# Hu and Zernike Moments for Sign Language Recognition

K. C. Otiniano-Rodríguez, G. Cámara-Chávez, D. Menotti
UFOP - Federal University of Ouro Preto
Computing Department
Ouro Preto, MG, Brazil
Email: {karlaotiniano,gcamarac,menottid}@gmail.com

***Abstract*—Sign Language is a complex way of communication in which hands, limbs, head, facial expressions and body language play an important role for understanding between deaf-and-dumb people without the use of sounds. In this paper, we propose two methods for Sign Language Recognition using the SVM classifier and features extracted from Hu and Zernike Moments. In the experiments, a comparison between the proposed methods using a database composed of 2040 images for recognition of 24 symbol classes is performed. The results obtained by the method using the Zernike moments features overcomes the ones obtained by the method using the Hu moments achieving an accuracy rate about $96\%$ which is comparable to the ones found in the literature, which holds that our proposal is promising.**

***Index Terms*—sign language, support vector machine(SVM), Hu moments, Zernike moments, principal component analysis (PCA)**

## I. Introduction

Sign Language is a complex way of communication in which hands, limbs, head, facial expression and body language are used to communicate a visual-spatial language without sound, mostly used between deaf-and-dumb people. Wherever communities of deaf people exist, sign languages are developed. Their complex spatial grammars are remarkably different from the grammars of spoken languages [1], [2]. Hundreds of sign languages are in use around the world and are at the cores of local deaf cultures. Some sign languages have obtained legal recognition at some extend, while others have no status at all.

There are several defined languages such as ASL (American Sign Language), BSL (British Sign Language), Auslan (Australian Sign Language) and LIBRAS(Brazilian Sign Language) [1]. As these languages are barely known outside of the deaf community, a barrier between them and commom languages are usually imposed.

Since the past decades in the world of Computer Science, specifically in the area of Image Processing and Computer Vision, several techniques have been developed to achieve an adequate recognition rate of sign language. Over the years and with the advance of technology, methods have been proposed in order to improve the data acquisition or their processing. In this direction, we can find methods employing classifiers such as Hidden Markov Models (HMM) [3], [4], [5], Artificial Neural Networks (ANN) [2], [6], *etc*. Usually they differs in the way the features are extracted.

In this paper, we propose two methods for Sign Language Recognition using Hu or Zernike Moments (separately) to extract features of the images (static signs of the alphabet). The classification/recogntion task is performed using a SVM (Support Vector Machine) classifiers. Experiments are performed using a pubilc database composed of 2040 images stating 24 symbols classes [7]. The obtained results show that the accuracy obtained by the method using the Zernike Moments are quite greater than the one obtained by the method using Hu Moments. Moreover, the accuracy obtained by the proposed method is comparable to the one found in [8], which holds that our proposal is promising.

The remainder of this paper is organized as follows. Section II shows sign language definition. In Section III, our proposed methods are introduced. The experiments are presented in Section IV, where the results are discussed. Finally, conclusion and future fork are presented in Section V.

## II. Sign Language

In order to understand better the problem, we present the definition of sign and the parameters that define it.

### A. Signs

The signs are composed by the combination of shape and movement of hands and the body or a point in space where these signs are made. In sign languages, we can find the following parameters that form the signs [1]:

### B. Sign parameters

A sign have five basic parameters.

- **Shape or Hands Configuration**: They are hand shapes that can be from dactylology (manual alphabet) or other shapes made by the dominant hand (right hand for right handed or left for lefties), or by both hands.
- **Orientation**: or Direction. The signs have a direction with respect to other parameters.
- **Location**: or Point of articulation. Is where the hand focuses predominantly set, *i.e.*, where the sign is performed, may touch any part of the body or be in a neutral space.

- **Motion**: It is the change in time of any of the three functions described above. Is the most complex feature.
- **Facial and/or body expression**: Are of fundamental importance for the understanding of the sign, and the intonation in sign language is done through facial expressions.

