 21 segmented hand parts, form the feature vector. The classifiers used in the experiment were Decision Forests, Neural Networks and SVM, and in all cases the SVM had the best results with an accuracy rate of 99.9%.

Zafrulla [13] developed a sentence verification system for an electronic game for deaf children called *CopyCat*. Experiment videos with Kinect were acquired in two steps: in the first the users were seated and in the second standing. The classifier used was the HMM and the feature vector was based on the position of the body joints supplied by the Kinect. The hit rates in the verification of sentences were 51.5% for seated users and 76.12% for standing users.

As can be seen, features extracted for classification vary broadly, depending on the available data (sensor, 2D, 3D); and HMM or SVM are the most used classifiers, with SVM achieving better results when comparison was available. Besides, most authors focus on the recognition of isolated words, letters, numbers or predefined sentences. Since sign languages are complex linguistic systems that have rules and structures such as: phonology, morphology, syntax and semantics, an efficient recognition system would need to be trained to recognize all the words in a given language [1], [14]. The use of vocabularies with single words for the recognition of sign languages is

[2]http://web.inf.ufpr.br/vri/databases/

[3]LeapMotion is a promising 3D sensor to be released in 2013 (http://www.leapmotion.com).

therefore inappropriate due to the large amount of data required for training the system [15], [16].

III. LIBRAS Hand Configurations Database

A video database (*LIBRAS-HC-RGBDS*) containing all 61 hand configurations of LIBRAS was acquired with the Kinect sensor, using specially developed acquisition software based on the OpenNI[4] library.

Five actors performed the predefined sequence of movements depicted in Fig.2 twice for each hand configuration, resulting in 610 videos (10 for each class). Besides the RGBD (red, green, blue, depth) channels, the videos also contain the skeleton of the actor, as generated by OpenNI during acquisition, allowing to track the actor's joints. Each video has a duration of 5 to 10 seconds and VGA resolution (640×480 pixels).

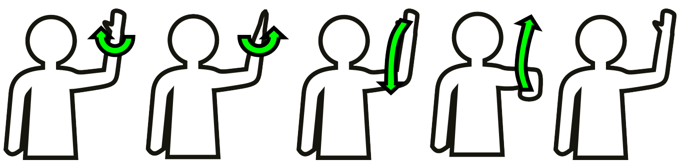

Fig. 2. Sequence of movements performed by actors in front of Kinect sensor.

The *LIBRAS-HC-RGBDS* database, as well as the acquisition and playback software, can be downloaded at http://web.inf.ufpr.br/vri/databases/. It was used to create two artificial hand configuration databases: one with 2D images (*DB_2D*) and one with 3D meshes (*DB_3D*) of the hand.

A. 2D Hand Configuration Database

Two frames, one with a frontal and one with a lateral view of the hand were manually extracted from each video of the *LIBRAS-HC-RGBDS*. Then the depth information was used to segment the hand, and all images were submitted to a manual correction process, to avoid artifacts and incorrect rotations that could compromise the generation of 3D meshes. The resulting 1220 images were scaled to a common size of 50×50 pixels, and compose the 2D LIBRAS hand configurations database (*DB_2D*). An example can be seen in Fig.3a.

B. 3D Hand Configuration Database

Both views of each hand configuration from *DB_2D* were used to create the 3D LIBRAS hand configurations database (*DB_3D*), containing 610 meshes of hand configurations, each corresponding to one video of the original *LIBRAS-HC-RGBDS*. The meshes were generated with the Shape from Silhouette method [17], as depicted in Fig.3.

This method generates a three-dimensional volume from the intersections of a group of silhouette images of a given object. It assumes that the images of the object are segmented (only the silhouette visible) and that all images are collected with the object in a static pose.

[4]OpenNI, the standard framework for 3D sensing, available at http://www.openni.org/.

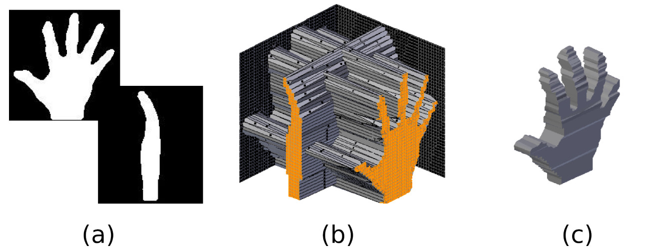

Fig. 3. Shape from Silhouette method. *DB_2D* frontal and lateral view (a), projections (b) and 3D mesh (c).

