,

Enhancing Gesture Dictionary of a Commercial Data Glove Using Complex Static Gestures and an MLP Ensemble

Ognjan Lužanin[*] - Miroslav Plančak
Faculty of Technical Sciences, Novi Sad, Serbia

This paper focuses on the enhancement of the static gesture dictionary of commercial data glove 5DT 5 Ultra, with the primary goal to improve its ergonomic features and usability for Mechanical CAD (MCAD). The standard gesture dictionary of this data glove is based on 16 joint-limit simple static gestures designed in the 1990s by the NASA Aimes laboratory. Although simple to learn and perform and easy to recognize, most of these gestures have poor ergonomic features and are non-intuitive in the symbolical sense. The authors addressed this problem and suggested improvements by eliminating 11 original simple static gestures and substituting them with new complex static gestures. Since the restructured gesture dictionary of 12 simple and complex static gestures imposed a problem of lower gesture recognition rate, this issue was approached using artificial intelligence. Namely, an ensemble of five multilayer perceptrons (MLPs) with backpropagation was used as gesture classifier. Bearing in mind that variable hand anatomies of different data glove users are one of the crucial factors impeding gesture recognition, two female and three male subjects participated in the gesture data acquisition to provide a total of 2400 static gestures which were used to train, validate and test the ensemble classifier. For each of the five member networks, a resampling of the data set was performed, aleviating the problem of variance. The results showed that the proposed restructuring of data dictionary can be efficiently supported by the ensemble-based gesture classifier.
Keywords: virtual reality, data glove, static gesture, artificial neural network

0 INTRODUCTION

Virtual reality technologies are widespread in today's production manufacturing. This results in a number of simulations which, besides the classical human-computer interface (HIC), also use multi-modal HIC. Multi-modal interaction puts a human in the center of the interface, focusing on extraction and interpretation of information gathered from hand gestures, speech, tactile feedback and other modalities of communication.

Since the second half of the 1970s when their use was pioneered by M. Krueger [1], hand gestures have been recognized as a natural and intuitive way of communicating not only with virtual environments, but with computers in general. However, despite of some three decades of development, the application of gestures in virtual environments is far from desired. As Poupyrev et al. noted [2], there exists no unified framework for virtual environment interaction, no desktop-style metaphor familiar to the majority of users, and no optimal interaction technique for all possible task and input devices in virtual environments. Furthermore, despite their original intention, gestures are often criticised as being imprecise, non-ergonomic and not self-revealing [3]. The area of gesture design, application and recognition is still in its infancy.

Amongst various definitions of the gesture found in general dictionaries and scientific papers, a definition by Turk [4] seems to be most comprehensive stating that a gesture is a meaningful body motion - i.e. physical movement of the fingers, hands, arms, head, face or body with the intent to convey information or interact with the environment [3]. A number of authors have contributed to the development of classification and taxonomy of gestures [3] to [9].

Discussion in this paper is confined to hand gestures. In addition, the distinction is made between hand gestures based on two fundamental properties: complexity and dynamics, which are required for proper gesture recognition. According to these criteria, hand gestures can be classified into:

- simple static gestures (postures) - in which finger configurations consist of either fully closed or fully open fingers,
- complex static gestures (postures) - in which fingers can be flexed at an arbitrary angle,

---

[*]Corr. Author's Address: Faculty of Technical Sciences, Institute of Production Engeneering., V. Perića Valtera 2, 21000 Novi Sad, Serbia, luzanin@uns.ac.rs

- simple dynamic gestures - in which either just a hand is moved or the fingers are moving with the hand in a fixed position,
- complex dynamic gestures - which involve the movement of fingers, as well as changes in location and orientation of a hand.

With this classification in view, the scope of this paper is limited to simple and complex static gestures.

One of the key advantages of virtual environments is the possibility of intuitive and natural interaction with the tasks in hand. As a highly articulated part of the human body, hand outperforms the most modern primary input devices with six degrees of freedom. In that respect, the data glove is no doubt the first true interface for direct manipulation [10]. The first data glove SayreGlove, was developed by Defanti and Sandine in 1977 [11]. However, more intensive research of their application in virtual environments began in the 1990s.