## III. Sign Language Recognition using Moments

In the proposed model is only used the hand shape parameter. Figure 1 shows the proposed model. The first stage consists in segmenting the hand. Many methods have been proposed for hand segmentation, for example, skin color is used to detect and segment hands [9], but unfortunately by itself is not a reliable modality. For simplification, we assume that we have a uniform background and clothes. Thus, the segmentation can be easily perform using a threshold. After that, the segmented hand (binary image) could be used as a mask, for extracting the Zernike moments, or as a hand binary image for detecting the boundaries and computing the Hu moments. Finally, these descriptors are used as input of our SVM classifier.

### A. Feature Extraction

*1) Hu Moments:* The feature extraction with Moments seeks for global features and invariants of an image. Invariant moments are statistical measures designed to remain constant after some transformations, such as object rotation, scaling and translation. Such statistical moments work directly with regions of pixels. The moments provide a generic representation of any object and are easily extractable [10].

The moments most commonly used are the seven invariant moments of Hu of order 2 and 3:

$$
\begin{aligned}
\phi(1) &= (\mu_{20}+\mu_{02}) \\
\phi(2) &= (\mu_{20}-\mu_{02})^2+4\mu_{11}^2 \\
\phi(3) &= (\mu_{30}-3\mu_{12})^2+(3\mu_{21}-\mu_{03})^2 \\
\phi(4) &= (\mu_{30}+\mu_{12})^2+(\mu_{21}+\mu_{03})^2 \\
\phi(5) &= (\mu_{30}-3\mu_{12})(\mu_{30}+\mu_{12})[(\mu_{30}+\mu_{12})^2-3(\mu_{21}+\mu_{03})^2] \\
&\quad +(3\mu_{21}-\mu_{03})(\mu_{21}+\mu_{03})[3(\mu_{30}+\mu_{12})^2-(\mu_{21}+\mu_{03})^2] \\
\phi(6) &= (\mu_{20}-\mu_{02})[(\mu_{30}+\mu_{12})^2-(\mu_{21}+\mu_{03})^2] \\
&\quad +4\mu_{11}(\mu_{30}+\mu_{12})(\mu_{21}+\mu_{03}) \\
\phi(7) &= (3\mu_{21}-\mu_{03})(\mu_{30}+\mu_{12})[(\mu_{30}+\mu_{12})^2-3(\mu_{21}+\mu_{03})^2] \\
&\quad +(\mu_{30}-3\mu_{12})(\mu_{21}+\mu_{03})[3(\mu_{30}+\mu_{12})^2-(\mu_{21}+\mu_{03})^2]
\end{aligned}
$$

*2) Zernike Moments:* The Zernike polynomials were first proposed in 1934 by Zernike [11]. Their moment formulation appears to be one of the most popular, outperforming the alternatives [12] (in terms of noise resilience, information redundancy and reconstruction capability). The pseudo-Zernike formulation proposed by Bhatia and Wolf [13] further improved these characteristics. However, here we study the original formulation of these orthogonal invariant moments.

Complex Zernike moments [14] are constructed using a set of complex polynomials which form a complete orthogonal basis set defined on the unit disc $((x^2+y^2)\leq 1)$. They are expressed as $A_{pq}$. Two dimensional Zernike moment:

