Optimization of Hand Gesture Command Vocabularies - A Multiobjective Quadratic Assignment Approach

J. P. Wachs, H. Stern, and Y. Edan

Abstract— In this research we attack the difficult problem of designing a hand gesture control vocabulary. A hand gesture vocabulary is used to command robotic manipulators to carry out specific tasks. We review and classify three approaches to this problem: Ad hoc, Rule based and Analytical. We believe this is the first conceptualization of the optimal hand gesture design problem in analytical form. To solve the problem, a non-trivial abstract representation of a natural mapping between gestures and commands and relationships between them is introduced. Three mathematical models are developed, which reflect the ergonomic and technical performance measures upon which a gesture control system is judged. The first, is a mathematical program using a quadratic assignment problem embedded within a heuristic search tree. The quadratic problem is used to select a gesture vocabulary (a command-gesture matching) through simulated annealing, The second, is a genetic algorithm representation, and the third is a multi-objective decision problem for which a 3D representation of the solution space is used to display candidate solutions, as well as, the Pareto optimal ones. A computational example is given for the design of a small command gesture vocabulary for the multiobjective procedure.

***Index Terms*— Hand gesture, optimal vocabulary, human interfaces, combinatorial optimization problems, quadratic assignment problem, multiobjective decision, genetic algorithm, heurisitcs.**

I. Introduction

In this research we are concerned with one-way communication in which a human commands a machine through gestures. The machine recognizes and responds to the gesture carrying out some defined action. Our gesture control vocabularies are in contrast other forms such as sign language for human-human communication. Thus, our domain of application is human- machine communication, as opposed to human-human communication. There are two major aspects of human-machine communication through the use of gestures. The first is human based, and the second is machine based. Human based aspects are learnability and memorability of the gestures [1] while the machine-based aspect involves pattern recognition algorithms. Gestures good for one may not be good for the other.

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 [2]. 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 [3]. Advantages and disadvantages of these methods are discussed in detail in [4]. Also, machine vision and haptic-based approaches for acquisition of gesture data are discussed in Shahabi et. al [5].

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. [6] provide a nice review of vision based hand gestures for Human Computer Interaction (HCI). Hand gestures are one but a few of the methods used in telerobotic control [7]. This type of communication provides an expressive, natural and intuitive way for humans to control robotic systems. One benefit [8] of such a system is that it is a natural way to send geometrical information to the robot such as: left, right, etc. Gestures may represent a single command, a sequence of commands, a single word, or a phrase, and may be static or dynamic. Human gesture serves three functional roles: semiotic, ergotic, and epistemic [9]. The semiotic function of a gesture is to communicate meaningful information. The structure of a semiotic gesture is conventional and commonly results from shared cultural experiences [10]. The good-bye gesture, the American Sign Language, the operational gestures used to guide airplanes on the ground, and even the vulgar ``finger", each illustrates the semiotic function of gesture. The ergotic function of gesture is associated with the notion of work. It corresponds to the capacity of humans to manipulate the real world, to create artifacts, or to change the state of the environment by ``direct manipulation". Shaping pottery from clay, wiping dust, etc. result from ergotic gestures. The epistemic function of gesture allows humans to learn from the environment through tactile experience. By moving your hand over an object, you appreciate its structure; you may discover the material it is made of, as well as other properties.

For now, we will not be concerned with multisemiotic activity, which are gestures that accompany other languages

such as oral. Our gestures, therefore, will be created as an independent semiotic language. Our design of a gesture vocabulary will be directed toward the control of a specific object such as a fixed robot, mobile robot or a motorized camera.

In this paper a global approach to hand gesture vocabulary design is proposed. Suppose one has a set of commands of size *n*, then the problem is how to select a subset of *n* gestures from a set of all possible gestures in order to meet some performance measures such as; usability, accuracy, robustness, etc. Of course, the set of all physically possible hand gestures is infinite, so we will limit this set to manageable finite number. Further, in order to make the problem tractable 2D static hand gesture poses with a finite countable number of configurations are considered. These configurations can be defined by the finger positions (extended, spread) and palm orientations (up, down sideways). From this large set of candidate gestures one can select a subset of size *n* and pair them up with the set of *n* predefined commands. Given a master set of gesture configurations of size *m* the number of subsets of size *n* can be quite large. Furthermore, there are n! possible one to one assignments of gestures to commands. The task of finding an optimal assignment of command-gesture parings may be approached as Quadratic Assignment Problem (QAP) [11]. An improved annealing scheme for the QAP [12] is embedded in meta heuristic for solving the optimal gesture vocabulary design. The methodology described in this work will be flexible enough to handle single and multi-task and user independent and independent systems. Indices obtained from ergonometrics studies will be used (extracted from users) representing physcophisical aspects of users as well as machine recognition accuracy

In section II to follow we discuss some issues concerning the design of a gestural command vocabularies. In section III a problem definition and notation is given. Section IV describes performance measures, and architecture for our system solution procedures. Description of the solution methodology appears in section V. In VI construction and estimation of input sets and matrices are presented**.** In VII, three solution procedures are offered; (a) a restricted mathematical program embedded in a tree heuristic, (b) a genetic algorithm and (c) a multiobjective decision problem including a small example to illustrate the methodology. The final section provides conclusions.

