Hand Gesture Vocabulary Design: A Multicriteria Optimization *

Helman I. Stern
Department of Industrial Engineering and Management
Ben Gurion University of the Negev
Be'er-Sheva 84105, Israel
helman@bgu.ac.il

Juan P. Wachs
Department of Industrial Engineering and Management
Ben Gurion University of the Negev
Be'er-Sheva 84105, Israel
juan@bgu.ac.il

Yael Edan
Department of Industrial Engineering and Management
Ben Gurion University of the Negev
Be'er Sheva 84105, Israel
yael@bgu.ac.il

Abstract - *A global approach to hand gesture vocabulary (GV) design is proposed which includes human as well as technical design factors. The human centered desires (intuitiveness, comfort) of multiple users are implicitly represented through indices obtained from ergonomic studies representing the psycho-physiological aspects of users. The main technical aspect considered is that of machine recognition of gestures. We believe this is the first conceptualization of the optimal hand gesture design problem in analytical form. The problem is formulated as a multicriteria optimization problem (MCOP) for which a 3D representation of the solution space is used to display candidate solutions, as well as Pareto optimal ones. A computational example is given for the design of a small robot command GV using the MCOP procedure.*

Keywords: Hand gesture, optimal vocabulary, human interfaces, multiobjective decision, multicriteria optimization, man-machine interaction, intuitive interfaces

1 Introduction

In this work we consider hand gesture representations of a vocabulary of commands, which are vision based. Although less accurate than encumbered gestures (digital gloves), they allow users to avoid restricting devices. Hand gestural input to an artificial recognition device takes on two forms; encumbered and unencumbered. The first includes digital gloves, hand markers, infrared tags or any other unnatural accessory placed in contact with the hand [9]. Unencumbered gestural input, referred to as vision based, can be achieved through visual capture devices such as color or infrared cameras making no physical contact with the hand [2]. Advantages and disadvantages of these methods are discussed in detail in [3]. Also, machine vision and haptic-based approaches for acquisition of gesture data are discussed in Shahabi et. al. [1]. The difficulties with data gloves are size of fit, cumbersome, tethered hands, long-term reliability, calibration problems, and cost. Vision based gesture recognition is susceptible to variable illumination, camera resolution and quality of feature extraction for recognition. Although less accurate for fine manipulative tasks vision based capture allows the user to be free of any restricting devices and is more natural. Pavlovic et al. [11] provide a nice review of vision based hand gestures for Human Computer Interaction (HCI).

Design of gesture interfaces is a virgin area of research. Examination of the literature reveals unstructured approaches. Current solution methods of *GV* design may be classified as ad-hoc [8], and rule-based [6]. The two main factors that should be considered in the design of a *GV* are: human (intuitiveness and comfort) and technical (accuracy). Most research has dealt with the machine aspects of a *GV*, focusing on recognition algorithms. A few researchers have considered the human and technical factors jointly. One of them [7], where human factors are considered, gives limited attention to technical aspects. The approach is heavy on human interpretation of rules. Matching of gestures to commands is done empirically through user response queries. In this paper, a new analytical method for *GV* design is proffered. The method considers both human and machine based factors.

In section 2 to follow the basic research problem is defined, along with notation, and the multiobjective performance measures. Section 3 contains the proposed methodology to solve the hand gesture vocabulary design problem, including the system architecture and data requirements. The multicriteria optimisation problem appears in section 4. In 5 a small example problem is solved to illustrate the methodology. The final section provides the conclusions.