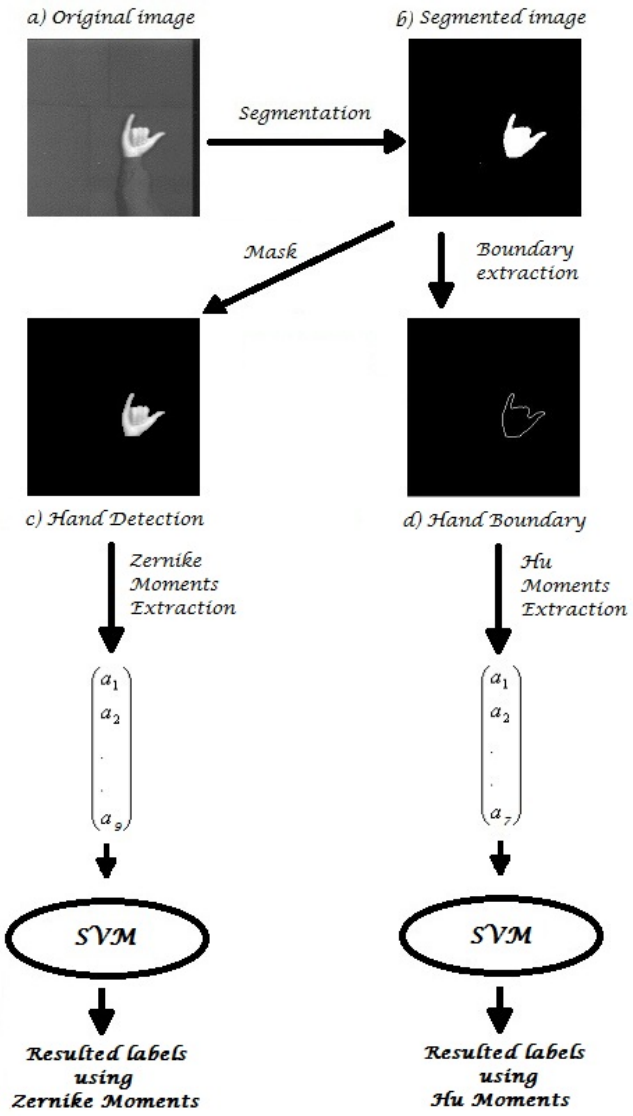


Fig. 1. Proposed model for Sign Language Recognition.

$$A_{mn}=\frac{m+n}{\pi}\int_x\int_y f(x,y)[V_{mn}(x,y)]^* dxdy \qquad (1)$$

where $x^2+y^2\leq 1$, $m=0,1,2,...,\infty$ and defines the order, $f(x,y)$ is the function being described and $^*$ denotes the complex conjugate. While $n$ is an integer (that can be positive or negative) depicting the angular dependence, or rotation, subject to the conditions:

$$m-|n|=even, |n|\leq m \qquad (2)$$

and $A_{mn}^*=A_{m,-n}$ is true. The Zernike polynomials [11] $V_{mn}(x,y)$ expressed in polar coordinates are:

$$V_{mn}(r,\Theta)=R(m,n)exp(jn\Theta) \qquad (3)$$

where $(r, \Theta)$ are defined over the unit disc, $j = \sqrt{-1}$ and $R_{mn}(r)$ is the orthogonal radial polynomial defined as:

$$R_{mn}(r) = \sum_{s=0}^{\frac{m-|n|}{2}} (-1)^s F(m, n, s, r), \quad (4)$$

where

$$F(m, n, s, r) = \frac{(m-s)!}{s!(\frac{m+|n|}{2} - s)!(\frac{m-|n|}{2} - s)!} r^{m-2s}, \quad (5)$$

where $R_{mn}(r) = R_{m,-n}(r)$.

Experiments were performed to obtain the best order, and they are shown in Section IV.

### B. Support Vector Machine

Support Vector Machines (introduced as a machine learning method by Cortes and Vapnik [15]). Furthermore, SVM have been successfully applied in many real world problems and in several areas: text categorization, handwritten digit recognition and object recognition, etc. The SVM have been developed as a robust tool for classification and regression in noisy and complex domains. SVM can be used to extract valuable information from data sets and construct fast classification algorithms for massive data.

Another important characteristic of the SVM classifier is to allow a non-linear classification without requiring explicitly a non-linear algorithm thanks to kernel theory.

In kernel framework data points may be mapped into a higher dimensional feature space, where a separating hyperplane can be found. We can avoid to explicitly compute the mapping using the kernel trick which evaluate similarities between data $K(d_t, d_s)$ in the input space. Common kernel functions are: linear, polynomial, Gaussian radial basis, gaussian with $\chi^2$ distance and triangular. In our experiments, we use a Gaussian kernel. The classification was performed using SVM. The library LIBSVM (A Library for Support Vector Machines) [16] was used in our implementation.

## IV. Experiments

The Gesture recognition database [7] consists of 2040 images of $248 \times 256$ pixels that represent 24 static signs in gray scale. The database is divided as follows:

- The signs A-F have 40 images for each class.
- The signs G-Y have 100 images for each class.