II. Issues Concerning the Design of Hand Gesture Command Vocabularies

There is the theory of universality of gestures which is the belief that some gestures have standard cross-cultural meanings [13]. In reality, gesture meanings are very culturally dependent. Within a society or a culture most gestures have standard meanings, but some gestures may have different meanings to different individuals [14]. Device control gestures, however, can be freely chosen to have specific meanings related to the particular device [15]. For example, there is no universal known gesture for “go to the home position” or ”open the robot gripper”. There has been virtually no research concerned with the issue of how to design an optimal gesture based control vocabulary. The first step is to decide what commands are to be included in the vocabulary such as “move left”, “increase speed”, etc. The second step is to decide how to express the command in gestural form i.e.; what physical vocabulary to use such as; waving the hand left to right or making a "V" sign with the first two fingers.

A. Gesture Designers

Gesture vocabularies can be overtly or inadvertently designed. The thumps up and down signs come to us from Roman times whereas, the OK sign is more recent. Both can be considered as inadvertently designed or naturally evolved (emblems is the current term). More complete sign vocabularies appeared in this manner without overt determination of the vocabulary by a designer. As for overtly designed vocabularies most researchers merely present a vocabulary of gestures and start from there. We can consider this as the “Centrist Approach” where, a single individual decides which gesture vocabulary should be used for all users. Alternatively, we can define a “Consensus Approach” where a group of users decide on a common vocabulary to express a given set of commands. At the lowest level is the “Customized Approach” where each individual defines his or her very own vocabulary. The centrist approach, where the individual developing the system decides on the vocabulary is most common. One may hypothesize that the consensus and customized approaches will be more comfortable, easier to remember and more natural to execute. The disadvantage is that the users will not consider other design factors such as the speed and accuracy of recognition.

In summary the three subjective approaches to designing a gesture vocabulary are: (a). Authoritarian (the designer decides on the commands and gestures for all users). (b) Custom (the user selects his/her own set of gestures), and (c) Consensus (multiple users decide jointly on a set of common gestures).

B. Overt Gesture Design Methods

Current methods of gesture vocabulary design may be classified as: (a) ad hoc and subjective, (b) design rule based, and (c) analytic.

One of the few works that explore the process of gesture design is that of Long, et. al. [1]. The application is that of a pen-based user interface where, gestures are drawn marks or strokes that cause a command to be executed. A gesture design tool, (quill), advises designers on how to improve their gestures, and suggests a method for evaluating gestures for pen based applications. In a more recent work [16] a procedure and a benchmark to find gestures based on nine usability heuristics are presented. However, the important factor of vision recognition was ignored.

1) Ad Hoc Methods

Ad hoc methods are the prime method of determining a gesture vocabulary. They are not hard to find. Many examples prevail in the literature. Most are of the centrist type whereby an individual constructs the vocabulary mostly with no mention of the method or rational for the choices made. See for example [8][17][18][19][20][21].

2) The Rule Based Approach

The work of Baudel and Beaudouin-Lafon [22] provides an example of the use of design rules. For example they provide guidelines such as: Favor ease of Learning, Use hand tension at the start of a dynamic gesture and relaxed position of the hand at the end. Another is that of Baudel et. al. [23] who provides a set of guidelines for designing a gestural command set although no mention is made on how these guidelines are actually implemented to generate the actual vocabulary. The application allows a user to give a lecture by navigating through a set of slides with data glove based gestural commands. Kjeldsen and Hartman [24], in a vision based computer interaction setting, present a set of constraints for control actions defined as: "the motion user makes to effect control". Stating that "the choice of such control movements is more art than science" they proceed to consider what are good control actions for different task types. Again the approach is rather intuitive as opposed to being scientifically based.

3) Analytical Methods

Analytical methods are scientific based, involving perhaps the use of human factors aspects, ergonomics, hand biomechanics, cognitive science, experimental statistics, machine recognition and mathematics. Although, there exist sporadic works applying these disciplines to the hand gesture problem, we have found no work using analytical methods for the complete design of a *GV*. Thus, we believe this is the first comprehensive analytical method for design of an optimal gesture vocabulary.

Any analytical method needs performance criteria in the form of a stated objective. We propose the objective be the best performance to control an inanimate object or device to carry out a task using a given vocabulary of control gestures. The best performance may be the minimal time within the constraints of accuracy or tolerance to carry out a task. For example move a robot from A to B in a maze without hitting its walls. Or pick up a cup of coffee without spilling its contents. In order to meet such an objective we postulate the following factors that a good *GV* should posses in order to affect minimal time performance.

1. Easy to remember - If the gesture is natural and intuitive it will increase the rate of recall.
2. Easy to learn – If a gesture is intuitive and can be presented with ease (physically) it will be easier to learn.
3. Physically easy to perform. The gesture should not be such as to cause strain on the operator while holding the pose or during a transition between poses.
4. The gesture corresponds to its meaning or intent. For example, extending one finger represents the number one.
5. Easy to recognize. Gestures should be sufficiently different so that there is enough discriminatory power between them for the recognition system to classify them.

