Scientific
Research
Publishing

Automatic Mexican Sign Language Recognition Using Normalized Moments and Artificial Neural Networks

Francisco Solís, David Martínez, Oscar Espinoza

Universitary Center UAEM of Teotihuacan Valley, Autonomous University of Mexico's State, Axapusco, México
Email: xiuhcoatl0@gmail.com

Abstract

This document presents a computer vision system for the automatic recognition of Mexican Sign Language (MSL), based on normalized moments as invariant (to translation and scale transforms) descriptors, using artificial neural networks as pattern recognition model. An experimental feature selection was performed to reduce computational costs due to this work focusing on automatic recognition. The computer vision system includes four LED-reflectors of 700 lumens each in order to improve image acquisition quality; this illumination system allows reducing shadows in each sign of the MSL. MSL contains 27 signs in total but 6 of them are expressed with movement; this paper presents a framework for the automatic recognition of 21 static signs of MSL. The proposed system achieved 93% of recognition rate.

Keywords

Mexican Sign Language, Automatic Sign Language Recognition, Normalized Moments, Computer Vision System

1. Introduction

Sign Language is the main communication way for unhearing community [1] [2]. It is good to know Sign Language is complete; it means that signers can express as much as oral speakers but there is an important limitation for signers because most people don't know Sign Language; this phenomenon produces numerous obstacles to deaf people.

Deaf community learns to live with numerous limitations in day by day interactions with society [3] [4] [5]. The problem of trying to produce some technology capable to impact and make some important beneficial changes is huge, because it is a complex problem [6]. First Sign Language is not universal; there are many sign languages, for

instance, American Sign Language, Mexican Sign Language, Taiwan Sign Language, Indian Sign Language, Chinese Sign Language, Japanese Sign Language, among others.

There are many Sign Languages because of geographical and cultural reasons [6]. All Sign Languages are expressed using postures and movements of fingers, hand, arms, fists and body, including face expressions as well.

All these dynamics generate a complex problem to analyze in the computer science context, then major of systems for sign language recognition have serious limitations.

It could be said that computer vision systems for sign language recognition can be classified as two main groups (G1 and G2) based on acquisition data method [7]. First one (G1) uses electronics and devices (like position and movement sensors, accelerometers) to capture accurate data of fingers, hands and/or arms [8]; second group (G2) uses computer vision systems [9].

G1 systems have some interesting advantages, as long as they don't use digital cameras, they don't depend of illumination conditions, and they neither need to express sign oriented to some particular spot. Besides G1 systems provide the most accurate features of position, orientation, movement and velocity of signs. Nevertheless this kind of system has a serious disadvantage due to the permanent physical contact with the sensors [10].

On the other hand G2 systems allow a more natural interaction since they don't require signers to be connected physically to the system, but this benefit causes a considerable loss in data accuracy. Probably the most complex task in this kind of systems is segmentation (segmenting each hand from the other, hands from face, or hands from background). For their particularities, G2 systems have important limitations; some of them have special background solid color, and some other signers use special clothes, gloves or special color markers in order to locate and segment hands. It's harder in these systems to calculate accurate data of position or movement of fingers, hands or some other parts of the body needed to recognize some sign, because most of systems try to solve the problem using Digital Image Processing (DIP) techniques [10]. Some of these systems use a depth sensor to improve hand shape segmentation from background; they use kinect or some other devices to capture depth information [11].

Besides the classification based on acquisition data methods (G1 and G2), sign language recognition systems could be categorized by captured data range (C1, C2 and C3). C1 systems have the shorter range, due to they concentrate on finger spelling (finger movements and/or hand orientations), that's why movement range of hands is small. C2 systems use a wider range in order to capture hands movements around upper body (these kinds of systems focus on sentences recognition) and finally C3 projects consider face expressions in order to consider a sign language recognition system more complete.

This paper focuses on performing an automatic Mexican sign language recognition using a computer vision system. The meaning of automatic means for this particular case the continuous recognition of isolated alphabets (finger spelling) in Mexican sign language. Details of this development are described throughout the text by a four sections structure. Section 2 describes computer vision system; Section 3 presents a brief description about geometric, central and normalized moments; in Section 4 Mexican

sign language database is introduced and Section 5 reports experimental results; in final section conclusions are presented.

2. Mexican Sign Language Database

Mexican sign language according to CONAPRED (National Council to Discrimination Prevention of Mexican Government) consists in 27 signs, of which 21 are static signs and the rest are dynamic signs. For the purpose of this work, 21 static signs are considered to develop a computer vision system to recognize isolated signs automatically, in Figure 1 can be seen all static signs of Mexican sign language.

