Gesture Recognition Performance Score: A New Metric to Evaluate Gesture Recognition Systems

Pramod Kumar Pisharady and Martin Saerbeck

Institute of High Performance Computing (IHPC)
Level 16-16 Connexis, 1 Fusionopolis Way
A*STAR, Singapore-138632
Email: pramodkp@mit.edu, Saerbeckm@ihpc.a-star.edu.sg

Abstract. In spite of many choices available for gesture recognition algorithms, the selection of a proper algorithm for a specific application remains a difficult task. The available algorithms have different strengths and weaknesses making the matching between algorithms and applications complex. Accurate evaluation of the performance of a gesture recognition algorithm is a cumbersome task. Performance evaluation by recognition accuracy alone is not sufficient to predict its successful real-world implementation. We developed a novel Gesture Recognition Performance Score ($GRPS$) for ranking gesture recognition algorithms, and to predict the success of these algorithms in real-world scenarios. The $GRPS$ is calculated by considering different attributes of the algorithm, the evaluation methodology adopted, and the quality of dataset used for testing. The $GRPS$ calculation is illustrated and applied on a set of vision based hand/ arm gesture recognition algorithms reported in the last 15 years. Based on $GRPS$ a ranking of hand gesture recognition algorithms is provided. The paper also presents an evaluation metric namely Gesture Dataset Score (GDS) to quantify the quality of gesture databases. The $GRPS$ calculator and results are made publicly available (http://software.ihpc.a-star.edu.sg/grps/).

1 Introduction

Successful research efforts in gesture recognition within the last two decades paved the path for natural human-computer interaction systems. Challenges like identification of gesturing phase, sensitivity to size, shape, and speed variations, and issues due to occlusion and complex backgrounds keep gesture recognition research still active. One ongoing goal in human-machine interface design is to enable effective and engaging interaction. For example, vision based gesture recognition systems can enable contactless interaction in sterile environments such as hospital surgery rooms, or simply provide engaging controls for entertainment and gaming applications. Other applications of gesture recognition systems include human-robot interaction, augmented reality, surveillance systems, behavior analysis systems, and smart phone applications. However current gesture recognition systems are not as robust as standard keyboard and mouse interaction.

Hand gestures are one of the most common category of body language used for communication and interaction. Hand gestures are distinguished based on temporal relationships, into two types; *static* and *dynamic* gestures. Static hand gestures (*aka* hand postures / hand poses) are those in which the hand position does not change during the gesturing period. Static gestures mainly rely on the shape and flexure angles of the fingers. In dynamic hand gestures, the hand position changes continuously with respect to time. Dynamic gestures rely on the hand trajectories and orientations, in addition to the shape and fingers flex angles. Dynamic gestures, which are actions composed of a sequence of static gestures, can be expressed as a temporal combination of static gestures [1].

1.1 Taxonomy of Gesture Recognition Systems

The initial attempts in hand gesture recognition utilized contact sensors that directly measure hand and/or arm joint angles and spatial position, using glove-based devices [2]. Later vision based non-contact methods developed. Based on feature extraction, vision-based gesture recognition systems are broadly divided into two categories, appearance-based methods and three dimensional (3D) hand model-based methods. Appearance-based methods utilize features of training image/video to model the visual appearance, and compare these parameters with the features of test image/video. Three-dimensional model-based methods rely on a 3D kinematic model, by estimating the angular and linear parameters of the model. Appearance based methods are the more widely used approach in gesture recognition with RGB cameras, whereas model based methods are more suitable for the use with new generation RGB-D cameras having skeletal tracking capability.

Mitra *et al.* [3] provided a survey of different gesture recognition methods, covering hand and arm gestures, head and face gestures, and body gestures. The hand gesture recognition methods investigated in the survey include Hidden Markov Models (HMM), particle filtering and condensation algorithms, Finite State Machines (FSM), and Artificial Neural Networks (ANN). Hand modeling and 3D motion based pose estimation methods are reviewed in [4]. An analysis of sign languages, grammatical processes in sign gestures, and issues relevant to the automatic recognition of sign languages are discussed in [5]. The review concluded that the methods studied are experimental and their use is limited to laboratory environments.

1.2 Performance Characterization in Gesture Recognition

A major cause which limits the utility of gesture recognition systems (hardware and software) in real-world applications is the lack of user's expertise to make the right choice of the algorithm, for a specific application in mind. Proper guidance on the type of gestures to be used and algorithms to recognize them is limited. In spite of the vast number of gesture recognition algorithms proposed in recent years, the availability of off-the-shelf gesture recognition softwares and standard

APIs remains limited. There exist no standards for hardware or software for gesture recognition systems.

The difficulty in predicting how a given algorithm perform on a new problem makes the performance characterization in computer vision challenging. Thacker *et al.* [6] provided a review of performance characterization approaches in computer vision. Performance characterization is referred as 'obtaining a sufficiently quantitative understanding of performance that the output data from an algorithm can be interpreted correctly'. The paper reviewed good practices in assessing the performance of essential stages such as sensing, feature detection, localization and recognition in computer vision systems. Some specific topics, face recognition, structural analysis in medical imaging, coding, optical flow, and stereo vision, are explored in depth. The evaluation methods explored for recognition performance characterization include true-false detection metric, receiver-operating characteristics, confusion matrix, and recognition rate. The paper concluded that accurate quantitative performance characterizations should be application specific and it is impossible to define one single measure applicable in all domains.