III. Problem Definition and Notation

The basic research problem here is to find an optimal hand *GV*. An optimal hand *GV* is defined as a set of gesture-command pairs, such that it will minimize the time τ for a given user (or users) to perform a specified task, or group of related tasks.

$$\underset{\forall G}{Min}\ \tau(GV) = \Psi(T,G,C,F,I,S,A) \tag{1}$$

Ψ is some function of the following factors:

$T = \{t_1, \ldots, t_n\}$, the set of tasks that can be performed in the current ontology.

$G = \{g_1, \ldots, g_m\}$, the set of all feasible hand gestures that can be evoked by the user.

$C = \{c_1,..,c_n\}$, the set of commands spanning all tasks in T.

$F_{n \times n}=\{f_{ij}\}$, the command transition matrix, or after normalization the stochastic matrix $P=\{p_{ij}\}$ of commands (where f_{ij} is the frequency of transition from command i to command j, and p_{ij} is the probability of using command i after command j).

$I_{n \times m}=\{a_{ik}\}$, the intuitiveness matrix, where a_{ik} is a measure of cognitive association between the gesture i and the command k.

$S_{m \times m}=\{s_{kl}\}$, the stress or fatigue matrix, where 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.

$\overline{S}_{m\times m}$ is the comfort matrix, some inverse function of S.

A = the recognition accuracy of a given subset of gestures (a scalar).

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.

$GV = \{(i, p(i) \mid i = 1,..,n\}$, the set of gesture-command pairs.

IV. Performance Measures and Architecture

A. Multi-Objective Performance Measure

The main performance measure is the completion time to

perform a task. However, since the measurement of task completion time involves the evaluation of the function, (1), and this function is unknown and time consuming to estimate from experimental data, we propose instead, analytical performance measures acting collectively as a proxy for completion time. These performance measures will represent, intuitiveness, comfort, and recognition accuracy, designated $Z_1(GV)$, $Z_2(GV)$ and $Z_3(GV)$, respectively. Each is shown as a function of the given gesture vocabulary, GV, in (2), (3), and (4) below. Let Γ represent the set of all feasible gesture vocabularies. Maximizing each of the measures over the set Γ defines a multiobjective decision problem. We note, that the analytical form of the objectives; $Z_1(GV)$, $Z_2(GV)$ and $Z_3(GV)$ are linear, quadratic and unknown, respectively. The objective functions $Z_1(GV)$ and $Z_2(GV)$ are human valued measures, while $Z_3(GV)$ is machine valued.

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:

$$Z_1(GV) = \sum_{(k,p(k))\in GV}^{n} a_{k,p(k)} \tag{2}$$

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.

$$Z_2(GV) = \sum_{[(i,p(i)),(j,p(j)]\in GV\times GV}^{n} f_{ij}\ \bar{s}_{p(i),p(j)} \tag{3}$$

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:

$$Z_3(GV) = A = \frac{(total\ gestures - gestures\ misclassified)}{total\ gestures}100 \tag{4}$$

B. Combined Performance Measure

$$Max\ Z(GV) = w_1 Z_1(GV) + w_2 Z_2(GV) + w_3 Z_3(GV) \tag{5}$$

w_i = the relative importance of factor $Z_i(GV)$.

Weights are used to reflect the relative importance of the three main objective ($Z_1(GV)$, $Z_2(GV)$, $Z_3(GV)$). The values of the weights can be found empirically, and involves a collection of experimental data through direct or indirect solicitation of decision maker's preferences. The use of the weights allows the three objectives to be mapped into the single valued objective, $Z(GV)$.

V. Description of the 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 measure.

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

(e) The weights are available (except in the case of problem P_4 in Section VII)

A. 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.

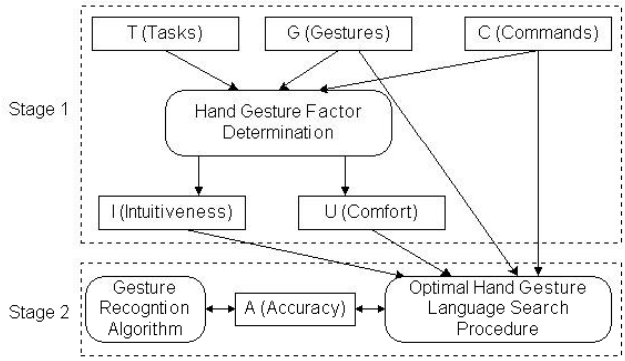


Fig. 1. Architecture of optimal hand gesture vocabulary methodology

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.

The inputs to the second stage are the calculated matrices; intuitiveness I, comfort U, and the sets command C, gesture G, and recognition value of the accuracy, A. This stage

employs a search procedure to find the best vocabulary *GV,* using three approaches: (a) a math program relaxed problem embedded within a heuristic search, (b) a genetic algorithm approach, and (c) a multiobjective decision problem. These three approaches are discussed in Section VII and all require the same inputs described in the next section.

VI. Construction and Estimation of Input Sets and Matrices (*T,C,F,G,I,U*)

The architecture of Stage 1, the Hand Gesture Factor Determination Stage, is shown in Figure 2.

