HAND-GESTURE BASED FILM RESTORATION

Attila Licsár
University of Veszprém, Department of Image Processing and Neurocomputing,H-8200 Veszprém, Egyetem u. 10, Hungary
Email: licsara@freemail.hu

Tamás Szirányi
Analogical Computing Laboratory, Computer. & Automation Research Institue,Hungarian Academy of Science,
H-1111 Budapest, Kende u. 13-17, Hungary
Email: sziranyi@sztaki.hu

Keywords: Human Computer Interaction, Gesture-based man-machine interface, Vision-based gesture recognition, Shape analysis, Fourier descriptor, Online interactive working and training, Film restoration

Abstract: We have developed a static hand-gesture recognition system for the Human Computer Interaction based on shape analysis. This appearance-based recognition uses modified Fourier descriptors for the classification of hand shapes. Usually systems use two phases: training and running phase under the recognition. A new method is shown that under the running phase of the system users can interactive modify and learn hand gestures by the gesture motion, so they could improve the efficiency of the system. With this interactive learning algorithm our system is able to adapt to similar gestures of other users or small changing of hand posture. We will show a gesture recognition application applying these methods for the controlling of old film restoration.*

1. INTRODUCTION

In this paper we demonstrate an effective human-computer interface for controlling the steps in the restoration of old films. The method is based on hand-gesture recognition and tracking.

In the information society the communication between user and computer has become very active research area. Decreasing the camera and computer prices, the vision-based systems are more available by everyone, so development of computer vision and analysis is a main point of the Human Computer Interaction.

In the spatial gesture model there are two main approaches: the 3D hand model-based and appearance-based methods. We have chosen appearance-based methods in our system because it's 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 and 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 operating in the virtual environment.

For recognizing static gestures in our system there exist lots of methods in appearance-based recognition. The majority of methods use parameters derived from image. In such cases, the model

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

parameters are derived from description of the shape. Systems can use one or more camera pictures (Utsumi, 1996). They include: edge-based contours (Cho, 1994), edges, image moments (Starner, 1995), image eigenvectors (Imagawa, 2000) or geometric moment description of hand shapes. Some other techniques use second order moments, Zernike methods (Schlenzig, 1994) which invariant to rotation of the shape. Other methods use orientation histograms (Freeman, 1995), 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, for example Fourier descriptor (Zahn, 1972) use only points of the contour. However, the Fourier descriptor is sensitive to starting point of the shape boundary.

In this paper we adopt a modified Fourier descriptor method (MFD) (Rui, 1998), which is already used in the field of the character recognition. We apply this method for the gesture-based man-machine interface.

This example-based system involves two phases: training and running phase. In the training phase, the user shows the system one or more examples of hand gestures. The system stores the Fourier coefficients of the hand shape and in the run phase the computer compares the current hand shape with each of stored shapes by coefficients. The best match gesture is selected by the nearest-neighbour method with the metric of MFD.

Next, using an interactive learning algorithm during the recognition phase, the system performance is increased for the recognition rate by user feedback. Here, the user checks the decision of the system on the screen and denotes if the decision is false. The organization of this paper is as follows. First, we describe features of MFD method then analyse 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 learning method.

2. MODIFIED FOURIER DESCRIPTOR

The first implementations of Fourier descriptor had some weakness, for example invariant to starting point of the shape contour.

Extending this traditional method to the modified Fourier descriptor (Rui, 1998), 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 to noise, when shape boundaries often contain local irregularities due to image noise. Another advantage of the MFD that feature vectors should be computed efficiently. It has a new distance measure to describing and comparing closed curves.

Some application like methods (Kohler) with Fourier descriptor use other classification algorithm (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 *nth* boundary points. The MFD is defined as the Discrete Fourier Transform (DFT) of previous complex sequence. The magnitude and phase angle 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 (Rui, 1998). 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 is defined as standard deviation of the previous two calculated sequences.

So, the classification method of MFD is fast because it computes a simple standard deviation. The system's processing speed is 12 samples/sec on a Celeron 733Mhz computer.