The sign J and Z are not used, because these signs have motion and the proposed model only works with static sign. Figure 2 shows some examples of Gesture recognition database.

In this work we conduct two experiments. In the first, different Zernike moments order were evaluated in order to find the the best order with low computational cost. In the second, we compare the Zernike moments obtain in the first experiments with Hu moments. In both experiments, we perform a cross validation with 10 folds.

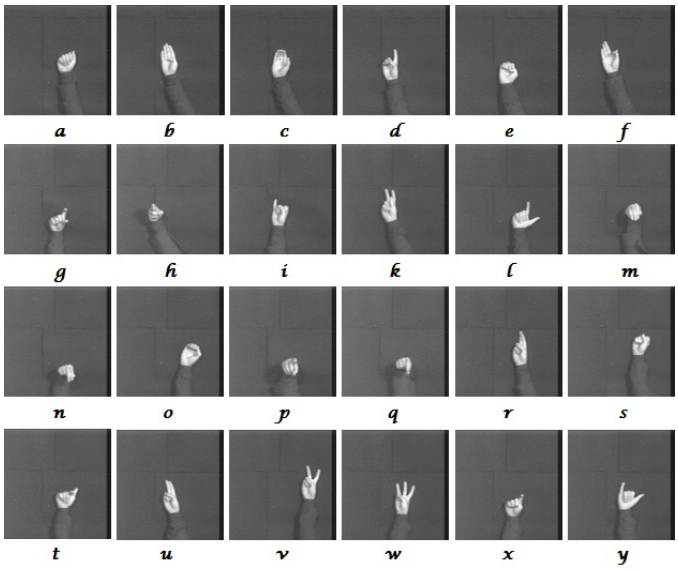


Fig. 2. Gesture recognition database: 24 static signs.

In the first experiment, different orders (from 5 to 10) were tested. In all cases, the accuracy obtained were over $90\%$. In Figure 3, we present the mean accuracy and standard deviation of the orders with best results, in this case the orders are 8 to 10. We can see, in all cases, the standard deviation remains with low value, this means that accuracy of Zernike moments stay stable.

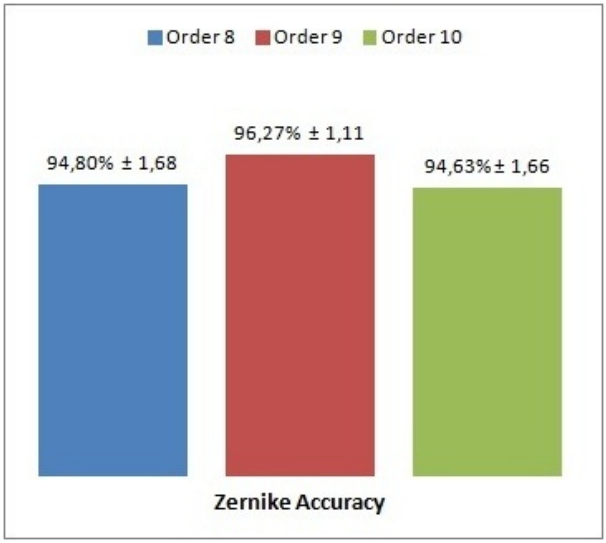


Fig. 3. Zernike Accuracies with order 8,9 and 10.

In the second experiment, a comparison between Hu and Zernike moments is done. Figure 4 shows the mean accuracy and standard deviation of both descriptors. The accuracy of Hu moments suffers more variations, we can see it through its standard deviation. As in the first experiment, Zernike