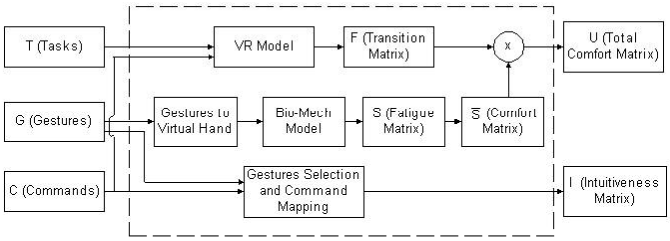


Fig. 2. Hand gesture factor determination stage

A. 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.

B. 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.

C. 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 labeled 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 gesture generator.

D. 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 (see Figure 3). 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. Using this information it is straightforward to construct an intuitiveness matrix, $I_{n\ x\ m}$. The entries of I are represented as a_{ik}.

$$a_{ik} = \frac{n_{ik}}{N_i} \quad ,i = 1,...,n \quad , \quad k = 1,...,m \tag{6}$$

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 ith command.

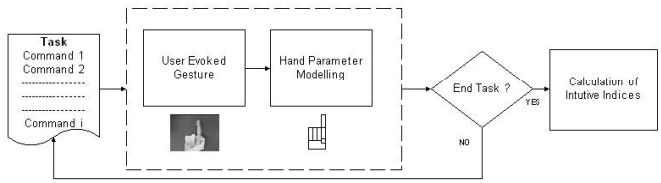


Fig. 3. The application for empirically determining the intuitive indices

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

The fatigue (or comfort) indices are determined through the use of a biomechanical model of the hand. A gesture and a binary vector representing the encoded vector are described in Figure 4. Each virtual gesture includes information of the finger and hand configuration and orientation. The user's gestures are encoded using a hand parameter model,

$\varphi = \{f, r, s\}$.

Let $f = \{f_1 , ..., f_5\}$ represent the finger flex states, starting from the thumb until the little finger.

$r = \{r_1 , ..., r_4\}$ represents the spreading states between fingers, starting from the space between thumb and index, until ring and little finger.

s = represents the status of the palm (up or down).

The finger states f_i can be straight or flexed (1 or 0), the spreading states r_i can be open or close (1 or 0) and the status of the palm can be down or up (1 or 0). Using this encoding method, the space of all possible gestures Q is $2^{|\varphi|}$.

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

element of the matrix S, $(\bar{s}_{ij} = \delta(s_{ij}))$.

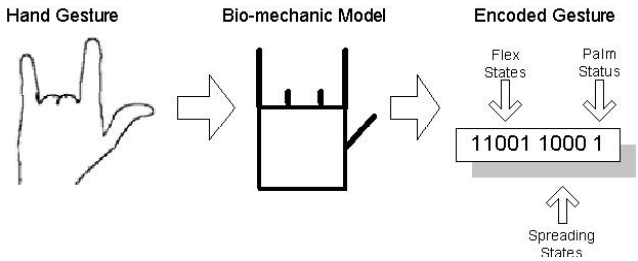


Fig. 4. Encoded gesture schema

*F. 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 $u_{ijkl} = f_{ij} \times \bar{S}_{kl}$ means the frequency of transition between commands i to j times the comfort of transition between gestures k to l when commands i and j are associated with gestures k and l respectively in a given GV. This product reflects the concept that the comfort measure of GV depends on the frequency of use of the gesture or a gesture pair transition.

VII. SOLUTION PROCEDURES

Recall the measure $Z(GV)$ is a weighted combination of, (a) $Z_3(GV)$, the recognition accuracy of a machine vision algorithm for the hand gesture set used, (b) $Z_2(GV)$, the comfort to perform a gesture or a transition between gestures and, (c) $Z_1(GV)$, the intuitiveness of representing each command by it associated gesture. Three solution approaches will be described.

(a) Relaxed math program embedded in a heuristic search, and

(b) A genetic algorithm approach, and

(c) Multiobjective Decision Problem

As all three will include the same gesture recognition algorithm this will be described first.

A. Gesture Recognition Algorithm

The hand gesture recognition process will be vision based comprised of 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 [25] 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 grayscale partition blocks will be created from a hand silhouettes. 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 (4). This result indicates the recognition ability of the system for the given set of hand gestures. Further details may be found in [26].

B. Math Programming Relaxed Problem P1:

Since the recognition accuracy is not dependent on the matched command-gesture pairs in the gesture vocabulary, but only on the set of gestures, we place the third objective function of (5) in the constraints to obtain the relaxed problem P_1.

Problem P_1

$$Max\ \bar{Z} = w_1 Z_1(GV) + w_2 Z_2(GV) \tag{7}$$

s.t.

$$Z_3(GV) \geq A_{\min} \tag{8}$$

where A_{min} , is the minimal acceptable accuracy.

The solution procedure will entail solving the relaxed problem P_2 that is P_1 without (9). The relaxed problem can be formulated as a quadratic integer assignment problem (QAP) [11]. Given a set $N = \{1,2,..,n\}$, whose indices represent commands or gestures, and $n \times n$ matrices; $F = (f_{ij})$, $U = (u_{kl})$, $I = (a_{ik})$:define problem $P_2(G_n)$ below.