2.1 GESTURE RECOGNITION SYSTEM by MFD

The Fourier descriptors are calculated from the boundary of the palm. The system uses restrictive background to localize the hand efficiently. Hand silhouettes or contours can be easily extracted from the hand image. 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 (Imagawa, 2000), which use the direction of the arm for the automatic forearm segmentation. From the image moments

(Horn, 1986), we can get the global direction and position of the hand.

Because the method invariant to orientation, to position and to size of the palm, we may use these parameters to control or manipulate virtual objects. Using these parameters system could get feedback from the user about the recognition.

We have tested the system with 9 gesture classes. The starting set of training was very small, usually only one, but the continuous supervised/unsupervised learning adaptively changed the class-parameters. Testing process was done by a set of 140 samples. The efficiency of the different recognition methods has been tested to find the best feature-detection.

The methods to select the most appropriate Fourier descriptors:

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

In Table 2-1 the above methods are demonstrated. For each methods the distances among the classes are counted. In case of Method 2, the distances are greater than in the other 2 cases.

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

Table 2-1: Comparing efficiency of selecting methods

Starting with a one-class/one-sample training, a limited number of classes can be appropriately recognized.

3. INTERACTIVE LEARNING SYSTEM

The conventional learning and recognition phases of the system will be extended by a new interactive learning 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, improving the efficiency. In this additional phase the user can modify the recognition strategy by modifying his gestures similarities among classes. The user feedback signal is (crucial shapes to be trained) to avoid false 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 very natural considering the human behavior. Under the recognition phase the system refreshes parameters if the decision is right, and so it's able to adapt to the gestures of the user. With this interactive learning algorithm our system is able to adapt to gestures of other users. In Figure 3-1 we see the algorithm of the interactive learning algorithm.

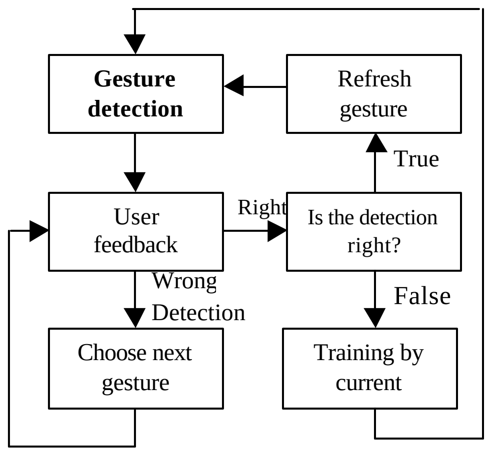

Figure 3-1: Interactive learning algorithm

These example-based systems usually have two phases: training and recognition phase. Our goal that under the recognition phases the system will be able to correct faulty detected gestures and it can interactively modify and train hand gestures with helping of the user's feedback. With this interaction, the recognition efficiency will be grown. In the natural languages, when people show rapid hand shaking, it usually means denial signal. Our method applies the same movement to get user feedback of negation. The user 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.

In the recognition phase the user shows a gesture and the system detects the closest gesture class. The user could see the result of the decision. 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. With continuously refreshing the static gesture the system is able to adapt to small changes of the 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 choose the next most-probable gesture-class and this gesture will appear on the screen. The user can see the result and could generate feedback again until the result will be accepted. In this case the system use supervised training and correct the false recognition. If in the Equation 3-1 Y is the parameter of the trained gesture and X is the parameter of the trainer gesture we can set the training method with w_1 and w_2.

$$Z = \frac{w_1 * X + w_2 * Y}{w_1 + w_2}$$

Equation 3-1: Parameter training method

In Figure 3-2 and Figure 3-3 we can see, the distance of the two gestures in time. In the Figure 3-2 are trainings with several weight, $w_1 = 30$ and w_2. = 1 and so on. In the Figure 3-3 we can see an example for the supervised training among two gestures. In this training it can be found that the distance is decreasing slowly the gradient of the curve is small, while the harder training at the 4. and 6. sec results in a deeper descent of the curve.