2 Problem Definition and Notation

The basic research problem here is to provide an analytical method to find an optimal hand *GV*. An optimal hand *GV* is defined as a set of gesture-command pairs that it will minimize the time for a given user(s), to perform a task(s). Since task completion time, as a function of *GV*, is unknown, we propose instead acting collectively as proxies, the three performance measures; intuitiveness, comfort, and recognition accuracy designated as *Z1*(*GV*), *Z2*(*GV*) and *Z3*(*GV*), respectively, each of which are to be maximized. The first two are human valued, while the third is machine valued. Intuitiveness of a *GV* is the sum total of the intuitiveness of each gesture-command pair in the vocabulary. Comfort is inversely related to strength needed to perform a gesture. Total comfort is equal to the comfort values of gestures and transitions between them, weighted by the frequencies of use. Accuracy, obtained from a gesture recognition algorithm, measures the percent of gestures successfully recognized. Weights can be used to reflect the relative importance of the criteria, but are difficult to determine. The MCOP shown below allows the decision maker to provide subjective preferences.

2.1 Multiobjective Performance Measures

Each of the performance measures is shown as a function of the given gesture vocabulary *GV* in (1), (2), and (3) below. Maximizing each of the measures over the set of all feasible *GV*s defines a multiobjective decision problem. We note, that the analytical form of the objectives; *Z1*(*GV*), *Z2*(*GV*) and *Z3*(*GV*) are linear, quadratic, and unknown, respectively. The objective functions *Z1*(*GV*) and *Z2*(*GV*) are human valued measures, while *Z3*(*GV*) is machine valued.

A vocabulary *GV* may be described in terms of an assignment function p where p(i)=j indicates that the command i is assigned to gesture j.

GV={(*i*, *p*(*i*) | *i* =1,..,n} the set of gesture-command pairs.

Intuitiveness of a gesture is the naturalness of expressing a given command. The intuitiveness of a gesture vocabulary is the sum total of the intuitiveness of each gesture-command pair in the vocabulary:

$$Z_1(GV) = \sum_{(k,p(k))\in GV}^{n} a_{k,p(k)} \qquad (1)$$

where, $a_{k,p(k)}$ is the intuitiveness of representing command *k* by its matched gesture *p(k)*.

Comfort is related to the strength needed to perform a gesture. Obviously there are gestures that are easier to perform than others. Even when some of them look comfortable in the beginning, after some time the user may feel fatigue and the fatigue measure is related to muscle forces, which causes finger and palm tensions. Total comfort is a scalar value equal to the weighted sum of the individual comfort values of the gestures (gesture transitions) weighted by the frequencies of use.

$$Z_2(GV) = \sum_{[(i,p(i)),(j,p(j)]\in GV\times GV}^{n} f_{ij}\ \bar{s}_{p(i),p(j)} \qquad (2)$$

Where f_{ij} is the frequency of transition from command i to command j, and s_{kl} is the physical difficulty of a transition between gesture *k* and gesture *l*. Note that s_{kk} is the fatigue of holding the same gesture.

$\bar{s}_{p(i),p(j)}$ is the comfort value of a transition from gesture *p*(*i*) to gesture *p*(*j*) (an inverse function of s_{kl}).

Accuracy is a measure of how well a set of gestures can be recognized. This is obtained from the confusion matrix, which is based on the classification results of a given recognition algorithm. The recognition accuracy (in percent) is:

$$Z_3(GV) = \frac{(total\ gestures - gestures\ misclassified)}{total\ gestures}100 \qquad (3)$$

3 Methodology

The proposed methodology will be developed under the following assumptions:

(a) The gestures are static postures.

(b) Each gesture cannot represent more than one command, and each command must be expressed by exactly one gesture.

(c) A simple biomechanical model will yield enough information to estimate the fatigue (comfort) measure.

(d) Intuitiveness will be based on a small empirical experiment.

3.1 Architecture

The optimal hand gesture vocabulary architecture (Figure 1) will include two stages: Stage 1 - Hand Gesture Factor Determination and Stage 2 - Optimal Gesture Vocabulary Search Procedure. Stage 1 is the determination of the human psychophysical factors, comfort and intuitiveness. Stage 2 is an optimal vocabulary search procedure incorporating the machine factor, accuracy.