Problem P_2: (G_n)

$$\max_{p \in \Pi_N} \bar{Z}(GV) = \sum_{i=1}^{n}\sum_{j=1}^{n} f_{ij}\ \bar{s}_{p(i)p(j)} + \sum_{i=1}^{n} a_{i\,p(i)} \tag{9}$$

Here, Π_N is the set of all permutations in N, *and t*he comfort cost of the pair of assignments (i, k) and (j, l) (assigning command i to gesture k and command j to gesture l) is $f_{ij}\ \bar{s}_{kl}$. The intuitiveness of the assignment (i, k) is a_{ik}. [1].

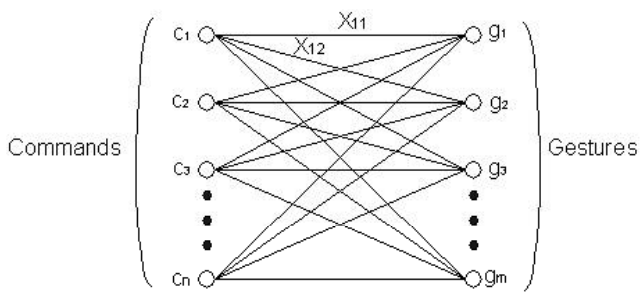


Fig. 5. Network representation underlying the quadratic assignment problem

By defining a set of integer 0,1 decision variables $\{x_{ij}\}$ a quadratic assignment problem QAP(G_n) can be formulated which is equivalent to Problem P_2. Here, G_n represents a gesture set of size n. A network representation of the problem is shown in Figure 5. However the following formulation is

[1] Here we assume $w_1=w_2=w_3=1$

more general as it is described for the master set of gestures G_m, of size m ($m>n$).

Problem P_3: QAP(G_m)

$$\max \overline{Z}(GV) = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{m} u_{ijkl} x_{ik} x_{jl} + \sum_{i}^{n}\sum_{j}^{m} a_{ij} x_{ij} \tag{10}$$

s.t.

$$\sum_{j=1}^{m} x_{ij} = 1, \qquad i = 1,..,n, \tag{11}$$

$$\sum_{i=1}^{n} x_{ij} \le 1, \qquad j = 1,...,m, \tag{12}$$

$$x_{ij} \in \{0,1\}, \qquad i = 1,...,n, \qquad j = 1,...,m., \tag{13}$$

where $U_{ijkl} = f_{ij}\overline{S}_{kl}$

Here, x_{ij} is a binary assignment variable equal to 1 if command i is assigned to gesture j, and zero otherwise. Constraint (11) insures that each command is paired with exactly one gesture. Constraint (12) insures that each gesture can be paired with no more than one command.. Since the decision variables are binary, a subset of exactly n gestures will be selected from the master set of size m.

A possible approach to the solution of the QAP is to find a proper linearization [26]. If the quadratic term of (9) disappears, then an ordinary linear assignment type problem is obtained. Also, the QAP can be relaxed to a (0, 1) linear integer program by introducing new binary variables $y_{ijkl}=x_{ij}x_{kl}$ and new constraints. Using the new variables, the quadratic term in the objective function can be replaced. In this work simulated annealing is used.

C. Heuristic Search Algorithm for Problem P1

An iterative technique can be outlined for solving Problem P_1. We start with P_3, the unconstrained P_1 problem. Once the quadratic integer problem P_3 is solved the gesture recognition is checked to see if the constraint (8) is satisfied. If so this process is finished, otherwise a new subset of gestures G of size n is selected from the master set G_m.

For the case where a master set of gestures G_m ($m>n$) is used, a subset G_n of size n must be selected. This may be done by using m gestures nodes in the network of Figure 5 and solving the associated problem $QAP(G_m)$. This determines the initial subset of gestures used in the feasibility algorithm described below. The following algorithm (FA) searches for a feasible solution to P_1, for a fixed accuracy level A_{min}. This provides an upper bound on $\overline{Z}(GV)$. If a feasible solution is found the threshold accuracy level may be increased to $A_{min} = A$ and FA run again.

ALGORITHM FA HEURISTIC SEARCH FEASIBLE ALGORITHM

1: input: G_m, G_n, A_{min}, bt
/* bt is the solution binary tree*/
2: $GV \leftarrow$ **Solve**(P_3:QAP(G_n))
3: [A, $\mathcal{M}$] ← **GetAccuracy**(G_n)
/* Using (4) obtain confusion matrix $\mathcal{M}$ */
4: **if** $A \ge A_{min}$ **then**
5: **return** $Z(GV)$
/* Use (7) and (4). This is a feasible solution*/
6: **else**
7: [i , j] ← **FindC$_{ij}$**($\mathcal{M}$)
/* Use (14) to get the most confused gestures*/
8: g ← **GetGesture**(i , G_n)
9: $G_n^1 \leftarrow G_n - \{g\}$
10: g_1 ← **MinSimilarity**(g, G_m -G_n)
11: $G_n^1 \leftarrow G_n \cup \{g_1\}$
12: bt ← **InsertLeaf**(bt, G_n^1)
13: **Call FA**(G_m, G_n^1,A_{min},bt)
14: g ← **GetGesture**(j , G_n)
15: $G_n^2 \leftarrow G_n - \{g\}$
16: g_2 ← **MinSimilarity**(g , G_m -G_n)
17: $G_n^2 \leftarrow G_n \cup \{g_2\}$
18: bt ← **InsertLeaf**(bt, G_n^2)
19: **Call FA**(G_m, G_n^1,A_{min},bt)

A binary tree of several QAP programs is formed (Figure 6), in which each problem has the same constraints and objective as Problem P_3 (10-13), except for the new set G_n.

