Hand Gesture Recognition based on Morphologic Features*

Francisco Flórez Revuelta, Juan Manuel García Chamizo, Francisco Ibarra Picó

Universidad de Alicante
Departamento de Tecnología Informática y Computación
Campus de S. Vicente del Raspeig
Ap. de Correos 99, E-03080, Alicante.
{florez,juanma,ibarra}@dtic.ua.es

Abstract

In this paper, we try to characterize the hand posture by means of the morphologic coefficient, measure that allows to determine its morphology, in order to later perform its classification by means of an artificial neuronal network model, the Growing Cell Structure. The monitoring of the neurons that are activated for the successive postures that the hand takes throughout the time allows us to determine the gesture that is being made. Finally, we present the results of experiments with a vocabulary of 7 gestures, in order to verify the efficiency of this method of classification and extraction of characteristics.

Keywords: hand gesture recognition, morphologic coefficient, growing neural networks, self-organizing networks.

1 Introduction

Sign language consists of both static and dynamic hand gestures. The former are characterized by the hand posture determined by a particular finger-thumb-palm configuration, while the latter are given by the spatial and morphological changes of the hand throughout the time.

Special glove-based devices [1] have been developed to extract hand posture, some of which also allow to recover hand position and orientation. These systems are annoying to

* Work partially funded by *CICYT, TAP98-0333-C03-03*.

use, reason why different vision-based techniques [2] have appeared to favour human-computer interaction.

In this paper we present a vision-based system to solve the problem of hand posture classification and the morphological changes of the hand throughout the time.

2 Preprocessing

To avoid the problem of segmentation of the hand moving in front of a complex background, in some systems the gestures are made dressing a coloured glove that allows to separate the hands from the rest of the scene [3]. In this work, users dress dark clothes in front of a dark background, method already used in the literature [4]. This way, by means of a threshold filter the hand can be easily segmented.

Later, we normalize the resulting image by locating its centroid and the point of the hand more distant to it. This point will correspond, in most of the cases, with the tip of a finger. Then, we proceed to the translation, scaling and rotation of the image (Figure 1).

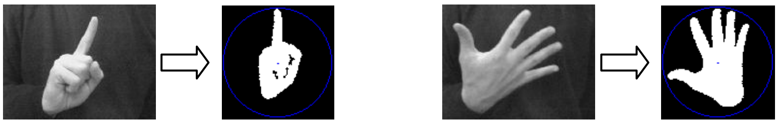

Figure 1. Application of preprocessing to several images.

3. Feature extraction

3.1. Morphologic coefficient

The Hausdorff Dimension (HD) has been used, mainly, in fractal studies [5]. One of the most attractive features of this measure when analysing images is its invariant properties under isometric transformations. However, one of the HD main problems is its difficult computation, and that is why, in general, aproximative box-counting methods are used.

Definition 1. A packing or semicover of a set S is a collection of sets $SC(S)=\{A_i/i=1..n\}$ verifying that $A_i \cap A_j=\emptyset\ \forall i\neq j$ and $\cup A_i \subset S$.

Definition 2. We call δ-semicover of a set S (δ-SC(S)) to a semicover of S formed by a finite collection of sets $\{A_i\}$ having a diameter of δ or less.

Definition 3. We call the Morphologic Coefficient of the semicover of a set S to

$$MC(S) = \lim_{\delta\to\infty} \frac{\log|\delta - SC(S)|}{-\log\delta} \tag{1}$$

or its discrete version, for some discrete values of δ (1,2,...,D):

$$MC(S) \approx \frac{1}{D}\sum_{i=1}^{D}\frac{\log|\delta_i - SC(S)|}{\log\delta_i} \tag{2}$$

The number of pieces that can be packed into a set by a δ-semicover depends basically on the morphology of the pieces and the set. So, from the calculation of several semicovers we can obtain a good estimation of the morphology of an object. In [6] the equivalence between the Hausdorff Dimension and the continuous Morphologic Coefficient of the semicover is demonstrated.

