

Pattern Recognition Letters

Authorship Confirmation

Please save a copy of this file, complete and upload as the “Confirmation of Authorship” file.

As corresponding author I, Seshan Srirangarajan, hereby confirm on behalf of all authors that:

1. This manuscript, or a large part of it, has not been published, was not, and is not being submitted to any other journal.
2. If presented at or submitted to or published at a conference(s), the conference(s) is (are) identified and substantial justification for re-publication is presented below. A copy of conference paper(s) is(are) uploaded with the manuscript.
3. If the manuscript appears as a preprint anywhere on the web, e.g. arXiv, etc., it is identified below. The preprint should include a statement that the paper is under consideration at Pattern Recognition Letters.
4. All text and graphics, except for those marked with sources, are original works of the authors, and all necessary permissions for publication were secured prior to submission of the manuscript.
5. All authors each made a significant contribution to the research reported and have read and approved the submitted manuscript.

Signature _____ Date _____

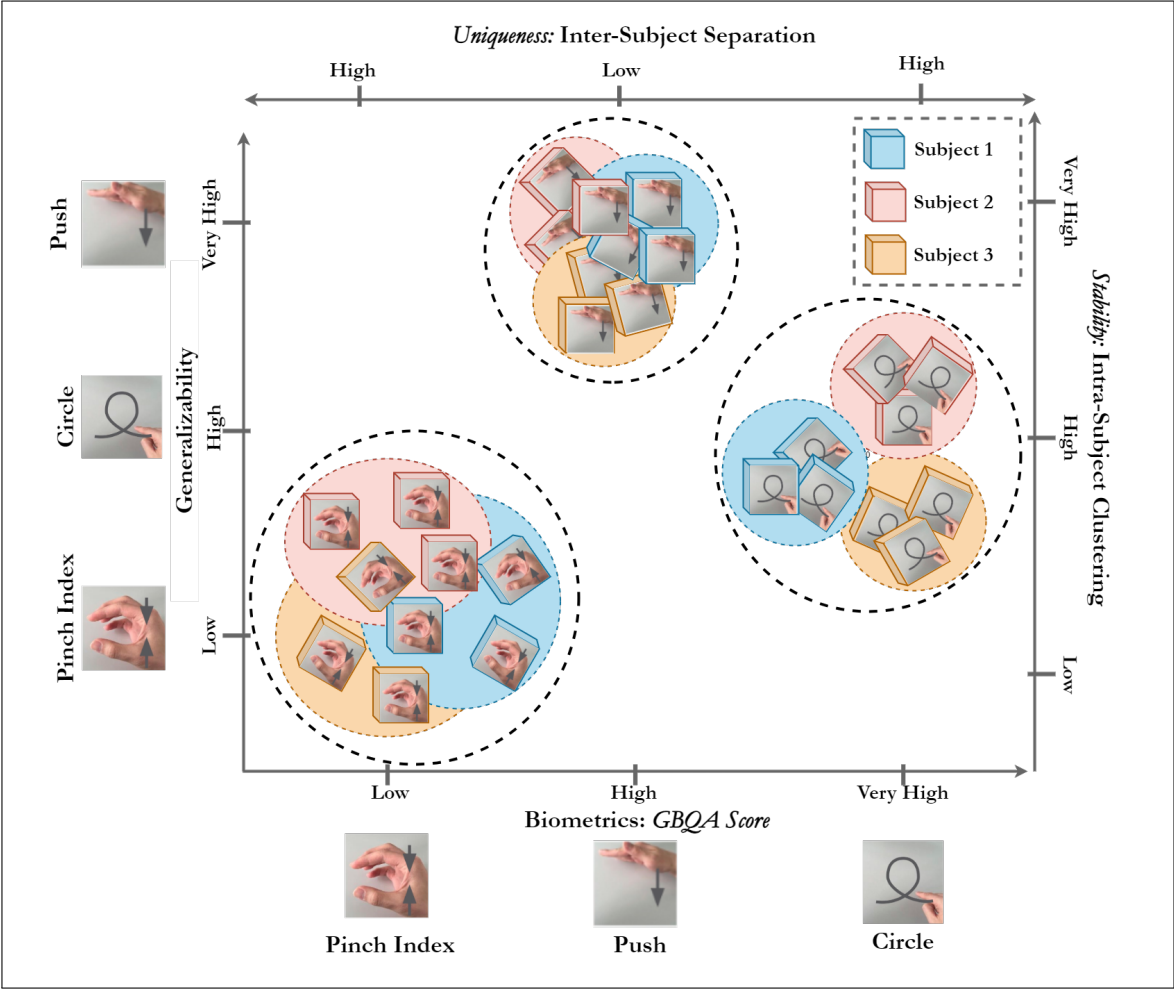
List any pre-prints:

Relevant Conference publication(s) (submitted, accepted, or published):

Justification for re-publication:

Graphical Abstract (Optional)

To create your abstract, please type over the instructions in the template box below. Fonts or abstract dimensions should not be changed or altered.