Up to date, numerous commercial and experimental models have been made, but regardless of the technical solution, data gloves have two basic applications, namely: (i) to allow real-time control of hand avatar and direct manipulation with virtual objects and (ii) to allow control of VR simulation using gestures. Modern commercial data gloves can be divided into (i) data gloves which measure fingure flexion and (ii) data gloves which register contacts between fingers/palms. Among the most popular commercial data gloves in academia and industry are Cyberglove (Immersion), 5DT 5/14 Ultra (Fifth Dimension Technologies), which measure fingure flexion, and Pinch Glove (Fakespace), which registers contacts. Due to a large number of sensors (18/22 – depending on the variant) and the developed software support, Cyberglove is presently a *de facto* industrial standard.

1 PROBLEM DEFINITION

Owing to its all-round characteristics and low price, 5DT 5 Ultra data glove (Fifth Dimension Technologiess) is very popular among industry professionals and researchers. The glove is equipped with five proprietary optical sensors, which allow it to measure finger flexure using one sensor per finger. Thus, signals from the sensors represent mean value of finger flexions at metacarpophanlangeal and proximal joints [12]. It also comes with a ready software support for the detection of the predefined set of 16 gestures. The authors have used this glove with an experimental VR desktop system, which is described in more detail in [12]. However, the glove is not suitable for use with MCAD without some important modifications:

- only a third of the predefined gesture set are ergonomically suitable for prolonged use, since they cause muscle fatigue,
- although thumb flexure is measured, none of the standard-dictionary gestures include the use of thumb, which not only prevents the simulation of grasping but also makes it impossible to simulate useful gestures which require the use of thumb,
- owing to the ergonomic issues and the principles of operation of the optical sensors, the gestures are sometimes misinterpreted or undefined,
- most of the gestures lack symbolical meaningfulness which is necessary for efficient use with MCAD applications.

Gestures supported by 5DT 5 Ultra are coded as the combination of binary states (open = 1/closed = 0) of the four fingers, excluding the thumb. Thus, it is possible to create $2^4=16$ gestures as combinations of open/closed fingers. The gestures are assigned numbers from 0 to 15, 0 representing the fist (all fingers closed) and 15 representing the open hand (all fingers open). Gesture recognition is based on boundary values. If the sensor reading for a particular finger is above the predefined boundary value, the finger is considered closed (flexed). Conversely, if the reading falls below the boundary value, the finger is open (no flexion). All readings which fall between the boundary values are considered as errors, and the gesture is undefined.

This simple method of gesture recognition works well in some situations but is not always reliable. This can be attributed to anatomical variations in various users, as well as to the disposition of optical sensors and their cross-coupling, which can often result in incorrect recognition of a gesture [14].

1.1 Biomechanical Aspects of Finger Movement

Instead of discussing hand degrees of freedom on a complex anatomy model, it is more

convenient to use simplified mechanical models proposed by [15] to [19]. For this purpose, the model by Beifang et al. [15] was chosen (Fig. 1).
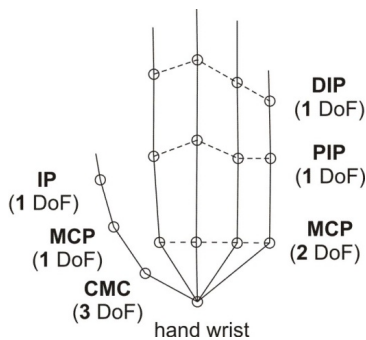

Fig. 1. *Simplified mechanical model of a human hand with joint designations [15]*

This model depicts finger bones as lines, while the joints are represented by dots. Hand movements are observed as combinations of bone rotations in various joints. Four fingers have three joints each - metacarpophalangeal (MCP), proximal interphalangeal (PIP) and distal interphalangeal joint (DIP). MCP allows two degrees of freedom (flexion/extension and abduction/adduction), while the other two joints (PIP, DIP) allow one degree of freedom each - (flexion/extension). DIP and PIP joints on the four fingers and IP thumb joint allow one degree of freedom each (flexion). MCP and CMC thumb joints, as well as MCP finger joints, allow two degrees of freedom each - flexion and abduction/adduction.

