An Analytic Approach for Optimal Hand Gestures

Helman Stern, Juan Wachs, and Yael Edan

Department of Industrial Engineering and Management,
Ben-Gurion University of the Negev, Beer Sheva, 84105, Israel
{helman, juan, yael)@bgu.ac.il

Abstract. This work presents an analytical approach to the design of a gesture vocabulary (GV) using multiobjectives for psycho-physiological and gesture recognition factors. A meta-heuristic approach is taken by decomposing the problem into two sub-problems: (i) finding the subsets of gestures that meet a minimal accuracy requirement, and (ii) matching gestures to commands to maximize the human factors objective. The result is a set of solutions from which a Pareto optimal subset is selected. An example solution from the Pareto set is exhibited using prioritized objectives.

Keywords: hand gesture vocabulary design, multiobjective optimization, gesture interfaces, hand gesture recognition, human-computer interaction.

1 Introduction

Hand gesture control systems require high learnability, usability, ergonomic design and comfort [1]. Unfortunately, most gesture interfaces are designed with recognition accuracy as the central focus. The selection of hand gestures that consider recognition accuracy as well as the ease of learning, lack of stress, cognitively natural, and ease of implementation is still an open research question.

Many applications can be criticized for their idiosyncratic choice of hand gestures or postures to control or direct computer-mediated tasks [2]. These choices where probably perfectly natural for the developer of the application but may be not for others, which may show the selection of gestures is dependent their cultural and social environment. In the American Sign Language (ASL), few signs are so clearly transparent that a non-signer can guess their meaning without additional clues [3]. In [4], however, it was found that people consistently used the same gestures for specific commands. Test subjects were not coached beforehand, indicating that there may be intuitive, common principles in gesture communication.

In this work the aim is to design a gesture vocabulary that is both intuitive and comfortable on the one hand, and can be recognized with high accuracy on the other. 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 gestural 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 gestural expression is defined here

as a "gesture vocabulary" (GV). This 2 step procedure is formulated as a multiobjective optimization problem (MOP).

In the next section the problem and its solution methodology are described. Section 3 describes a Confusion Matrix Derived Solution Method for gesture subset selection, followed by command-gesture matching, and selection of pareto optimal multiobjective solutions on sections 4 and 5, respectively. In section 6 an example is solved to illustrate the procedure. Section 7 provides conclusions.

2 Solution Methodology

An optimal hand gesture vocabulary, 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 task, (or collection of tasks). However, since the task completion time, as a function of GV, has no known analytical form we propose three different performance measures as proxies: intuitiveness $Z_1(GV)$, comfort $Z_2(GV)$ and recognition accuracy $Z_3(GV)$. Maximizing each of the measures over the set of all feasible GVs defines a MOP. For the MOP, a set of Pareto solutions can be used to aid the decision maker to select the GV according to his own preferences. Because finding the solutions the MOP requires a large amount of computation time, an analytical methodology is proposed in which the MOP is relaxed to a dual priority objective problem where recognition accuracy is considered of prime importance, and the human performance objectives are secondary. The optimal GV methodology architecture is comprised of four modules (Fig. 1). In Module 1 human psycho-physiological input factors are determined. In Module 2 gesture subsets, subject to machine gesture recognition accuracy are determined. Module 3 constitutes a command - gesture matching procedure, and Module 4 finds the set of Pareto solutions.