Ward *et al.* [7] proposed a set of specific performance metrics for action recognition systems, highlighting the failure of standard evaluation methods borrowed from other related pattern recognition problems. The metrics attempted to capture common artifacts such as event fragmentation, event merging and timing offsets. They extended the standard confusion matrix notion to include eight new error categories. The new metrics are evaluated on a limited set of three algorithms (string matching, Hidden Markov Models, and decision trees).

In this paper we focus on performance characterization in gesture recognition. A new metric called Gesture Recognition Performance Score ($GRPS$) is proposed which considers a wide range of factors for performance evaluation of gesture recognition algorithms (Section 2). Based on $GRPS$ the gesture recognition algorithms are ranked. $GRPS$ is calculated by considering three groups of factors, i) the algorithm performance, ii) evaluation methodology followed, and iii) the quality of the dataset utilized (how challenging the dataset is?) to test the algorithm. $GRPS$ predicts the possibility of an algorithm to be successful in its real-world implementation. It helps the algorithm designer to follow the best practices to make the algorithm effective in real applications. Based on the proposed scoring strategy, we ranked hand gesture and posture recognition algorithms published in the last 15 years and provided a list of 10 top-performing algorithms in each category. Both $GRPS$ calculator and algorithm rankings are publicaly available (http://software.ihpc.a-star.edu.sg/grps/). The dataset evaluation components of the $GRPS$ are utilized to rank a list of publicly available hand gesture and posture datasets (Section 3). The paper also discusses possible improvements of the proposed metric and its customization for other related pattern recognition tasks (Section 4).

2 The Gesture Recognition Performance Score

Evaluation of a gesture recognition algorithm with recognition accuracy alone is not sufficient to predict its success in real-world applications. Factors such as the number of classes the algorithm can recognize, its person independence, and its robustness to noise and complex environments are also to be considered while evaluating gesture recognition algorithms. Multi-problem benchmarks exist for the performance comparison of hardware and software components like CPUs and compilers. Such application based benchmarks provide a better measure of the real-world performance of a given system. We derived the factors (Table 1) affecting the effectiveness of a gesture recognition system from a survey[1] of algorithms reported in the past 15 years. The proposed *GRPS* is based on the factors listed.

2.1 Components of *GRPS*

We considered 14 component factors in the calculation of *GRPS* (Table 1). The components are divided into three groups based on the factor they depend on (algorithm, methodology, and dataset). Fourteen index scores are calculated from the 14 components and the *GRPS* is calculated as the weighted average of these index scores. The different levels of weight assignment are shown in Fig. 1. The description of each of these components and calculation of index scores are provided in the following subsections.

Accuracy Index The *GRPS* is proposed due to the limited expressiveness of recognition accuracy of the algorithm about its effectiveness in real world applications. However we considered the recognition accuracy as one component in the *GRPS*, together with other factors affecting the recognition accuracy. The accuracy index X_1 of *GRPS* is calculated from the reported recognition accuracy of the algorithm (1).

$$X_1 = \frac{Number\ of\ correctly\ classified\ samples}{Total\ number\ of\ samples} \times 100 \tag{1}$$

Spotting Index The spotting index X_2 of the *GRPS* provides credit to algorithms which can spot (detect) gestures. X_2 is a binary variable representing whether the algorithm has the capability to spot gestures ($X_2 = 1$) or not ($X_2 = 0$).

Class Index The number of classes a recognition algorithm can discriminate between is a major factor in multi-class pattern recognition. Algorithm which can recognize more number of classes are to be given higher performance scores as

[1] We are in the process of publishing a detailed survey on the topic of gesture recognition.

Table 1. Different components of the *GRPS*

No.	Component	Depends on*	Deciding factor
1	Accuracy index (X_1)	Algorithm	Recognition accuracy of the algorithm
2	Spotting index (X_2)	Algorithm	Ability of the algorithm to spot gestures
3	Class index (X_3)	Method./Data.	Number of classes considered
4	Subjects index (X_4)	Method./Data.	Number of subjects in the test set
5	Samples index (X_5)	Method./Data.	Number of test samples per class per subject
6	Complexity index (X_6)	Algorithm	Computational complexity of the algorithm
7	Cross validation index (X_7)	Methodology	Cross validation or not
8	Dataset index (X_8)	Dataset	Public or private dataset
9	Availability index (X_9)	Methodology	System availability
10	Background index (X_{10})	Dataset	Complex or simple background
11	Noise index (X_{11})	Dataset	Presence of other human in the background
12	Scale index (X_{12})	Dataset	Variation in scale/ size considered or not
13	Lighting index (X_{13})	Dataset	Variation in lighting considered or not
14	Extensibility index (X_{14})	Algorithm	Online or offline learning

*Components depend on factors such as algorithm performance, evaluation methodology, and quality of dataset. The factors shown as Method./Data. depend on methodology for algorithm evaluation (Section 2.2) and on dataset itself for dataset evaluation (Section 3).