The task set *T*, gesture set *G* and command set *C* are fixed inputs to the first stage. Note, that *C* is determined by *T*. Given a set of tasks, the union of all commands used to perform all tasks constitutes *C*. The objectives of this stage are to establish associations between commands and gestures based on user intuitiveness (intuitiveness matrix) and to find the comfort matrix based on command transitions and fatigue measures.

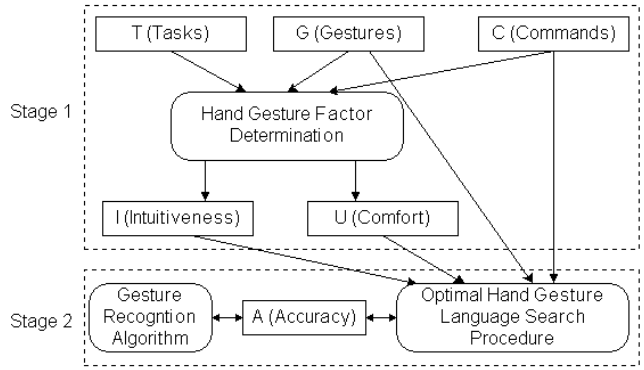


Figure 1. Architecture of optimal hand gesture vocabulary methodology

The inputs to the second stage are the calculated matrices; intuitiveness *I*, comfort *U*, and the command set *C*, gesture *G*, and recognition value of the accuracy, *A*. This stage employs a search procedure to find the best vocabulary *GV*, is formulated as a multiobjective decision problem. The architecture of Stage 1, the Hand Gesture Factor Determination Stage, is shown in Figure 2.

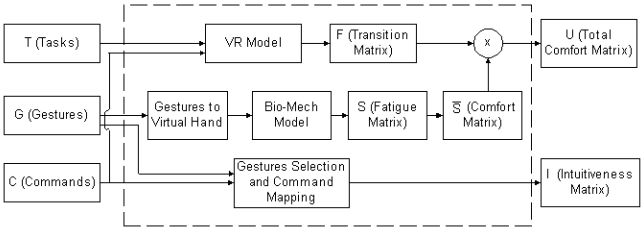


Figure. 2. Hand gesture factor determination stage

3.2 Task and Command Sets (T, C)

The task set can be single element or multiple elements (multi-tasks) set. For each task t_i , a set of C_i commands are defined. For a multi-task set $T=\{t_1,..,t_n\}$ the command set is the set of common commands $C = \bigcup_{i=1,,n} C_i$. For example for a 'place' task with commands C_1={'left', 'right', 'up', 'down', 'backward', 'forward'} and for a 'pick' task with commands C_2={'up', 'down', 'backward', 'forward', 'open', 'close'}, a new task (multi-task) 'pick & place' will include the command set *C*={'left', 'right', 'up', 'down', 'backward', 'forward', 'open', 'close'} which is the union of the previous two command sets.

3.3 Command Transition Matrix (F)

To estimate the frequency of command usage for the set of selected tasks *T* it is necessary to carry out experiments using a real or virtual reality robotic model. For a command set *C* of size *n* a matrix F_{nxn} is constructed, where f_{ij} represents the frequency that a command c_j is evoked, given that the last command was c_i. This measure is significant in the sense that it is hypothesized that; (a) an optimal hand gesture vocabulary will pair high frequency commands to gestures that are easy to perform (low fatigue); and (b) the physical ease of movement between gestures will be paired with high frequency command transitions.

3.4 Set of Master Gestures (G)

Since the set of all possible gestures is infinite, we first establish a set of plausible gesture configurations. To create the set of all plausible hand pose gestures there are two possible approaches; (a) visual capture of gesture images, or (b) creation of synthetic gestures. For small hand gesture databases, real hand gestures images may captured and labelled with the configuration parameters that characterize that gesture; For large gesture sets (thousands of gestures) a tedious effort is required which may be overcome by the use of a synthetic (virtual) gesture generator. One possible way is to generate the configurations by specifying finger positions (extended, spread) and palm orientations (up, down sideways).

