IDIAP RESEARCH REPORT

Evaluation Protocols and Comparative Results for the Triesch Hand Posture Database

Sébastien Marcel [a]

IDIAP–RR 02-50

November 2002

submitted for publication

Dalle Molle Institute for Perceptual Artificial Intelligence • P.O.Box 592 • Martigny • Valais • Switzerland

phone +41 − 27 − 721 77 11
fax +41 − 27 − 721 77 12
e-mail secretariat@idiap.ch
internet http://www.idiap.ch

[a] Dalle Molle Institute for Perceptual Artificial Intelligence (IDIAP)

IDIAP Research Report 02-50

Evaluation Protocols and Comparative Results for the Triesch Hand Posture Database

Sébastien Marcel

November 2002

submitted for publication

Abstract. Research efforts in the design of image-based gestural interfaces have steadily increased over the last few years. Numerous approaches have been investigated to recognize gestures such as facial expressions, hand gestures or hand postures. Nevertheless, there exist no reference databases and no standards for the evaluation and the comparison of developed algorithms in gesture recognition and especially in hand posture recognition. This document proposes an evaluation protocol for a benchmark hand posture database, namely the Triesch hand posture database. This document provides also comparative results on this database.

1 Introduction

Research efforts in the design of image-based gestural interfaces have steadily increased over the last few years. Gestural interfaces based on images are the most natural advanced man-machine interfaces. Thus, recognizing gestures becomes a challenging problem in computer vision and pattern recognition.

Numerous approaches have been investigated to recognize gestures such as facial expressions, hand gestures or hand postures. Nevertheless, there exist no reference databases and no standards for the evaluation and the comparison of developed algorithms in gesture recognition and especially in hand posture recognition. We believe that these are two important issues to advance research results. Common reference databases will allow different research labs to evaluate their methods on identical problems and common evaluation procedures will allow the direct comparison of performance.

The section 2 introduces the reader to hand posture recognition. The section describes also a state-of-the-art machine algorithm, namely Multi-Layer Perceptron (MLP), for use in pattern recognition, i.e hand posture recognition. Section 3 proposes an evaluation protocol for a benchmark hand posture database, namely the Triesch hand posture database. Finally, this document provides comparative results on this database and concludes.

2 Hand Posture Recognition

2.1 Problem Description

The recognition of hand postures is a difficult task because the hand is an object highly deformable which has 29 degrees of freedom. The first step to hand posture recognition consists in representing the image of the hand in a multi-dimensional feature space. Many feature extraction techniques have been investigated. Simplest features are based on the gray-scale image [8], but more accurate features could also be used. Size functions [14][5], orientation histograms [4] or signatures [2] are efficient and fast to compute. Steerable filters [3] and Zernike's moments [7][12] are also efficient and have some invariant properties in scale or rotation, but require more computing. The second step is the recognition task itself. Numerous approaches have been investigated such as eigen-vectors [9], Gabor wavelets and Elastic Graph Matching [13], maximum of likelihood [10] or Neural Networks [8].

The aim of the next section is not to propose a new approach for hand posture recognition but to introduce a state-of-the-art machine learning algorithm, namely Multi-Layer Perceptron (MLP), and to describe its application on a hand posture recognition task.

2.2 MLPs for Hand Posture Recognition

A Multi-Layer Perceptron (MLP) is a particular architecture of Artificial Neural Networks. Artificial Neural Networks are learning machines used in many classification problems. A good introduction to machine learning algorithms can be found in [1, 6].

We will assume that we have access to a training dataset of l pairs $(\mathbf{x}_i, y_i)$ where $\mathbf{x}_i$ is a vector containing the pattern, while y_i is the class of the corresponding pattern often coded respectively as 1 and -1.

An MLP is composed of layers of non-linear but differentiable parametric functions. For instance, the output $\hat{y}$ of a 1-hidden-layer MLP can be written mathematically as follows

$$\hat{y} = b + \mathbf{w} \cdot \tanh\left(\mathbf{a} + \mathbf{x} \cdot \mathbf{V}\right) \tag{1}$$

where the estimated output $\hat{y}$ is a function of the input vector $\mathbf{x}$, and the parameters $\{b, \mathbf{w}, \mathbf{a}, \mathbf{V}\}$. In this notation, the non-linear function tanh() returns a vector which size is equal to the number of hidden units of the MLP, which controls its capacity and should thus be chosen carefully, by cross-validation for instance.