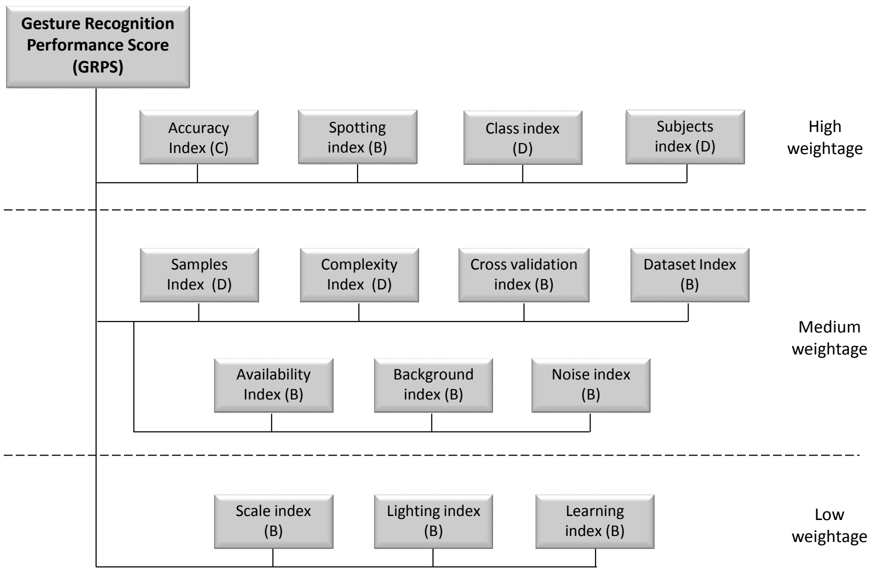

Fig. 1. Components of the *GRPS*. The components are divided into three levels and are given weightage in the ratio 4 : 2 : 1 (top to bottom) for the calculation of *GRPS* (Section 2.2). **B** - Binary, **C** - Continuous, **D** - Discrete.

those algorithms have better versatility. The class index X_3 of $GRPS$ represents the number of classes the algorithm can handle; the number of classes considered while testing the algorithm. X_3 varies in a non-linear and saturating manner with respect to the number of classes. The value of X_3 saturates at large values of the number of classes. A scaled sigmoidal logistic function (2) is used for the calculation of X_3.

$$X_3 = 2 \times \left(\frac{1}{1 + e^{-4lc}} - 0.5 \right) \tag{2}$$

where c is the number of classes and l represents the slope of the logistic function at the origin. The parameter l is calculated using (3).

$$l = \frac{1}{N_c^{max}} \tag{3}$$

where N_c^{max} is the maximum of the number of classes reported in the literature surveyed (Table 2).

Table 2. Parameters* in the calculation of $GRPS$

Parameter	**Value**	
	Hand gesture	Hand posture
Maximum of the number of classes (N_c^{max})	**120** [8]	**30** [9]
Maximum of the number of subjects (N_s^{max})	**75** [8]	**40** [1]
Maximum of the number of test samples (N_t^{max})	**80** [10]	**100** [11]

*The parameters for hand gestures and postures are identified separately (by reviewing the literature), to make the comparison using $GRPS$ precise.

The number of classes considered for calculation of $GRPS$ is discrete and finite. Identifying continuous blends of discrete gestures is out of scope for this study. Typical applications of gesture recognition systems only require a finite number of classes (for example 26 gestures are needed to represent alphabets in English language). Fig. 2 shows the logistic function utilized (corresponding to N_c^{max}=120, the maximum number of dynamic gesture classes reported in the literature [8]). The selection of parameter l as the inverse of maximum number of classes N_c^{max} is intuitive as the component X_3 achieves a value of 0.964 when number of classes is 120. This assignment provides future researchers the space to consider more number of classes[2].

Subjects Index Gesture recognition algorithms are trained using gestures performed by one or more subjects. In order to ensure the person independence and

[2] Considering larger number of classes will not increase the score much. This is reasonable as the number of gestures used in interaction applications is limited.

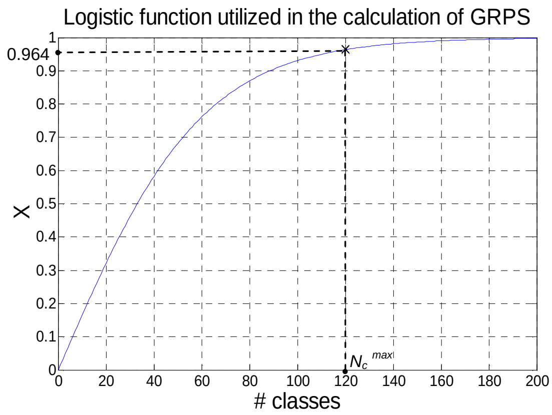

Fig. 2. Logistic function (2) utilized in the calculation of class index X_3 of $GRPS$. The slope of the logistic function at origin is tuned according to the maximum number of classes N_c^{max} (= 120 here). The component X_3 attains a value 0.964 when the number of classes is 120, providing space to consider more number of classes. The logistic functions for calculation of subjects index (X_4) and samples index (X_5) are tuned in a similar manner.

