Hand Posture Recognition in a Body-Face centered space

Sébastien Marcel

France Telecom CNET

2 avenue Pierre Marzin

22307 Lannion, FRANCE

sebastien.marcel@cnet.francetelecom.fr

ABSTRACT

We propose to use a neural network model to recognize a hand posture in an image. Hand gestures are segmented using a space discretisation based on face location and body anthropometry.

Keywords

Neural networks, hand posture recognition, discrete space.

INTRODUCTION

LISTEN is a real-time computer vision system which detects and tracks a face in a video image [2]. In this system, faces are detected, within skin color blobs, by a modular neural network. This paper deals with a LISTEN based system using hand posture recognition to execute a command. In order to detect the intention of the user to issue a command, "active windows" are investigated in the body-face space. When a skin color blob enters an "active window", hand posture recognition, using a specific neural network for each hand posture, is triggered.

THE BODY-FACE SPACE

We map over the user a body-face space based on a "discrete space for hand location" [4] centered on the face of the user as detected by LISTEN.
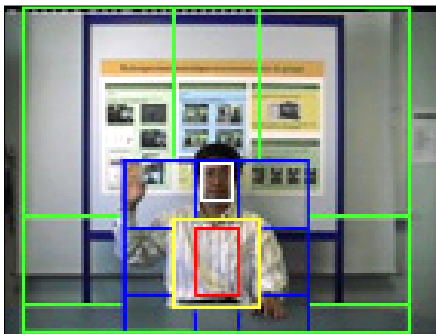

Figure 1. Body-face space

The body-face space (Figure 1) is built using the anthropometric body model expressed as a function of the total height (Figure 2) itself calculated from the face height.
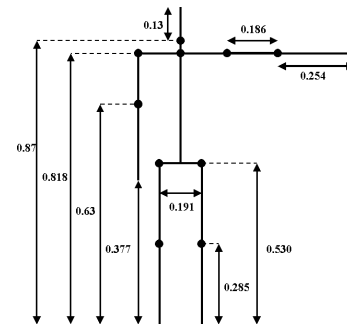

Figure 2. Anthropometric body model

In our images, faces and hands have small sizes. Therefore, hand posture recognition becomes a hard task.

THE NEURAL NETWORK MODEL

Neural networks, such as discriminant models [5] or Kohonen features maps models [1], have previously been applied to hand posture recognition. In this work, we propose to use a neural network model already applied to face detection: the constrained generative model (CGM) [3] (Figure 3).
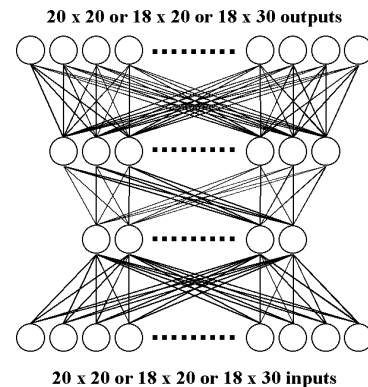

Figure 3. Constrained generative model

The goal of the constrained generative learning is to closely fit the probability distribution of the set of hands using a non-linear compression neural network and non-hand examples. Each hand example is reconstructed as itself and each non-hand example is constrained to be reconstructed as the mean neighbourhood of the nearest hand example. Then, the classification is done by measuring the distance between the examples and the set of hands.

RESULTS ON OUR DATABASE

A small set of hand postures was selected: A, B, C, Five, Point and V. A database of thousands various images with both uniform and complex backgrounds was built. The window sizes for each posture are: 20x20 for A, 18x20 for C and Five, and 18x30 for B, Point and V (Figure 3). Then images are tested at different position and scale in order to frame the hand posture.
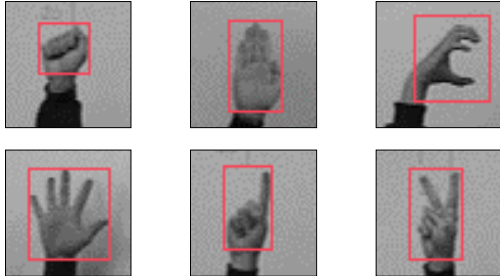
Figure 3. Images form our test database

Most of the database was used for training and the remainder was used for testing. Although the images with complex backgrounds are more difficult to learn, the CGM achieves a good recognition rate and a small false alarm rate (Table 1 and 2).

Postures	Number of images	Mean detection rate	Mean false alarm rate
Table 1 : Mean results on our test database with uniform backgrounds			
A,B,C,V	241	93.8 %	1 / 11,111
A to V	382	93.4 %	1 / 11,904

Postures	Number of images	Mean detection rate	Mean false alarm rate
Table 2 : Mean results on our test database with complex backgrounds			
A,B,C,V	165	74.8 %	1 / 18,867
A to V	277	76.1 %	1 / 14,084

The false alarm rate is of prime necessity to evaluate the performance of the system, and must be small comparing to the number of tests in a image. At the present time, the false alarm rate on various images containing no hands is around 1 false alarm for 27777 tests.
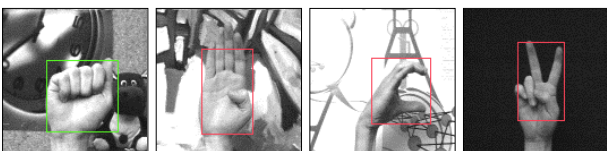
RESULTS ON A BENCHMARK DATABASE

Figure 4. Images form Jochen gallery

The CGM was also tested on the Jochen hand posture gallery [6]. This database contains 128x128 grey-scale images of 10 hand signs performed by 24 persons against uniform light, uniform dark and complex backgrounds. We only tested the A, B, C and V postures (Figure 4).

Postures	Number of images	Mean detection rate	Mean false alarm rate
Table 3 : Mean results on Jochen gallery with uniform backgrounds			
A,B,C,V	191	93.7 %	1 / 25,641

Postures	Number of images	Mean detection rate	Mean false alarm rate
Table 4 : Mean results on Jochen gallery with complex backgrounds			
A,B,C,V	96	84.4 %	1 / 15,873

The CGM applied to hand posture recognition gives satisfactory recognition results (Table 3 and 4) on this benchmark database. These results can be improved by adding more non-hand examples in order to lower the false alarm rate.

CONCLUSION

Work is in progress to integrate the hand posture detection and the body-face space in a new LISTEN based system.

A future work is to supply LISTEN with a hand gesture recognition kernel based on stroke detection and motion analysis.

REFERENCES

[1] Boehm, K. and Broll, W. and Sokolewicz, M., Dynamic Gesture Recognition Using Neural Networks: A Fundament for Advanced Interaction Construction. *SPIE, Conference Electronic Imaging Science and Technology* (1994).

[2] Collobert, M., Feraud, R., Le Tourneur G., Bernier, O., Viallet, J.E., Mahieux, Y. and Collobert, D. LISTEN: A System for Locating and Tracking Individual Speakers. *2nd Int. Conf. on Automatic Face and Gesture* (1996), 283-288.

[3] Feraud, R. PCA, Neural Networks and Estimation for Face Detection. *NATO ASI*, Face Recognition: from Theory to Applications (1997), 424-432.

[4] McNeill, D. Hand and Mind: What gestures reveal about thought. *Chicago Press* (1992).

[5] Murakami, K. and Taguchi, H. Gesture recognition using Recurrent Neural Networks. *CHI*'91, 237-242.

[6] Triesch, J. and Malsburg, C. Robust Classification of Hand Postures against Complex Backgrounds. *2nd Int. Conf. on Automatic Face and Gesture* (1996), 170-175.