| Sign | Zernike Moments | | Hu Moments | |
|---|---|---|---|---|
| | Precision | Recall | Precision | Recall |
| A | 0.85 | 0.98 | 0.80 | 9.96 |
| B | 0.98 | 0.98 | 0.98 | 0.98 |
| C | 0.83 | 0.96 | 0.98 | 1.00 |
| D | 0.93 | 0.98 | 0.98 | 0.87 |
| E | 0.85 | 0.91 | 0.50 | 0.93 |
| F | 0.93 | 0.96 | 0.98 | 0.98 |
| G | 0.95 | 0.94 | 0.99 | 0.97 |
| H | 0.89 | 0.91 | 0.85 | 0.86 |
| I | 0.95 | 0.94 | 0.96 | 0.99 |
| K | 0.99 | 1.00 | 1.00 | 1.00 |
| L | 1.00 | 1.00 | 1.00 | 1.00 |
| M | 0.99 | 0.96 | 0.93 | 0.87 |
| N | 0.93 | 0.94 | 0.94 | 0.95 |
| O | 0.96 | 0.94 | 0.98 | 1.00 |
| P | 0.96 | 0.92 | 0.80 | 0.91 |
| Q | 0.94 | 0.96 | 0.89 | 0.88 |
| R | 1.00 | 0.98 | 0.99 | 0.99 |
| S | 0.98 | 0.98 | 1.00 | 0.81 |
| T | 0.98 | 0.98 | 0.94 | 0.77 |
| U | 1.00 | 0.99 | 0.96 | 0.99 |
| V | 1.00 | 1.00 | 1.00 | 1.00 |
| W | 1.00 | 1.00 | 1.00 | 1.00 |
| X | 0.98 | 0.95 | 0.70 | 0.91 |
| Y | 1.00 | 1.00 | 1.00 | 1.00 |

TABLE I
PRECISION AND RECALL OF ALL SIGNS

moments achieves good results (over $96\%$), better than Hu moments, with a low standard deviation.

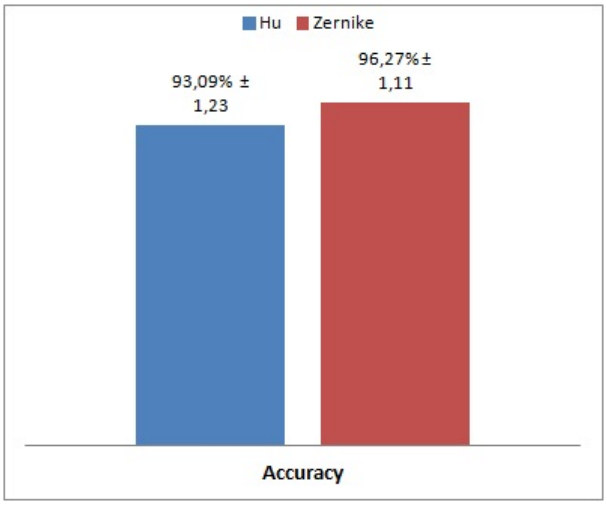


Fig. 4. Hu and Zernike accuracies.

We also evaluate both type of moments using precision and recall. The precision represents the portion of real positive items that were correctly classified among all items classified as positive. The recall represents the amount that was classified with success, in other words, how many items were correctly classified as a positive class. In Figure 5, we show the precision and recall of the signs with more difficult classification. Sign E has the lowest recognition rate because it is very similar to sign A. They only differ on the position of the thumb, see Figure 2. Even though this small position variation, Zernike moments get a better precision and recall (0.85 and 0.91, respectively). This behavior remains similar in the other three signs.

Table I shows the precision and recall for all signs. We can see that A, E ,H and P signs have the worst results for Hu as well as for Zernike descriptor. In Q, T and X signs, we can observe that Zernike moments obtain better results than Hu moments. However, there are signs, such as K, L, V, W and Y, that were totality recognized with both descriptors.

## V. CONCLUSIONS

In this paper, we proposed, implemented and tested two methods for Sign Language Recognition using the SVM classifier and features extracted from Hu and Zernike Moments. From the experiments, we concluded that Zernike moments obtained slightly better results ($\approx 96\%$) than Hu Moments ($\approx 93\%$). Nonetheless, both methods achieve promising results.

As future work, we plan to study some feature reduction selction method such PCA in order to improve the results obtained. Moreover, we plan to perform more tests on other databases in order to verify how robust are the methods using Zernike and Hu moments.

## VI. ACKNOWLEDGEMENTS

The authors would like to thank FAPEMIG, CAPES and CNPq for the financial support.