generality of the algorithm, the testing is to be done with the data from multiple persons. The subjects index X_4 of $GRPS$ represents the number of persons from the test data of the algorithm is acquired. A logistic function (4) similar to that for class index is utilized for the calculation of subjects index.

$$X_4 = 2 \times \left(\frac{1}{1 + e^{-4ms}} - 0.5 \right) \tag{4}$$

where s is the number of subjects in the test data. The desired slope m of the logistic function at the origin is calculated by (5).

$$m = \frac{1}{N_s^{max}} \tag{5}$$

where N_s^{max} is the maximum of the number of subjects reported in the literature surveyed (Table 2).

Samples Index The number of samples in the test data of the algorithm is another important factor which decides the reliability of reported recognition performance. Successful recognition of more number of samples with variations shows the algorithm's generality and robustness. The logistic function (6) is utilized to extract the samples index X_5 of the $GRPS$. X_5 considers the number of samples per class per subject in the test data.

$$X_5 = 2 \times \left(\frac{1}{1 + e^{-4nt}} - 0.5 \right) \tag{6}$$

where t is the number of test samples/ class/ subject. The slope n of the logistic function at origin is given by (7).

$$n = \frac{1}{N_t^{max}} \tag{7}$$

where N_t^{max} is the maximum of the number of test samples/ class/ subject reported in the literature surveyed (Table 2).

Complexity Index The computational complexity of the gesture recognition algorithm with respect to the number of classes is a major factor which decides its success in real world implementation. For simplicity we only consider seven categories of worst case complexities (Table 3). The complexity index X_6 is calculated as the inverse of the complexity class number C_N (8).

$$X_6 = \frac{1}{C_N} \tag{8}$$

Table 3. Complexity classes considered in the calculation of *GRPS*

Complexity class number (C_N)	**Complexity type**
1	Constant
2	Logarithmic
3	Liner
4	Quadratic polynomial
5	Cubic polynomial
6	Higher order (>3) polynomial
7	Exponential or higher

Cross Validation Index The cross validation index X_7 of the *GRPS* provides credit for algorithms tested through cross validation. X_7 is a binary variable representing whether the reported results are average accuracies on cross validation ($X_7 = 1$) or not ($X_7 = 0$).

Dataset Index Sharing data and code is important for replication of systems and the community needs to build on the work of others to make advancements, as in the case of any other scientific discipline [6]. The availability of public (downloadable) gesture datasets was limited till the year 2007 and has been increased recently. Publishing the dataset used to test the algorithm helps other researchers to verify the results and to utilize the database for their own research.

Reporting the performance of the algorithm by testing it using a publicly available dataset increases the authenticity of reported results. Dataset index X_8 of the *GRPS* provides credit to algorithms tested using publicly available datasets. X_8 is a binary variable representing whether the algorithm is tested using publicly available dataset ($X_8 = 1$, credit is also given if the authors of the paper published the dataset used), or whether the algorithm is tested using a dataset private to the authors ($X_8 = 0$).

Availability Index The availability index X_9 of *GRPS* provides credit to publicly available algorithms. Making the algorithm available helps other researchers to recreate the study and evaluate the results objectively. X_9 is a binary variable representing whether the source code (or binaries) of the algorithm is available for download ($X_9 = 1$) or not ($X_9 = 0$). This component is included to motivate researchers and developers to make their algorithm available to the community, in spite of the current limited availability of testable gesture recognition algorithms.

Background Index Backgrounds in real visual scenes are complex. To ensure success in real world application, developers of gesture recognition algorithms should consider complex and cluttered backgrounds with the gesture patterns to be recognized. To provide better ranking to algorithm which can handle complex backgrounds[3], the background index X_{10} is included in the *GRPS* ($X_{10} = 1$ if the algorithm is tested with complex background data, $X_{10} = 0$ otherwise).

Noise Index Practical use of gesture recognition systems may need its implementation in crowded places or in places where humans other than the gesturer are present. The noise index X_{11} of the *GRPS* provides credit ($X_{11} = 1$) to algorithms which are tested using samples with noises such as full or partial human body, and hands or faces of other human in the background.

Scale Index The size and scale of the gesture varies with relative position of the sensor with respect to gesturer. Algorithms having robustness against size and scale variations of the gestures are given higher credit in the scale index (binary) X_{12} of the *GRPS*. Algorithms which are tested using size/ scale variations of the gestures have $X_{12} = 1$ whereas $X_{12} = 0$ for other algorithms.

Lighting Index The practical use of gesture recognition systems requires its operation in indoor and outdoor environments with various lighting conditions. The robustness of the algorithm against lighting variations is another important factor which decides its success in real-world implementation. The lighting index

[3] The complexity due to the presence of other objects is considered in background index. The complexity due to the presence of other human (which is more challenging due to skin colored backgrounds) is considered in noise index.

(binary) X_{13} of the $GRPS$ provides credit ($X_{13} = 1$) to algorithm which can handle lighting variations in the scene.