The four fingers and the thumb are characterized by different flexibility and interdependence of movements. Due to its special position, the thumb is most independent in its movements, as was experimentally proven by Häger-Ross and Schieber [20]. Extrinsic muscles which enable thumb movements mostly act independently, thus its relative independence in comparison with movements of other fingers. The index finger can perform flexion and extension independent from other four fingers. Flexion and extension of the middle finger are limited. Also, full flexion of the middle and small finger hinders extension of the small finger [21]. Due to connections between hand and forearm, extension of the ring finger depends on the neighbouring fingers. The small finger has its own extrinsic muscles which allow it relative independence of

movements. Thanks to a couple of long extensor muscles, the small finger, just like the index finger, can extend independently when all other fingers are fully flexed.

2 CRITICAL ASSESSMENT OF THE STANDARD 5DT GESTURE DICTIONARY

Gestures of the 5DT 5 Ultra glove (Fig. 2) are based on the gesture dictionary proposed in 1993 by the NASA Ames Research Center. These static gestures were originally intended for glove-based navigation tasks in virtual environments and can be designated as joint-limit postures because they use only configurations of either fully open or fully closed fingers. The advantages of joint-limit approach are simplicity of software support required for gesture recognition and ease of learning. On the other hand, they have several disadvantages.

Firstly, some of these gestures cause hand fatigue even in cases when no protracted use is required, causing a distinct ergonomic problem. Secondly, most of them are devoid of any symbolical meaning which devalues them from the cognitive aspect, making them less suitable for a natural, logical contextual application, which should be the primary advantage of gestures in virtual environments. Gesture number 4 (Fig. 2) can be taken as an example. In his study on the functional anatomy of the hand, Tubiana [21] and [22] reports that full flexions of the middle and small finger completely impede extension of the ring finger, which indicates that gesture 4 is not only ergonomically inadequate but is also impractical from the anatomic point of view.

Besides the gestures, shown in Fig. 2 are also their ergonomic ratings. Filled circle denotes good ergonomic features, white circle stands for bad ergonomy, while the semi-filled circle denotes average ergonomic features of the particular gesture.

3 MODIFICATION OF THE STANDARD 5DT GESTURE DICTIONARY

We addressed the ergonomic and cognitive issues by restructuring the predefined gesture dictionary, eliminating some simple gestures and substituting them with the complex ones.
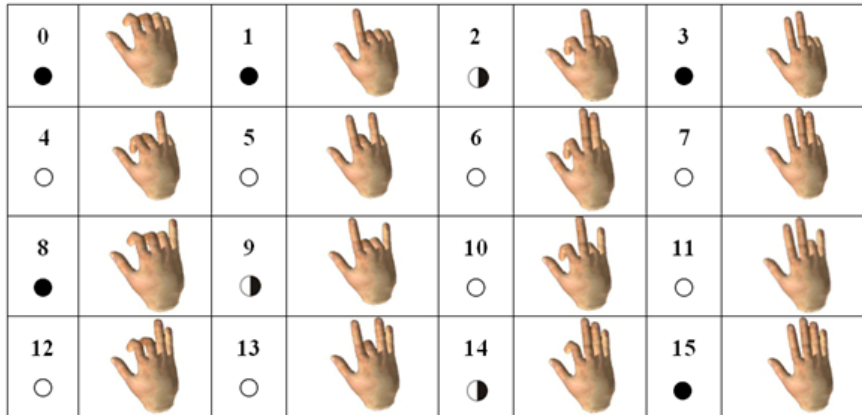
Fig. 2. *Standard gesture dictionary of 5DT 5 Ultra data glove*

Bearing in mind the ergonomic features and iconic and metaphoric meaningfulness, the existing gesture dictionary was restructured in order to allow the elimination of:

- finger configurations which involve full extension of the middle finger with the index and ring finger fully flexed,
- finger configurations which involve full extension of the ring finger with the middle finger or middle and index finger fully flexed,
- finger configurations which involve full flexion of the index, middle and ring finger with the small finger and thumb fully extended.

