Scientific Research

Static Digits Recognition Using Rotational Signatures and Hu Moments with a Multilayer Perceptron*

Francisco Solís[1], Margarita Hernández[1], Amelia Pérez[1], Carina Toxqui[2]

[1]Autonomous University of Mexico State, Toluca, Mexico
[2]Polytechnic University of Tulancingo, Tulancingo, Mexico
Email: jfsolisv@uaemex.mx

Abstract

This paper presents two systems for recognizing static signs (digits) from American Sign Language (ASL). These systems avoid the use color marks, or gloves, using instead, low-pass and high-pass filters in space and frequency domains, and color space transformations. First system used rotational signatures based on a correlation operator; minimum distance was used for the classification task. Second system computed the seven Hu invariants from binary images; these descriptors fed to a Multi-Layer Perceptron (MLP) in order to recognize the 9 different classes. First system achieves 100% of recognition rate with leaving-one-out validation and second experiment performs 96.7% of recognition rate with Hu moments and 100% using 36 normalized moments and k-fold cross validation.

Keywords

Sign Language Recognition, Rotational Signatures, Hu Moments, Multi-Layer Perceptron

1. Introduction

Sign Language is the basic way for communication of deaf community, and consists of a set of signs represented by different static hand shapes (dynamic in some cases) [1]. Sign Language is not universal; there are several alphabets such as American Sign Language (ASL), Taiwan Sign Language (TSL), Mexican Sign Language

---

*This paper presents two experiments to recognize 9 digits (from 1 to 9) of ASL, each one uses its own database, first system applies a rotational-correlation operator, the second one implements normalized geometric moments and Hu moments. Both methods achieve 100% of recognition rate.

(MSL), German Sign Language (GSL), Persian Sign Language (PSL) and so on. Sign Languages change due to different customs and cultures that characterize each region or country.

The whole set of movements that people make in order to express some idea in Sign Language, generally includes facial gestures, corporal postures, arms and hand expressions, among others [2]; this is why Sign Language Recognition is too complex. This leads most of the works to limit the problem with some predefined conditions.

Several papers focus on static hand gestures which contain mainly information about static alphabet signs of a certain Sign Language [3]-[5], reporting from 92% to 98% of recognition rate approximately. Working only with this data, the dimension of the problem is reduced; nevertheless, due to the freedom degrees of human hand, recognizing static signs of any language is a complex task.

In Sign Language, a lot of research work has been done over last decades. In general terms, Sign Language Recognition could be classified in two main categories: those using computer vision systems and those using electronic sensors attached to the fingers, hands or articulations [6]. First ones allow expressing signs in a more natural way because signer doesn't need to be physically connected to the system by wires, batteries or some other hardware, nevertheless with this systems, signs are much complex to analyze because it's hard to get accurate data of fingers, like position, orientation and so on, from 2D images. In the other hand, electronic systems that use motion sensors or similar hardware can generate precise data of hand gesture with or without CCD's or digital cameras, but these systems do not allow signers to express themselves freely, because hands and fingers are connected to some hardware.

2. Databases

In order to see some visual dynamics of rotational-correlation signatures of static signs, for this work was selected a set of 9 digit signs of American Sign Language (ASL). Five signs (1 - 5) face the dorsum to the digital camera, the remaining four (6 - 9) face the palm to the sensor.

These datasets don't use electronic hardware attached to the hand or fingers, neither special color markers nor gloves, in order to analyze a more natural way of the signs. Signs were captured by EOS 1100D digital Canon camera with CMOS sensor.

Figure 1 shows a complete version of the 9 static signs, this dataset was named "Database01" and was used to analyze the rotational-correlation signatures, in order to see if this technique can represent the extended fingers. Database01 has 5 versions per sign or digit; files have 2848 width with 4272 height pixels of image size, each pixel has 24 bits of depth. The images of this database were stored in the RGB space color.

Figure 2 shows pictures from dataset named "Database02", which has 3 versions per sign. The images displayed were cropped from the original ones, in order to get a better look of the gestures.