Extensibility Index Gesture recognition algorithm which can be trained online for new gesturers have flexibility and better utility compared to algorithm which is to be trained offline. The extensibility index (binary) X_{14} of $GRPS$ consider this factor. It provides higher score ($X_{14} = 1$) to algorithm which can be trained online, than algorithm which is to be trained offline ($X_{14} = 0$).

2.2 Calculation of *GRPS*

There are 14 indices in the $GRPS$ as detailed in Section 2.1, which collectively decides the overall effectiveness of a gesture recognition algorithm. The influence of different indices of the $GRPS$ in the effectiveness of the algorithm are different. To consider the different levels of influence of the components, three levels of weightage are given to the 14 $GRPS$ indices (Fig. 1). The $GRPS$ is calculated as the weighted mean of the 14 indices in three levels (9-12).

$$GRPS_{c1} = w_1 \times \sum_{i=1}^{i=4} X_i \tag{9}$$

$$GRPS_{c2} = w_2 \times \sum_{i=5}^{i=11} X_i \tag{10}$$

$$GRPS_{c3} = w_3 \times \sum_{i=12}^{i=14} X_i \tag{11}$$

$$GRPS = \frac{GRPS_{c1} + GRPS_{c2} + GRPS_{c3}}{n_1 \times w_1 + n_2 \times w_2 + n_3 \times w_3} \times 100 \tag{12}$$

where,

X_i i^{th} index of $GRPS$,

w_1, w_2, w_3 the three level weights of the components, = 4, 2, and 1 respectively,

n_1, n_2, n_3 the number of indices in $GRPS_{c1}$, $GRPS_{c2}$, and $GRPS_{c3}$, = 4, 7, and 3 respectively.

The ideal (maximum possible) value of $GRPS$ is 100. The weights w_1, w_2 and w_3 are selected as per the division of GRPS components into three levels with high, medium, and low weightages (Fig. 1, refer Section 4 for a discussion on weight selection).

2.3 Online Web-portal for *GRPS* Calculation and Algorithm Ranking

A web-portal (http://software.ihpc.a-star.edu.sg/grps/) (Fig. 3) is created to provide the gesture recognition researchers and algorithm developers the facility to calculate the $GRPS$ of their algorithm online. The users are prompted to input the values of 14 components of the $GRPS$ to calculate the corresponding index scores and the $GRPS$. In addition to the GRPS calculator tool, the portal provides a list[4] of ten top-ranked hand gesture and posture recognition algorithms. Table 4 provides the list of ten top-ranked algorithms at the time this paper is submitted for publication.

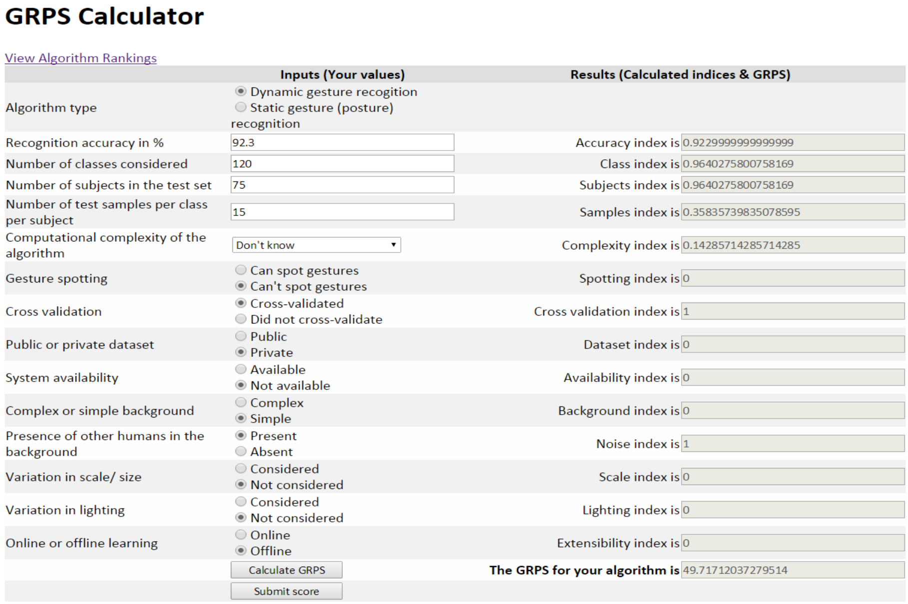

Fig. 3. A screen shot of the GRPS web-portal showing the score calculator.

3 The Gesture Dataset Score

We propose a score namely Gesture Dataset Score (GDS) to evaluate quality of publicly available gesture datasets (Table 5). The score quantifies how challenging a dataset is. GDS is calculated using the dataset depended components of GRPS. The components used are class index (X_3), subjects index (X_4), samples index (X_5), background index (X_{10}), noise index (X_{11}), scale index (X_{12}), and lighting index (X_{13}). GDS is calculated using (13)-(15). The class, subjects, and samples indices are calculated based on corresponding maximum numbers

[4] The list will be maintained and updated regularly. The portal provides authors of research papers a provision to submit their GRPS score and paper details to be included in the ranking list.