Reconstruction with the silhouette method using two input images resulted in meshes that do not resemble very much a real hand mesh, especially due to hard corners and high number of polygons (Fig.4a). In order to approximate the meshes to the output expected of a 3D sensor or a more sophisticated reconstruction technique, the meshes were smoothed using two algorithms: *Decimate* and *Dual Contouring*.

The *Decimate* algorithm [18] is mainly focused on polygonal reduction, and although it reduces the number of polygons of the object (Fig.4b) it still leaves irregular/unnatural features at the corners, as can be seen in Fig.4c.

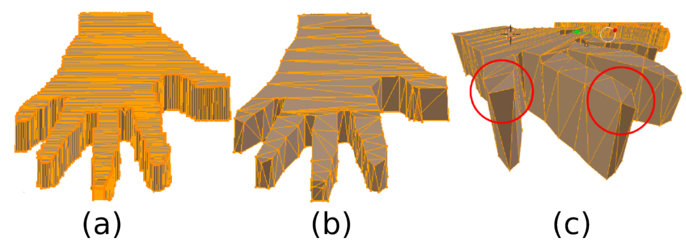

Fig. 4. Original 3D object (a), object after decimate *Decimate* (b) and irregular corners (c).

The *Dual Contouring* algorithm [19] also reduces the number of polygons, but produces smoother and more regular meshes, as seen on Fig.5. Therefore it was chosen to smooth all meshes before feature extraction, resulting in 610 hand configuration meshes that simulate the result of a 3D sensor.

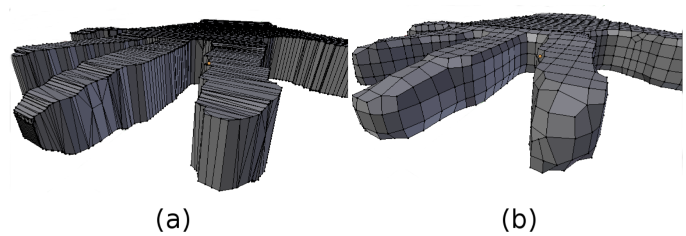

Fig. 5. Polygonal reorganization, original (a) and reorganized mesh after Dual Contouring (b).

IV. Proposed Method

A. 2D Feature Extraction

In order perform a comparison between 2D and 3D classification, following classic features were extracted from the

DB_2D database using the OpenCV[5] library: seven Hu Moments [20]; eight Freeman Directions [21]; Horizontal and Vertical Histogram Projections [22]. Features extracted from both views of the same hand were used together, so that the feature vectors had 14, 16, 100 and 100 elements, respectively. The features were also normalized for each classification test.

B. 3D Feature Extraction

One of the main reasons for using 3D meshes is that it allows the extraction of features that are invariant to lighting conditions, translation, rotation and scale. The Spherical Harmonics [23] is a 3D mesh feature extracting method that meets these requirements, and its choice was based on a comparative study between mesh descriptors and manual classification developed by Clark *et.al.* [24]. That study compared human made classification to three descriptors: D2 shape distribution, Spherical Harmonics and Surface Partitioning Spectrum, and concludes that the Spherical Harmonics was the most effective due to the low false negatives rate.

As depicted in Fig.6, the Spherical Harmonics method takes the 3D mesh of an object, rasterizes it into a 64^3 voxels volume (1); the volume is then decomposed into 32 spherical functions (2); each function is then decomposed as the sum of its first 16 harmonic components, analogous to a Fourier decomposition (3). Each of the 16 decompositions generates a signature (4); and finally the combination of all signatures generates the feature set of the 3D mesh (5).