An MLP can be trained by gradient descent using the backpropagation algorithm [11] to optimize any derivable criterion, such as the *mean squared error* (MSE):

$$\mathrm{MSE} = \frac{1}{l}\sum_{i=1}^{l}(y_i - \hat{y}_i)^2 \ . \tag{2}$$

In this hand posture recognition approach, an MLP is trained (for each posture) to classify an input to be either the given posture or not. The input of the MLP is a feature vector corresponding to the hand image. The output of the MLP is either 1 (if the input corresponds to a posture) or -1 (if not). The MLP is trained using both hand images and non-hand images. As non-hand images, we used patterns randomly selected from background images.

Finally, the decision to accept or reject an input depends on the score obtained by the corresponding MLP which could be either above (accept) or under (reject) a given threshold, chosen on a separate validation set to optimize a given criterion. When no validation set is available, the threshold has to be selected a priori on the training set for instance.

3 Database and Protocol

3.1 The Jochen Triesch Hand Posture Database

Some results on hand gesture recognition were already published using the Jochen Triesch hand posture database [13][8]. This database consists of 10 hand signs performed by 24 persons against three backgrounds. Images were recorded in 8-bit grey-scale and are 128x128 pixels. For each person the 10 postures were recorded in front of uniform light, uniform dark and complex background giving 720 images of which 2 were lost. The filename of all images is made first by six letters corresponding to the identity of the person (*bfritz* for instance), second by the letter corresponding to the hand posture (*a* for example) and finally by a digit for the background (1, 2 or 3). *bfritza1.pgm* is an example of such an image file.

3.2 Evaluation Protocol

In this section, we propose an evaluation protocol with two configurations for the hand posture recognition task using the Jochen Triesch database. Each configuration consists in splitting the 718 images into two subsets: a train set and a test set. The train set is used to train the parameters of the model and the test set is used to evaluate the performance of the model.

3.2.1 Triesch Configuration

This configuration was initially proposed by [13]. The images of three persons against light and dark background (60 images) were used for the **train set**. For each posture, 6 training images are available. The remaining pictures constitute the **test set**. The pictures of the three subjects taken against complex background enter into the test set.

3.2.2 Martigny Configuration

This new configuration is an extension of the Triesch configuration. The images of three persons (the same as in the Triesch configuration) against light, dark and complex background were used for the **train set**. For each posture, 9 training images are available. The remaining pictures constitute the **test set**.

The Jochen Triesch hand posture gallery[1] is provided with a set of protocol files and manually located hand postures.

[1] The database and the protocol are available at http://www.idiap.ch/∼marcel/Databases/main.html.

Table 1: Summary of results published on the test set of the Jochen Triesch database (Triesch Configuration only)

Light Background ($b = 1$)			
	p	$T_{p,1}$	$RR_{p,1}$ ($R_{p,1}$)
EGM	all postures	210	94.3% (198)
Dark Background ($b = 2$)			
	p	$T_{p,2}$	$RR_{p,2}$ ($R_{p,2}$)
EGM	all postures	208	93.9% (194)
Complex Background ($b = 3$)			
	p	$T_{p,3}$	$RR_{p,3}$ ($R_{p,3}$)
EGM	all postures	239	86.2% (206)
CGM	A, B, C, V only	96	84.4% (81)
Light and Dark Background ($b = 2 + 3$)			
	p	$T_{p,2+3}$	$RR_{p,2+3}$ ($R_{p,2+3}$)
CGM	A, B, C, V only	167	93.7% (156)

3.3 Performance Measures

The goal of the hand posture recognition is to identify the hand posture in a given image. The recognition rate is computed as follows for each type of hand postures and for each background type:

$$RR_{p,b} = R_{p,b}/T_{p,b} \tag{3}$$

where p is the hand posture type (A, B, C, D, G, H, I, L, V, Y) and b is the background type (1, 2, 3). $R_{p,b}$ is the number of recognized postures p over background b. $T_{p,b}$ is the total number of hand postures p over background b.

4 Experimental Comparison on the Triesch Database

4.1 State-of-the-art Results

To our knowledge, only two authors already published results on the Triesch hand posture database. The first author [13] use Elastic Graph Matching (EGM) and Gabor wavelets. The second author [8] use a constrained Neural Network (CGM). Results of these two methods are presented in table 1.

The recognition rate obtained by the EGM method is presented as the average over all postures against the three backgrounds. The CGM method was, unfortunately, tested on posture A, B, C and V only. Also, authors didn't distinguished light and dark background, and published results on an average of both.

4.2 MLP-based Hand Posture Recognition

We have trained an MLP for each hand posture using the training set of the Triesch database on both configurations.

4.2.1 Training the Models

For each posture, the training database is composed of a hand training set (6 images for Triesch configuration and 9 images for Martigny configuration) and a non-hand training set. The hand training set is enlarged by shifting (8 directions and 4 pixel shifts), scaling (2 scales) the original hand bounding box.