3.5 Matrix of Intuitive Indices (I)

The intuitive index is a measure of how "natural" it is for a user to express a command with a particular gesture. These indices are determined empirically. For each command c_i a user is queried to select or display the gesture that he/she "cognitively" associates the most with the command. If desired, gesture command pairs that do not meet some minimum threshold for intuitiveness (e.g. having the user point up for a scroll down or move down command) can be a priori thrown out by setting the associated a_{ik}. value to zero.

Using this information it is straightforward to construct an intuitiveness matrix, $I_{n\,x\,m}$. The entries of *I* are represented as a_{ik}. and are determined by (4).

$$a_{ik} = \frac{n_{ik}}{N_i} \quad , i = 1,...,n \quad , \quad k = 1,...,m \tag{4}$$

Where,

n_{ik} = the number of users that selected gesture g_k to express command c_i

N_i = the number of trials for the i^{th} command.

Note that the values of a_{ik} lie between 0 and 1. For a given command c_i, gestures with larger values of a_{ik} represent more intuitive associations.

3.6 Fatigue and Comfort Matrices (S, $\bar{S}$)

The fatigue (or comfort) indices are determined through the use of a biomechanical model of the hand. The biomechanical hand gesture model is used to find $S_{m\,x\,m}$, whose common element s_{ij} represents the physical difficulty of performing a transition from gesture *i* to gesture *j*. The comfort matrix $\bar{S}$ is achieved by applying an inverting function to each element.

3.7 Total Comfort Matrix (*U)*

Let the coefficients u_{ijkl} be the entries of a square matrix U of size n^2, such that u_{ijkl} is on row $(i-1)n+k$ and the column $(j-1)n+l$. An entry $u_{ijkl} = f_{ij} \times \bar{S}_{kl}$ represents the frequency of transition between commands i to j times the comfort of a k to l transition when commands i and j are paired with gestures k and l, respectively in a given GV. This product reflects the concept that the total comfort measure of GV depends on the frequency of use of a gesture or a gesture pair transition. Note, that the diagonal terms represent the total comfort of using a gesture repeatedly to carry out the same command.

3.8 Gesture Recognition Algorithm

The hand gesture recognition process will be vision based, and contains two sequential tasks; (a) extracting relevant features from the raw image of a gesture, and (b) using those image features as inputs to a classifier. Such an algorithm is described in [4] where the segmentation consists of the extraction of the hand gestures from the background using grayscale cues. The evoked gesture will lie on a uniform black background, and greyscale partition blocks will be created from a segmented hand silhouette. Using some metric, these features will be compared to clusters already created using the fuzzy c-means algorithm. The classification results in a confusion matrix. From the confusion matrix, the recognition accuracy $Z_3(GV)$ is computed using (3). This result indicates the recognition ability of the system for the given set of hand gestures. Further details may be found [5].

4 The Multicriteria Optimization Problem

The multicriteria optimization problem (MCOP) shown below allows the decision maker to provide subjective preferences.

$$\max Z_1(GV) = \sum_{i}^{n}\sum_{j}^{n} a_{ij} x_{ij} \tag{5}$$

$$\max Z_2(GV) = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} u_{ijkl} x_{ik} x_{jl} \tag{6}$$