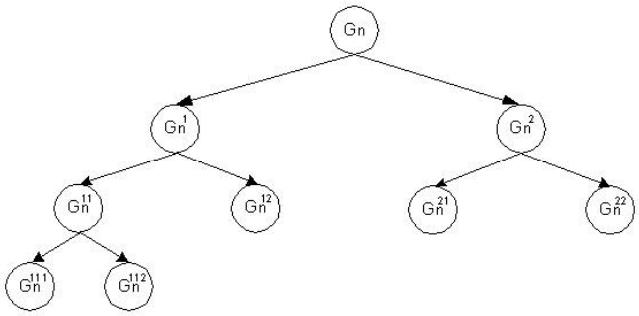


Fig. 6. Solution tree

The solution of the root problem will yield to recognition accuracy A, which can be higher or less than variable recognition accuracy threshold A_{min}. If it is higher then problem P_3 is feasible. If it is desired to increase the accuracy threshold then A_{min} is set to A and the FA problem is resolved. If it is not higher, the confusion matrix is studied, and the two gestures with the maximum confusion values are candidates to change. The confusion value between a pair of gestures gesture i and j is defined as:

$$C_{ij} = \frac{n_{ij}}{N} \tag{14}$$

where,

C_{ij} = the level of confusion between gesture i and j.

n_{ij} = the number of times gestures i is recognized as gesture j.

N = the total number of gestures.

This gives rise to two sub-problems. The left-child problem and the right-child problem, where the first and the second confused gestures are replaced by new ones, to form, two new $QAP(G_n)$ problems. Reasonable good solutions to the QAP, in Problem P_2, can be obtained through Simulated Annealing [12]. This branching process continues where each of the two new problems will give rise to two more problems. Suppose the two new problems G_n^1, G_n^2 are constructed from G_n. Let i', j' be the pair of the most confused gestures determined by $Max\ C_{ij}=C_{i'j'}$. Then in G_n^1 gesture i' is retained, and replaced by the most dissimilar unused gesture j'' found in the master set. It follows from this construction that the tree obtained is binary. The exit condition is when the recognition accuracy of one of the sub-problems is higher than the specified accuracy level A_{min}, then the gesture-command mapping is the selected solution. Figure 7 is a flow chart of this procedure.

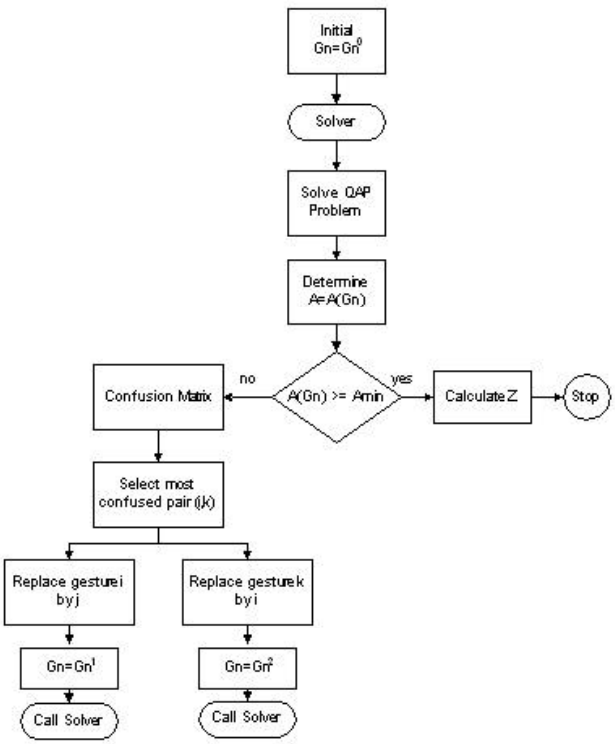


Fig. 7. Flowchart of heuristic search algorithm for problem

D. Genetic Algorithm Approach

The genetic algorithm (GA) uses the principles of biological evolution and was developed by Holland [28], GAs have been used for the solution of large combinatorial optimization problems [27]. A population of initial feasible solutions is created randomly. Each individual is encoded as a chromosome of length n which represents the associations between commands $C=\{c_1,\ldots, c_n\}$ and gestures $G=\{g_1,\ldots, g_n\}$. Figure 8 illustrates the chromosome representation.

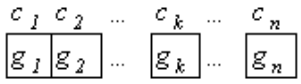


Fig. 8. A chromosome representation

The value of the gene g_k varies from 1 to m (where the master gesture set G_m is of size m), and represents the gesture to be paired with the command c_k. Each chromosome encoding represents a GV, of size n. The fitness of a chromosome is computed using (5).