## REFERENCES

[1] LIBRAS, "Brazilian sign language," http://www.libras.org.br/, last visit: March 10, 2012.
[2] P. W. Vamplew, "Recognition of sign language gestures using neural networks," *Australian Journal of Intelligent Information Processing Systems*, vol. 5, pp. 27–33, 1996.
[3] G. Fang, W. Gao, and D. Zhao, "Large vocabulary sign language recognition based on hierarchical decision trees," in *International Conference on Multimodal Interfaces (ICMI)*, 2003, pp. 125–131.
[4] J. Zieren and K.-F. Kraiss, "Robust person-independent visual sign language recognition," in *Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA)*, 2005, pp. 1–8.
[5] C. Vogler, H. Sun, and D. Metaxas, "Framework for motion recognition with applications to american sign language and gait recognition," in *Workshop on Human Motion*, 2000.
[6] L. Vargas, L. Barba, and L. Mattos, "Sistema de identificacin de lenguaje de señas usando redes neuronales artificiales," *Revista Colombiana de Fsica*, vol. 42, no. 2, pp. 222–226, 2010.
[7] T. Moeslund, "Gesture recognition database," http://www.prima.inrialpes.fr/FGnet/data/12-MoeslundGesture/database.html, last visit: March 10, 2012.
[8] H. Birk, T. B. Moeslund, and C. B. Madsen, "Real-time recognition of hand alphabet gestures using principal component analysis," in *Scandinavian Conference on Image Analysis (SCIA)*, 1997, pp. 261–268.

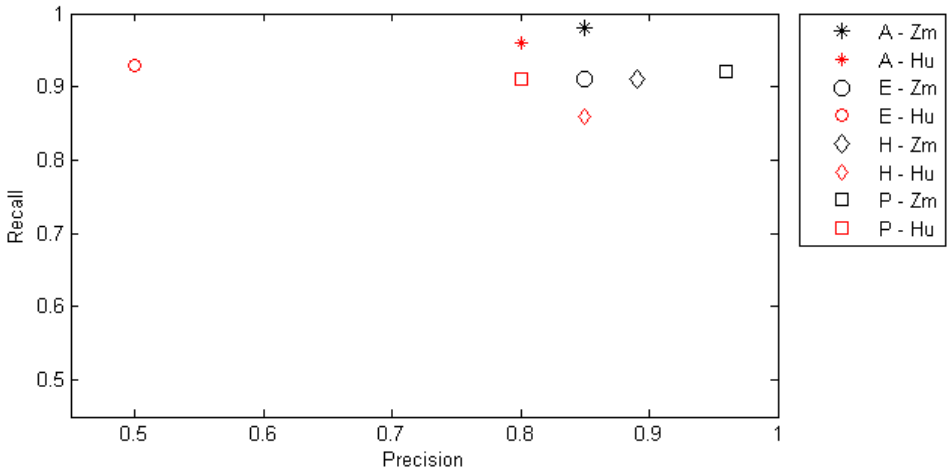


Fig. 5. Precision and Recall of A,E,H and P sings.


[9] M. Jones and J. Rehg, "Statistical color models with applications to skin detection," *International Journal of Computer Vision*, vol. 46, no. 1, pp. 81–96, 2002.

[10] M. K. Hu, "Visual pattern recognition by momentos invatiants," *IRE Transactions on Information Theory*, vol. 8, no. 1, pp. 179–187, 1996.

[11] F. Zernike, "Diffraction theory of the cut procedure and its improved form, the phase contrast method," *Physica*, vol. 1, pp. 689–704, 1934, beugungstheorie des Schneidenverfahrens und seiner verbesserten Form, der Phasenkontrastmethode.

[12] C. Teh and R. T. Chin, "On image analysis by the method of moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, p. 1988, 496–513.

[13] A. B. Bhatia and E. Wolf, "On the circle polynomials of zernike and related orthogonal sets," *Proceedings of the Cambridge Philosophical Society*, vol. 50, no. 1, pp. 40–48, 1954.

[14] M. R. Teague, "Image analysis via the general theory of moments," *Journal of the Optical Society of America*, vol. 70, no. 8, pp. 920–930, 1979.

[15] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[16] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.