An Analysis of Features for Hand-Gesture Classification

Thiago R. Trigo and Sergio Roberto M. Pellegrino
Electronic Engineering and Computer Department
Aeronautics Technological Institute (ITA)
São José dos Campos, Brazil
{trigo, pell}@ita.br

Abstract— the human-computer interaction, also known as HCI, depends mostly on physical devices. The goal of this work is the evaluation and analysis of methods which allows the user to interact to machines using a hand gesture based natural language. Here we present some approaches which are used in HCI systems based on hand gesture and a new proposal that uses geometric shape descriptors for hand gesture classification. The results analysis shows that this new proposal beats some limitations of other known HCI methods.

Keywords-hand-gesture classification;shape descriptors;k-curvature, invariant moments.

I. Introduction

The human-computer interaction, also known as HCI, depends mostly on physical devices. Normally, people use mouses and keyboards, but this kind of interaction can be unnatural for humans, who are used to express their ideas, their feelings and their wishes through voice, corporal gestures, facial expression, hands gestures, and so on. Therefore, an interface which allows one of those kinds of interaction between human and machines would be more natural and instinctive for users.

If we consider interactions concerning to object manipulation, then the interaction based on hand gestures seems to be an attractive solution. In the most popular HCI systems based on hand gesture, the user needs to wear special gloves which measure the hand pose and the joint angles [1]. The problem of this kind of technique is that once the user has to wear a glove, the system becomes invasive, besides the fact of special gloves being expensive. Due to this, a hand-gestures interface based on computer vision appears as a reasonable option to reach a more natural human-computer communication.

Since the hands are capable to produce great number of gestures (thanks to its flexibility and fingers), we defined a gesture alphabet containing a reduced number of gestures to be recognized. The Figure 1 shows the defined alphabet.

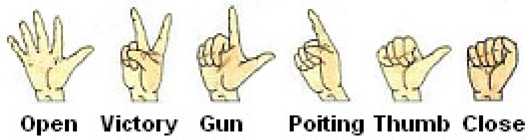

Figure 1. Gesture alphabet

This paper presents a new technique to classify hand gestures which could be used in HCI systems. Next section shows some usual approaches for HCI systems based on hand gestures. In section three, we present some limitations of these approaches and we propose a new one. In section four we conclude this paper by analyzing the produced results and appointing to the future directions.

II. Related Works

The literature shows diverse ways to use hand gestures for HCI and many of them extract some features from the captured gesture to classify it. For example, we can see in [2] [3] the use Hu invariant moments, other approaches studies the use of k-curvature algorithm [4], [5], [6], moreover [7] and [8] use template matching in their solutions. All of these approaches were made to work in specific situations, and they have limitations, either in number of possible gestures or in robustness due to variations, like scale variations. Next subsections make a short introduction of main approaches.

A. Hu Invariant Moments

The most commom use of moments is in statistics and mechanics; they characterize the distribution of random variables and bodies by their spatial distribution of mass. Considering binary images as two-dimensional density distribution functions, then moments can be used to extract some properties that have analogies in statistics and mechanics [9]. For a binary image, a moment, m_{pq}, of order p + q, is defined as:

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

From equation 1, Hu derived a set of seven values which are invariant to translation, scale and rotation [10]. These seven values are used as feature vector for each gesture. In some cases, the classification is given by the best similarity betwen distance metrics from new gestures and reference vectors [12], in others a more robust classifier is used, like Artificial Neural Network (ANN) [2] or Supported Vector Machine (SVM) [3].

B. *K-curvature*

From the binary image, a boundary region can be extracted, this boundary region is represented by a list of boundary points P(i) = (x(i), y(i)) and the k-curvature is measured at each point. The k-curvature is an angle $\alpha(i)$ between the vectors [P(i-k), P(i)] and [P(i), P(i+k)], where k is a constant. Figure 2 shows how the algorithm finds the features. The main idea is that points P belonging to fingertips will have k-curvature angles inside a specific interval. With this kind of approach is possible to find fingertips, and to know how many fingers there are in the gesture.

C. *Template Matching*

This approach try to find small parts in image which match a template image, in other words, it try to find some specific pattern in an image. Figure 3 shows an example of templates and an image where the templates are trying to be found.

III. Proposed Approach

The approaches described previously have some limitations. Hu invariant moments are sensitive to morphological deformations and it is inevitable produce hand gestures with variations in its morphology, moreover, hands of different people have different morphologies and it could affect directly the invariant moments performance.

The k-curvature has two easily identifiable problems. First, the definition of constant k, a selected value that could work well in some cases, but it could fail in cases of scale variances. Second, the algorithm can indicate how many fingers there are in the produced gesture, but the method can produce false positives answers in the cases where different gestures has the similar number of curves (fingers) , e.g. gesture Victory and Gun in Figure 1

The template matching approach is invariant to rotation and translation, but it will misclassify if scale variations are introduced. Based on these observations, this article proposes the uses of different features to classify the hand gestures.

