A Holistic Framework for Hand Gestures Design

Juan Wachs[1], Helman Stern[2] and Yael Edan[2]

[1]Department of Computer Science
Naval Postgraduate School, 700 Dyer Road
Monterey, CA, 93943-5001, USA
jpwachs@nps.edu
[2]Department of Industrial Engineering and Management,
Ben-Gurion University of the Negev, Beer Sheva, 84105, Israel
{helman,yael}@bgu.ac.il

Abstract

Hand gesture based interfaces are a proliferating area for immersive and augmented reality systems due to the rich interaction provided by this type of modality. Even though proper design of such interfaces requires accurate recognition, usability, ergonomic design and comfort. In most of the interfaces being developed the primary focus is on accurate gesture recognition.

Formally, an optimal hand gesture vocabulary (GV), can be defined as a set of gesture-command associations, such that the time τ to perform a task is minimized over all possible hand gestures in our ontology. In this work, we consider three different cost functions as proxies to task completion time: intuitiveness $Z_1(GV)$, comfort $Z_2(GV)$ and recognition accuracy $Z_3(GV)$. Hence, we can establish that $Max(Z_{ij})(GV)$: i=1,2,3) over all GV's is our multiobjective problem (MOP).

Because finding the solutions to the MOP requires a large amount of computation time, an analytical methodology is proposed in which the MOP is converted to a dual priority objective problem where recognition accuracy is considered of prime importance, and the human performance objectives are secondary.

This work, as opposed to previous research done by the authors, is focused on two aspects: First,a modified cost function for an enhanced simulated annealing approach is explained and implementation issues are discussed. Second, a comparative study is performed between hand gesture vocabularies obtained using the methodology suggested, and vocabularies hand picked by individuals.. The superiority of our method is demonstrated in the context of a robotic vehicle control task using hand gestures.

1. Introduction

The majority of human machine interfaces for everyday device control aims for affordable prices while mimicking realistic natural interactions. This type of interface can be activated by voice, face, hand and body posture recognition algorithms. Most of them are designed to achieve a high recognition performance while allowing the user to interact with the systems similar to interactions with another human. Human-robot interaction was exploited in [1] in the context of ambient intelligence (intelligence algorithms involving measurement, transmission, modeling, and control of environmental information) for human detection and gesture recognition. Hand detection and pose recognition was achieved in [2] through an infra-red time-of-flight range camera. The author's interface system was able to recognize 7 DoFs of a human hand with a 2-3 Hz frame rate. In [3] a method for recognizing hand gestures using depth image data acquired from active vision hardware was suggested. The authors are able to recognize different static poses while tracking the hand in real time. The authors were motivated by the development of an interface to control home appliances. A notable work in home appliance control was done in [4]. The authors propose an universal remote control system based merely on hand gestures. In this system, the user first selects the device to be controlled by pointing it with his hand. Then, the user operates it through 10 predefined basic hand motions. Marcel [5] developed a system that combined face tracking with hand gestures recognition based on face location and body anthropometry. This system was capable of recognizing a five gesture vocabulary in uniform and cluttered environments, however no applications were suggested for such an interface. Dynamic gesture recognition to drive mobile phone applications was developed by [6] based on accelerometers attached to the mobile phone. The authors present a proof of concept of their system through a "navigate and select" application, such as Google Earth. Wireless communication is also used by [7] for voice and real-time continuous sign language recognition. The authors implemented their system in a post-wearable PC in the domain of ubiquitous computing applications

which allowed the users to freely move with their portable terminal while naturally interacting with the embedded-ubiquitous environment. A hybrid system capable of using information from faces and voices to recognize people's emotions was developed in the PHYSTA project [8]. In [9] a system was developed capable to incrementally learn to recognize affective states from body postures for human-robot interaction.

Two types of hand gesture interfaces have been distinguished in human-machine interaction according to their objective: the first is designed to cope with the challenge of hand gesture recognition with a high accuracy and speed, and the other is focused on the ergonomic aspects of the hand gesture vocabulary design. In all the research presented above enormous efforts were invested in achieving the first objective, (technical focused), but the second objective was not addressed. Understanding the user's physiologic and cognitive needs is one of the key tasks associated with an efficient and natural hand gesture based interface. To tackle that task, machine vision and analysis techniques have to be developed, while, at the same time, psychological and linguistic analyses of hand gestures must be considered. An example of the second type of hand gesture interfaces can be found in [10], where intuitive hand gestures are selected in a fashion that allows the user to act more naturally since no cognitive effort is required in mapping function keys to robotic hand actions. This system, like others, is based in navigation control. The author's selection of gestures can be criticized since their interpretation of an intuitive gesture-command association may not suit others' cognitive perception of intuitiveness. This issue was addressed in [11] where it was found that people consistently used the same gestures for specific commands. In particular they found that people are also very proficient at learning new arbitrary gestures. In [12] it was found that test subjects used very similar gestures for the same operations. All these may indicate that there may be intuitive, common principles in gesture communication. A notable work discussing an ergonomic based approach for hand gestures design is presented in [13], where the comfort associated with specific gestures when they are performed rapidly and repeatedly is considered. The authors conclude that designers of gesture languages for computer input should minimize the use of those hand gestures associated with upper extremity discomfort. Also in [14] a similar conclusion is achieved. The authors used a biomechanics based objective function to reflect comfort in the framework of a hand gesture based interface.