$$\max Z_3(GV) \tag{7}$$

$s.t$

$$\sum_{j=1}^{m} x_{ij} = 1, \qquad i = 1,..,n, \tag{8}$$

$$\sum_{i=1}^{n} x_{ij} \le 1, \qquad j = 1,..,n, \tag{9}$$

$$x_{ij} \in \{0,1\}, \qquad i = 1,...,n, \qquad j = 1,...,n., \tag{10}$$

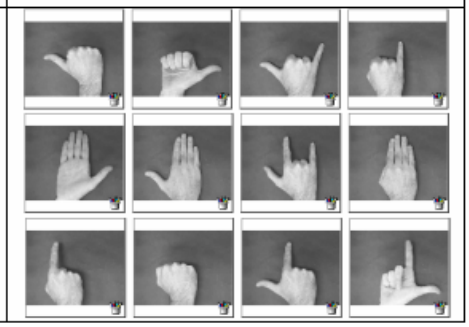
In (6) maximizing u_{ijkl} tends to pair high frequency use commands with less stressful gestures. (where $u_{ijkl} = f_{ij}\bar{S}_{kl}$). The binary variable x_{ij} =1 represents an assignment of gesture j to command i, and 0 otherwise. Constraints (8) and (9) insure that each command i is assigned a unique gesture, and each gesture j is assigned to no more than one command, respectively. Here, the index $a_{i,j}$ represents the intuitiveness of representing command i by its matched gesture j, f_{ij} is the frequency of transition from command i to command j, and $\bar{S}_{kl}$ is the comfort of the transition between gestures k and l. To evaluate $Z_3(GV)$, a recognition algorithm must be called, and solved for the particular GV represented by the 0-1 assignment variables. When there is more than one non-commensurable objective function to be maximized, solutions exist for which the performance in one cannot be improved without sacrificing performance in at least one other. Such solutions are Pareto optimal points [10], and the set of all such points form the Pareto frontier. A solution x^* is a Pareto point iff does not exist another solution y such that $f_i(y) \ge f_i(x^*) \quad \forall i = 1,..., D$, and $f_i(y) < f_i(x^*)$ for some i, where f_i is the ith objective function.

Given that the gesture set is of size m and the command set of size n, there are m!/((m-n)!n!) different gestures sub sets. For each subset of size $|G_n|$ the total number of command-gesture matching is $|G_n|$!.

5 Example Problem

A small example of 12 gestures and 8 robot commands (see Table I) is considered. For this problem the size of the GV solution space is 495. Considering a GV has 8! possible matching, the solution space is ~ $20x10^6$.

Table I. Hand Gesture Vocabulary

Commands	Gestures
LEFT RIGHT FORWARD BACK FAST SLOW START STOP	

For each gesture subset we select the best command-gesture matching by solving a quadratic program comprised of a quadratic objective (5) + (6) subject to the constraints (8), (9), and (10). To evaluate $Z_3(GV)$, a recognition algorithm must be called, and solved for the particular GV represented by the optimal 0-1 assignment variables. The optimal assignment variables are also used to obtain the intuitiveness, $Z_1(GV)$, and comfort, $Z_2(GV)$, performance values. Each result can be viewed as a point in 3D space, whose coordinates are; intuitiveness, comfort, and accuracy, allowing the decision maker to select the desired solution based on his internalised priorities (see Figure.3) To aid the decision maker we also provide the

Pareto optimal points shown in Table II, and in Figure .3 in bold.

Table II. Optimal Pareto points for the MCOP example

Pareto Pts	Accuracy(%)	Intuitiveness(%)	Comfort(%)
1	100	68.92	95.87
2	96.25	100	69.87
3	95.41	76.79	100