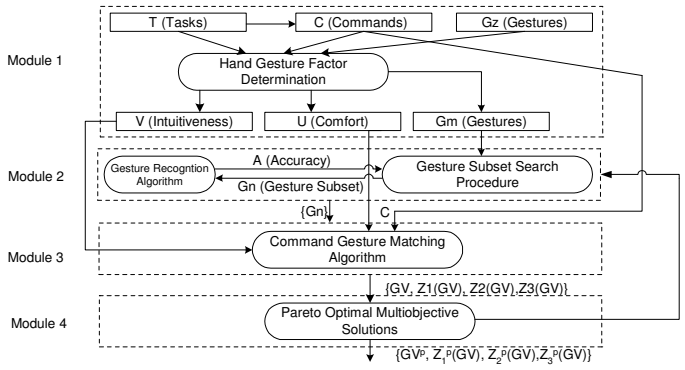

Fig. 1. Architecture of optimal hand gesture vocabulary

The task set T, the large gesture master set G_z and the set of commands C are the input parameters to Module 1. The object of Module 1 is to find; (a) V the user intuitiveness (direct and complementary) matrix, (b) U a comfort matrix based on command transitions and fatigue measures, and (c) to reduce the large set of gestures,

to the master set G_m. The V, U, and G_m, matrices values were determined through an experimental study [5]. For Module 2, the necessary inputs are G_m, and a recognition algorithm to determine the gesture recognition accuracy, A. This module employs an iterative search procedure to find a set of feasible gesture subsets, satisfying a given pre specified acceptable accuracy level. For this, a meta- heuristic search called "confusion matrix derived solution" (CMD) is proposed. This method is initiated by finding the accuracy associated with G_m. A confusion matrix $\mathcal{C}_m$ is created for the G_m problem from which the recognition accuracies associated with various gesture subsets, G_n is estimated. The set of gestures G_n that exhibit accuracies A_{min} are feasible solutions that can be approximated from $\mathcal{C}_m$. In Module 3 each of the gesture subsets G_n , found in Module 2, are matched with the set of given commands, C, such that the human measures are maximized. This matching problem is represented as a quadratic integer assignment problem (QAP). The matching solutions constitute a collection of GV's among which the Perato subset {GV*} is found in Module 4.

3 Confusion Matrix Derived Solution Method (CMD)

The CMD method consists of three phases: (i) train the recognition algorithm for the gestures in G_m, and let $\mathcal{C}_m$ be the resulting confusion matrix. Without loss of generality we assume $\mathcal{C}_m$ is square The confusion matrix is obtained directly from the partition result of the training set using a supervised FCM optimization procedure, [6], (ii) find a submatrix $\mathcal{C}_n$ from $\mathcal{C}_m$ with the highest recognition accuracy whose corresponding G_n meets the minimum accuracy constraint, and (iii) Repeat (ii) until a given number of solutions are found.

Let G_k be a set of gesture indices ($k \leq n$). Let $\mathcal{C}_{m-k}$ be a reduced confusion matrix after deleting the set of rows and columns defined by G_k. Let N be the number of solutions requested. The CMD algorithm obtains N solutions (or all the solutions with associated accuracy above a given minimum allowed A_{min} if less than 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. Let j be the current solution number. The first time that this algorithm is called, $G_n = \phi$ and j=0.

The CD Routine (G_k , j, A_{min})

1. Let the number of gestures $k=|G_k|$. Let $n=|C|$ be the number of commands
2. Repeat (n-k) times: (a) Find the least confused gesture *i'* (break ties arbitrary) in the confusion matrix $\mathcal{C}_{m-k}$ using $\underset{i=1,\ldots,m-k}{arg\,max}\{C_{ii}\} = C_{i'i'}$. (b). $G_{k+1} = G_k \cup i$. (c) Remove the corresponding column and row *i* from C_m.
3. Find $\mathcal{C}_n^j$ according to the indices in G_n and calculate $A(G_n)$
4. If $A(G_n) \geq A_{min}$ then $G_n^j = G_n$ is a feasible solution, and added to$\{G_n\}$.
5. Stop

The CMD Algorithm(N, A_{min})

1. Initialization: $G_k = \phi$, j=0,
2. G_n^j=CD (G_k , j, A_{min})

3. Calculate A using $\mathcal{C}_n$, If $G_n^j = \phi$ then exit
4. Remove the most confused gesture *i* from G_n^j, and the column and row *i* from $\mathcal{C}_m$
5. G_n^{j+1}=CD(G_n^j,j, A_{min})
6. If G_n^{j+1}belongs to the feasible solution subset. (a) Remove the highest confused gesture *k,* from G_n^j . (b) Restore the column and row *k* from $\mathcal{C}_m$. (c) Go to 4.
7. If $A \geq A_{min}$ add G_n^{j+1} to the feasible solution subset. Restore $\mathcal{C}_m$ to the original
8. If j<N and $A \geq A_{min}$, return to 4.