Database02 is used to perform the 2nd experiment by using Hu moments, original images were of 1280 height with 720 width pixels; color is represented by RGB channels.
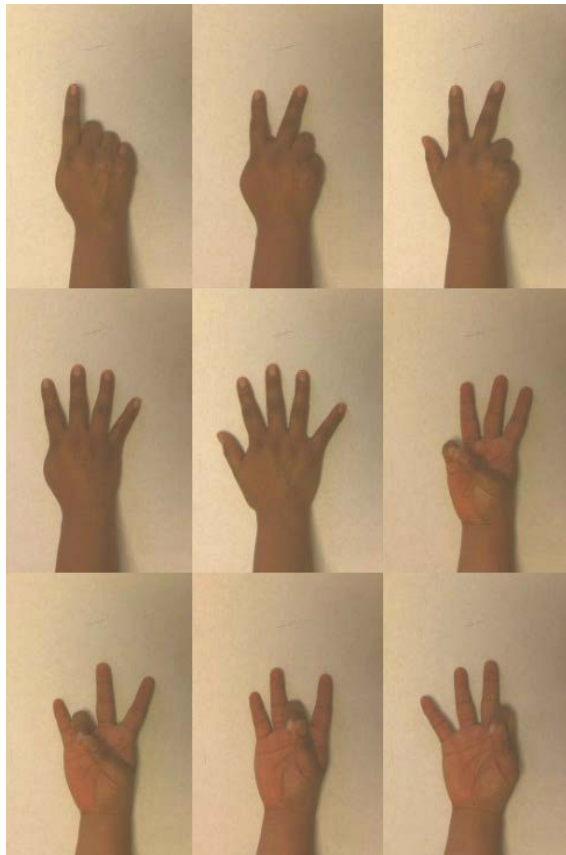
3. Rotational-Correlation Operator with Minimum Distance

Correlation can be expressed as

$$g(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x+i, y+j) h(i, j) \mathrm{d}i\mathrm{d}j, \tag{1}$$

this operator can be used to generate a rotational-correlation signature [7] when *h* is a rotated version of *f* which contains a binary object. Function *g* could be represented by a single scalar via the maximum correlation value, if *f* rotates 360 times with $\theta = 1, 2, 3, \cdots, 360°$ then *f* could be characterized by a set of 360 maximum correlation values, normalizing these data it's obtained a rotational-correlation signature.

As shown in **Figure 3**, binary objects can be described by rotational signatures. This technique has some valuable interesting properties for this particular problem. Rotational-correlation signatures are scale and rotation invariants besides don't require a continuous border along the hand shape perimeter.

The scheme of first system is shown in **Figure 4** where it can be watched original image of digit 3 version 1 at the beginning, then original RGB image is transformed into HSI space color and Saturation channel was experimentally selected to achieve the best segmented image, this last image is binarized and finally the rotational-correlation signature is calculated with 360 values.

Nine static signs that represent digits from one to nine respectively, captured with digital camera avoiding the use of gloves or special color markers

Figure 1. "Database01".

Static signs from 1 to 9 respectively, above pictures were captured with a digital camera
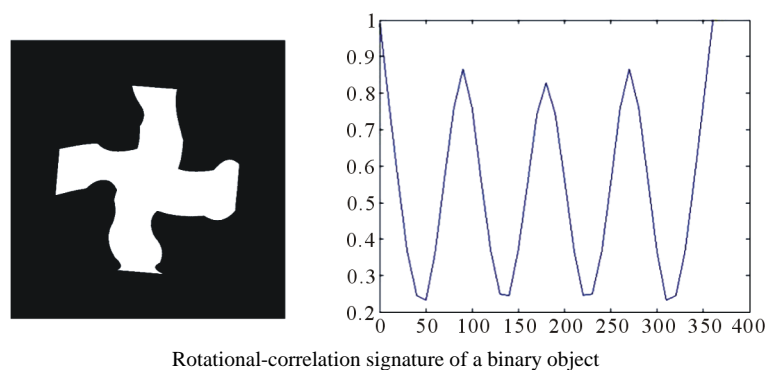
Figure 2. "Database02".

Rotational-correlation signature of a binary object

Figure 3. Rotational-correlation signature.