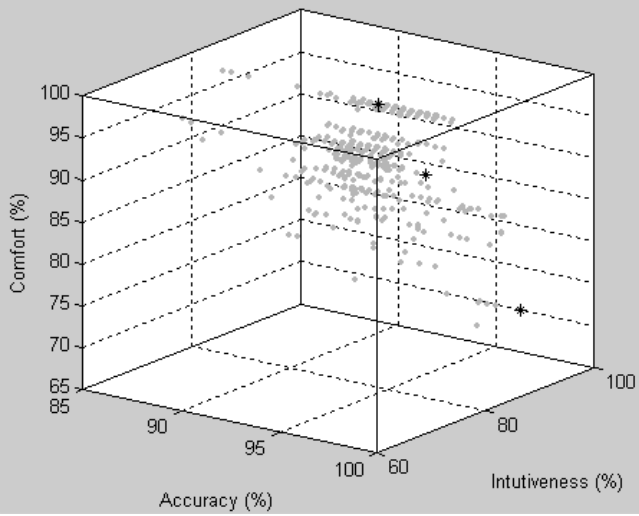


Figure. 3. 3D plot of *GV* solutions

6 Conclusion

In this research a rigorous formulation and solution methodology to the *GV* design problem is proffered. Two aspects drive the need for such a method; *GV* design research is presently an ad-hoc procedure, and gesture interfaces are needed to fill the need for more natural intuitive communication with non-human devices such as computers and robots. Of the three design methods, ad-hoc, rule-based and analytical, we believe this is the first conceptualization of the optimal hand *GV* design problem in analytical form. A MCOP is developed, which reflects the ergonomic and technical performance measures upon which a *GV* control system is judged. By posing the optimal *GV* design problem as a MCOP, solutions can be presented as 3D representations, including Pareto optimal ones. This allows the designer to have an overview of possible solutions and select one based on his/her preferences. Calculating the entire Pareto set for larger problems requires an approach such as an evolutionary multicriteria procedure. Although, our methodology does require effort to obtain human ergonomic and cognitive indices, it provides a structure for replacement and expansion. More accurate fatigue or intuitiveness indices can easily replace old data by updating the gesture knowledge database. This effort will not be lost as it can provide a database for subsequent studies.

7 Acknowledgment

This project was supported by the Ministry of Defense MAFAT Grant No. 1102 and partially supported by the Paul Ivanier Center for Robotics Research & Production Management, Ben-Gurion University of the Negev.

References

[1] C. Shahabi, L. Kaghazian, S. Mehta, A. Ghoting, G. Shanbhag, M. McLaughli, "Analysis of Haptic Data for Sign Language Recognition," in *9th Intl Conf. Human Computer Interaction*, New Orleans, Aug. 2001.

[2] F. Quek, "Unencumbered Gestural Interaction," *IEEE Trans. MultiMedia*, vol. 3, no. 4, pp. 36-47, 1996.

[3] J. J. LaViola, "Whole-Hand and Speech Input In Virtual Environments," Master's Thesis, CS-99-15, Brown University, Department of Computer Science, Providence, RI, 1999.

[4] J. P. Wachs, H. Stern, Y. Edan, "Real-Time Hand Gesture Telerobotic System Using the Fuzzy C-Means Clustering Algorithm," *WAC 2002*, Orlando, Florida, U.S.A, 2002.

[5] J. P. Wachs, H. Stern, Y. Edan, "Parameter Search for an Image Processing Fuzzy C-Means Hand Gesture Recognition System" in *Proceedings of IEEE International Conference on Image Processing*, vol. 3, Barcelona, Spain, pp. 341-345, 2003.

[6] K. Abe, H. Saito, and S. Ozawa, "Virtual 3-D Interface System via Hand Motion Recognition from Two Cameras," *IEEE Trans. Systems, Man and Cybernetics, Part A*, vol. 32, no. 4, pp. 536-540, Jul. 2002.

[7] M. Nielsen, M. Storring, T. B. Moeslund, and E. Granum, "A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for Man-Machine Interaction," Technical Report CVMT 03-01, CVMT, Aalborg University, March, 2003.

[8] T. Agrawal and S. Chaudhuri, "Gesture Recognition Using Position and Appearance Features," in *Proc. IEEE Int. Conf. Image Processing*, Barcelona, Spain, 2003.

[9] T. G. Zimmerman and J. Lanier, "A Hand Gesture Interface Device," in *Proc. ACM SIGCHI/GI*, pp. 189-192, 1987.

[10] V. Pareto. Manuel, *D' Economie Politique*. Marcel Giard, Paris, 2nd Edition, 1927.

[11] V. Pavlovic, R. Sharma, and T. Huang, "Visual Interpretation of Hand Gestures for Human Computer Interaction: A Review," *IEEE PAMI*, vol. 19. pp. 677-695, 1997.