Open access • Proceedings Article • DOI:10.1109/ICPR.2002.1048206

Supervised training based hand gesture recognition system

Attila Licsár, Tamás Szirányi

Institutions: Hungarian Academy of Sciences

Published on: 11 Aug 2002 - International Conference on Pattern Recognition

Topics: Gesture recognition, Sketch recognition, Three-dimensional face recognition, Intelligent character recognition and Feature (machine learning)

Supervised training based hand gesture recognition system

Attila Licsár
University of Veszprém, Department of Image Processing and Neurocomputing,H-8200 Veszprém, Egyetem u. 10, Hungary
licsara@www.knt.vein.hu

Tamás Szirányi
Analogical & Neural Computing Laboratory, Computer & Automation Research Institute, Hungarian Academy of Sciences, H-1111 Budapest, Kende u. 13-17, Hungary
sziranyi@sztaki.hu

Abstract

We have developed a hand gesture recognition system, based on the shape analysis of static gestures, for Human Computer Interaction purposes. Our appearance-based recognition uses modified Fourier descriptors for the classification of hand shapes. As always found in literature, such recognition systems consist of two phases: training and recognition. In our new practical approach, following the chosen appearance-based model, training and recognition is done in an interactive supervised way: the adaptation for untrained gestures is also solved by hand signals. Our experimental results with three different users are reported. In this paper, besides describing the recognition itself, we demonstrate our interactive training method in a practical application.[*]

1. Introduction

In the information society the communication between the user and the computer has become a very active research area. In this paper we will demonstrate a practical method to improve the performance of our gesture recognition system. In the spatial gesture model there are two main approaches: 3D hand model-based and appearance-based methods. We have chosen appearance-based methods in our system because it is simple and efficient for two-dimensional desktop applications.

Appearance-based systems have two categories: motion-based and posture-based recognition. Posture-based recognition not only handles location of the hand but also recognizes the shape features of the hand. Motion-based systems use dynamic gestures, while posture-based systems use static hand gestures. Our system recognizes static hand postures because dynamic parameters of the hand e.g. position or movements are used for controlling purpose in the virtual environment.

There are numerous methods in appearance-based recognition for recognizing static gestures. The majority of methods use parameters derived from image. In such cases, the model parameters are derived from the description of the shape. Systems can use one or more camera pictures [1]. They include: edge-based contours [2], edges, image moments [3][4], image eigenvectors [5], or geometric moment description of hand shapes. Some other techniques use second order moments (like Zernike methods [6]), which are invariant to the rotation of the shape. Another method uses orientation histograms [7], which are invariant to lighting conditions and represent summarized information of small patch orientations over the whole image. Geometric moment description is not invariant to rotation and the invariance of other moment-based methods is restricted. We need a method for contour classification where the parameters are invariant to translation, rotation, and scaling. The disadvantage of invariant moments is its high computational cost because features are computed using the entire region. Boundary-based methods, such as Fourier descriptor [8] use only contour-points. However, the Fourier descriptor is sensitive to the starting point of the shape boundary.

2. Overview of our gesture recognition system

In our system we adopt a modified Fourier descriptor method [9] (MFD), which is already used in the field of character recognition. We apply this method for the gesture-based man-machine interface. The example-based system involves two phases: training and running. In the training phase, the user shows to the system one or more examples of hand gestures. The system stores the Fourier coefficients of the hand shape and in the running phase the computer compares the current hand shape with each of stored shapes by coefficients. The best match gesture is

[*] *Hand Recognition Demo can be downloaded from* http://www.knt.vein.hu/staff/licsara/

selected by the nearest-neighbor method with the distance metric of MFD. Our goal is to enable the system to correct faulty detected gestures and to modify and train hand gestures with the help of the user's feedback during the recognition phase. With this interaction, the recognition efficiency will be grown.

The organization of this paper is as follows. First, we describe features of MFD method, and then analyze the parameters of the gesture recognition system. Next, we will discuss the interactive learning method and examine the recognition rate of the system with and without interactive learning. Finally, we present our practical application with this interactive training method.