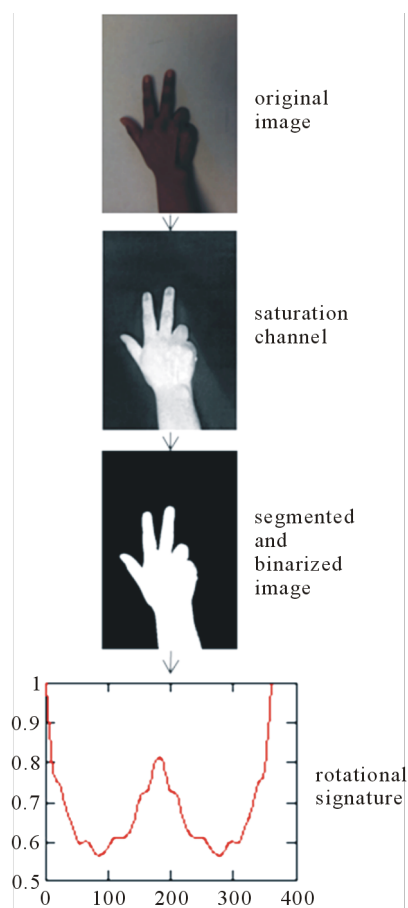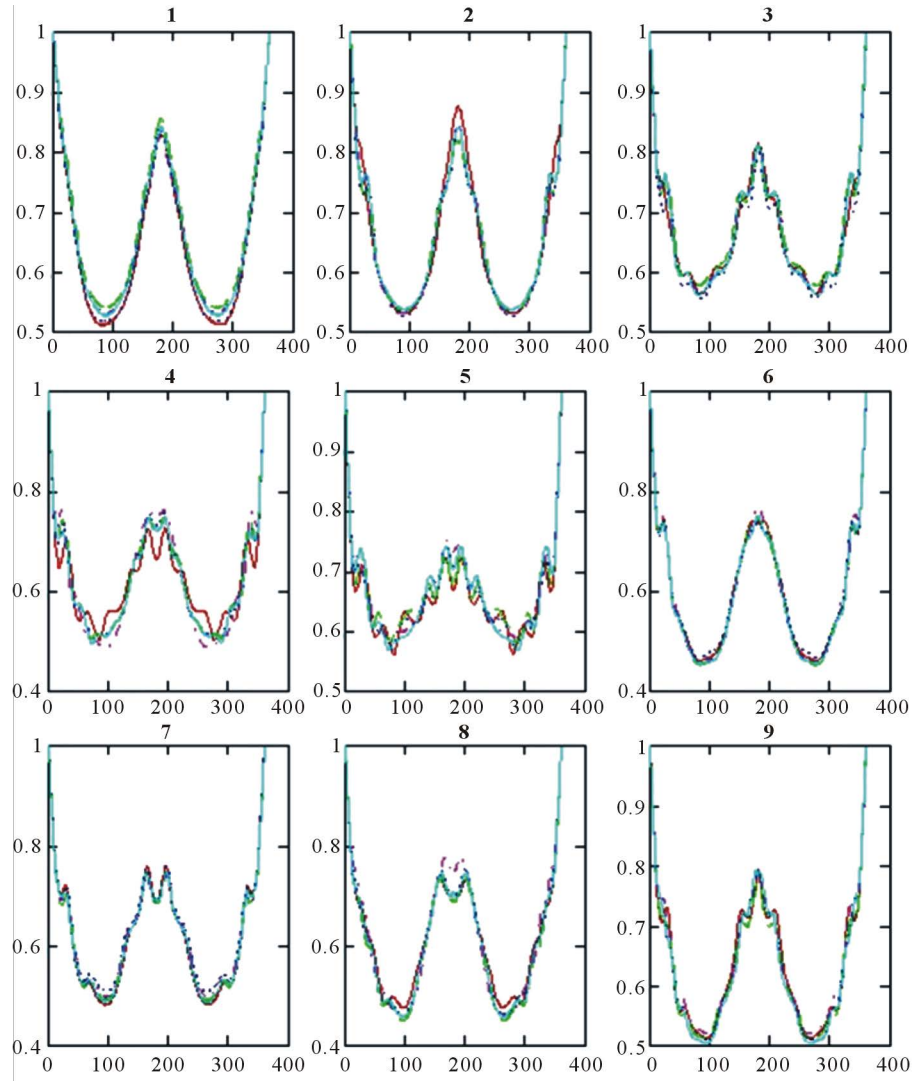Scheme of the first system using rotational signatures

Figure 4. First system framework.