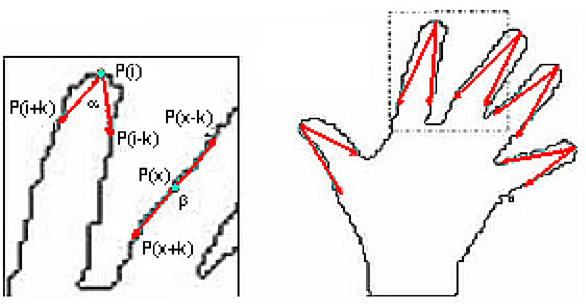

Figure 2. The K-curvature idea [5]

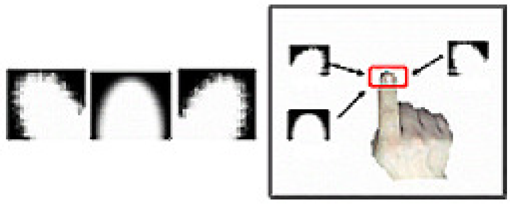

Figure 3. Template matching example [7]

A. *Shape Descriptors and Others Features*

Some geometric shapes descriptors were experimented, they are:

- Aspect ratio: it is a measure of the elongation of a boundary profile [11]. It is defined as:

$$\text{aspect ratio} = (\text{major axis})/(\text{minor axis}) \quad (2)$$

- Circularity: it indicates the object similarity to a circumference [12], it is defined as:

$$\text{circularity} = (4*\pi*\text{area})/(\text{perimeter})^2 \quad (3)$$

- Spreadness: it indicates how the object is spread [12], it is defined as:

$$\text{spreadness} = (\mu_{20} + \mu_{02})/(\mu_{00}*\mu_{00}) \quad (4)$$

Where μ_{ij} represents the central moments, for more details about central moments see [10].

- Roundness: it is sensitive to the elongation of a boundary profile. The roundness is equal to 1 for a circle and is less for any other shape [12], it is defined as:

$$\text{roundness} = (4*\text{area})/((\pi*(\text{major axis})^2) \quad (5)$$

- Solidity: it describes the ruggedness of a boundary profile. The solidity is also equal to 1 for a region that has no concavities in its boundary and is less for particles with indentations, it is given by:

$$\text{solidity} = (\text{area})/((\text{convex area}) \quad (6)$$

- Number of fingers: this feature indicates the number of fingers found in image, it´s obtained from k-curvature algorithm.

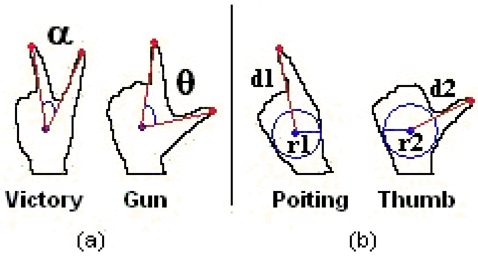

Figure 4. New features. (a) Angle between two fingers. (b) Relation "distance-radius"

is calculated when the number of fingers is equal to two. It is an angle calculated between the vectors formed by fingertips and the center of the hand. In Figure 4 (a) is possible to notice that the angle α of the Victory gesture is different from the angle θ of Gun gesture.

- Relation "distance-radius": when the number of fingers is equal to one, this feature calculated. This feature is a division of the distance from fingertip to the center of the hand by the radius. Figure 4 (b) shows the distance from fingertip to the center of the hand, and the radius in Pointing and Thumb gesture.

IV. Classification and Results

The classification was conducted with the standard multilayer perceptron artificial neural network from Weka[1]. It is well-known by its information classification capacity, basically it is fed with some known patterns and, after a training step, it will be able to classify the patterns using a set of features. In this work, the pattern used to feed the network is the vector of features from hand gestures. To do the training step was used a database with 1051 instances distributed as shown in Table 1.

TABLE 1. Database distribution

Gesture	Number of Instances
Open	200
Victory	162
Gun	208
Pointing	184
Thumb	165
Close	132

All the images of the database were colleted with a webcam and they were segmented manually. For each gesture, the collected samples vary in translation, rotation and scale, as shown in Figure 5.

A. *Experiments and Results*

Seven experiments were defined, they differ in their features vectors. We defined three groups of features, and in each experiment the feature vector contains one or more groups of features. The Table 2 shows these experiments.

TABLE 2. Experiments

Experiment Number	Feature Group		
	Invariant Moments	*K-curvature*	*Geometric Shape Descriptors*
1	X		
2		X	
3			X
4	X	X	
5	X		X
6		X	X
7	X	X	X

[1] Weka is an open source software which contains a collection of machine learning algorithms, it can be downloaded at http://www.cs.waikato.ac.nz/ml/weka/

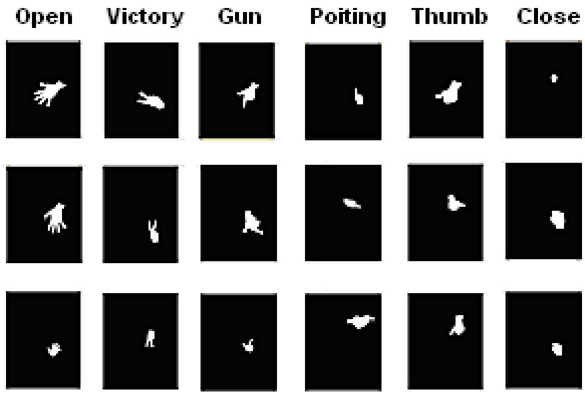

Figure 5. Gesture Samples

Invariant moments group contains the seven moments described in [10]; K-curvature group contains the features: number of fingers, angle between two fingers, and relation "distance-radius"; and Geometric shape descriptors group contains aspect ratio, circularity, spreadness, roundness and solidity.