3.2. Feature extraction from the image

It has been observed that the greatest discrimination between different postures takes place in the remotest points to the centroid, reason why we have thought to use a system that allow to extract morphologic characteristics of the image depending on the position of the points into the hand. For it, the morphologic coefficient of different regions (rings) of the image has been calculated (Figure 2).

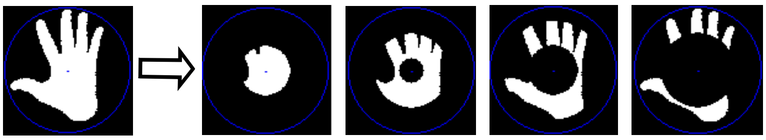

Figure 2. Subregions of the image around the centroid.

We build the feature vector with the results of the different semicovers, priors to the calculation of the Morphologic Coefficient. Then, the feature vector is

$$(SC^{2}_{0,20}, SC^{4}_{0,20}, \ldots, SC^{18}_{0,20}, SC^{2}_{10,30}, \ldots, SC^{18}_{0,20}, SC^{2}_{20,40}, \ldots, SC^{18}_{20,40}, SC^{2}_{30,50}, \ldots, SC^{18}_{30,50}) \tag{3}$$

where $SC^{t}_{r1,r2}$ represents the t-semicover (number of pieces of size t x t that fit) of the portion of the image (normalized hand) whose points are into a distance $r / r_1 \le r \le r_2$.

4. Classification of hand postures

In other works, there have been used different methods to classify hand postures: statistical classifiers [4,7], template matching [7,8] or neural networks [4]. In our case a neuronal network model called Supervised Growing Cell Structures has been used [9].

The Growing Cell Structures is a self-organizational network that performs non-supervised learning. This network does not have a predetermined structure but from an initial number of neurons allows to add or to remove neurons and/or synaptic connections in those zones of the input space in which it is necessary.

As a result of this learning, and due to the addition and removal of neural cells, the neural map has not got a predetermined topology and, moreover, it can be divided into several submaps, feature that makes possible, in an unsupervised way, to partition the data in clusters of mutually similar items.

5. Gesture recognition

5.1. Gestures as sequences of activated neurons

Tracking of the different hand postures that a gesture takes, is going to be made following the sequence of winning neurons that are successively activated, so that if n_i is the winning neuron for the posture i of the gesture, the gesture will be $g=\{n_1,n_2,..,n_k\}$.

To simplify this sequence, if two consecutive images activate the same neuron (the hand posture is not significantly modified) it will be considered once. This way, the gesture $g=\{n_1,n_2,..,n_{k'}\}$ where k' represents the different winning neurons.

5.2. Establishing a distance measure between gestures

To classify the gestures, we calculate the distance between symbol (neuron) chains of different lengths. One of the most usual distance is the Weighted Levenshtein Distance (WLD), that indicates the minimum number of replacements r(i), insertions i(i) and deletions d(i) of symbols that have to be made in a chain to be converted in other one:

$$WLD(g_1, g_2)=\min\{p_1 \cdot r(i) + p_2 \cdot i(i) + p_3 \cdot d(i)\} \quad (4)$$

where p_1, p_2 and p_3 are the probabilities for a symbol to be replaced, inserted or removed.

This distance has been modified to introduce the information we have about the distance between symbols (neurons) instead of the different probabilities. Then,

$$WLD'(g_1, g_2)=\min\{d_1 \cdot r(i) + d_2 \cdot i(i) + d_3 \cdot d(i)\} \quad (5)$$

where d_1, d_2 and d_3 are the distances in the pattern space between the neurons that are replaced, inserted or removed.

6. Experiments and results

In order to perform the experiments we designed a vocabulary of 7 gestures, some of them of similar characteristics (Figure 3**).** At the time of the training of the system, a user made each one of the gestures of the vocabulary 5 times. These gestures were taken at the rate of 24 frames per second, with a 320x200 resolution and 256 grey levels.