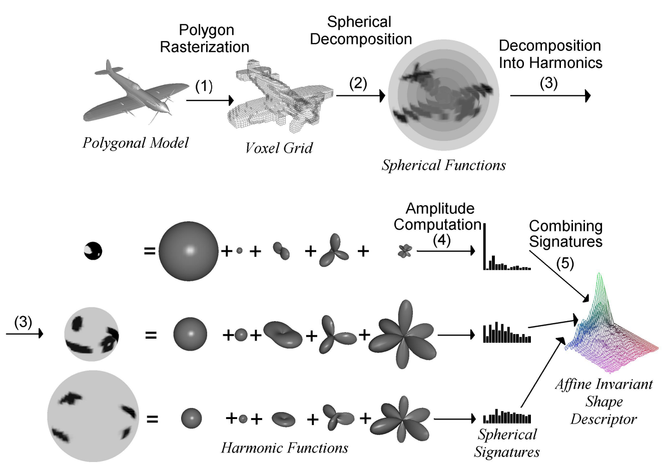

Fig. 6. Spherical Harmonics Computation [25].

A tool[6] developed by Kazhdan was used to compute the features for the *DB_3D* database, resulting in the 610 feature sets used in the classification step. Each feature vector is composed of 544 real values: 16 harmonic components and 1 alignment parameter for each of the 32 spherical functions.

V. Experimental Results

This section describes three experiments undertaken to evaluate hand configuration recognition of the LIBRAS sign language. The first two experiments evaluate the classification performance of the proposed method on the *DB_2D* and *DB_3D* databases, respectively. The third experiment uses a reduced set of hand configurations in order to compare the developed method to a similar experiment reported in [5].

The Support Vector Machine (SVM) algorithm, implemented in the LibSVM toolset [26], was used to classify the LIBRAS hand configurations, according to best practices recommended by [27]. Two SVM kernels were tested: Linear and RBF, achieving very similar results (inside the standard deviation margin), with the Linear performing $\approx 1\%$ better than RBF.

For each experiment 10 data sets were generated. Each set is a random separation of the database with 70% of the instances for training and 30% for testing. The SVM average hit rate for all 10 data sets (Rank 1 and Rank 3) are reported. Thus, every hand configuration was tested 30 times.

A. 2D Features Classification

TABLEI shows the Rank 1 and Rank 3 classification results for the 2D database (*DB_2D*) classified with the SVM Linear and RBF Kernels, using the descriptors presented in Subsection IV-A. The rotation invariant descriptors (Hu and Freeman) did not perform well, and best results were obtained with a combination of horizontal and vertical histogram projections.

TABLE I
Average hit rate of 10 SVM executions using *DB_2D*.

	Feature	Linear kernel		RBF kernel	
		Hit (%)	σ (%)	Hit (%)	σ (%)
Rank1	Horizontal Projection	82.33	1.26	83.42	1.09
	Vertical Projection	76.03	2.64	74.78	2.17
	Vertical + Horizontal	88.15	2.40	88.96	1.45
	HU Moments	57.06	3.67	56.73	3.01
	Freeman Directions	10.65	1.90	6.57	4.91
Rank3	Horizontal Projection	97.93	0.90	98.36	0.48
	Vertical Projection	94.89	2.25	94.61	0.89
	Vertical + Horizontal	98.58	0.69	98.75	0.68
	HU Moments	81.73	2.49	82.66	3.21
	Freeman Directions	24.45	2.34	13.26	9.99

It is necessary to emphasize that this result is mainly due to the “artificially perfect” acquisition of the *DB_2D* database. Under real conditions, acquisition of segmented and noise free images from video sequences of LIBRAS conversation is subject to the classic segmentation problems encountered by most of the authors (illumination problems, occlusion, rotation and scale variance).

The importance of this result is twofold: first it provides a base classification rate for a carefully suited hand configuration images database, which cannot be easily obtained under real-time conditions, against which the 3D method can be compared. Secondly, it shows that simple characteristics, that can be easily extracted from a 3D mesh, allow for a good classification.

B. 3D Features Classification

The 3D features extracted with the Spherical Harmonics method, as described in Section IV-B, were also classified

[5]http://opencv.willowgarage.com.
[6]Available at http://www.cs.jhu.edu/~misha/HarmonicSignatures/.

using SVM. TABLEII shows the hit rate for Rank 1 and Rank 3. Considering that the 3D features are invariant and could be acquired with similar quality in real time, the 96.83% hit rate for Rank 3 can be considered very good.

