Hand Posture Recognition in a Body-Face centered space

Sébastien Marcel, Olivier Bernier

France-Telecom CNET DTL/DLI
2 avenue Pierre Marzin, 22307 Lannion, France
{sebastien.marcel, olivier.bernier}@cnet.francetelecom.fr

Abstract. We propose a model for image space discretisation based on face location and on body anthropometry. In this body-face space, a neural network recognizes hand postures. The neural network is a constrained generative model already applied to face detection.

1 Introduction

LISTEN is a real-time computer vision system which detects and tracks a face in a video image [1]. 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 [2]. In order to detect the intention of the user to issue a command, "active windows" are defined 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.

2 The Body-Face Space

The aim of this section is to introduce a technique of spatial segmentation, the body-face space, in order to assist in hand posture recognition.

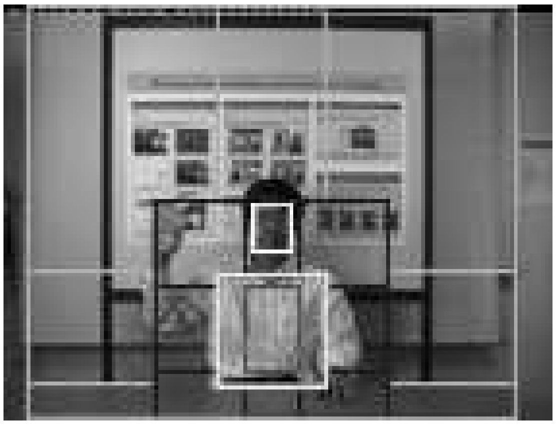

Fig. 1. The body-face space in a full body user image

We map over the user a body-face space based on a "discrete space for hand location" [3] centered on the face of the user as detected by LISTEN.

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

Given the body-face space, we can define "active windows" as areas of interest where hand postures are detected. An "active window" can be placed where hand gestures are usually executed, for example at the left (right) of the body-face space for deictic gesture or at the left (right) of shoulders for command gestures.

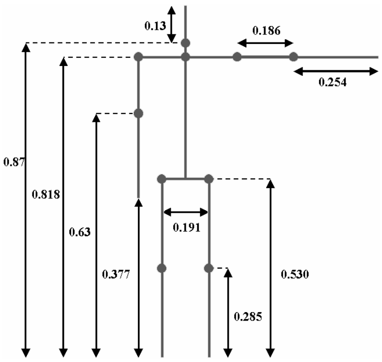

Fig. 2. Anthropometric body model

When a skin color blob enters an "active window", hand posture recognition is triggered. In order to see the whole body, our images are taken with a large field of view. Therefore, faces and hands have small sizes (15x20 pixels for faces and 18x30 pixels for hands) and hand posture recognition becomes a hard task. Nevertheless, LISTEN is able to detect such faces using neural networks. Consequently, we apply the same approach for hand posture recognition.

3 The Neural Network Model

Neural networks, such as discriminant models [4] or Kohonen features maps models [5], 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) [6] (Figure 3).

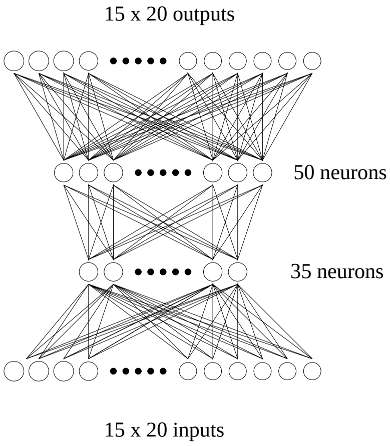

Fig. 3. Constrained generative model

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.

After learning, the Euclidean distance between inputs and outputs of the neural network is smaller for hand examples than for non-hands examples. Using this distance, and given a distance threshold estimated on a generalisation set, the classification of the input image is possible.

In order to frame the hand posture, images are tested by the specific CGMs at different position and scale.

4 Results on our Database

Our goal is to recognize hand postures used in command gestures. A small set of hand postures was selected: A, B, C, Five, Point and V. A database of thousands various hand posture images with both uniform and complex backgrounds was built. The window sizes for each posture (corresponding to the number of inputs for each CGM) are: 20x20 for A, 18x20 for C and Five, and 18x30 for B, Point and V (Figure 4). Most of the database was used for training (80 %) and the remainder was used for testing (20 %).