Table 4. Top-ranked algorithms and their $GRPS$

Rank	Hand gesture recognition		Hand posture recognition	
	Work	$GRPS$	Work	$GRPS$
1	[8]	56.57	[1]	67.06
2	[12]	41.20	[13]	53.08
3	[10]	40.31	[11]	51.94
4	[14]	37.49	[15]	43.85
5	[16]	35.92	[17]	42.28
6	[18]	34.79	[19]	39.95
7	[20]	34.50	[21]	38.06
8	[22]	34.40	[9]	36.44
9	[23]	30.98	[24]	34.14
10	[25]	28.99	[26]	31.66

available in the dataset (which may be different from the number actually used to test an algorithm).

$$GDS_{c1} = z_1 \times \sum_{i=3}^{i=5} X_i \tag{13}$$

$$GDS_{c2} = z_2 \times \sum_{i=10}^{i=13} X_i \tag{14}$$

$$GDS = \frac{GDS_{c1} + GDS_{c2}}{m_1 \times z_1 + m_2 \times z_2} \times 100 \tag{15}$$

where,
X_i i^{th} index of $GRPS$,
z_1, z_2 the two level weights of the components = 2 and 1 respectively,
m_1, m_2 the number of indices in GDS_{c1} and GDS_{c2}, = 3 and 4 respectively.

The ideal (maximum possible) value of GDS is 100. Table 5 provides the GDS based rank list of gesture datasets. The score varied from 53.87 to 78.32. On comparison with the competence of algorithms the datasets are more competitive and challenging considering the high values of GDS.

4 Discussion

The proposed Gesture Recognition Performance Score evaluates gesture recognition algorithms more effectively than the evaluation using recognition accuracy alone. For example Patwardhan and Roy [45] reported a recognition accuracy of 100% for their algorithm. However the experiments are conducted on an 8 class private dataset, collected from only one subject, without considering complex

Table 5. List of publicly available hand gesture databases and their *GDS*

Rank	Name, Year	Works	*GDS*
1	ChaLearn gesture gata, 2011	[27, 10, 28, 29, 11, 30]	78.32
2	MSRC-12 Kinect gesture dataset, 2012	[31]	72.71
3	ChaLearn multi-modal gesture data, 2013	[32]	70.93
4	NUS hand posture dataset-II, 2012	[1]	69.37
5	6D motion gesture database, 2011	[33]	66.91
6	Sebastien Marcel interact play database, 2004	[34, 22]	65.10
7	NATOPS aircraft handling signals database, 2011	[35]	62.44
8	Sebastien Marcel hand posture and gesture datasets, 2001	[13, 36–38]	61.73
9	Gesture dataset by Shen *et al.*, 2012	[39]	59.94
10	Gesture dataset by Yoon *et al.*, 2001	[14]	57.82
11	ChAirGest multi-modal dataset, 2013	[40]	56.73
12	Sheffield Kinect Gesture (SKIG) Dataset, 2013	[41]	55.21
13	Keck gesture dataset, 2009	[42]	54.98
14	NUS hand posture dataset-I, 2010	[43]	54.19
15	Cambridge hand gesture data set, 2007	[44]	53.87

backgrounds/ noises, and without a cross validation. The *GRPS* rated the algorithm with a score of 20.77. The algorithm by Ramamoorthy *et al.* [18] received a score of 34.79, even though it provided only 81.71% recognition accuracy. The higher *GRPS* of the algorithm is due to the experiments with more number of subjects (5) and test samples (14/ class/ subject), in extreme testing conditions (complex backgrounds, lighting variations), and by considering external noises (face of the posturer, other human in the background).

The maximum values of reported *GRPS* are 56.57 and 67.06 for gesture and posture recognition algorithms respectively. This points out the scope for improving current gesture recognition systems and the testing methodology followed. For example the person independence of algorithms is to be improved and the algorithm testing is to be conducted in environments outside laboratory, to enhance its performance in complex scenarios. The different factors to be considered while evaluating the performance of a gesture recognition system are listed in the paper, motivating researchers to develop and test new algorithms using competitive methodology, in challenging environments.

4.1 Selection of Weights

The weights w_1, w_2 and w_3 in the *GRPS* calculation (12) are selected based on the preferences given to the different *GRPS* indices. The three weights are selected such that $w_1 > w_2 > w_3$ which gives high, medium, and low weightages to the three classes of indices (Fig. 1). The reported results are achieved with $w_1 = 4, w_2 = 2$, and $w_3 = 1$. Our experiments have shown that there is no major changes in the algorithm comparison and ranking with variations in weights, provided the rule $w_1 > w_2 > w_3$ is followed.

4.2 Possible Improvements

The effectiveness of proposed *GRPS* in comparing gesture recognition algorithms can be improved by including objective measures of problem size (or problem difficulty) in the *GRPS* calculation. For example including measures of interclass similarity and speed of gestures will help to give credit to algorithms which are discriminative (which can discriminate classes with higher interclass similarity), and which can recognize gestures in spite of its high speed. Another possible improvement of the *GRPS* is its modification by considering different levels of noises, scale and lighting variations to refine on its components X_{11}, X_{12} and X_{13} respectively.