TABLE II
AVERAGE HIT RATE OF 10 SVM EXECUTIONS USING *DB_3D*.

	Linear kernel		RBF kernel	
	Hit (%)	σ (%)	Hit (%)	σ (%)
Rank1	86.06	2.06	85.68	1.65
Rank3	96.83	1.31	96.67	1.27

As can be seen in Fig.1, LIBRAS has various similar hand configurations, that even humans might confuse during a conversation. In this case the context of the conversation and other global parameters (face expression, motion) contribute to the correct understanding of the signal. This also highlights the importance of the Rank 3 classification, which is very useful on a context-aware system.

Observing the confusion matrix of the classification (TABLEIII), it is possible to observe that there are some similar meshes, especially for hand configurations between 18 and 28, which are confused by SVM. Fig.7 shows two 3D meshes corresponding to hand configurations 20 and 21. The visual and structural similarity (polygonal organization) between them is great, leading to classification errors.

TABLE III
PARTIAL CONFUSION MATRIX FOR 10 RUNS OF SVM USING *DB_3D*. SHOWN ARE HAND CONFIGURATIONS 17 TO 29.

	17	18	19	20	21	22	23	24	25	26	27	28	29
17	22	0	0	3	1	0	1	0	1	2	0	0	0
18	1	18	9	1	0	0	0	0	0	1	0	0	0
19	0	4	17	0	1	0	0	0	4	0	0	2	0
20	6	0	0	8	9	6	0	0	0	0	0	0	0
21	2	1	0	15	7	3	2	0	0	0	0	0	0
22	0	0	0	0	3	24	1	0	0	0	0	0	0
23	3	0	0	3	5	1	18	0	0	0	0	0	0
24	0	3	5	0	0	0	0	13	0	0	0	9	0
25	0	0	0	1	0	0	0	0	23	3	0	0	0
26	0	0	0	0	0	0	0	0	3	27	0	0	0
27	0	0	1	0	0	2	0	0	0	0	27	0	0
28	0	0	1	0	0	0	0	6	0	0	1	17	5
29	0	0	0	0	0	0	0	1	0	0	0	3	25

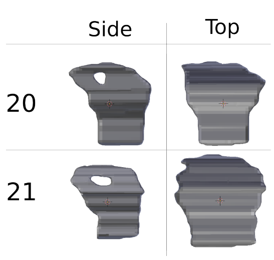

Fig. 7. Side and top view of two similar LIBRAS hand configuration meshes.

Furthermore, the lack of detail in the 3D mesh caused by the silhouette technique omits information that could distinguish these hand configurations. This problem can be corrected with the use of more accurate meshes from another sensor or technique, improving the classifier hit rate.

Another classification test was performed to analyze the impact of polygonal reorganization and smoothing of the meshes (*Dual Contouring*, Subsection III-B). The test was realized under identical conditions (SVM linear and RBF kernels and default parameters), but the meshes resulting from the Silhouette method were not submitted to a polygonal reorganization filter. The results shown in TABLEIV let us conclude that the suavization, performed to create a more realistic mesh, also produced (slightly) better classification results.

TABLE IV
AVERAGE HIT RATE OF 10 SVM EXECUTIONS WITH *DB_3D* AND NO MESH PREPROCESSING.

	Linear kernel		RBF kernel	
	Hit (%)	σ (%)	Hit (%)	σ (%)
Rank1	86.01	1.49	84.64	2.29
Rank3	96.88	1.00	95.46	1.12

C. *Digits classification*

Comparing results with literature is very difficult, due to a multitude of factors: use of different sign languages, incompatible acquisition methods or feature extraction, different objectives (recognition of complete signals or sentences and not hand configurations). In such cases a direct comparison of the hit rate is not coherent.

The most similar work classifying hand configurations was described by Keskin [5], and uses artificially generated hand configurations based on a 3D model, with joint locations extracted as features. The result presented by the author is very good (a hit rate of 99.9%), however he only used hand configurations relative to the 10 digits of the American Sign Language (totaling 10 classes), whilst the LIBRAS has 61 hand configurations (aka. classes).