Following the above stated guidelines, a total of eleven gestures were eliminated from the standard 5DT gesture dictionary (2, 4, 5, 6, 7, 9, 10, 11, 12, 13 and 14) (Fig. 2). Of the remaining five gestures, three of them were left in their original form (0, 8 and 15), while the two remaining gestures were modified. Thus, gesture designated 1, was modified into gesture G03, while gesture 3 became gesture G05. Also added were five novel gestures, G04, G06, G08, G09 and G10 (Fig. 3).

The modified gesture dictionary comprises twelve static gestures and, according to conditional classification proposed by LaViola [8], belongs to small gesture dictionaries. This is an advantage regarding the efficiency of gesture recognition. Should the dictionary require an extension at any point, this can be solved with a context-dependent gesture recognition where a single gesture can be attributed with several meanings, depending on the suggested context of the application.

4 DEVELOPMENT OF ANN-BASED GESTURE RECOGNITION SUPPORT

In order to enhance ergonomic features of the 5DT 5 Ultra a data glove, the original gesture dictionary was modified by adding complex static gestures. This means that it now features gestures with partially flexed fingers (gestures G6, G7, G8, G9 and G10) (Fig. 3).

On the other hand, the departure from the joint-limit concept deteriorated the gesture recognition rate, so in order to compensate for that, new gesture recognition support was designed and implemented.

A report on two main stages of development - the gesture data acquisition and the training and testing of artificial neural network (ANN) gesture classifier is also presented.

4.1 Gesture Data Acquisition

Five subjects, two females and three males, age 14 to 44, took part in the experiment, i.e. gesture data acquisition. None of them had medical history affecting the upper limbs, whereas they differed in finger skills and general hand flexibility - from below-average to excellent. Only one of the subjects had previous experience with the data glove. The subjects had 45 seconds to perform each of the twelve gestures, wearing left-handed 5DT 5 Ultra data glove. They were instructed to make small, controlled finger movements while performing static gestures in order to generate noise which would, in a normal situation, be the result of hand
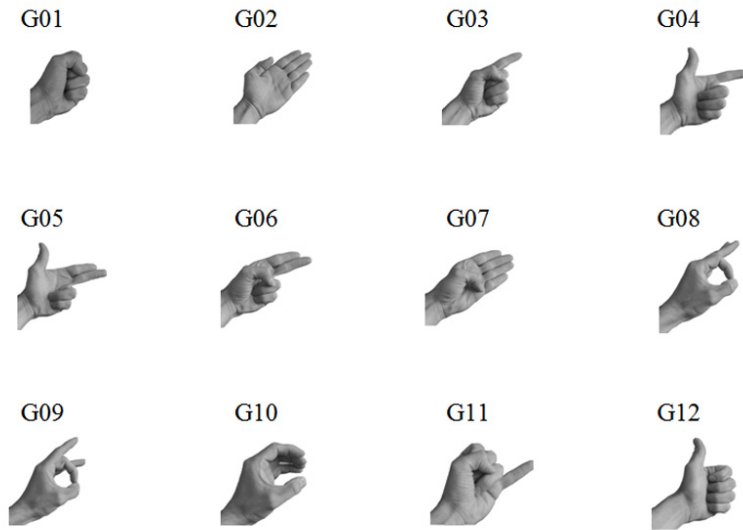
Fig. 3. *Modified gesture dictionary containing simple and complex static gestures*

fatigue and gesture inconsistency. A custom made software application monitored input from the data glove on a USB port, with the update frequency of 5Hz, and recorded gesture data into comma-separated-value files (CSV) (one for each gesture, for each subject). Each record consisted of six fields, the first five being the sensor readings, while the sixth field was gesture designation (1 to 12). The five subjects were represented with 40 records per gesture, thus 5x40x12=2400 gestures were used in total to train and test the ANN classifier.