Previously, in [15] and [16] we have shown a methodology for the design of a gesture vocabulary that is both intuitive and comfortable on the one hand, and can be recognized with high accuracy, on the other. A two-step procedure for solving the gesture vocabulary design problem was introduced. This procedure was formulated as a multiobjective optimization problem (MOP). The first step is to decide on a task-dependent set of commands 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 gesture form i.e., what physical expression to use, such as waving the hand left to right or making a "V" sign with the first two fingers. The association (matching) of each command to a gesture expression is defined here as a "gesture vocabulary" (GV). In this paper, the gesture-command matching algorithm based on simulated annealing is discussed for the first time, and new experiments comparing "human hand picked" vocabularies with automated hand gesture vocabulary design are presented.

In the next section the GV design problem is defined. This is followed in section 3 by a description of the main methodology; comprised of hand gesture factor determination, gesture subset selection, command-gesture matching, and selection of pareto optimal multiobjective solutions. In section 4, the extended simulated annealing approach to solve the optimal gesture-command association is presented. Section 5 compared our automated approach with human selected GVs. Section 5 provides conclusions.

2. Problem Statement

A suitable definition of a hand gesture vocabulary (GV) is the set of gesture-command pairs that minimizes the time τ required for a user/s to perform a task/s. The number of commands is fixed, and determined by the given task. The set of gestures G_n is obtained from a large set of postures; a "master-set" of gestures, denoted by G_m.
Three performance measures are used as proxies for the task completion time τ. intuitiveness $Z_1(GV)$, comfort $Z_2(GV)$ and recognition accuracy $Z_3(GV)$. The first two measures are related to ergonomic side, while the last is strictly related to the technological aspect.

The problem is to find a GV that maximizes the proxies, achieving a minimal performance time, over all feasible gesture vocabularies, Γ. This multi-objective problem (MOP) is complex given that the performance time is not a well-behaved function on the proxies manifolds. Moreover, there exist conflicting solutions where all the objectives cannot be maximized simultaneously. This can be overcome by allowing the decision maker to select the best GV according to his own preferences.

$$\underset{GV \in \Gamma}{Max}\ Z_1(GV), Max\ Z_2(GV), Max\ Z_3(GV) \qquad (1)$$

Let us define Z1, the intuitiveness, of the GV as the naturalness of expressing a given command with a gesture. We recognize two types of intuitiveness: direct and complementary.

Let p be defined as an assignment function where p(i)=j indicates that the command i is assigned to gesture j. Consequently, the direct intuitiveness, $a_{i,p(i)}$ is expressed by the strength of the association between command i and its matched gesture p(i). Following the same concept, complementary intuitiveness, $a_{i,p(i),j,p(j)}$ is the level of association expressed by the matching of complementary gesture pairs (p(i), p(j)) to complementary command pairs (i,j). The total intuitiveness is shown in (2).

$$Z_1(GV) = \sum_{i=1}^{n} a_{i,p(i)} + \sum_{i=1}^{n}\sum_{j=1}^{n} a_{i,p(i),j,p(j)} \quad (2)$$

Let us define Z_2 as the Stress/Comfort needed to perform a gesture. Obviously, there are gestures that are easier to perform than others. Total stress is a scalar value equal to the sum of the individual stress values to hold the postures, and to perform transitions between them, weighted by duration and frequency of use.
Thus s_{kl} is the physical difficulty of a transition between gestures k and l. The duration to reconfigure the hand between gestures k and l is represented by d_{kl}. The symbol f_{ij} stands for the frequency of transition between commands i and j. The value K is a constant and is used to convert stress into its inverse measure comfort.

$$Z_2(GV) = K - \sum_{i=1}^{n}\sum_{j=1}^{n} f_{ij}\, d_{p(i)p(j)} s_{p(i),p(j)} \quad (3)$$

Accuracy is a measure of how well a set of gestures can be recognized. To obtain an estimate of gesture accuracy, it is necessary to train a gesture recognition system on a set of sample gestures for each gesture in G_n. The number of gestures classified correctly and misclassified is denoted as Tg and Te, respectively. The gesture recognition accuracy is denoted by (4).

$$Z_3(GV) = \left[\left(T_g - T_e\right)/T_g\right]100 \quad (4)$$

3. Main Methodology

One approach to solve (1) is find the performance measures over the set of all feasible GVs (complete enumeration). This approach is untenable, for even reasonable size vocabularies, thus, a dual priority objective optimization, (where recognition accuracy is considered of prime importance, and the human performance objectives are secondary) is proposed as a more tractable approach. Lets combine the intuitive and comfort objectives into one objective $\bar{Z}$ using weights w_1 w_2, and let A_{min}, be the minimum acceptable accuracy. Then we obtain:

$$Max\bar{Z}(GV) = w_1 Z_1(GV) + w_2 Z_2(GV) \quad (5)$$

$$GV \in \Gamma$$

$$s.t.\quad Z_3(GV) \geq A_{min} \quad (6)$$

The architecture of the solution methodology is comprised of four modules (Fig. 1). In Module 1 human psycho-physiological input factors are determined. In Module 2 gesture subsets, satisfying (6) are determined; Module 3 constitutes a command - gesture matching procedure. Finally, the set of Pareto optimal solutions is found in Module 4.