In order to obtain similar conditions, a “scaled down” experiment was performed. 10 classes of the LIBRAS hand configuration were selected, referring to the 10 digits. According to Pimenta and Quadros [3], the hand configurations used to represent digits in LIBRAS are expressed in the TABLEV (see also Fig.1).

TABLE V
LIBRAS HAND CONFIGURATIONS RELATIVE TO THE DIGITS.

Digit	0	1	2	3	4	5	6	7	8	9
Hand Config	22	2	38	51	55	31	25	9	7	25

Since hand configuration 25 is used for two digits (6 and 9), it was necessary to include a tenth class, which was chosen to be hand configuration 60.

The same classification procedure used on the previous experiments was used for the digits hand configurations, with the default SVM parameters and Linear and RBF kernels. Results presented in TABLEVI and are similar to Keskin’s work, showing that the proposed method is a viable alternative.

TABLE VI
AVERAGE OF 10 SVM EXECUTIONS IN A 10 CLASSES PROBLEM USING *DB_3D*.

	Linear kernel		RBF kernel	
	Hit (%)	σ (%)	Hit (%)	σ (%)
Rank1	98.33	2.23	98.33	1.66
Rank3	100	0.00	100	0.0

VI. DISCUSSIONS AND CONCLUSION

Sign language recognition is socially important and much demanded because it facilitates inclusion of deaf people into daily life activities. Most methods proposed to date focus on the recognition of words, sentences, digits or letters. This work followed a different approach, by proposing a recognition method for a global parameter of the LIBRAS sign language: the hand configuration.

A database containing 610 videos of 5 actors performing, twice, all 61 hand configurations of the LIBRAS was created and is available for further research use. The videos also contain depth and skeleton information, as provided by the Kinect and OpenNI library.

The use of 3D meshes for classification of LIBRAS hand configuration was proved to be a viable alternative with hit rates above 96% on Rank 3. The great advantage of using 3D meshes is the ability to identify details of the hand and fingers, regardless of the position and orientation of the hand in space.

During the execution of a sign, the reconstruction of a 3D mesh allows the extraction of normalized features by correcting parameters such as orientation, position and scale of the 3D object, or by using a features invariant to these parameters (such as the Spherical Harmonics). In addition the use of 3D models avoids problems of lighting and occlusion, typical in two-dimensional images.

Although the 3D meshes used in the *DB_3D* database were artificially generated, it is predictable that technology advance (new devices and techniques) will allow the capture of real time 3D scenes in a near future, enabling the real-time construction of the 3D hand mesh.

Finally, future research should use a more powerful acquisition method, that generates better meshes for all frames of a video. A feature selection can reduce the complexity of the SVM and thus increase classification rates. And in order to advance in the goal of LIBRAS recognition, context and movement information must be added to the hand configuration.

REFERENCES

[1] D. R. Antunes, C. Guimaraes, L. S. Garcia, L. E. S. Oliveira, and S. Fernandes, "A framework to support development of sign language human-computer interaction: Building tools for effective information access and inclusion of the deaf," in *Proc. Fifth Int Research Challenges in Information Science (RCIS) Conf*, 2011, pp. 1–12.

[2] C. Guimaraes, D. R. Antunes, D. F. G. Trindade, R. A. L. Silva, and L. S. Garcia, "Structure of the brazilian sign language (libras) for computational tools: Citizenship and social inclusion," in *Organizational, Business, and Technological Aspects of the Knowledge Society*. Springer Berlin Heidelberg, 2010.

[3] R. M. d. Q. Nelson Pimenta, *Curso LIBRAS 1*, 4th ed. Vozes, 2010.

[4] I. Oikonomidis, N. Kyriazis, and A. Argyros, "Efficient model-based 3d tracking of hand articulations using kinect," in *BMVC 2011*. BMVA, 2011.

[5] C. Keskin, F. Kira, Y. E. Kara, and L. Akarun, "Real time hand pose estimation using depth sensors." in *ICCV Workshops*. IEEE, 2011, pp. 1228–1234.

[6] M. Shantz and H. Poizner, "A computer program to synthesize american sign language," *Behavior Research Methods*, vol. 14, pp. 467–474, 1982, 10.3758/BF03203314.

[7] H. Wang, M. C. Leu, and C. Oz, "American sign language recognition using multi-dimensional hidden markov models," *Journal of Information Science and Engineering*, vol. 22, no. 5, 2006.