3. Gesture recognition by MFD

The first implementations of Fourier descriptor had some weaknesses, for example, invariance to starting point of the shape contour. As this traditional method has been extended to the modified Fourier descriptor, it is invariant to transition, reflection, scaling of shapes, as well as the starting point used in defining boundary sequence. The MFD is robust against noise when shape boundaries often contain local irregularities due to image noise. Another advantage of the MFD is that feature vectors should be computed efficiently. It has a new distance measure for describing and comparing closed curves. Some applications, like methods [10] with Fourier descriptor, use other classification algorithms (e.g. neural network), but this simple metric, based on features of Fourier coefficients, is quite fast and reliable: cc. 80% recognition rate for 7-8 gestures. For the calculation of the MFD we need a complex sequence from the x and y coordinates of the n^{th} boundary points. The MFD is defined as the Discrete Fourier Transform (DFT) of the previous complex sequence. The magnitude and phase of Fourier descriptor coefficients of the compared boundaries are related to each other. From the features of the Fourier transform and orientations of the major axes of the two shapes we can calculate two sequences [9]. It is easy to see that these sequences will be constant if the two compared gestures belong to the same class. The distance measure for magnitude and phase angle are defined as the standard deviation of the previous two calculated sequences. So, the classification method of MFD is fast because it computes a simple standard deviation.

The Fourier descriptors are calculated from the boundary of the palm. The system uses restrictive background to localize the hand efficiently. We avoided applying complex background because the projection of the film happens in a dark room for better vision. Since forearm features do not contain important information the perfect and consequent segmentation of arm and forearm is important. The problem of automatic segmentation has solved by other systems [5], which use the direction of the arm for automatic forearm segmentation. From the image moments [11], we can get the global direction and position of the hand. Since the method is invariant to orientation, position and size of the palm, we may use these parameters to control or manipulate virtual objects. Using these parameters the system can get feedback from the user about the recognition.

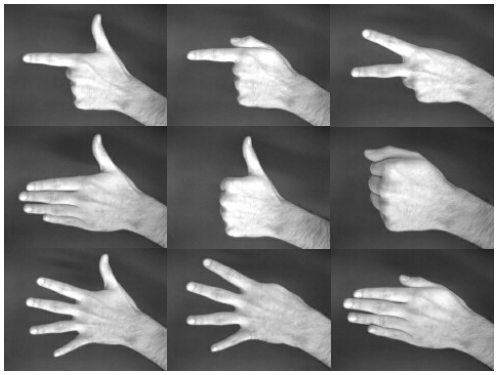

Figure 3-1: Some of our hand gestures used in our system

We have tested the system with 9 gesture classes (Figure 3-1). The starting set of training was very small, usually only one but the continuous supervised/unsupervised training adaptively changed the class-parameters. Testing process has been done by a set of 140 samples. The efficiency of the recognition methods has been tested to find the best feature-detection. Methods for selecting the most appropriate Fourier descriptors are as follows:

- (Method 1.) The first 3 coefficients are used in the comparison;
- (Method 2.) Coefficients which are greater than a threshold (here *11.0*) in case of at least one class;
- (Method 3.) Starting from the lowest spatial frequencies where each coefficient is greater than the threshold.

We demonstrate the above methods in Table 3-1. For each method the distances among the classes are counted. In case of Method 2, the distances are greater than in the other 2 cases. Starting with a one-class/one-sample training, a limited number of classes can be appropriately recognized.

Table 3-1: Comparing efficiency of selecting methods

Used method	Recognition rate	Average distance
Method 1.	97%	0.731
Method 2.	98%	1.008
Method 3.	97%	0.74

4. Interactive learning system

The conventional learning and recognition phases of the system will be extended by a new interactive training algorithm, which will improve the efficiency of the recognition methods. In the recognition phase the system can correct faulty detected gestures and interactively can modify and teach hand gestures with the user feedback. In this additional phase the user can modify the recognition strategy by modifying his gestures' clustering among classes. The user's feedback signal (crucial shapes to be trained) is a rapid gesture moving or shaking because under a normal interaction the user doesn't apply rapid moving. After the rapid moving or shaking, the system modifies the decision and it chooses the next most probable gesture. This gesture parameter will be trained by the actual gesture parameter grabbed from the camera. This way of interaction is quite natural considering the human behavior. Under the recognition phase the system refreshes parameters if the decision is right, so it is able to adapt to the gestures of the user. With this interactive training algorithm our system is able to adapt to gestures of other users. In Figure 4-1 we see the algorithm of our training method.