Hence, the **hand training set** contains 990 patterns ($6 * P$) for Triesch configuration and 1485 patterns for Martigny configuration.

The extended number of pattern P is computed such that $P = A * B$, i.e. the shifted and scaled hand patterns. $A =$ number of shifts $* 8 + 1$ is the total number of shifts, in 8 directions, including the original frame, for each scale. $B =$ number of scales $* 2 + 1$ is the total number of scales, in 2 directions (sub-scaling and over-scaling), including the original scale. The **non-hand training set** is different for each posture.

#P	width x height	Triesch Conf #non-hand	Martigny Conf #non-hand
A	20 x 20	1114	1684
B	20 x 40	1034	1587
C	27 x 28	1455	2180
D	20 x 38	1046	1571
G	30 x 22	1458	2229
H	31 x 22	1452	1982
I	20 x 32	1474	2227
L	34 x 40	1319	2016
V	24 x 42	1016	1538
Y	34 x 27	1017	1561

Table 2: Windows size and number of non-hand patterns for each posture.

The representation used to code input images is based directly on the gray-scale hand image. Thus, as the input of an MLP is fixed, a window, in which the posture should belong, has to be selected.

The table 2 contains for each posture the selected input window and the number of non-hand patterns. The window size was selected on the training images as the average of bounding box size.

#P	#M	Triesch Conf #nhu	Martigny Conf #nhu
A	400	110	70
B	800	120	80
C	756	120	90
D	760	100	110
G	660	50	70
H	682	110	120
I	640	70	80
L	1360	110	120
V	1008	90	90
Y	918	110	110

Table 3: Input dimension and number of hidden units for each MLP.

For each MLP, the “optimal” number of hidden units was selected by a 3-fold cross-validation on both configurations. Then, each MLP was re-trained with its own number of hidden units on all the training set.

The table 3 contains the input dimension (#M) and the number of hidden units (#nhu) of each MLP on both configurations.

Table 4: Results on the test set of the Jochen Triesch database (Light background, $b = 1$) using MLP on both configurations.

		Triesch Conf	Martigny Conf
p	$T_{p,1}$	$RR_{p,1}$ $(R_{p,1})$	$RR_{p,1}$ $(R_{p,1})$
A	21	95.2% (20)	61.9% (13)
B	21	95.2% (20)	100% (21)
C	21	52.4% (11)	61.9% (13)
D	21	85.7% (18)	90.5% (19)
G	21	57.1% (12)	85.7% (18)
H	21	85.7% (18)	100% (21)
I	21	95.2% (20)	95.2% (20)
L	21	95.2% (20)	85.7% (18)
V	21	90.5% (19)	90.5% (19)
Y	21	80.9% (17)	47.6% (10)
all	210	74.76% (157)	81.9% (172)

4.2.2 Testing the trained Models

For each hand posture P of size $\mathcal{W}_P$x$\mathcal{H}_P$, the corresponding MLP ($\mathcal{M}_P$) is tested over the test set of hand posture images P. The test image is scanned at different scales (the image is sub-sampled) and locations with a fixed window of size $\mathcal{W}_P$x$\mathcal{H}_P$. Then, the sub-image within this window is extracted and forwarded into the MLP $\mathcal{M}_P$. If the output of the MLP is above zero threshold then the sub-image is classify as a hand posture. Zero has been selected a priori to be the threshold, because no validation set is available, and because zero is at half distance between training targets (-1 and $+1$). Finally, at the end of the scanning, the window with the highest score is chosen. This window is considered as a good recognition if its bounding box matches the true bounding box of the posture at least at 70%.

Results using this MLP approach are provided in table 4, 5 and 6 for the two configurations of the proposed protocol. These results are not as good as the results obtained by the EGM or the CGM. Anyway, results show that the global recognition rate is better using the Martigny configuration than using the original Triesch configuration. Indeed, the Martigny configuration uses hand postures with complex background while the Triesch configuration does not. It is important to compare a recognition algorithm on these two protocols to emphasize its robustness on complex background. Again, the purpose of this paper was not to present a new state-of-the-art method but to illustrate with an example the use of this protocol.