Research Highlights (Required)

To create your highlights, please type the highlights against each `\item` command.

It should be short collection of bullet points that convey the core findings of the article. It should include 3 to 5 bullet points (maximum 85 characters, including spaces, per bullet point.)

- Systematic identification of issues and biases in cross-subject and cross-session hand gesture recognition.
- *GBQA*: A novel framework to estimate biometric characteristics within a hand gesture class.
- *GBQA* operates over biases prevailing in cross-subject and cross-session hand gesture recognition experiments, and hence does not require explicit user identity information (labels).
- Introduction of direct and relative bias measurement strategies for quantification in the *GBQA* framework.



Biometric Characteristics of Hand Gestures through Joint Decomposition of Cross-Subject and Cross-Session Biases

Aman Verma^a, Gaurav Jaswal^c, Seshan Srirangarajan^{a,b,**}, Sumantra Dutta Roy^a

^aDepartment of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi 110016, India

^bBharti School of Telecommunication Technology and Management, Indian Institute of Technology Delhi, New Delhi 110016, India

^ciHub and HCI Foundation, Indian Institute of Technology Mandi, Himachal Pradesh, 175005, India

Article history:

Keywords: Hand Gestures, Biometric Characteristics, Cross-Subject/Session Biases, Generalizability

ABSTRACT

Hand gestures are natural in human-computer interfaces for their use in gesture-based control, as well as personalized devices with privacy preservation and device personalization. However, some hand gestures bear biometric traits, whereas some others generalize well across subjects. The ‘entanglement’ between these concepts is in gestures with tight intra-subject clusters (in a feature space), and low inter-subject separation. This puts forth a key requirement: *quantifying the biometric ‘goodness’ of a gesture so as to segregate between the gestures prudent for authentication and recognition-level generalization.* We propose a biometric characterization framework that estimates how uniquely and with how much variability, any unseen/seen subject would perform a particular gesture. We leverage a theoretical understanding of personality and stability biases present in cross-user (CU) and cross-session (CS) gesture recognition experiments. Structuring upon the biases that influence performance in these experiments, we derive mathematical relationships that quantify the biometric goodness of a gesture. In order to disentangle bias for identity understanding, we introduce two strategies based upon ‘bias mitigation’ and ‘bias intensification’. We empirically validate the proposed framework through experiments on three datasets. The proposed framework is generic and does not require user identity information. Additionally, this can operate over any existing hand-gesture recognition pipelines.

© 2025 Elsevier Ltd. All rights reserved.

1. Introduction

Hand gestures are one of the most common way for humans to communicate, express emotions and instruct. On the other hand, these gestures also encapsulate biometric identity by the virtue of uniqueness in motion patterns across subjects [1, 2]. Employing hand gestures as a biometric modality is a key step towards personalized human-computer interfaces e.g., identity adaptive device control [3]. The ease in resetting and collecting hand gesture templates, allows it to also become an

anywhere-anytime authentication modality [4]. Although this modality offers tangible benefits, one major concern is yet to be addressed: *Which set of hand gestures are to be used for authentication purposes?* Alongside, it is also equally important to categorize gestures suitable for gesture recognition tasks. These somewhat-entangled requirements require systematic study and quantification of all aspects that influence biometric integrity and generalizability across various cross-subject and cross-session settings.