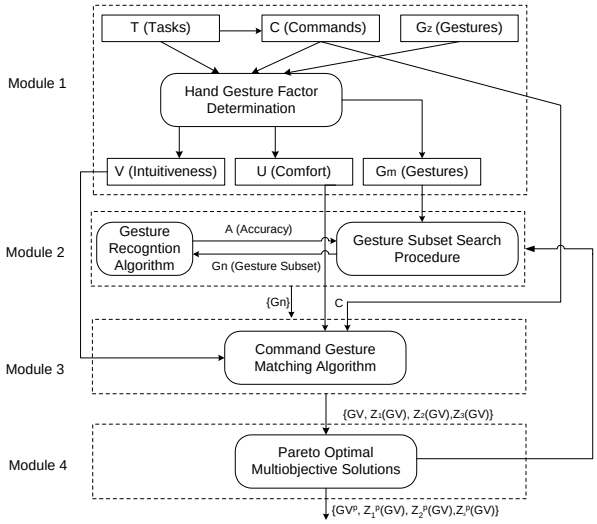


Figure 1. Architecture of optimal hand gesture vocabulary solution procedure

3.1. Module 1: Hand Gesture Factor Determination

The input parameters to the Module 1is the task set T, a large gesture master set G_z and the set of commands C. The procedure to obtain the intuitiveness V, comfort U, and gesture G_m, matrices is explained in [14]. For each task t_i , a set C of c_i commands are defined, as the union of all the task commands. Given the sequence of commands needed to complete a task the command transition matrix (F) is computed. The f_{ij} entries in F, represent the frequency that a command c_j is evoked given that the last command was c_i.

Since the set of all possible gestures is infinite, we established a set of plausible gesture configurations based on an articulated model including finger positions (extended, spread), palm orientations (up, down

sideways), and wrist rotations (left, middle, right) as the primitives, see Figure 2. Moreover, the gesture set is further reduced by considering the normalized popularity of gesture among the users. This final set is called the Gesture Master Set (G_m).

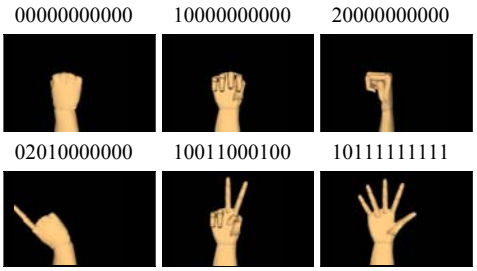


Figure 2. Articulated hand gesture model

Once the gesture set is reduced, the intuitiveness matrix I can be obtained. The entries of this matrix a_{ik} represent the naturalness of using gesture i for command k. In the same fashion, the complementary intuitive matrix (I') is attained, where the entry a_{ijkl} express the naturalness of matching up a pair of complementary commands (i, j) with a pair of complementary gestures (k,l). Denote V=[I, I'] as the set of matrices including both the direct and complementary matrices.

The fatigue (or comfort) indices are arranged in a matrix (S) where the element s_{ij} represents the physical difficulty of performing a transition from gesture i to gesture j. An entry u_{ijkl} in the comfort matrix (U) is defined as K-f_{ij} x s_{kl} where the last term represents the frequency of transition between commands i to j times the stress of a command transition k to l given that i and j are paired with gestures k and l, respectively.

3.2. Module 2: Gesture Subset Selection

The inputs of Module 2 are the reduced master set of gestures G_m, and a recognition algorithm to determine A. An iterative search procedure to find a set of gesture subsets $\{G_n\}$ is used in this module, to satisfy a given accuracy (6). The subset search procedure is based on the properties of the confusion matrix of the multi gesture recognition algorithm, and is called: Confusion Matrix Derived Solution Method (CMD) [14].

The CMD method consists of three steps: (i) train the recognition algorithm for the gestures in G_m , and let C_m be the resulting confusion matrix. The confusion matrix is obtained directly from the partition result of the training set using a supervised FCM optimization procedure, [17], (ii) find a submatrix C_n from C_m such that the recognition accuracy is highest, equal or below A_{min} (6), and (iii) repeat (ii) until a given number of solutions are found.