4 Command – Gesture Matching

Given a single set of gestures G_n found from Module 2, the gesture-command matching can be represented as a quadratic integer assignment problem (QAP) where the objective is a weighted sum of the intuitiveness and comfort measures. A x_{ij} binary assignment variable is defined which equals to 1 if command i is assigned to gesture j, and zero otherwise. Linear constraints insure that each command is paired with exactly one gesture, and that each gesture is paired with exactly one command. A simulated annealing approach from [7] is adopted to find the QAP matching solution. For each subset G_n found on Module 2, the QAP is solved repeatedly by varying the comfort and intuitive weights, w_1 and $w_2$such that, $w_1+w_2=10$. This results in a set of GV solutions corresponding to each G_n in $\{G_n\}$.

5 Pareto Optimal Multiobjective Solutions

Let each of the $\mathcal{N}$ solutions (gesture subsets G_n) from Module 2, have $\mathcal{M}$ associated solutions. This results in a total of $\mathcal{N}$x$\mathcal{M}$ candidate GV's, each may be represented as a point in 3D space, (Z_1,Z_2,Z_3). The total set of multiobjective candidate solutions is then $\{Z_1(GV), Z_2(GV), Z_3(GV): GV=\{1,\ldots, \mathcal{N}x\mathcal{M}\}$. From this set is possible to find a set of Pareto solutions. 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.

6 Illustrative Example

To determine the feasibility of the approach, a robotic arm control task using hand gestures was used. The task includes fifteen 'navigational' (directional) commands to control the direction of movement of the robot, and additional functions to interact with the objects in the environment. The gestures are extracted from a master set of 23 postures, and matched to the 15 commands (see Fig. 22).

The CMD algorithm was used to generate five solutions, where the minimum acceptable accuracy was A_{min}=98.33%. Once the gesture subsets were found, it was possible to match the commands to gestures in such a way that the psycho-physiological measures are maximized by solving the binary integer quadratic assignment problem QAP(G_n).

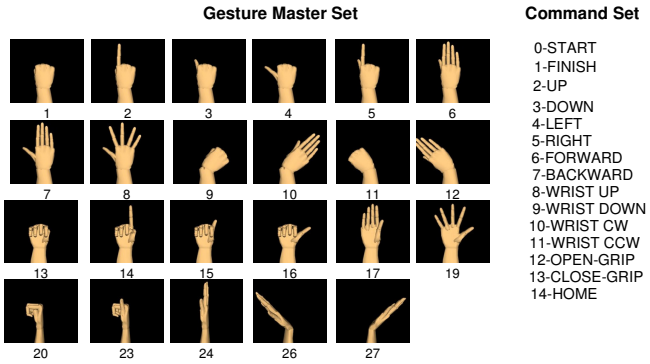

Fig. 2. Gesture master set and command set for the robotic arm task

The intuitiveness and the comfort measures were scaled by weights (from 0 to 10, in steps of 1, such that $w_1+w_2=10$) which reflects the importance of each factor on the solution. For each pair of weights (w_1, w_2) and a gesture subset G_n, a GV solution is obtained and associated values of Z_2 and Z_3. The plots in Fig. 3 show the intuitiveness versus comfort trade offs for each solution G_n.