4.2 Ensemble Design, Training and Testing

The gesture recognition task was fulfilled using an ensemble of neural networks, i.e. multilayer perceptrons (MLPs) with one hidden layer and back propagation learning. Simulations were performed in the neural networks module of *Statistica 7*. Instead of using just one neural network for the task, a set of ten MLPs was formed, trained, validated and tested in order to select the best five MLPs which formed an *ensemble*. Ensembles improve performance since averaging across different MLPs lowers the expected variance, i.e. the sensitivity of MLPs to the choice of the data set which would otherwise cause variations in classification error. Logistic and linear transfer functions were used for the neurons in the hidden and output layer, respectively.

Train and test errors as well as the number of neurons in the hidden layer for the five constituent MLPs of the ensemble, are given in Table 1. Member networks are designated by *MLP*, while the ensemble is designated by *Output*. The summary statistics for classification results obtained by the ensemble are given in Table 2. The data represent average values scored by the five member networks (MLPs).

Table 1. *General data for the five member MLPs and the ensemble*

Member network / ensemble	Train error	Test error	Inputs	Hidden
MLP 5:5-18-12:1	0.343628	0.37812	5	18
MLP 5:5-30-12:1	0.202739	0.30542	5	30
MLP 5:5-28-12:1	0.192906	0.37029	5	28
MLP 5:5-29-12:1	0.209057	0.39992	5	29
MLP 5:5-30-12:1	0.195123	0.31809	5	30
Output 5:[5]:1	0.228690	0.35443	5	5

Table 2. *Classification results for the ensemble*

	G01	G02	G03	G04	G05	G06	G07	G08	G09	G10	G11	G12
Total	200	200	200	200	200	200	200	200	200	200	200	200
Correct	193	199	196	199	196	191	200	190	175	200	200	185
Wrong	7	1	4	1	4	9	0	10	25	0	0	15
Unknown	0	0	0	0	0	0	0	0	0	0	0	1
Correct (%)	96.5	99.5	98.0	99.5	98.0	95.5	100	95.0	87.5	100	100	92.5
Wrong (%)	3.5	0.5	2.0	0.5	2.0	4.5	0.0	5.0	12.5	0.0	0.0	7.0
Unknown	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5

5 DISCUSSION OF RESULTS

As can be seen from Table 1, the five member networks differ in the number of neurons in the hidden layer with the minimum of 18 (MLP 5:5-18-12:1) and the maximum of 30 neurons (MLP 5:5-30-12:1, networks 2 and 5). Network 2 tested best, scoring an error of 0.30542, while network 4 had the largest test error of 0.39992. The overall error of the ensemble (Output 5:[5]:1) is the mean value of test errors for the five member networks and equals 0.35443.

The classification results for individual gestures (Table 2) reveal that gestures G7, G10 and G11 were classified with 100% accuracy. As expected, the ensemble performed worst in the case of two very similar gestures, G8 and G9, and gesture G12, which is very similar to G01. In addition, a single case of gesture G12 was unclassified and thus labeled as unknown gesture. Its class could not be decided by voting due to a disagreement between the five member networks.

In all, gesture recognition using the five-membered ensemble yielded satisfactory results showing that the modified gesture dictionary can be successfully used without significant deterioration of recognition accuracy.

6 CONCLUSIONS

The restructuring of the standard static gesture dictionary of 5DT 5 Ultra data glove successfully tackled the issue of ergonomy and symbolic meaning. Improvements were made by introducing complex static gestures which not only improved the ergonomic features of the gesture dictionary, but also imposed a distinct symbolic framework which helps users to pinpoint the function of particular gestures without extensive training. The problem of gesture recognition was efficiently solved using an ensemble of five MLPs.

Ongoing investigation is aimed towards designing a system for static gestures recognition which would be flexible enough to allow efficient introduction of novel static gestures, as well as new users of data glove. Multilayer perceptrons with backpropagation are not suitable for this task primarily due to the fact that they require lengthy retraining with old and new data sets in case of any modifications. For that reason, the system shall be based on a probabilistic neural network (PNN) which shall be trained on a clustered data set to allow the necessary reduction of network complexity and the increase of processing speed.

7 REFERENCES

[1] Krueger, M., Gionfriddo, T., Hinrichsen, K. (1985) VIDEOPLACE: An Artificial Reality, Proceedings of the 1985 ACM Conference on Human Factors in Computing Systems (CHI'85), ACM Press, p.35-40.