The GA selects matching pairs of individuals according to proportional selection, and children chromosomes are generated based on a cross over operation. The cross over operation consists of taking two parent chromosomes and performing a single point crossover. After crossover a "bad" individual may be generated. Those are individuals that have the same gesture for different commands. Such individuals will be given low fitness values to avoid their propagation. After the mutation operation the next generation is obtained. The processes of crossover and mutation are used in an attempt to explore and exploit the solution space. Mutation consists of occasionally introducing random bit changes using a mutation probability value. The process is run for a few hundred generations (iterations) or when the algorithm converges and hopefully by then, the fittest genes will dominate, and a local maximum for $Z(GV)$ is found. Genetic algorithms can succeed where traditional methods may fail. However, since GA's are of probabilistic in nature the optimal solution is not guaranteed.

E. Multi-Objective Approach

In the previous approach the weights were given by the decision maker to map the three multi objectives into one metric. Here we consider all three objectives separately as the multi-objective decision problem, designated problem P_4.

Problem P_4

$$\max Z_1(GV) = \sum_i^n \sum_j^m a_{ij} x_{ij} \tag{15}$$

$$\max Z_2(GV) = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^m \sum_{l=1}^m u_{ijkl} x_{ik} x_{jl} \tag{16}$$

$$\max Z_3(GV) \tag{17}$$

s.t.

$$\sum_{j=1}^m x_{ij} = 1, \qquad i = 1,..,n, \tag{18}$$

$$\sum_{i=1}^{n} x_{ij} \le 1, \qquad j = 1,...,m, \tag{19}$$

$$x_{ij} \in \{0,1\}, \qquad i = 1,...,n, \qquad j = 1,...,m., \tag{20}$$

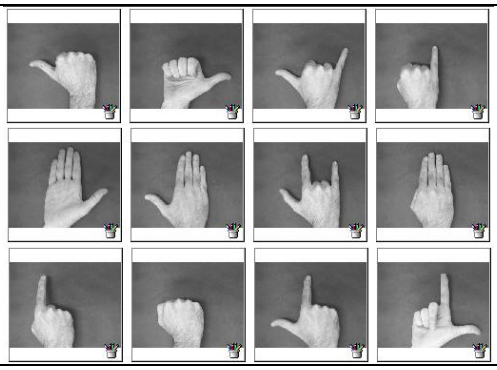
where $U_{ijkl} = f_{ij}\bar{S}_{kl}$

This problem is related to the Multiobjective quadratic assignment problem [29] were it not for the third objective $Z_3(GV)$ which has no explicit analytical form. To evaluate $Z_3(GV)$ the recognition algorithm must be called, and solved for the particular solution 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 performance on one cannot be improved without sacrificing performance in at least one other. Such solutions are called Pareto optimal points [30], 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. As an illustration of the solution procedure we consider a small example of set of twelve gestures and eight commands (see Table I).

Table I. Hand Gesture Vocabulary

Commands	Gestures
LEFT RIGHT FORWARD BACK FAST SLOW START STOP	

For this problem we devise a system to display the three objective values in a 3D plot, which will allow the decision maker to examine different solutions and select the one he desires, based on his internalized weighted priorities for each of the objectives. (see Figure 9) To aid the decision maker we also provide the Pareto points

For each command-gesture set, (C, G_n) the optimal 1-1 mapping is obtained using simulated annealing to solve the quadratic program P_3. Thus 495 GVs are acquired. For each GV, the three objective measures: Intuitiveness ($Z_1(GV)$), Comfort ($Z_2(GV)$), and Accuracy ($Z_3(GV)$) are evaluated (see previous section). Each GV is thus represented as a point in a 3D space whose coordinates are; intuitiveness, comfort, and accuracy. After applying this algorithm to the example, three Pareto points were obtained (see Table II). The Pareto points are displayed in bold in Figure 9.