Figure 5 shows nine graphs corresponding from digit 1 to 9 respectively, signs 1 and 2 are similar and sign 5 is the most dynamic one. Using this technique with minimum distance the recognition rate achieves 100% in a leave-one-out scheme.

4. Hu Moments with MLP

Invariant moments describe geometrically objects from digital images by values that tend to maintain constant despite changes in translation, rotation, scale and others. Low order moments give some geometrical information

695

Each graph represents five rotational-correlation signatures from one to nine respectively

Figure 5. Rotational-correlation signatures.

about the object, such as, area, mass center or mass distribution; if moments were normalized then they could be interpreted as statistical measures such as mean or variance. Geometric moments are expressed as follows:

$$m_{pq} = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} x^p y^q f(x, y)\, \mathrm{d}x \mathrm{d}y. \tag{2}$$

Central moments are defined as

$$\mu_{pq} = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} (x - x_c)^p (y - y_c)^q f(x, y)\, \mathrm{d}x \mathrm{d}y$$

where

$$x_c = \frac{m_{10}}{m_{00}}, \quad y_c = \frac{m_{01}}{m_{00}}$$

which are invariant to translation. Normalized moments have a scale factor, they can be expressed by central moments as
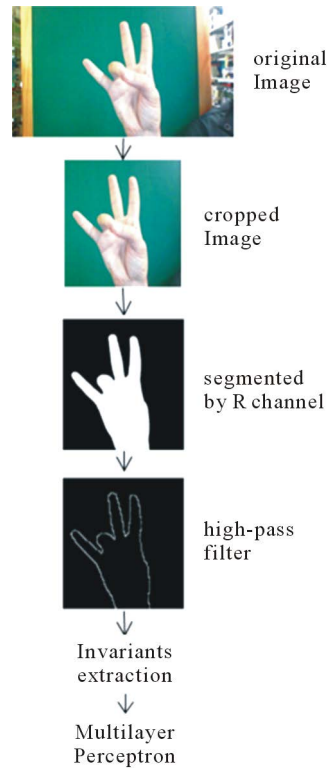
$$v_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\omega}}, \tag{3}$$

where

$$\omega = \frac{p+q}{2} + 1$$

Hu derived his seven invariant moments to rotation [8] and they are given as

$$\phi_1 = m_{20} + m_{02}$$

$$\phi_2 = (m_{20} - m_{02})^2 + 4m_{11}^2$$

$$\phi_3 = (m_{30} - 3m_{12})^2 + (3m_{21} - m_{03})^2$$

$$\phi_4 = (m_{30} + m_{12})^2 + (m_{21} + m_{03})^2$$

$$\phi_5 = (m_{30} - 3m_{12})(m_{30} + m_{12})\left((m_{30} + m_{12})^2 - 3(m_{12} + m_{03})^2\right)$$

$$+ (3m_{21} - m_{03})(m_{21} + m_{03})\left(3(m_{30} + m_{12})^2 - (m_{21} - m_{03})^2\right)$$

$$\phi_6 = (m_{20} - m_{02})\left((m_{30} + m_{12})^2 - (m_{21} + m_{03})^2\right) + 4m_{11}(m_{30} + m_{12})(m_{21} + m_{03})$$

$$\phi_7 = (3m_{21} - m_{03})(m_{30} + m_{12})\left((m_{30} + m_{12})^2 - 3(m_{12} + m_{03})^2\right)$$

$$- (m_{30} - 3m_{12})(m_{21} + m_{03})\left(3(m_{30} + m_{12})^2 - (m_{21} + m_{03})^2\right)$$

Figure 6 shows the scheme of second system starting with the sign 7 version 1 original image that is cropped

original Image

cropped Image

segmented by R channel

high-pass filter

Invariants extraction

Multilayer Perceptron

Scheme of the second system using invariant moments and artificial neural networks