The CMD algorithm obtains $\mathcal{N}$ solutions (or all the solutions with associated accuracy above a given minimum allowed A_{min} if less than $|\mathcal{N}|$). Each iteration of the CMD algorithm generates a new solution by excluding each time a different gesture, from the subset of gestures of the current solution, and adding a new gesture from the master set. The number of solutions $|\mathcal{N}|$, is determined by the number of GV's that we want to consider based on the three measures Z_1, Z_2 and Z_3. This is usually specified by the decision maker.

3.3. Module 3: Command-Gesture Matching

The inputs to the third module are the intuitiveness V and comfort U matrices, the command set C, and the subset of gestures G_n. The purpose of this module is to match the set of gestures G_n to the set of commands, C, such that the human measures are maximized. The resulting gesture-command assignment constitutes a gesture vocabulary, GV.

Given a single set of gestures $G_n \in \mathcal{N}$ found from module 2, the gesture-command matching can be represented as a quadratic integer assignment problem (QAP) [8] and is formulated in (7)-(10).

$$\max \bar{Z}(G_n^*) = w_2 \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} u_{ijkl} x_{ik} x_{jl} + w_1 \left[\sum_{i}^{n}\sum_{j}^{n} v_{ij} x_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} v_{ijkl} x_{ik} x_{jl} \right] \quad (7)$$

$$\sum_{j=1}^{n} x_{ij} = 1, \quad i = 1,..,n, \quad (8)$$

$$\sum_{i=1}^{n} x_{ij} = 1, \quad j = 1,...,n, \quad (9)$$

$$x_{ij} \in \{0,1\}; \quad i = 1,...,n, \quad j = 1,...,n, \quad (10)$$

Let x_{ij} be the binary assignment variable. x_{ij} is equal to 1 if command i is assigned to gesture j, and zero otherwise. Equation (8) constraints each command to be matched with exactly one gesture. Equation (9) constrains each gesture to be matched with exactly one command. An enhanced simulated annealing is adopted to solve the QAP and it will be described in Section 4.

For each subset G_n found on Module 2, the QAP is solved by varying the weights such that w1+w2=10. This results in a set of GV solutions corresponding to each G_n in $\mathcal{N}$.

3.4 Module 4: Pareto Optimal Multiobjective Solution

Each of the $\mathcal{N}$ solutions (gesture subsets G_n) from Module 2, can result in $\mathcal{M}$ derived solutions. Each combination of the weights, for a given G_n, results in a new solution GV.

Thus a total of $\mathcal{N}$x$\mathcal{M}$ candidate GV's solutions are expected. Each of these solutions may be represented as a point in 3D space, (Z_1,Z_2,Z_3). The total set of multiobjective candidate solutions is then $\{Z_1(GV), Z_2(GV), Z_3(GV)$: GV=$\{1,\ldots, \mathcal{N}x\mathcal{M}\}$.

A set of Pareto solutions exists for this 3D manifold surface. A Pareto solution is one that is not dominated by any other solution. That is, a Pareto solution is one in which one cannot increase one performance measure without decreasing at least one of the others. The Pareto solutions offer a reduced set of candidate solutions from which a decision maker can select the GV that meets his/her internal preferences.

4. Solving the QAP by annealing

A simulated annealing scheme is used to solve the QAP since it provides improved solutions for several of the largest combinatorial problems available in literature and requires low computational effort [18]. The core idea of this approach is defining a "smart" strategy to find certain uphill steps to avoid convergence in local minima. This means: (a) move from the current solution to a neighboring one efficiently, (b) compute the change in the objective function, (c) if the objective function is improved by the step accept it, otherwise (d) accept the step with a probability P(accept)= $e^{-\delta/kT}$.

Where δ is the perturbed solution, T is a value representing the absolute temperature (in the analogy used to simulate energy levels in cooling solids) and k is Boltzmann's constant.

4.1. Annealing formulation for the QAP

The objective solution presented in [16] is given by:

$$\min \sum_{i=1}^{n}\sum_{a=1}^{n} B_{ik}x_{ik} + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} C_{ijkl}x_{ik}x_{jl} \qquad (11)$$

$$\sum_{j=1}^{n} x_{ik} = 1, \quad k = 1,..,n, \qquad (12)$$

$$\sum_{j=1}^{n} x_{jl} = 1, \quad l = 1,...,n, \qquad (13)$$

$$x_{ik} \in \{0,1\}; \quad i = 1,...,n, \quad k = 1,...,n, \qquad (14)$$

Where B_{ik} is the cost of assigning facility i to location k and the cost by the double assignment of i to k and j to l is represented by C_{ijkl}. This value can also be seen as the flow F_{ij} between facilities i and j times the distance D_{kl} between locations k and l.

For notation simplicity we will denote any feasible solution by a permutation p of the integers from 1 to n where $p(i)$ represents the chosen location for facility i.