4.3 Customization for Specific Applications and Other Recognition Tasks

The *GRPS* measure could be customized for the evaluation of gesture recognition algorithms for specific applications, by adjusting the weights of its constituent indices. For example the accuracy index (X_1) could be given higher weightage to the class index (X_3) in the case of vision system for doctor-computer interaction in a surgery room, whereas X_3 could be given higher weightage to X_1 in the case of vision system for a social robot operating in a supermarket.

The proposed performance score could be extended to other recognition tasks like face, object, and action recognition with necessary modifications in the constituent components. For example the presence of complex backgrounds is not relevant for the evaluation of a face recognition algorithm (as robust face detection algorithms are available), and the number of subjects is not applicable for object recognition.

5 Conclusion

The quantitative performance characterization of pattern recognition systems is a challenging task. We took an initial step in this direction and proposed novel evaluation methods for gesture recognition algorithms and gesture datasets. The proposed scores provided ranking for both algorithms and datasets. We are currently preparing a detailed survey of gesture recognition algorithms with qualitative comparison of the ranked algorithms. The quantitative comparison using *GRPS* will be supported by testing the top ranked algorithms under same conditions to extract reliable scientific conclusions on gesture recognition systems.

Acknowledgement. The authors would like to thank Mr. Joshua Tan Tang Sheng for helping in the implementation of online web-portal for the calculation of *GRPS*.

References

1. Pisharady, P.K., Vadakkepat, P., Loh, A.P.: Attention based detection and recognition of hand postures against complex backgrounds. International Journal of Computer Vision **101** (2013) 403–419
2. Dipietro, L., Sabatini, A.M., Dario, P.: A survey of glove-based systems and their applications. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews **38** (2008) 461–482
3. Mitra, S., Acharya, T.: Gesture recognition : A survey. IEEE Transactions on Systems, Man, and Cybernetics-Part C: Application and Reviews **37** (2007) 311–324
4. Erol, A., Bebis, G., Nicolescu, M., Boyle, R.D., Twombly, X.: Vision-based hand pose estimation: A review. Computer Vision and Image Understanding **108** (2007) 52–73
5. Ong, S.C.W., Ranganath, S.: Automatic sign language analysis: A survey and the future beyond lexical meaning. IEEE Transactions on Pattern Analysis and Machine Intelligence **27** (2005) 873–891
6. Thacker, N.A., Clark, A.F., Barron, J.L., Beveridge, J.R., Courtney, P., Crum, W.R., Ramesh, V., Clark, C.: Performance characterization in computer vision: A guide to best practices. Computer Vision and Image Understanding **109** (2008) 305–334
7. Ward, J.A., Lukowicz, P., Gellersen, H.W.: Performance metrics for activity recognition. ACM Transactions on Intelligent Systems and Technology **02** (2011) 6:01–6:23
8. Lichtenauer, J.F., Hendriks, E.A., Reinders, M.J.T.: Sign language recognition by combining statistical dtw and independent classification. IEEE Transactions on Pattern Analysis and Machine Intelligence **30** (2008)
9. Teng, X., Wu, B., Yu, W., Liu, C.: A hand gesture recognition system based on local linear embedding. Journal of Visual Languages & Computing **16** (2005) 442–454
10. Lui, Y.M.: A least squares regression framework on manifolds and its application to gesture recognition. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2012. (2012) 13–18
11. Keskin, C., Kirac, F., Kara, Y., Akarun, L.: Randomized decision forests for static and dynamic hand shape classification. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2012. (2012) 31–46
12. Yang, M.H., Ahuja, N., Tabb, M.: Extraction of 2d motion trajectories and its application to hand gesture recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **24** (2002) 1061–1074
13. Triesch, J., Malsburg, C.: A system for person-independent hand posture recognition against complex backgrounds. IEEE Transactions on Pattern Analysis and Machine Intelligence **23** (2001) 1449–1453
14. Yoon, H.S., Soh, J., Bae, Y.J., Yang, H.S.: Hand gesture recognition using combined features of location, angle, and velocity. Pattern Recognition **34** (2001) 1491–1501
15. Huang, D.Y., Hu, W.C., Chang, S.H.: Gabor filter-based hand-pose angle estimation for hand gesture recognition under varying illumination. Expert Systems with Applications **38** (2011) 6031–6042
16. Chen, F.S., Fu, C.M., Huang, C.L.: Hand gesture recognition using a real-time tracking method and hidden markov models. Image and Vision Computing **21** (2003) 745–758