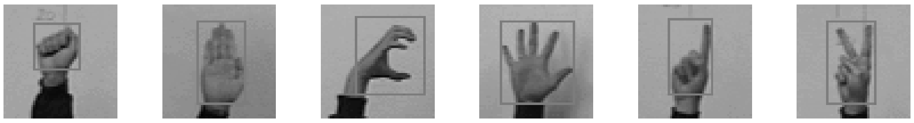

Fig. 4. Examples of hand postures from our test database

Although the images with complex backgrounds are more difficult, the CGM achieves a good recognition rate and a small false alarm rate (Table 1).

Table 1. Mean results on our test database with uniform and complex backgrounds

Uniform backgrounds			
Hand Postures	Number of images	Mean detection rate	Mean false alarm rate
A,B,C,V	241	93.8%	1/11,100
A to V	382	93.4%	1/12,000
Complex backgrounds			
Hand Postures	Number of images	Mean detection rate	Mean false alarm rate
A,B,C,V	165	74.8%	1/18,900
A to V	277	76.1%	1/14,000

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 27800 tests.

5 Results on a Benchmark Database

The CGM was also tested on the Jochen hand posture gallery [7]. 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 5). The CGM applied to hand posture recognition gives satisfactory recognition results (Table 2) on this benchmark database. These results can be improved by adding more non-hand examples in order to lower the false alarm rate.

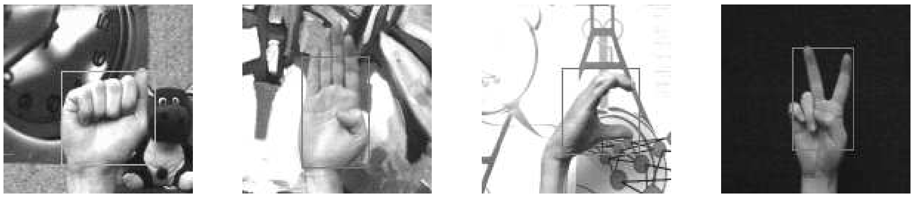

Fig. 5. Examples of hand posture from Jochen Triesch Gallery

Table 2. Mean results on Jochen Triesch gallery with uniform and complex backgrounds

Uniform backgrounds			
Hand Postures	Number of images	Mean detection rate	Mean false alarm rate
A,B,C,V	191	93.7%	1/25,600
Complex backgrounds			
Hand Postures	Number of images	Mean detection rate	Mean false alarm rate
A,B,C,V	96	84.4%	1/15,900

6 Conclusion

We have recently integrate the hand posture detection and the body-face space in a new LISTEN based system running in real-time at 12Hz. Currently, when an unknown hand posture is presented, there is no strategy to choose the neural network to use. All the neural networks are used and the best response identify the hand posture. Future work is to find a strategy for choosing the CGM to use and to supply this system with a hand gesture recognition kernel based on stroke detection and motion analysis.

References

1. Collobert, M. and Feraud, R. and Le Tourneur G. and Bernier, O. and Viallet, J.E. and Mahieux, Y. and Collobert, D.: LISTEN: A System for Locating and Tracking Individual Speakers. 2nd Int. Conf. on Automatic Face and Gesture Recognition (1996) 283–288
2. Baudel, T. and Braffort, A.: Reconnaissance de gestes de la main en environnement réel. Actes de Informatique'93 - L'interface des mondes réels et virtuel (1993) 207–216
3. McNeill, D.: Hand and Mind: What gestures reveal about thought. Chicago Press (1992)
4. Murakami, K. and Taguchi, H.: Gesture recognition using Recurrent Neural Networks. CHI'91, 237-242
5. 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)
6. Feraud, R.: PCA, Neural Networks and Estimation for Face Detection. NATO ASI, Face Recognition: from Theory to Applications (1997), 424-432
7. Triesch, J. and Malsburg, C.: Robust Classification of Hand Postures against Complex Backgrounds. 2nd Int. Conf. on Automatic Face and Gesture (1996), 170-175