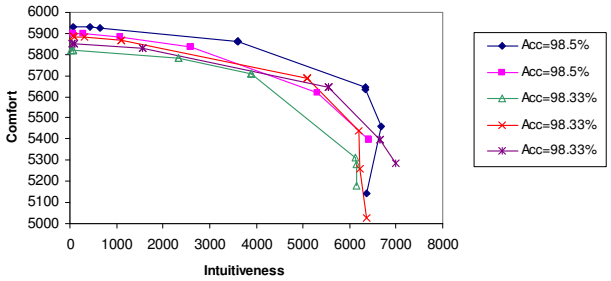

Fig. 3. Intuitiveness vs. comfort families (5 curves)

Thirteen Pareto points (non dominated solutions) were found from all the solutions. The decision maker can select a solution from the Pareto set according to his/her own preferences. Fig. 4 shows one of the Pareto solutions selected by considering accuracy, intuitiveness, and comfort as the 1st, 2nd and 3rd priorities, respectively. The solution contains many complementary gesture-command pairings. For example, the left and right commands are represented by wrist flips. Also, the command closing and opening the gripper are represented by closing and opening the fist.

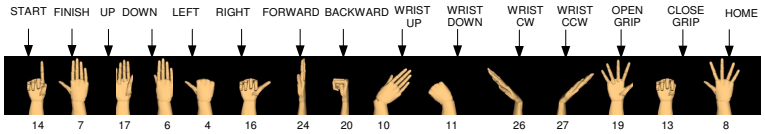

Fig. 4. A GV selected by the decision maker from the Pareto solutions

7 Conclusions

A two stage decomposition approach is suggested for solving the optimal hand gesture problem. The first stage finds a feasible subset of gestures from the master set, subject to a recognition accuracy threshold, A_{min}. The second stage finds a set of gesture vocabularies, each obtained by finding the best match between commands and gestures so that a weighted sum of the total intuitiveness and comfort are maximized. A confusion matrix derived (CDM) solution method is used to solve the first stage problem by searching for the best gesture candidates from a master set of gestures G_m. The CMD method is an approximation method for determining subsets of gestures and their associated accuracies. It requires using the supervised fuzzy c-means optimization procedure only once, and uses values from the confusion matrix to approximate the recognition accuracy of the subsets.

The CMD was used to obtain five initial gesture subsets G_n with A_{min}=98.33%, for a robotic arm task. For each G_n a set of 55 associated GVs were obtained which constituted a set of candidate solutions. From this set thirteen Pareto points were obtained and offered to the decision maker to select the GV according to his/her own preferences. This approach resulted in several 'good' solutions from a large set multiobjective decision problem search space of $6.4*10^{17}$ solutions.

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

References

1. Baudel T., and Beaudouin-Lafon M. 1993. Charade: remote control of objects using freehand gestures, Communications of the ACM, 36(7): 28-35.
2. Baudel T., Beaudouin-Lafon M, Braffort A. and Teil D. 1992. An interaction model designed for hand gesture input, Technical Report No. 772, LRI, Université de Paris-Sud, France.
3. Klima E. S. and Bellugi U. 1974. Language in another mode. Language and brain: developmental aspects, Neurosciences research program bulletin. 12(4): 539-550.
4. Hauptmann, A.G. and McAvinney, P. 1993. Gestures with speech for graphic manipulation, International Journal of Man-Machine Studies, 38(2): 231-49.
5. Stern H., Wachs J., Edan Y. 2006. Human factors for design of hand gesture Human - Machine Interaction. Proceedings of 2006 IEEE International Conference on Systems, Man, and Cybernetics, pp. 4052-4056.
6. Wachs J., Stern H., Edan Y. 2005. Cluster Labeling and Parameter Estimation for the Automated Setup of a Hand-Gesture Recognition System. IEEE Transactions on Systems, Man and Cybernetics. Part A. 35:(6), pp. 932-944.
7. Connolly D. T. 1990. An improved annealing scheme for the QAP, European Journal of Operational Research, 46: 93-100.