Table II. Optimal Pareto points for the multi-objective problem

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

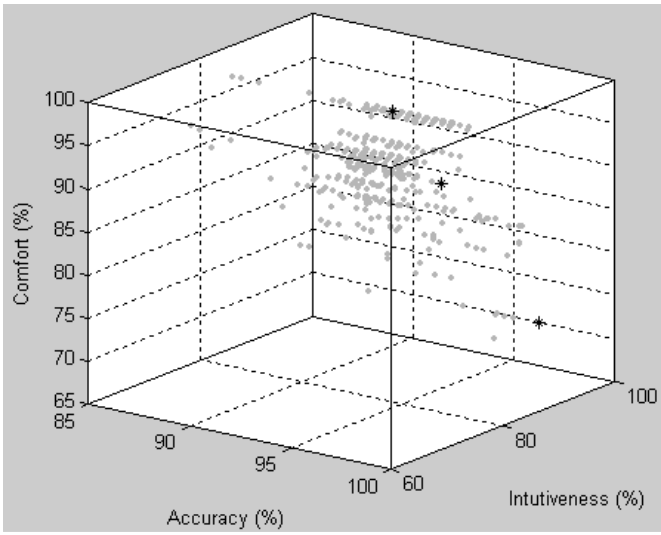


Fig. 9. 3D Tool to assist the decision maker

VIII. Conclusion

In this research we attack the difficult problem of designing a hand gesture control vocabulary. A hand gesture vocabulary is used to command robotic manipulators to carry out specific tasks. We review and classify three approaches to this problem: Ad hoc, Rule based and Analytical. We believe this is the first conceptualization of the optimal hand gesture design problem in analytical form. To solve the problem, a non-trivial abstract representation of a natural mapping between gestures and commands and relationships between them is introduced. Three mathematical models are developed, which reflect the ergonomic and technical performance measures upon which a gesture control system is judged. The first, is a mathematical program using a quadratic assignment problem embedded within a heuristic search tree. The quadratic problem, is used to select a gesture vocabulary (a command-gesture matching) through simulated annealing, The second, is a genetic algorithm representation, and the third is a multi-objective decision problem for which a 3D representation of the solution space is used to display candidate solutions, as well as, the Pareto optimal ones. A computational example is given for the design of a small command gesture vocabulary for the multiobjective procedure.

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] A. C. Long, J. A. Landay, and, L. A Rowe, "Implications for a gesture design tool," in *CHI 1999 ACM Conference on Human Factors in Computing Systems*, *CHI Letters*, vol. 1, no.1, pp. 40-47.

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

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

[4] 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.

[5] 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.

[6] 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.

[7] A. Katkere, E. Hunter, D. Kuramura, J. Schlenzig, S. Moezzi and R. Jain, "ROBOGEST: Telepresence Using Hand Gestures," Tech. Rep. VCL-94-104, University of California, San Diego, 1994.

[8] D. Kortenkamp, E. Huber, and R. P. Bonasso, "Recognizing and Interpreting Gestures on a Mobile Robot," in *AAAI96*, 1996.

[9] C. Cadoz, *Les realites virtuelles*, Dominos, Flammarion, 1994.

[10] D. McNeill, *Hand and Mind – What Gestures Reveal About Thought*. The University of Chicago Press. Paperback Edition, 1995.

[11] T. C. Koopmans and M. J. Beckmann, "Assignment problems and location of economic activities," *Econometrica*, no. 25, pp. 53-76, 1957.

[12] D. T. Connolly, "An improved annealing scheme for the QAP," *European Journal of Operational Research*, no. 46, pp. 93-100, 1990.

[13] R. Aboudan, and G. Beattie, "Cross-cultural similarities in gestures. The deep relationship between gestures and speech which transcends language barriers," *Semiotica*, no. 111, pp. 269-94, 1996.

[14] D. Archer, "Unspoken Diversity: Cultural Differences in Gestures." Special Issue on "Visual Sociology," *Qualitative Sociology,* vol. 20, no. 1, pp..3-137, 1997.

[15] C. Cohen. (1999, Feb. 10). *A Brief Overview of Gesture Recognition*, [Online]. Available: http://www.dai.ed.ac.uk/CVonline/local_copies/cohen/gesture_overview.html

[16] M. Nielsen, M. Störring, T. B. Moeslund, and E. Granum, "A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI," in *5th Int. Workshop on Gesture and Sign Language based Human-Computer Interaction (GW2003)*, Genova, Italy, 2003.

[17] S. Waldherr, S. Thrun, R. Romero, and D. Margaritis, " Template-based recognition of pose and motion gestures on a mobile robot," in *Proceedings of the AAAI Fifteenth National Conference on Artificial Intelligence*, pp. 977-982, 1998.

[18] M. Becker et al, "GripSee: A gesture-controlled robot for object perception and manipulation," *Autonomous Robots*, vol. 6, no. 2, pp. 203-221, Apr. 1999.

[19] T. Agrawal and S. Chaudhuri, "Gesture recognition using position and appearance features," in *Proc. IEEE Int. Conf. Image Processing*, Barcelona, Spain, 2003.

[20] 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.

[21] C. W. Ng and S. Ranganath, "Real-time gesture recognition system and application," *Image and Vision Computing*, vol. 20, no. 13-14, pp. 993-1007, Dec. 2002.

[22] T. Baudel, M. and Beaudouin-Lafon, "Charade: Remote Control of Objects using FreeHand Gestures," *Communications of the ACM*, vol. 36, no. 7, pp. 28-35, 1993

[23] T. Baudel, M. Beaudouin-Lafon, A. Braffort, D. and Teil, "An interaction model designed for hand gesture input," LRI Res. Rep. 772, Sept. 1992.

[24] R. Kjeldsen and J. Hartman, "Design Issues for Vision-based Computer Interaction Systems," in *Proc. of the Workshop on Perceptual User Interfaces*. Orlando, Florida, USA, 2001.

[25] 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.

[26] 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.

[27] G. Finke, R. E. Burkard, and F. Rendl, "Quadratic assignment problems," *Annals of Discrete Mathematics*, no. 31, pp. 61-82, 1987.

[28] J. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, 1975.

[29] J. D. Knowles, and D. W. Corne, "Instance Generators and Test Suites for the Multiobjective Quadratic Assignment Problem," in *Proc. Evolutionary Multi-Criterion Optimization (EMO 2003) Second International Conference*, Faro, Portugal, pp 295-310, 2003.

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