Table 5: Results on the test set of the Jochen Triesch database (Dark background, $b = 2$) using MLP on both configurations.

p	$T_{p,2}$	Triesch Conf $RR_{p,2}$ ($R_{p,2}$)	Martigny Conf $RR_{p,2}$ ($R_{p,2}$)
A	21	95.2% (20)	90.5% (19)
B	21	95.2% (20)	95.2% (20)
C	21	71.4% (15)	95.2% (20)
D	21	66.6% (14)	85.7% (18)
G	21	42.8% (9)	42.9% (9)
H	21	42.8% (9)	42.9% (9)
I	21	90.5% (19)	85.7% (18)
L	21	100% (21)	100% (21)
V	20	100% (20)	90% (18)
Y	21	100% (21)	81% (17)
all	209	80.4% (168)	80.8% (169)

Table 6: Results on the test set of the Jochen Triesch database (Complex background, $b = 3$) using MLP on both configurations.

p	$T_{p,3}$	Triesch Conf $RR_{p,3}$ ($R_{p,3}$)
A	24	62.5% (15)
B	24	70.8% (17)
C	24	25% (6)
D	24	58.3% (14)
G	24	16.7% (4)
H	23	30.4% (7)
I	24	79.2% (19)
L	24	83.3% (20)
V	24	83.3% (20)
Y	24	87.5% (21)
all	239	59.8% (143)

p	$T_{p,3}$	Martigny Conf $RR_{p,3}$ ($R_{p,3}$)
A	21	71.4% (15)
B	21	76.2% (16)
C	21	52.4% (11)
D	21	81% (17)
G	21	47.6% (10)
H	21	38.1% (8)
I	21	61.9% (13)
L	21	85.7% (18)
V	21	95.2% (20)
Y	21	76.2% (16)
all	210	68.57% (144)

5 Conclusion

In this paper, we have proposed an evaluation protocol for a benchmark hand posture database. The suggested benchmark database is the Triesch hand posture database. Two hand posture recognition methods using this database have been already published. This paper describes also a Multi-Layer Perceptron approach for hand posture recognition to illustrate the use of the proposed evaluation protocol. We encourage researchers to use this database as a reference and to use the proposed protocol to compare their algorithms in hand posture recognition and to advance research results.

Acknowledgments

The author wants to thank the Swiss National Science Foundation for supporting this work through the National Center of Competence in Research (NCCR) on "Interactive Multimodal Information Management (IM2)". This work was also funded by "France Telecom R&D" and by the European Face and Gesture Working Group "FG-NET" through the Swiss Federal Office for Education and Science (OFES).

The author thanks also Jochen Triesch for giving us the right to use and to distribute his database.

References

[1] C. Bishop. *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford, 1995.

[2] Ulrich Brckl-Fox. Real-time 3d interaction with up to 16 degrees of freedom form monocular video image flow. In *International Workshop on Automatic Face and Gesture Recognition*, pages 172–178, June 1995. Zurich, Switzerland.

[3] William T. Freeman and Edward H. Adelson. The design and use of steerable filters. In *IEEE Transaction Pattern Analysis and Machine Intelligence*, volume 13 of *9*, pages 891–906, september 1991. MIT - Media Lab and Dept of Brain and Cognitive Sciences.

[4] W.T. Freeman and M Roth. Orientation histograms for hand gesture recognition. In *International Workshop on Automatic Face and Gesture Recognition*, pages 296–301, june 1995. Zurich, Switzerland.

[5] P. Frosini. Measuring shapes by size functions. In *Proceedings SPIE on Intelligent Robotic Systems*, 1991. Boston.

[6] S. Haykin. *Neural Networks, a Comprehensive Foundation, second edition*. Prentice Hall, 1999.

[7] E. Hunter, J. Schlenzig, and R. Jain. Posture estimation in reduced-model gesture input systems. In *International Workshop on Automatic Face and Gesture Recognition*, pages 290–295, june 1995. Zurich, Switzerland.

[8] S. Marcel. Hand posture recognition in a body-face centered space. In *Conference on Human Factors in Computing Systems CHI'99*, pages 302–303, 1999. Extended Abstracts.

[9] Baback Moghaddam and Alex Pentland. Maximum likelihood detection of faces and hands. In *International Workshop on Automatic Face and Gesture Recognition*, pages 122–128, June 1995. Zurich, Switzerland.

[10] D. Rubine. *The automatic recognition of gestures*. PhD thesis, Carnegie Mellon University, 1991.

[11] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. In D. E. Rumelhart and James L. McClelland, editors, *Parallel Distributed Processing*, volume 1. MIT Press, Cambridge, MA., 1986.

[12] J. Schlenzig, E. Hunter, and R. Jain. Vision based hand gesture interpretation using recursive estimation. In *Proceedings of the 28thAsilomar Conference on Signals, Systems and Computer*, 1994.

[13] Jochen Triesch and Christoph Malsburg. Robust classification of hand posture against complex backgrounds. In *Proceedings of the second International Conference on Automatic Face and Gesture Recognition*, pages 170–175, october 1996. Killington, Vermont, USA.

[14] Claudio Uras and Alessandro Verri. Hand gesture recognition form edge maps. In *International Workshop on Automatic Face and Gesture Recognition*, pages 116–121, June 1995. Zurich, Switzerland.