[2] Poupyrev, I., Billinghurst, M., Weghorst, S., Ichikawa T. (1997) A Framework for testbed for studying manipulation techniques for immersive virtual reality. *Proc. of the ACM Symposium on Virtual Reality Software and Tech.*, p.21-28.

[3] Cerney, M., Vance, J. (2005) Gesture Recognition in virtual environments: A review and framework for future development. Iowa State University *Human Computer Interaction Technical Report*, ISU-HCI-2005-01.

[4] Turk, M. (2002) Gesture recognition, *Handbook of virtual environments, design, implementation and applications*, K.Stanney (Ed.), Lawrence Erlbaum Associates, 2002, ISBN: 080583270X, p. 223-239.

[5] Bowman, D., Kruijff, E., LaViola, J., Poupyrev, I. (2005) 3D User interfaces – theory and practice, *Addison-Wesley*, ISBN: 0-201-75867-9.

[6] Förster, J., Lang, H., Rinker, B., Santos, F. (2005) Gestenerkennung/Gesture-Recognition, *Report based on the Lecture on User Interfaces.*

[7] Funck, S., Fuchs, S. (2005) Development of a gesture-intuitive man-machine interface based on video-supported hand gesture recognition, Technical Report (in German), *TUD-FI02-05*, Dresden Technical University, ISSN 1430-211x.

[8] LaViola, J. (1999) A Survey of hand posture recognition techniques and technology, *Technical Report CS-99-11*, Department of Computer Science, Brown University, Providence.

[9] Nielsen, M., Störring, M., Moeslund, T., Nielsen, E.,G. (2003) A Procedure for developing intuitive and ergonomic gesture interfaces for man-machine interaction, *Technical Report CVMT 03-01*, ISSN 1601-3646, CVMT, Aalborg University.

[10] Hand, C. (1997) A Survey of interaction techniques, *Computer Graphics Forum*, vol. 16, no. 5, 1997, p. 269-281.

[11] Defanti, T., Sandin, D. (1977) Final Report to the National Endowment of the Arts, University of Illinois at Chicago.

[12] Fifth dimension technologies (2004), 5DT data glove ultra series, *User's Manual.*

[13] Lužanin, O., Plančak, M. (2008) Virtual reality technologies in virtual manufacturing - current trends and applications, *Journal of Technology of Plasticity*, vol. 33, no. 1-2, ISSN:0354-3870, p. 103-111.

[14] Kahlesz, F., Zachmann, G., Klein, R. (2004) Visual-fidelity dataglove calibration. *Computer Graphics International (CGI)*, June 16-19, Crete, Greece, IEEE Computer Society Press.

[15] Beifang, Y., Frederick, C., Harris, Jr., Ling W., Yusong, Y. (2005) Real-Time natural Hand Gestures, *Computing in Science and Engineering*, vol. 7, no. 3, p. 92-97.

[16] ElKooura, K., Singh, K. (2003) Handrix: Animating the Human Hand. *Eurographics/SIGGRAPH Symposium on Computer Animation.*

[17] Lee, J., Tosiyasu, K.. (1995) Model-Based Analysis of Hand Posture, *IEEE Computer Graphics and Applications*, vol. 15, i. 5, p. 77-86.

[18] Lin, J., Ying, W., Huang, T. (2000) Modeling the constraints of human hand motion, *Workshop on Human Motion (HUMO'00)*, p. 121-126.

[19] Marcel, S. (2002) Gestures for multi-modal interfaces: A Review, *IDIAP-research Report 02-34.*

[20] Häger-Ross, C., Schieber, M.H. (2000) Quantifying the independence of human finger movements: comparisons of digits, hands, and movement frequences, *The Journal of Neuroscience*, Nov. 15, 2000, 20(22), p. 8542-8550.

[21] Tubiana, R. (2005) Movements of the fingers, *Medical problems of Performing Artists*, vol. 20, no. 4, p. 187-192.

[22] Tubiana, R., Charmagne, P. (2005) Functional anatomy of the hand, *Medical Problems of Performing Artists*, vol. 20, no. 4, p. 183-187.

Lužanin, O. - Plančak, M.