Simulated annealing (SA) starts the searching process from a random permutation of the facilities. A neighborhood move is achieved by exchanging the pair of facilities i and j and evaluating the relative change in the objective function using the formula:

$$\delta = B_{ip(j)} + B_{jp(i)} - B_{ip(i)} - B_{jp(j)} + 2\sum_{h\neq i,j}\left[\left(F_{jh} - F_{ih}\right)\left(D_{p(i)p(h)} - D_{p(j)p(h)}\right)\right] \qquad (15)$$

The moves that improve the objective function (11) (i.e. δ≤0) are accepted while uphill steps (δ≥0) are accepted with a probability P(accept)= $e^{-\delta/T}$ by drawing a random number x from a uniform distribution [0,1] and accepting the exchange if x≤ $e^{-\delta/T}$. In our scheme, we try to maximize our cost function given by (7). In this context, a neighborhood move consists of exchanging two commands i and j and the relative change in the objective function. The marginal change δ is obtained by the contribution obtained by the exchange of the pair of gestures, minus the value associated to the current matching. When other command-gesture associations are affected by the exchange, the marginal change must be calculated for each of the remaining associations with respect to each of the pair exchanged.

Let δ_I, δ_S, δ_{IC} denote the particular contribution of the exchange of commands on the intuitiveness, the stress and the complementary intuitiveness respectively.

Let h_i be the scaling factor for each the intuitiveness, the stress and the complementary intuitiveness.

Let k_j be the weights assigned by the decision maker reflecting the importance of each term.

Let η(i,j) be a function that is equal to one if the commands i,j are complementary, otherwise zero.

$$\delta_I = h_1k_1(a_{r,p(s)} + a_{s,p(r)} - a_{r,p(r)} - a_{s,p(s)}) \qquad (16)$$

$$\begin{aligned}\delta_S = {} & -h_2k_2((s_{r,r} - s_{s,s})(f_{p(s),p(s)}d_{p(s),p(s)} - f_{p(r),p(r)}d_{p(r),p(r)}) + \\ & (s_{r,s} - s_{s,r})(f_{p(s),p(r)}d_{p(s),p(r)} - f_{p(r),p(s)}d_{p(r),p(s)}) + \\ & \sum_{k\neq r,s}(s_{k,r} - s_{k,s})(f_{p(k),p(s)}d_{p(k),p(s)} - f_{p(k),p(r)}d_{p(k),p(r)}) + \\ & (s_{r,k} - s_{s,k})(f_{p(s),p(k)}d_{p(s),p(k)} - f_{p(r),p(k)}d_{p(r),p(k)})\end{aligned} \qquad (17)$$

$$\begin{aligned}\delta_{IC} = {} & 2h_3k_3(\eta(r,s)k_1(a_{s,r,p(r),p(s)} - a_{s,r,p(s),p(r)}) \\ & \sum_{k\neq r,s}(\eta(k,s)(a_{k,s,p(k),p(r)} - a_{k,s,p(k),p(s)}) + \\ & \eta(k,r)(a_{k,r,p(k),p(s)} - a_{k,r,p(k),p(r)})))\end{aligned} \qquad (18)$$

Hence the relative change is evaluated as $\delta=\delta_I+\delta_S+\delta_{IC}$.
In our application we used $h_1= h_3=1$ and $h_2=0.001$.

4.2. Neighborhood structure

The next potential solution for our particular neighborhood structure is chosen based on a pseudo-random method, such that the pair exchange follows the sequence:

(1,2),(1,3),……,(1,n),(2,3),…,(n-1,n),
(1,2)……

The exchange is accepted according to the result of δ as described in Section 8.1. The probability of exchange is regulated by the temperature, which in turn drops after each attempted pair exchange. The temperature changes between initial and final values T_0 and T_f, respectively, according to (19):

$$T_{n+1} = T_n/(1+\beta T_n), \quad where \ \ \beta << T_0 \tag{19}$$

The cooling scheme is controlled by specifying a number of steps M using (20).

$$\beta = (T_0 - T_f)/MT_0T_f \tag{20}$$

Four trials were performed in the gesture-command matching problem where M was 6,000.

T_0 and T_f can be obtained using (19) by finding the maximum and minimum positive values of δ when running the neighborhood search for 1000 iterations.

$$\begin{aligned} T_0 &= \delta_{\min} + 1/10(\delta_{\max} - \delta_{\min}) \\ T_f &= \delta_{\min} \end{aligned} \tag{21}$$

The optimal temperature is the one that corresponds to the exchange of gestures that yields the maximum in (7). To avoid being trapped in a local maximum when consecutive marginal changes δ are rejected, we proceed in the following manner:

a) Accept the next negative contribution δ.
b) Set the optimal temperature to the current one.
c) Cooling is stopped (β=0).

This procedure was implemented for the optimal command-gesture matching problem, and we found that all the results obtained for 15 problems of a robotic arm example were global optima. Moreover, this approach was originally applied by [16] using the contribution δ as expressed in (15) and they reported solutions within 1% of the best known solutions for *n*=50 and 100.

5. Experiments and Results

A robotic vehicle control task using hand gestures is used to test the procedure explained in the previous chapters.

5.1. The Pareto Set of Solutions

Eight 'navigational' (directional) commands to control the direction of movement of the robot were chosen. From a master set of 22 postures, sets of 8 gestures are extracted and matched to the 8 commands (see Fig 3). The commands used were: start, finish, left, right, forward, backward, fast and slow.