The database was divided in two parts, 75% of it was used in the training step, and the rest of it was used to test the performance of the generated model. Table 3 shows the results obtained in the test performance of each experiment.

TABLE 3. Experiment Results

Experiment Number	% Correctly Classified Instances
1	35.3612
2	80.9886
3	98.8593
4	84.7909
5	99.2395
6	98.8593
7	98.8593

From results shown in Table 3 we can notice that the experiments with results above 90% - 3, 5, 6 and 7 - have the geometric shape descriptors group in their feature vector. The best classification result is produced by experiment 5 which uses invariant moment group and geometric shape descriptor group. But experiment 3, which produce the second best result, only uses the geometric shape descriptor group, if we consider that the processing time increases with the number of feature extracted, then the best performance would be from that which uses the smallest number of features in its feature vector. Thus, the experiment nº 3 presents the best performance.

The database used samples of the gestures with variation in scale, translation and rotation and these variation probably would produce a great number of misclassification in the described other approaches, but the ANN classifier using only geometric shape descriptor was capable to have a good performance in terms of classification time and accuracy.

V. CONCLUSIONS

This paper analized the use o invariant moments, k-curvature features and template matching to classify hand gestures for HCI. Isolatedly, each of these features presents limitations, to solve them this paper proposes the use of shape descriptors. The conducted experiments combined these features in groups and then used them to feed a multilayer perceptron. The results show that the group composed only by the geometric descriptors had the best performance considering time and accuracy.

The next step of this research is to do an attribute selection, in other words, we intend to analyze features individually and try to compound new groups, for example, a group that contains the first invariant moment and the aspect ratio. Other future step is to apply those feature vectors to different classifiers and compare their results with the multilayer perceptron result.

REFERENCES

[1] K. G. Derpanis, “A Review of Vision-Based Hand Gesture,” Department of Computer Science, York University, February, 2004.

[2] Premaratne, P.; Nguyen, Q., "Consumer electronics control system based on hand gesture moment invariants," Computer Vision, IET , vol.1, no.1, pp.35-41, March 2007.

[3] Yun Liu; Zhijie Gan; Yu Sun, "Static Hand Gesture Recognition and its Application based on Support Vector Machines," Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2008. SNPD '08. Ninth ACIS International Conference on , vol., no., pp.517-521, 6-8 Aug. 2008.

[4] Segen, J.; Kumar, S., "Human-computer interaction using gesture recognition and 3D hand tracking," Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on , vol., no., pp.188-192 vol.3, 4-7 Oct 1998.

[5] Truyenque, M. A. Q., "Uma Aplicação de Visão Computacional que Utiliza Gestos da Mão para Interagir com o Computador," Master´s thesis, Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, Rio de Janeiro, Brazil, March 2005.

[6] Taehee Lee; Hollerer, T., "Handy AR: Markerless Inspection of Augmented Reality Objects Using Fingertip Tracking," Wearable Computers, 2007 11th IEEE International Symposium on , vol., no., pp.83-90, 11-13 Oct. 2007.

[7] Jong-Min Kim; Woong-Ki Lee, "Hand Shape Recognition Using Fingertips," Fuzzy Systems and Knowledge Discovery, 2008. FSKD '08. Fifth International Conference on , vol.4, no., pp.44-48, 18-20 Oct. 2008

[8] Oka, K.; Sato, Y.; Koike, H., "Real-time fingertip tracking and gesture recognition," Computer Graphics and Applications, *IEEE* , vol.22, no.6, pp. 64-71, Nov/Dec 2002.

[9] R.J. Prokop and A.P. Reeves, "A Survey of Moment-Based Techniques for Unoccluded Object Representation and Recognition," CVGIP - Graphical Models and Image Processing, vol. 54, no. 5, pp. 438 - 460, 1992.

[10] Ming-Kuei Hu, "Visual pattern recognition by moment invariants," Information Theory, IRE Transactions on , vol.8, no.2, pp.179-187, February 1962.

[11] T. B. Kirk, D. Panzera, R. V. Anamalay, Z. L. Xu, Computer image analysis of wear debris for machine condition monitoring and fault diagnosis, Wear, Volumes 181-183, Part 2, 10th International Conference on Wear of Materials, March 1995, Pages 717-722, ISSN 0043-1648, DOI: 10.1016/0043-1648(95)90188-4.

[12] Ribeiro, H. L., "Reconhecimento de Gestos Usando Segmentação de Imagens Dinâmicas de Mãos Baseada no Modelo de Mistura de Gaussianas e Cor de Pele," Master´s thesis, Universidade de São Paulo, Escola de Engenharia de São Carlos, São Carlos, Brazil, 2006.