All signs in Figure 1 are in gray scale due to they represent the red channel from original RGB images. Experimentally red channel performs better than intensity (intensity from HSI model) at least to present work. Dynamic signs "j", "k", "ñ", "q", "x" and "z" are not considered in this research.

3. Computer Vision System

This paper focuses on automatic sign language recognition for isolated signs, in other words the purpose is to recognize finger spelling from Mexican sign language using a digital camera. For this work a digital IP camera was selected by its accessibility using Matlab, images from this particular camera are not good enough, due to illumination, Mexican sign language recognition in [12] was achieved using flash to reduce shadows, nevertheless digital camera used in present work doesn't have one. Therefore an illumination arrangement based on reflectors was employed.

In order to reduce shadows and improve segmentation process four LED reflectors were placed to 5.5 inches from camera, each one emits 700 lumens and has 120 degrees opening angle. The effect of this illumination is desirable due to generated shadow by each reflector is illuminated by the others and this is important for finger spelling because of this light configuration reduces vertical and horizontal shadows (see Figure 2).

Besides reducing shadows, reflectors improve segmentation of hand with background. A solid green background was selected to contrast the hand shape. Color gloves or special color markers were not required thanks to background (see Figure 3).
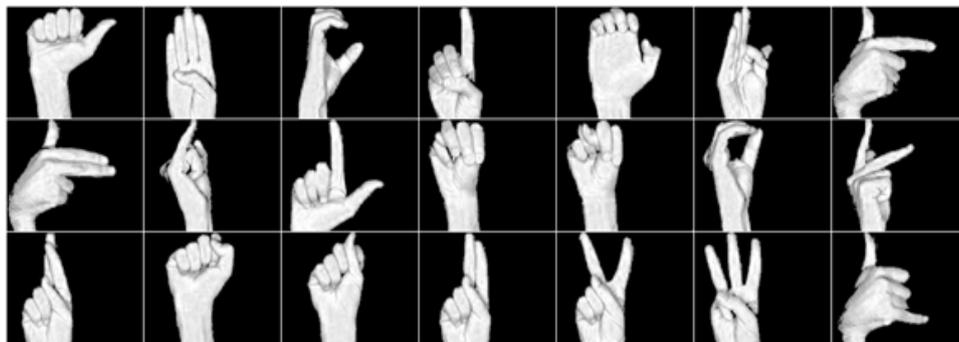
Figure 1. Mexican sign language. From left to right and from up to down, above pictures represent the signs: "a", "b", "c", "d", "e", "f", "g", "h", "I", "l", "m", "n", "o", "p", "r", "s", "t", "u", "v", "w" and "y" respectively.

Figure 2. Computer vision system (digital camera in the middle and four LED reflectors placed at corners).

Figure 3. Solid green background to segment hand shape.

The computer vision system described above provides appropriate images for finger spelling recognition; nevertheless images need to be processed to reduce computing costs. Normalized moments were selected experimentally in order to represent each image by some descriptors.

4. Normalized Moments

Moments can be computed from digital images. A digital image can be described as a bidimensional function $I(x, y)$, where spatial variables x and y represent some intensity level. Geometric moments of $(p+q)$ order can be defined in discrete way as: [13]

$$m_{pq} = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} x^p y^q I(x, y), \tag{1}$$

where M and N represent rows and columns of image I, these moments can be interpreted (at least their lower orders) as geometric measures. Moment m_{00} represents the area of a binary image, some relations between geometric moments can be interpreted also,

$$\overline{x} = \frac{m_{10}}{m_{00}} \tag{2}$$

and

$$\overline{y} = \frac{m_{01}}{m_{00}} \tag{3}$$

determine centroid of the image. Possibly geometric moments have an intuitive interpretation, but there are others which have significant properties, such as central moments defined as

$$\mu_{pq} = \sum_x \sum_y (x - \overline{x})^p (y - \overline{y})^q I(x, y),$$ (4)

these moments are important due to they generate invariant descriptors to translation. Central normalized moments are more sophisticated because they are invariant to scale and translation transformations, normalized moments are defined as

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}},$$ (5)

where

$$\gamma = \frac{p + q}{2} + 1 \forall p + q \geq 2$$ (6)

scale invariance is achieved by the normalization factor μ_{00}^{γ}. These moments are important for this work, because finger spelling presents translation and scale transformations. Normalized moments were used to represent each image captured from computer vision system.