In the unsupervised training there is a usual problem that two or more classes could get to close to each other, resulting in a 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.

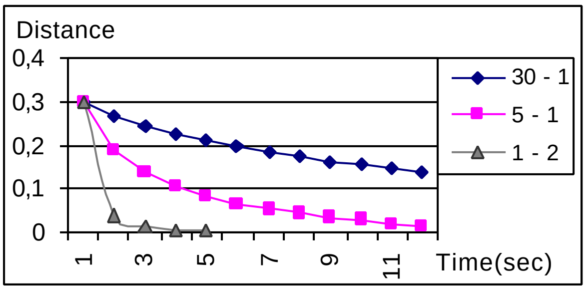

Figure 3-2: Training with several weighting value

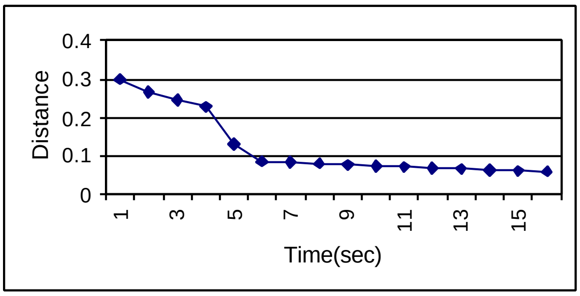

Figure 3-3: Supervised training

In Table 3-1 we summarize the efficiency of our method. First, the trainer is testing the method, correcting a little bit to get 99% rate. Then a new user is working at 86% rate, correcting the machine-parameters in the automatic interaction, and resulting 95% recognition rate.

Users	Recognition rates	
	Unsupervised learning	Supervised learning
Trainer user	98%	99%
New user	86%	95%

Table 3-1: Testing with a new user on the old training 9 gesture classes, 140 training samples

4. APPLICATION FOR THE OLD FILM RESTORATION

Our gesture-recognizer and unsupervised interactive trainer system is applied in our film-restoration project. Restoration of old films is an expensive process, and it needs several ways of human-interaction (lighting, contour, color, defects, noises, synchrony). Since the quality of the mainly automatic process must be checked very quickly in an overview process, including the staff of different people (lighting and color adjusting technical experts, aesthetes), the role of a quick and effective hand-gesture based controlling is crucial.

First, a video format is completed from the original film-frames in digital form. This digital video format is used as a helping support when the main instructions control the automatic (and very slow) restoration process. During this instruction-generation a report is stored parallel with the frames about the necessary tasks.

The hand-based interaction can run 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.

This interactive free-hand menu system also helps in the post-processing stage to record the remaining errors or quality needs. The system generates an output video (at some lower resolution than the final output of the restored film) with the automatically proposed and processed modifications, and the interactions may help again to make a report about the remained possible modifications. In this iterative process we disassociate the human expertise for quality-control and the automatic machine-process.

<u>Communicative gestures</u>:

Rapid movement	Effect of gesture
Rapid hand rotating	wrong gesture detection
Rapid hand shaking	Undo last operation
Rapid hand missing	Restart

<u>What is the meaning of the gestures(Figure 4-1)</u>:

- **Tracking**: positioning and tracking the cursor
- **Select**: drawing a continuous line
- **Start:** next frame, or go forward in frames if continuous
- **Back:** previous frame, or go backward in frames if continuous
- **Stop:** Stop the film
- **Cut:** Signing a cut in the process
- **Reference frame:** Define a key-frame
- **Good frame:** frame without error
- **Defected frame**: frame with error

In our program the reference point of the hand gives a virtual cursor (Figure 4-2), 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 refresh 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 cursor to alarm the user for the mistake of posture.