[8] Y. Quan, "Chinese sign language recognition based on video sequence appearance modeling," in *Proc. 5th IEEE Conf. Industrial Electronics and Applications (ICIEA)*, 2010, pp. 1537–1542.

[9] P. Marcotti, L. B. Abiuzi, P. M. S. R. Rizol, and C. H. Q. Forster, "Interface para reconhecimento dalngua brasileira de sinais," *XVIII Simpsio Brasileiro de Informtica na Educao - SBIE - Mackenzie*, 2007.

[10] Q. Wang, X. Chen, L. Zhang, C. Wang, and W. Gao, "Viewpoint invariant sign language recognition," in *Proc. IEEE Int. Conf. Image Processing ICIP 2005*, vol. 3, 2005.

[11] T. Starner, J. Weaver, and A. Pentland, "Real-time american sign language recognition using desk and wearable computer based video," vol. 20, no. 12, pp. 1371–1375, 1998.

[12] K. F. Li, K. Lothrop, E. Gill, and S. Lau, "A web-based sign language translator using 3d video processing," in *Network-Based Information Systems (NBiS) Conference*, sept. 2011, pp. 356 –361.

[13] Z. Zafrulla, H. Brashear, T. Starner, H. Hamilton, and P. Presti, "American sign language recognition with the kinect," in *Proceedings of the 13th international conference on multimodal interfaces*, ser. ICMI '11. New York, NY, USA: ACM, 2011, pp. 279–286.

[14] S. M. Peres, F. C. Flores, D. Veronez, and C. J. M. Olguin, "Libras signals recognition: a study with learning vector quantization and bit signature," in *Proc. Ninth Brazilian Symp. Neural Networks SBRN '06*, 2006, pp. 119–124.

[15] P. Dreuw, D. Stein, T. Deselaers, D. Rybach, M. Zahedi, J. Bungeroth, and H. Ney, "Spoken language processing techniques for sign language recognition and translation," *Technology and Dissability*, vol. 20, no. 2, pp. 121–133, 2008.

[16] P. Dreuw and H. Ney, "Visual modeling and feature adaptation in sign language recognition," *Voice Communication (SprachKommunikation), 2008 ITG Conference on*, pp. 1 –4, oct. 2008.

[17] K. M. Cheung, S. Baker, and T. Kanade, "Visual hull alignment and refinement across time: A 3d reconstruction algorithm combining shape-from-silhouette with stereo," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2003.

[18] W. J. Schroeder, J. A. Zarge, and W. E. Lorensen, "Decimation of triangle meshes," *SIGGRAPH Comput. Graph.*, vol. 26, no. 2, pp. 65–70, Jul. 1992.

[19] T. Ju, F. Losasso, S. Schaefer, and J. Warren, "Dual contouring of hermite data," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 339–346, Jul. 2002.

[20] M. Hu, "Visual pattern recognition by moment invariants," *Information Theory, IRE Transactions on*, vol. 8, no. 2, pp. 179 –187, february 1962.

[21] H. Freeman, "Computer processing of line-drawing images," *ACM Comput. Surv.*, vol. 6, pp. 57–97, March 1974.

[22] M. H. Glauberman, "Character recognition for business machines," *Electronics*, vol. 29, pp. 132–136, 1956.

[23] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz, "Rotation invariant spherical harmonic representation of 3d shape descriptors," in *Proceedings of the 2003 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, ser. SGP '03, 2003, pp. 156–164.

[24] D. E. R. Clark, J. R. Corney, F. Mill, H. J. Rea, A. Sherlock, and N. K. Taylor, "Benchmarking shape signatures against human perceptions of geometric similarity," *Computer-Aided Design*, vol. 38, no. 9, pp. 1038–1051, 2006.

[25] M. Kazhdan and T. Funkhouser, "Harmonic 3d shape matching," in *ACM SIGGRAPH 2002 conference abstracts and applications*, ser. SIGGRAPH '02. New York, NY, USA: ACM, 2002, pp. 191–191.

[26] C. Chang and C. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.

[27] C. Hsu, C. Chang, and C. Lin, "A practical guide to support vector classification," 2010.