5. Proposed System

In order to reduce computational costs, a black frame is always displayed to establish a ROI (Region Of Interest). This frame allows reducing computational costs by cropping original image. Cropped RGB image is reduced using one single channel thus 3D RGB matrix is reduced to a 2D matrix representing the red channel in gray scale. Later background is eliminated by an experimental threshold and hand shape is translated to an upper location, translation is the last digital image process. Forty two normalized central moments are computed for each image ($p = 0, 1, \cdots, 7$ and $q = 2, 3, \cdots, 7$) with experimentally orders ($p + q$) selected, this moments are used to generate descriptors of each image which are going to be introduced to a Multi-Layer Perceptron (MLP) to recognize the pattern (sign). MLP has forty two input neurons (each neuron for each normalized central moment) and twenty one output neurons (each neuron for each alphabet sign). Three versions per sign were captured and used to train the MLP. All this process described was programed using Matlab (GUI can be seen in **Figure 4(a)**).

Figure 4(b) shows a GUI developed for Automatic Mexican Sign Language Recognition using Matlab software, each frame processed last 0.4518 sec using a laptop with an i7-3630QM CPU @ 2.40 GHz and Windows 8 this system achieve 93% of recognition rate.

Figure 5 shows proposed framework to recognize Mexican Sign Language, starting with sign expression (a), in this case alphabet "f", using four reflectors and a green background to perform segmentation; then (b) a ROI (Region Of Interest) is displayed as a black frame and cropped. Red channel was selected (c) to represent each sign in order to reduce computational costs. Translation was made to improve recognition rate using an experimental threshold to point the starting hand shape. This step segment hand shape from green

(a)   (b)
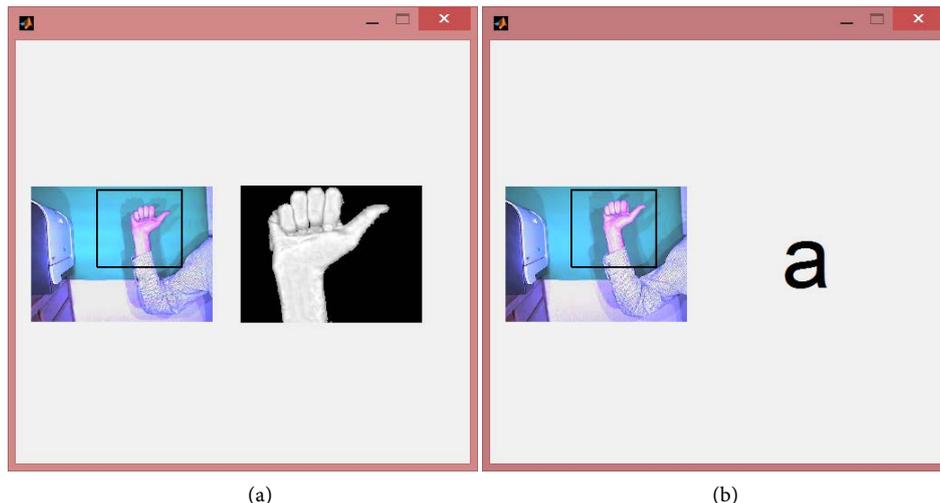
Figure 4. (a) System developed for training the MLP; (b) System developed to recognize automatically 21 signs from Mexican Sign Language.
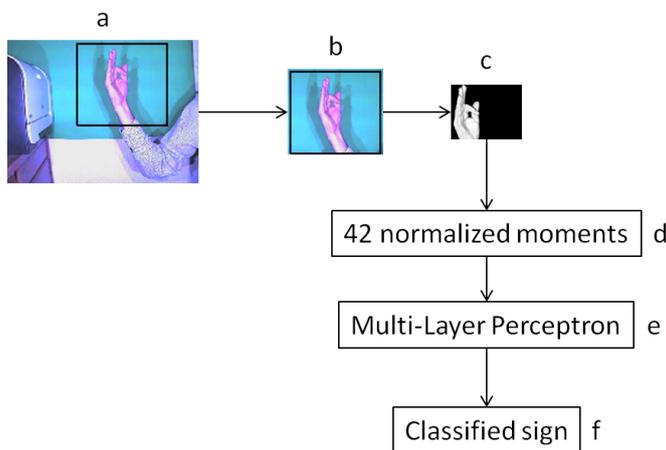
Figure 5. Framework proposed to recognize Mexican Sign Language. a—Sign expressed inside black frame; b—Cropped RGB image; c—Cropped and translated sign represented in gray scale of red channel; d—42 normalized moments computed per frame; e—Pattern recognition model and f—Sign classification by MLP.

background. Forty two normalized moments (d) was computed per frame. A Multi Layer Perceptron (e) was used as pattern recognition model. Finally (f) sign is classified and displayed as text.