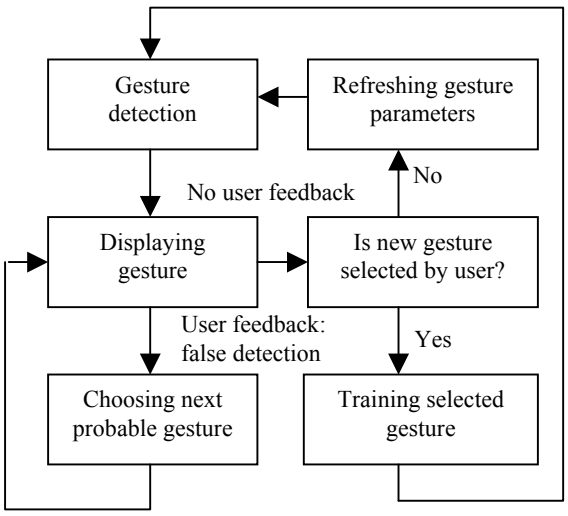

Figure 4-1:Interactive training algorithm

In the natural languages, when people show rapid hand shaking, it usually means denial signal. Our method applies the same movement to get user's feedback of negation. The user's feedback can also be hand shaking or rotating. These signals can be extracted from the variations of the palm position and orientation. When the variations of the position or the rotation are greater than a threshold value, the system recognizes a negation signal.

If the detection is true then parameters of the detected gesture continuously will be refreshed by the parameters of the current gesture with a predefined small weight. Continuously refreshing the static gesture, the system is able to adapt to small changes of gestures. For example, when the user is tired and cannot show standard gestures, the system may learn it. If the detection is wrong, the user indicates it by rapid hand shaking. In this situation the system chooses the next most-probable gesture-class and this gesture will appear on the screen. The user can see the result and can generate feedback again until the result is accepted. In this case the system uses supervised training and corrects the false recognition.

$$Z = \frac{w_1 * X + w_2 * Y}{w_1 + w_2} \qquad (1)$$

Where in the above equation Y is the parameter of the trained gesture and X is the parameter of the trainer's gesture. We can set the training method by w_1 and w_2. In Figure 4-2 we can see the distance of the two gestures in time, with several weights, $w_1 = 30$ and $w_2 = 1$ and so on. In this training the distance decreases slowly since the gradient of the curve is small, while the harder training at the *1*. and *3*. sec results in a deeper descent of the curve.

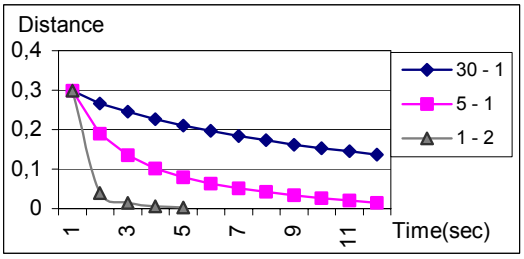

Figure 4-2: Training with several weighting value

In unsupervised training the problem is that two or more classes may get too close to each other, resulting in an unstable recognition. For this reason, the program continuously detects the distances among the classes, and it makes alert when the trained pattern is inconsistent or it is close to be unstable.

In Table 4-1 we summarize the efficiency of our method. First, the trainer tests the method, correcting a little bit to get 99% rate. Then new users work at 76%-86% rates, correcting the machine-parameters in the automatic interaction, resulting in 92%-95% recognition rates.

Table 4-1:Testing with a new user on the old training (9 gesture classes, 140 test images)

Users	Unsupervised learning	Supervised learning
Trainer user	98%	99%
New user #1	86%	95%
New user #2	76%	92%
New user #3	82%	94%

5. Practical application

Our gesture-recognizer and supervised interactive trainer system is applied in a practical application. Restoration of old films is an expensive process and it needs several ways of human-interaction (lighting, contour, color, defects, noises, synchrony). The hand-based interaction can be used in the pre-processing phase: registration of places for hard error corrections or enhancement, notching the key-frames or reference frames. This method supports both the frame-by-frame and frame-sequence operations. Using gestures we are able to track the position of the hand, drawing a continuous line on the screen, browsing among image sequences (start, stop, back), signing defected or key frames (reference frames, cuts).