1.1. Gesture Biometric Quantitative Assessment (GBQA): Quantifying the Entangled Relationship between Biometric Traits and Generalizability

The existing literature does not have a consensus and well-defined mathematical framework over the definitions of biomet-

^{**}Corresponding author

e-mail: eez227536@ee.iitd.ac.in (Aman Verma),

gaurav@tih.iitmandi.ac.in (Gaurav Jaswal), seshan@ee.iitd.ac.in (Seshan Srirangarajan), sumantra@ee.iitd.ac.in (Sumantra Dutta Roy)

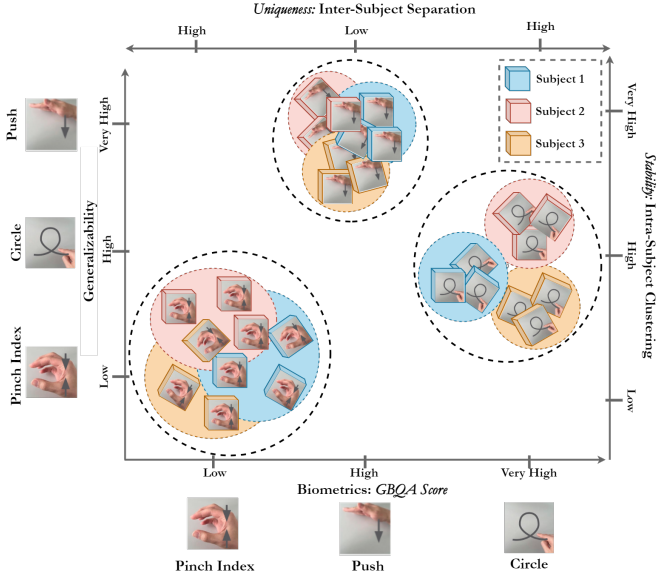


Fig. 1. An example illustrating a feature space representation over samples of three gestures (‘Circle’, ‘Push’ and ‘Pinch Index’) from three subjects. The feature space representation shows per-subject clusters for each gesture (orange, pink and cyan clusters, inside each dotted circle of an individual gesture). The figure outlines the relationships between biometric properties (the proposed *GBQA* Score), generalizability and the operating parameters: *uniqueness* and *stability*. Gestures having good biometric characteristics encompass high inter-subject separation and intra-subject clustering (the ‘Circle’ gesture). The ‘Push’ gesture illustrates the entanglement between good biometric characteristics (tight intra-subject clusters), and generalizability (low inter-subject separation).

ric traits and generalizability. For instance, biometric goodness of hand gestures can be accredited to the complexity involved in the gesture [5]. However, this does not guarantee inter-subject separation and intra-subject clustering. Similarly, [6] defines biometrics from a trust perspective. This includes fairness, explainability and privacy alongside performance, but does not quantify these in any way. It has also been observed that definitions of biometric goodness often coalesce with the ones from biometric quality (ISO/IEC 29794-1), which quantify the number of artifacts per template, and not uniqueness. Previous work [7] suggests inverse relationships between generalization and personalization, but at the same time generalizability borrows some provisions from biometrics (personalization) such as intra-subject clustering. An example is a coarse-grained gesture such as ‘Push’. This performs well for recognition tasks [8], and carries good biometric traits as well [9]. This illustrates the fact that generalization and biometrics have an entangled relationship. Fig. 1 illustrates some of these issues, and more.

Biometric Characterization: In order to differentiate between personalized and generalized gesture set, [7] leverages subject distances in feature space. Personalization score formulation of this work does not account for inter-subject distances, which in-turn is essential to quantify uniqueness. Furthermore, [7] considered inverse relationship between generalization and personalization. However, both of these require intra-subject clustering. This highlights the requirement to disentangle biometric details from generalization parameters. In [10], the authors attempt to define biometric goodness of online signature

templates using *distinctiveness*, *complexity* and *repeatability*. However, the defined parameters rely on identity information (labels), or require explicitly matching scores. In the literature, biometric quality is also regarded as a characterization of biometric goodness at the template level or the subject level. For a query, [11] computes variation from the entire population, followed by unsupervised classification into three quality groups. Quality-based formulations do not exactly encode biometric details. [11] does not propose a mechanism to assess if the low quality templates are originating because of biometric variations or artifacts. In [12] the theoretical bounds for a masterface attack to compromise a face recognition system are derived. These estimates parametrize over inter-subject distances to capture uniqueness but require a large number of identities to achieve generalizable results. [13] characterizes biometrics from a sphere-packing problem perspective. A major drawback is that it considers uniform identity clusters. However, intra-subject clustering differs significantly across subjects. Our recent work [14] considers both uniqueness and intra-subject clustering, but does not generalize beyond seen training subjects. A common shortcoming in [7, 10, 12, 13, 14] is that they require explicit identity information.