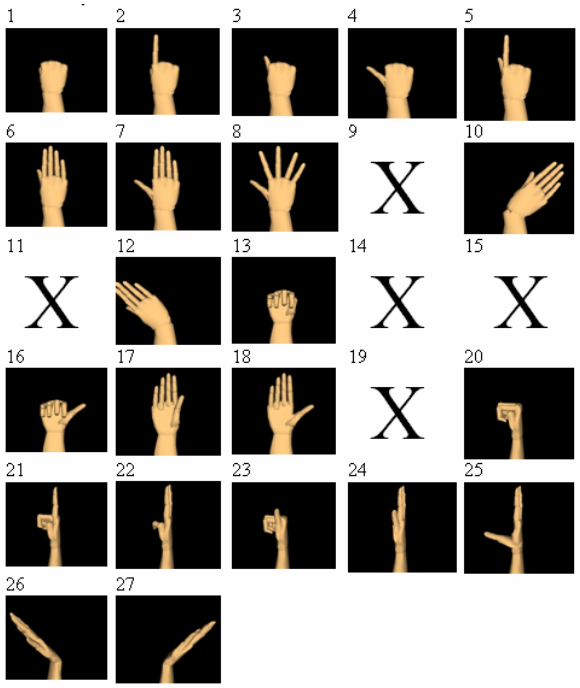


Figure 3. Gesture master set and command set for the robotic vehicle task

The algorithm generated eight solutions, where the minimal acceptable accuracy was set to 96.25 percent. Each of these solutions produced a set of 11 GV candidates. (a total of 88 GV's from eight different subsets of gestures G_n, and 11 weight combinations). The plots in Fig. 4 show the intuitiveness versus comfort trade offs for each G_n and its associated accuracy $A(G_n)$. Associated GV solutions are connected together, forming a curve for a given G_n. This family of curves is shown in a space orthogonal to the recognition accuracy coordinate (Figure 4). From this set of solutions a Pareto set of 8 GV's was obtained.

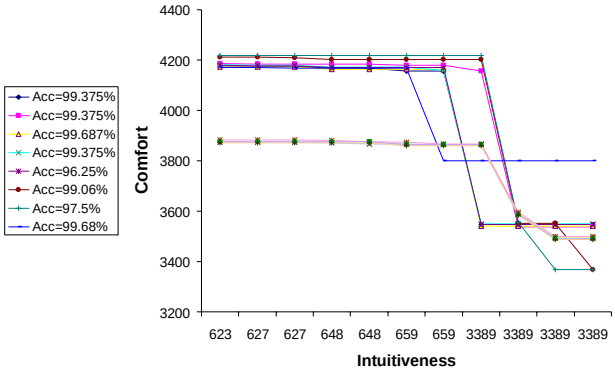


Figure 4. Intuitiveness vs. comfort families of 8 curves

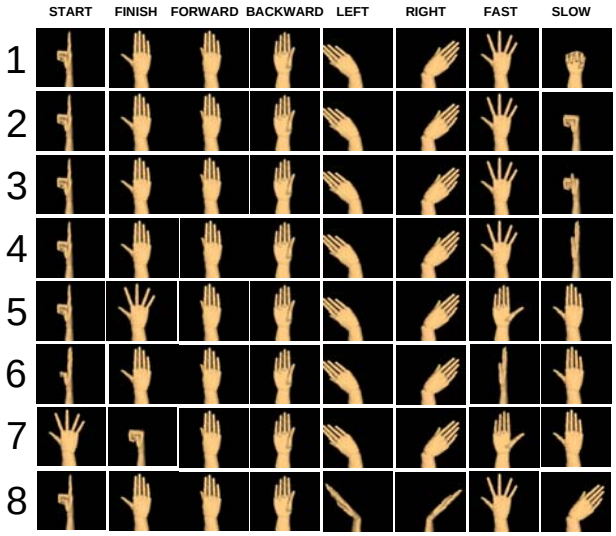


Figure 5. Pareto front GV solutions

5.2. Human Vs Computer Hand GV selection

In this section we aim to determine whether the automated methodology is better than a "hand-picked" method according to ergonomic and technical parameters. This issue was addressed through two small experiments in the context of a robotic arm "pick & place" task.

In the experiment used to obtain the natural association between commands and gestures the user was presented with a sequence of commands required to perform the "pick and place" task. The user manipulated a hand model until it was configured to represent the desired gesture, matching the displayed command. One by one all the commands were presented and their respective gesture matched according to the user desires. Once this data was collected for 35 users, two experiments were conducted.

In the first one, we investigated whether the automated system could find better associations than those provided by subjective experiments. Given a GV selected by a user, is it possible to find better gesture-command associations?