In our software the reference point of the hand gives a virtual cursor (Figure 5-1), while the recognized gesture generates the command, since the contours of hand gesture are rotation and shift invariant. In the interaction process we see the pictogram of the recognized hand gesture around the cursor, resulting in a continuous feedback about the position and the proposed command. In case of mistake or misunderstanding the rotation/shake of the hand cancels the command and the next probable command (proposed hand gesture) is processed. In this way the method continuously refreshes the gesture-parameters in the unsupervised (accepted cases and fine parameter modifications of slightly different class-features) or supervised (shaking and modifying) training. When there is a danger of mixing (or getting too close to) the different gesture-classes during the unsupervised training due to a forgetful gesture-series, there is a change in the shape of the plotted cursor to alarm the user for the mistake of posture.

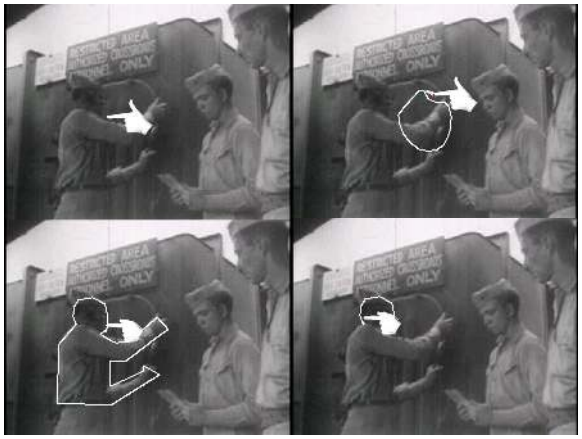

Figure 5-1: Signing the feature-areas in sample images

6. Conclusion

The above work has shown that

- Interactive training is user-friendly and user-independent;
- Gestures classes can be trained from a limited number of training sets (it also works with solo training set);
- Unsupervised training can be continuously run to follow the slight changes in the gesture styles;
- Supervised training is possible to recognize and correct the possible overlap among the different classes;
- We have tested the supervised training system with three users and found that the performance of recognition has increased significantly as experimental data shows above.

7. References

[1] A. Utsumi, T. Miyasato, F. Kishino and R. Nakatsu, "Hand Gesture Hand Gesture Recognition System Using Multiple Cameras", In *ICPR'96, Proceeding of ICPR*, 1996.

[2] K. Cho and S. M. Dunn, "Learning shape classes", in *IEEE Tran. on Pattern Analysis and Machine Intelligence*, vol. 16, 1994, pp. 882-888.

[3] T. Starner and A. Pentland, "Visual Recognition of American Sign Language Using Hidden Markov Models", In *Proc. Int'l Workshop on Automatic Face and Gesture Recognition*, Zurich, 1995.

[4] F.L. Alt, "Digital Pattern Recognition by Moments", JACM, 9, 2, April 1962, pp. 240-258.

[5] K. Imagawa, R. Taniguchi, D. Arita, H. Matsuo, S. Lu, S. Igi, "Appearance-based Recognition of Hand Shapes for Sign Language in Low Resolution Image", *Proc. of 4th Asian Conference on Computer Vision*, 2000, pp. 943-948.

[6] J. Schlenzig, E. Hunter, and R. Jain, "Vision-Based Hand Gesture Interpretation Using Recursive Estimation", *Proc. 28th Asilomar Conf. Signals, Systems, and Computer*, 1994.

[7] W. Freeman and M. Roth, "Orientation histograms for hand gesture recognition," In *International Workshop on Automatic Face and Gesture Recognition*, 1995.

[8] C.T. Zahn and R.Z. Roskies, "Fourier descriptors for plane closed curves", *IEEE Trans. on Computers C21*, 1972, pp. 269-281.

[9] Y. Rui, A. She, T.S. Huang, "A Modified Fourier Descriptor for Shape Matching in MARS", *Image Databases and Multimedia Search*, 1998, pp165-180.

[10] Kohler, "Vision Based Hand Gesture Recognition Systems", http://ls7-www.cs.uni-ortmund.de/research/gesture/vbgr-table.html.

[11] B.K.P. Horn, *Robot Vision*, MIT Press, Cambridge, Massachusets, 1986.