17. Zhou, R., Junsong, Y., Zhengyou, Z.: Robust hand gesture recognition based on finger-earth movers distance with a commodity depth camera. In: In Proceedings of ACM Multimeida. (2011)
18. Ramamoorthy, A., Vaswani, N., Chaudhury, S., Banerjee, S.: Recognition of dynamic hand gestures. Pattern Recognition **36** (2003) 2069–2081
19. Licsar, A., Sziranyi, T.: User-adaptive hand gesture recognition system with interactive training. Image and Vision Computing **23** (2005) 1102–1114
20. Lai, K., K.J., Ishwar, P.: A gesture-driven computer interface using kinect. In: IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI). (2012) 185–188
21. Van den Bergh, M., Carton, D., De Nijs, R., Mitsou, N., Landsiedel, C., Kuehnlenz, K., Wollherr, D., Van Gool, L., Buss, M.: Real-time 3d hand gesture interaction with a robot for understanding directions from humans. In: IEEE International Symposium on Robot and Human Interactive Communication (IEEE RO-MAN). (2011)
22. Just, A., Marcel, S.: A comparative study of two state-of-the-art sequence processing techniques for hand gesture recognition. Computer Vision and Image Understanding **113** (2009) 532–543
23. Frolova, D., Stern, H., Berman, S.: Most probable longest common subsequence for recognition of gesture character input. IEEE Transactions on Cybernetics **43** (2013) 871–880
24. Ge, S.S., Yang, Y., Lee, T.H.: Hand gesture recognition and tracking based on distributed locally linear embedding. Image and Vision Computing **26** (2008) 1607–1620
25. Shin, M.C., Tsap, L.V., Goldgof, D.B.: Gesture recognition using bezier curves for visualization navigation from registered 3-d data. Pattern Recognition **37** (2004) 1011–1024
26. Zhao, M., Quek, F.K.H., Wu, X.: Rievl: Recursive induction learning in hand gesture recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **20** (1998) 1174–1185
27. Guyon, I., Athitsos, V., Jangyodsuk, P., Hamner, B., Escalante, H.: Chalearn gesture challenge: Design and first results. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2012. (2012) 1–6
28. Malgireddy, M.R., Inwogu, I., Govindaraju, V.: A temporal bayesian model for classifying, detecting and localizing activities in video sequences. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2012. (2012) 43–48
29. Di, W., Fan, Z., Ling, S.: One shot learning gesture recognition from rgbd images. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). (2012)
30. Mahbub, U., Imtiaz, H., Roy, T., Rahman, M., Ahad, M.: A template matching approach of one-shot-learning gesture recognition. Pattern Recognition Letters **http://dx.doi.org/10.1016/j.bbr.2011.03.031** (2012)
31. Simon, F., Helena, M.M., Pushmeet, K., Sebastian, N.: Instructing people for training gestural interactive systems. In: International Conference on Human Factors in Computing Systems, CHI, ACM (2012) 1737–1746
32. Escalera, S., Gonzlez, J., Bar, X., Reyes, M., Lopes, O., Guyon, I., Athistos, V., Escalante, H.: Multi-modal gesture recognition challenge 2013: Dataset and results. In: Proceedings of the 15th ACM International Conference on Multimodal Interaction (ICMI), Sydney, Australia (2013)

33. Chen, M., AlRegib, G., Juang, B.H.: 6dmg: A new 6d motion gesture database. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). (2011)
34. Just, A., Bernier, O., Marcel, S.: Hmm and iohmm for the recognition of mono- and bi-manual 3d hand gestures. In: Proceedings of the British Machine Vision Conference (BMVC). (2004)
35. S., Y., D., D., D., R.: Tracking body and hands for gesture recognition: Natops aircraft handling signals database. In: Proceedings of the 9th IEEE Conference on Automatic Face and Gesture Recognition (FG 2011), Santa Barbara, CA (2011) 500–506
36. Triesch, J., Malsburg, C.: Robust classification of hand postures against complex backgrounds. In: Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, 1996, Killington, VT, USA (1996) 170–175
37. Triesch, J., Malsburg, C.: A gesture interface for human-robot-interaction. In: Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, 1998, Nara, Japan (1998) 546–551
38. Marcel, S.: Hand posture recognition in a body-face centered space. In: Proceedings of the Conference on Human Factors in Computer Systems (CHI). (1999)
39. Shen, X.H., Hua, G., Williams, L., Wu, Y.: Dynamic hand gesture recognition: An exemplar-based approach from motion divergence fields. Image and Vision Computing **30** (2012) 227–235
40. Ruffieux, S., Lalanne, D., Mugellini, E.: Chairgest: A challenge for multimodal mid-air gesture recognition for close hci. In: Proceedings of the 15th ACM on International Conference on Multimodal Interaction (ICMI). (2013)
41. Liu, L., Shao, L.: Learning discriminative representations from rgb-d video data. In: Proceedings of International Joint Conference on Artificial Intelligence (IJCAI). (2013)
42. Zhuolin, J., Davis, L.S.: Recognizing actions by shape-motion prototype trees. In: IEEE International Conference on Computer Vision (ICCV), 2009. (2009) 444–451
43. Pisharady, P.K., Vadakkepat, P., Loh, A.P.: Hand posture and face recognition using a fuzzy-rough approach. International Journal of Humanoid Robotics **07** (2010) 331–356
44. Kim, T.K., Wong, S.F., Cipolla, R.: Tensor canonical correlation analysis for action classification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007. (2007) 1–8
45. Patwardhan, K.S., Roy, S.D.: Hand gesture modelling and recognition involving changing shapes and trajectories, using a predictive eigentracker. Pattern Recognition Letters **28** (2007) 329–334