Figure 6. Second system framework.

and segmented in the Red channel (experimentally selected from RGB and HSI space colors), after been binarized, the sign is enhanced by Laplacian high-pass filter, then 7 Hu moments and 36 normalized moments were extracted, finally a Multilayer Perceptron performs classification in two tests, first with 7 Hu moments where system achieves 96.29% using k-fold crossvalidation supported by Weka Pattern Recognition software, second test achieving 100% of recognition rate with a Multilayer Perceptron and k-fold cross validation using 49 normalized invariants (see Equation (3)) by combining the values of p, $q = 0$, 1, $\cdots$, 6.

5. Conclusions

Two systems were presented for recognition of static signs (digits from one to nine) using two datasets (one per system). First system tested "Database01" with rotational-correlation signatures; this technique is a binary object descriptor and can be used for recognize some static signs, nevertheless some signatures of different classes look similar although with minimum distance the similar classes can be separated numerically achieving indeed 100% of recognition rate.

Second system used "Database02" with invariant moments' extraction, 7 Hu moments and 49 normalized moments to achieve a performance of 96.29% and 100% of recognition rate.

One of the most challenging tasks in Sign Language Recognition is the segmentation phase that enhances the hand shape; that's why many authors use gloves or color markers; some other works like this one use a special solid background color and there are other strong limitations about this field. For better segmentation results each database used its own preprocessing: first system changed the color space from RGB to HSI and used Saturation channel to get experimentally a solid hand shape while second system segmented the hand shape in the RGB space color by using the Red channel.

Although the two systems achieve 100% of recognition rate, invariant moments use fewer descriptors (7 or 49) than rotational-correlation signatures (360 in this case); therefore invariant moments allow cheaper computational costs.

Acknowledgements

References

[1] Oz, C. and Leu, M. (2011) American Sign Language Word Recognition with a Sensory Glove Using Artificial Neural Networks. *Engineering Applications of Artificial Intelligence*, **24**, 1204-1213. http://dx.doi.org/10.1016/j.engappai.2011.06.015

[2] Elons, S., Abull, M. and Tolba, F. (2013) A Proposed PCNN Features Quality Optimization Technique for Pose-Invariant 3D Arabic Sign Language Recognition. *Applied Soft Computing*, **13**, 1646-1660. http://dx.doi.org/10.1016/j.asoc.2012.11.036

[3] Karami, A., Zanj, B. and Kiani, A. (2011) Persian Sign Language (PSL) Recognition Using Wavelet Transform and Neural Networks. *Expert Systems with Applications*, **38**, 2661-2667. http://dx.doi.org/10.1016/j.eswa.2010.08.056

[4] Rajam, S. and Balakrishnan, G. (2012) Recognition of Tamil Sign Language Alphabet Using Image Processing to Aid Deaf-Dumb People. *Procedia Engineering*, **30**, 681-686. http://dx.doi.org/10.1016/j.proeng.2012.01.938

[5] Munib, Q., Habeeb, M., Takruri, B. and Al-Malik, A. (2007) American Sign Language (ASL) Recognition Based on Hough Transform and Neural Networks. *Expert Systems with Applications*, **32**, 24-37. http://dx.doi.org/10.1016/j.eswa.2005.11.018

[6] Al-Rousan, M., Assaleh, K. and Tala'a, A. (2009) Video-Based Signer-Independent Arabic Sign Language Recognition Using Hidden Markov Models. *Applied Soft Computing*, **9**, 990-999. http://dx.doi.org/10.1016/j.asoc.2009.01.002

[7] Urcid, G., Padilla-Vivanco, A., Cornejo, A. and Baez, J. (2004) Correlation-Based Rotational Operator Signature of Planar Binary Objects. *Proceedings of SPIE*5558, *Applications of Digital Image Processing XXVII*, 2 November 2004, 87. http://dx.doi.org/10.1117/12.558744

[8] Hu, M. (1962) Visual Pattern Recognition by Moment Invariants. *IRE Transactions on Information Theory*, **IT-08**, 179-187.

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or Online Submission Portal.