6. Conclusion

Automatic Mexican sign language recognition is a complex task; the results reported in this work don't match the requirement in real life. An artificial computer vision is proposed using four LED reflectors and a network camera. Mexican sign language can be recognized by digital image processing using red channel to reduce computational costs. Establishing a ROI and a uniform green background was enough to get high recognition rate; using a background helps to avoid the use of special color markers or

gloves. Normalized central moments can be used to represent each frame properly in order to recognize signs using a Multi Layer Perceptron. The system achieves 93% of recognition rate and less than 0.5 sec of frame rate. For future work the system should be able to improve frame rate reducing computational costs. This framework recognizes isolated Mexican Sign Language alphabets so it's necessary to develop a system able to recognize dynamic signs which is a more complex task.

Acknowledgements

References

[1] Li, K., Zhou, Z. and Lee, Ch. (2016) Sign Translation Modeling and Scalable Solution to Continuous Sign Language Recognition for Real-World Applications. *ACM Transactions on Accessible Computing*, **8**, 7-23. http://dx.doi.org/10.1145/2850421

[2] Wang, H., Chai, X., Hong, X., Zhao, G. and Chen, X. (2016) Isolated Sign Language Recognition with Grassman Covariance Matrices. *ACM Transactions on Accessible Computing*, **8**, 14-21. http://dx.doi.org/10.1145/2897735

[3] El-Gayyar, M.M., Ibrahim, A.S. and Wahed, M.E. (2016) Translation from Arabic Speech to Arabic Sign Language Based on Cloud Computing. *Egyptian Informatics Journal*. (Article in Press) http://dx.doi.org/10.1016/j.eij.2016.04.001

[4] Nguyen, T., Huong, T., Vu, T., Le, T. and Vu, S. (2015) Static hand Gesture Recognition for Vietnamese Sign Language (VSL) Using Principle Components Analysis. 2015 *International Conference on Communications, Management and Telecommunications (ComManTel)*, DaNang, 28-30 December 2015, 138-141.

[5] Wu, J., Tian, Z., Sun, L., Estevez, L. and Jafari, R. (2015) Real-Time American Sign Language Recognition Using Wrist-Worn Motion and Surface EMG Sensors. *IEEE* 12th *International Conference on Wearable and Implantable Body Sensor Networks* (*BSN*), Cambridge, 9-12 June 2015, 1-6.

[6] Inoue, K., Shiraishi, T., Yoshioka, M. and Yanagimoto, H. (2015) Deph Sensor Based Automatic Hand Recognition Extraction by Using Time-Series Curve and Its Application to Japanese Finger-spelled Sign Language Recognition. *Procedia Computer Science*, **60**, 371-380. http://dx.doi.org/10.1016/j.procs.2015.08.145

[7] Li, Y., Chen, X., Tian, J., Zhang, X., Wang, K. and Jang, J. (2010) Automatic Recognition of Sign Language Subwords Based on Portable Accelerometer and EMG Sensors. *ICMI-MLMI'*10, Beijing, 8-10 November 2010, 1, 7.

[8] Sun, Ch., Zhang, T. and Xu, Ch. (2015) Latent Support Vector Machine Modeling for Sign Language Recognition with Kinect. *ACM Transactions on Intelligent Systems and Technology*, **2**, 1-20. http://dx.doi.org/10.1145/2629481

[9] Raheja, J., Mishra, A. and Chaudhary, A. (2016) Indian Sign Language Recognition Using SVM. *Pattern Recognition and Image Analysis*, **26**, 434-441. http://dx.doi.org/10.1134/S1054661816020164

[10] Kausar, S. and Javed, M. (2011) A Survey on Sign Language Recognition. *Frontiers of In-*

formation Technology, 95-98. http://dx.doi.org/10.1109/fit.2011.25

[11] Trigeiros, P., Ribeiro, F. and Paulo, L. (2014) Vision-Based Portuguese Sign Language Recognition System. *Advances in Intelligent Systems and Computing*, **275**, 605-617. http://dx.doi.org/10.1007/978-3-319-05951-8_57

[12] Solís, F., Toxqui, C. and Martínez, D. (2015) Mexican Sign Language Recognition Using Jacobi-Fourier Moments. *Engineering*, **7**, 700-705. http://dx.doi.org/10.4236/eng.2015.710061

[13] Hu, M. (1962) Visual Pattern Recognition by Moment Invariants. *IRE Transactions on Information Theory*, **49**, 179-187.