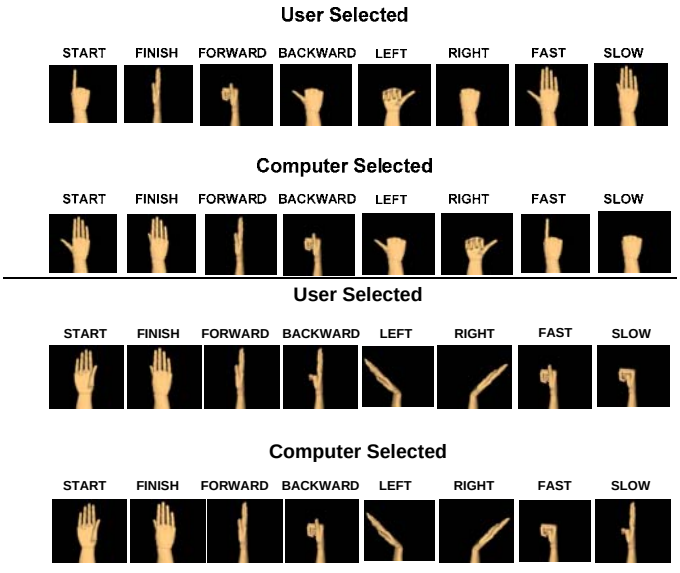
We used the results of the intuitiveness experiment to extract eight GV's from eight different users out of a set of 35 users. We selected those users that selected gestures that belonged to the reduced gesture set. We supplied each G_{n1} to the automated system and obtained new gesture-command associations. The results are compiled in Table 1.

Table 1. Human Vs Computer Hand GV selection

GV	Comfort (Z1)		Intuiteveness (Z2)		Accuracy (Z3)
	Human	Comp.	Human	Comp.	
1	3625	3625	2960	2960	95.30%
2	3661	**3807**	24	**3296**	87.50%
3	3617	3617	2706	2706	97.10%
4	3621	**3626**	**2854**	2851	99.00%
5	3569	3569	3488	3488	91.80%
6	3628	**3631**	2697	**2973**	93.70%
7	3615	3615	2524	2524	90.90%
8	3683	**3815**	552	**3334**	91.50%

The "Comfort" and "Intuitiveness" measures of eight GVs were obtained using a subjective test (Human) and the automated method (Comp). There is only one column for the Accuracy measure since both comparisons were based on the same subset of gestures, and the recognition accuracy is only a function of the gestures used, and not of their associations.

In all the GVs compared, the automated method performed better (GV2, GV6 and GV8) (see Figure 1) or equal (GV1, GV3, GV5 and GV7). There was only one case (GV4) where the GV selected by the user was more intuitive than the one selected by the automated approach; however this GV was lower in comfort.



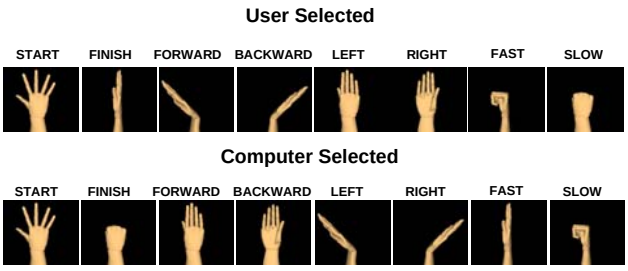


Figure 6. Three dominating solutions GV2, GV6 and GV8.

The GV's generated automatically differed from the human selected ones by at least three gesture-command matchings (GV 6) and at most eight gesture-command matchings (GV 2).

In the second experiment, we compared eight GVs obtained from the Pareto front, from the generated solutions by our methodology, to the same eight GVs created by eight users tested in the previous experiment.

The results are summarized in Table 2(a) and (b).

Table 2. Solutions found on the Pareto frontier

GV	Comfort Z1	Intuiteveness Z2	Accuracy Z3
1'	3546	3389	99.38%
2'	3549	3383	99.38%
3'	3548	3380	96.25%
4'	3552	3376	99.06%
5'	3541	3157	99.69%
6'	3556	3151	97.50%
7'	3539	3142	99.38%
8'	3801	3020	99.69%

Note that all the solutions found through the automated procedure are superior to those suggested by the user in two measures out of three (at least): Accuracy and Intuitiveness; except for two solutions GV3' and GV6', which were significantly less intuitive. The eight solutions are presented in Figure 5.

Both examples presented in this chapter show that hand gesture vocabularies obtained by the automated system have higher or equal ergonomic and technical measures than those proposed by the user in most cases.

6. Conclusions

Proper design of hand gesture-based human-machine interfaces requires accurate recognition, ergonomic design and comfort. Unfortunately, in most interfaces developed, efforts are focused primarily in accurate recognition of the gestures, which is a technical consideration only. In this work, we considered three different cost functions as proxies to task completion time: intuitiveness $Z_1(GV)$, comfort $Z_2(GV)$ and recognition accuracy $Z_3(GV)$. We established that the set of optimal hand gesture vocabularies can be accurately formulated as a maximization of the individual measures (Z1,Z2 and Z3) in a multiobjective problem fashion. The solutions are obtained by the Pareto points and the final solutions are chosen by the decision maker according to his preferences over the three objectives.