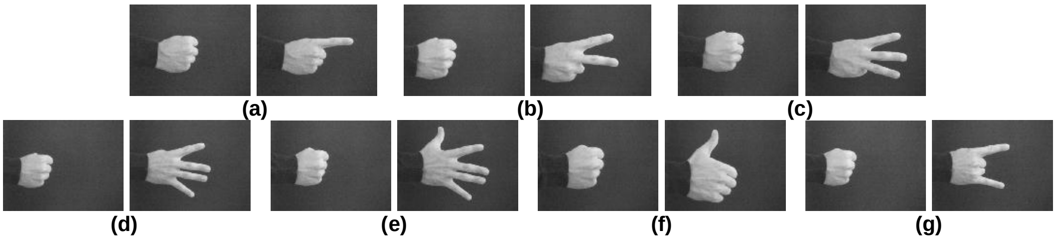

Figure 3. Initial and final frames of the seven gestures of the vocabulary: (a) one, (b) two, (c) three, (d) four, (e) open_hand, (f) up and (g) horns.

Later, each one of these sequences was preprocessed to obtain the different binary images from the hand. Their centroid was calculated and their size was normalized to a circle of radius 50 pixels. We built the feature vectors of the images of the hand of all the introduced postures (a total of 1500 vectors) and performed the training of the Growing Cell Structure setting the different parameters with a final number of neurons equal to 250.

After the neural network is trained we recognize each one of the introduced gestures, looking at the sequence of neurons that are activated, storing the sequence to make later comparisons of new gestures by means of the modified Levenshtein Distance.

Once finished this process of learning and construction of the models base, we come to the recognition of different gestures made by the user. In Table 1 we can see the good results obtained in the recognition. There are only some problems in the most similar gestures (*Two* and *Horns*).

Gesture	Success rate
One	9/10
Two	8/10
Three	10/10
Four	10/10
Open_hand	10/10
Up	10/10
Horns	8/10

Table 1. Results of the hand gesture recognition.

7. Conclusions

In this work, we present a new approach to gesture recognition. We introduce the morphological coefficient as a good measurement to build a feature vector that allows a good representation of the morphology of the hand. We use the Growing Cell Structure as a classifier, basically because the particular behaviour of this self-organizing neural network makes possible the learning of all the postures in an unsupervised way.

We have also tested the system with very good results, because in almost all the cases the gesture is correctly recognised.

References

[1] D. J. Sturman. Whole-hand Input. Ph.D. Thesis, Media Arts and Sciences Section, School of Architecture and Planning, Massachusetts Institute of Technology, 1992.

[2] M. Kohler. A Survey of Video-based Gesture Recognition -Stereo and Mono Systems-. Technical Report 693, Fachbereich Informatik, Universität Dortmund, August 1998.

[3] T. E. Starner. Visual Recognition of American Sign Language Using Hidden Markov Models. M.Sc. Thesis, Media Arts and Sciences Section, School of Architecture and Planning, Massachusetts Institute of Technology, 1995.

[4] D. S. Banarse. Hand Posture Recognition with the Neocognitron Network. M.Sc. Thesis, University College of North Wales, Bangor, December 1993.

[5] S. G. Hoggar. *Mathematics for Computer Science*. Cambridge University Press, 1993.

[6] F. Ibarra Picó. Memorias asociativas ortogonales y sus aplicaciones en análisis de textura. Institut de Cultura Juan Gil-Albert. ISBN: 84-7784-258-2, 1997.

[7] R. Watson. A Survey of Gesture Recognition Techniques. Technical Report TCD-CS-93-11, Department of Computer Science, Trinity College, Dublin, 1993.

[8] W. T. Freeman and C. D. Weissman. Television control by hand gestures. IEEE International Conference on Automatic Face and Gesture Recognition, Zurich, 1995.

[9] B. Fritzke. Growing cell structures – a self-organizing network for unsupervised and supervised learning. *Neural Networks*, 7(9):1441-1460, 1994.