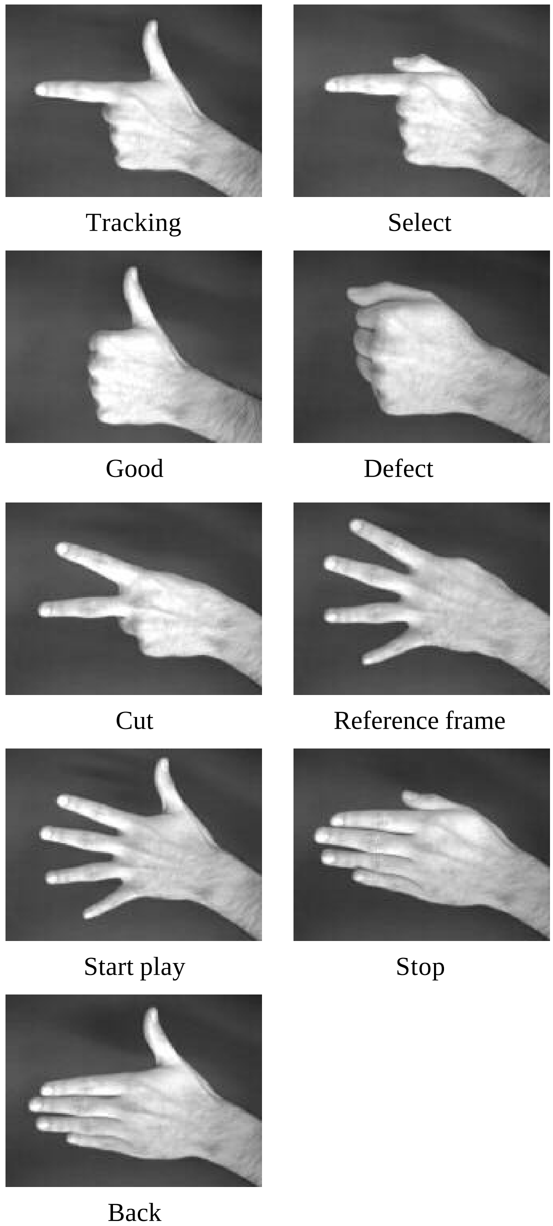

Figure 4-1: Hand signs for the application

Figure 4-2: Signing the feature-areas in sample images

5. CONCLUSION

The above work has shown that

- A satisfactory dictionary of hand-gestures can be generated for film-restoration control;
- Gestures classes can be trained from a limited number of training set (it also works with solo training sets);
- Unsupervised training can be continuously run to follow the slight changes in the gesture styles;
- Unsupervised training is possible to recognize and correct the possible overlap among the different-classes;
- Training is quite fast and user-independent;
- The training and interactive tracking system is quite robust;
- The system generates alert when user is slubberer.

REFERENCES

Alt, R., 1962. Digital Pattern Recognition by Moments. In *JACM.*

Cho, D., 1994. Learning Shape Classes. In *IEEE Trans. Pattern Analysis and Machine Intelligence*.

Freeman, R., 1995. Orientation Histogram for Hand Gesture Recognition. In *Proc. Int'l Workshop on Automatic Face and Gesture Recognition*, Zurich.

Horn, 1986. Robot Vision. In MIT Press, Cambridge.

Imagawa, T., A., M., L., I., 2000. Appearance-Based Recognition of Hand Shapes for Sign Language in Low Resolution Image. In *ACCV'2000. Proceedings of the Fourth Asian Conference on Computer Vision.*

Kohler. Vision Based Hand Gesture Recognition Systems http://ls7-www.cs.uni-dortmund.de/research/gesture/vbgr-table.html

Rui, S., H., 1998. A Modified Fourier Descriptor for Shape Matching in MARS

Starner, P., 1995. Visual Recognition of American Sign Language Using Hidden Markov Models. In *Proc. Int'l Workshop on Automatic Face and Gesture Recognition,* Zurich.

Utsumi, M., K., N., 1996. Hand Gesture Hand Gesture Recognition System Using Multiple Cameras. In *ICPR'96, Proceeding of ICPR.*

Zahn, R., 1972. Fourier Descriptors for Plane Closed Curves. In *IEEE Trans. on Computers*.