Associating a subset of gestures with commands was presented as a binary integer quadratic assignment problem which was solved by simulated annealing. The first contribution of this work is to present the modified cost function for the enhanced simulated annealing. The second contribution was a comparative study between hand gestures vocabularies obtained using the methodology suggested, and vocabularies obtained by user hand selections.

Two experiments were carried out to show the superiority of our method. In the first one we showed that the automated system can find better or equal associations than those provided from subjective experiments (using the same subset of gestures). In the second experiment, we compared the previous eight GVs selected by the user to those obtained through the Pareto front, from the generated solutions by our methodology. All the solutions found through the automated procedure were superior to those suggested by the user in two measures out of three (at least). These results indicate, in a quantitative fashion, the importance of considering technical and ergonomic aspects for a successful development and design of hand gesture interface systems.

7. Acknowledgements

This research was partially supported by the Paul Ivanier Center for Robotics Research & Production Management at Ben-Gurion University of the Negev.

8. References

[1] Kubota, N. Tomioka, Y. "Evolutionary robot vision for human tracking of partner robots in ambient intelligence" *IEEE Congress on Evolutionary Computation,* 2007. pp.1491-1496.

[2] Breuer, P. Eckes, C, Müller, S. "Hand Gesture Recognition with a Novel IR Time-of-Flight Range Camera: A Pilot Study." In *Proc. of International Conference on Computer Vision/Computer Graphics Collaboration Techniques*. Lecture Notes in Computer Science. Berlin: Springer, 2007. pp. 247-260.

[3] Xia Liu Fujimura, K. "Hand gesture recognition using depth data." in *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition,* 2004. pp. 529- 534.

[4] J.H. Do, H. Jang, S. H. Jung, J. J. Z. Bien. "Soft remote control system in the intelligent sweet home" in IEEE Conference on Intelligent Robots and Systems, 2005. pp. 3984- 3989.

[5] Marcel, S.; Bernier, O.; Viallet, J.-E.; Collobert, D. "Hand gesture recognition using input-output hidden Markov models." in *Proceedings of Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000. pp. 456 - 461

[6] Majoe, D. Schubiger, S. Clay, A. Arisona, S.M. "SQEAK: A Mobile Multi Platform Phone and Networks Gesture Sensor." *In the 2nd International Conference on Pervasive Computing and Applications, 2007. ICPCA 2007*. pp. 699-704

[7] J.H. Kim and K.-S. Hong. "Multi-Modal Recognition System Integrating Fuzzy Logic-based Embedded KSSL Recognizer and Voice-XML" in 2006 IEEE International Conference on Fuzzy Systems. 2006, pp. 956-961.

[8] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J." Emotion recognition in human-computer interaction." *IEEE Signal Processing Magazine*. 2001: 18 (1), 32–80.

[9] Berthouze, N. Fushimi, T. Hasegawa, M. Kleinsmith, A. Takenaka, H. Berthouze, L. "Learning to recognize affective body postures" in IEEE International Symposium on Computational Intelligence for Measurement Systems and Applications, 2003. pp. 193- 198.

[10] Pook P.K, Ballard D.H.. "Teleassistance: A gestural sign language for teleoperation," *Proceedings of Workshop on Gesture at the User Interface, International Conference on Computer-Human Interaction CHI 95*, Denver, CO, USA. 1995

[11] Hauptmann, A.G. and McAvinney, P. Gestures with speech for graphic manipulation, International Journal of Man-Machine Studies, 1993. 38(2): 231-49.

[12] Wolf C. G. and Morrel-Samuels P. The use of hand-drawn gestures for text editing, International Journal of Man-Machine Studies, 1987. 27: 91-102.

[13] Rempel D, Hertzer E, Brewer R. "Computer Input with Gesture Recognition: Comfort and Pain Ratings of Hand Postures." *Human-Computer Interaction International* 2003, Vol III, Crete, Greece.

[14] Kölsch, Mathias; Beall, Andrew C.; Turk, Matthew. "An Objective Measure for Postural Comfort." *Human Factors and Ergonomics Society Annual Meeting Proceedings, Communications* , 2003. pp. 725-728(4)

[15] Wachs J. Optimal Hand Gesture Vocabulary Design Methodology for Virtual Robotic Control. PhD Dissertation, Ben Gurion University of the Negev, Israel. 2007

[16] Stern, H., Wachs, J. P. and Edan, Y. "Designing Hand Gesture Vocabularies for Natural Interaction by Combining Psycho-Physiological and Recognition Factors," *Int. J of Semantic Computing. Special Issue on Gesture in Multimodal Systems*. 2008, Accepted

[17] Wachs J., Stern H., Edan Y. "Cluster Labeling and Parameter Estimation for the Automated Setup of a Hand-Gesture Recognition System." *IEEE Transactions on Systems, Man and Cybernetics. Part A*. 2005 35:(6), pp. 932-944.

[18] Connolly D. T. "An improved annealing scheme for the QAP," *European Journal of Operational Research*, 1990, 46: 93-100.