We take a step-back from conventional authentication, and put forth an attempt to formulate a mathematical framework for quantifying inherent biometric traits and generalizability, within a gesture class. The proposed framework subsumes the two basic requirements in any biometric setting [10, 6, 14] namely, (i) *Uniqueness*: high separation between gestures of different subjects, and (ii) *Stability*: high consistency in motion patterns of gestures for the same subject. Subsequently, we also acknowledge that the generalizability of a gesture recognition model is well characterized through performance in cross-subject/session/training bias evaluation. To be specific, a generalized gesture recognition model should be robust against variability across subjects and sessions. Hence, both biometrics suitability as well as generalizability have relationships with similar parameters - *uniqueness* and *stability*. Therefore, we examine the performance failures in cross-subject and cross-session evaluation (hereafter, *CU* and *CS* experiments). From our analysis, we identify the biases that account for biometric characteristics or conversely, affect generalizability. Based on this, we derive proportionality relationships for the proposed *gesture biometric quantification assessment (GBQA)* score. The *GBQA* score accounts for biometric characteristics within a hand gesture. The proposed framework provides a systematic understanding of gesture recognition with respect to inter-subject separation and intra-subject clustering. Since the *GBQA* score is calculated over a cross-subject/session/training gesture-recognition paradigm, it can be transformed into a measure of generalizability as well. At the same time, the *GBQA* score formulation also eliminates the direct requirement of user identity information (labels). Such a measure can supplement the existing gesture recognition and authentication frameworks for determining a gesture set prudent for the respective tasks. We summarize the major contributions of this work as follows:

- Systematic theoretical decomposition of performance obtained in *CU* and *CS* gesture recognition experiments.

- Mathematical formulation for deriving biometrics (the *GBQA* score) and generalizability from the obtained decomposition. To the best of our knowledge, this is the first study which employs an understanding of different types of biases, for quantifying biometric characteristics.
- For estimating the *GBQA* score, introduction of direct and relative measurement strategies using bias control. Specifically, we introduce gesture pruning strategy to disentangle biometric details from generalization parameters. The proposed method does not require explicit identity information (labels).
- Quantitative and qualitative validation of the proposed method on three datasets with different acquisition modalities and set of gestures. Our proposed method either outperforms or attains comparable results with state-of-the-art biometric characterization methods.

The organization of the rest of the paper is as follows. Section 2 introduces the idea of biometric measurements within a gesture recognition framework. The section identifies cross-subject and cross-session variability and biases, and presents the *GBQA* formulation. Section 3 discusses *GBQA* measurements and corresponding experimental results. Section 4 concludes the paper.

2. The *GBQA* Framework

2.1. Biometrics from Gesture Recognition?

A gesture recognition framework offers a medium to identify gesture clusters (Fig. 1), which indirectly encode subject specific-details within a cluster. As a matter of practical viability, there is a large volume of literature on gesture recognition, and similarly pre-trained models are easily accessible. Fig. 1 depicts inter-dependencies between *uniqueness* and *stability* over biometric characteristics and generalizability. If identity details are preserved in the feature space representation (Fig. 1: the ‘Circle’ gesture), this corresponds to high biometric characteristics. One can also infer that high generalizability (Fig. 1: the ‘Push’ Gesture) also induces high intra-subject clustering, which corresponds to *stability* criterion of biometrics.

2.2. Classification of Biases

This section identifies the role of *CU* (cross-subject) inter-subject variations, and *CS* (cross-session) intra-subject clustering. We show that both biometric characteristics and generalizability are influenced by similar parameters, though specific relationships vary. This motivates the use of cross-subject/session/training settings in formulating a framework for quantification of biometric characteristics (‘*GBQA*’). A tabular description of this classification is given in the supplementary material (refer Section 1, Table S1).

2.2.1. Cross-Subject (*CU*) Biases

The performance drop of a gesture recognition system in a cross-subject (*CU*) setting can be attributed to two types of biases: (a) *CU* personality bias (P_{Bias}) and (b) *CU* gesture understanding bias ($G_{Bias-CU}$). This classification also seeds our *CU*

decomposition and bias parameter framework shown in Fig. 2. (a) ***CU* Personality Bias (P_{Bias})**: This originates because of variation in gesture patterns across different subjects. One expects the training feature space to differ from the testing one (with different subjects). In a cross-validation exercise over all subjects, the overall performance drop of a gesture recognition system reveals the *average variance of subjects from the training set i.e., uniqueness*.

(b) ***CU* Gesture Understanding Bias ($G_{Bias-CU}$)**: The other possible source of a performance drop is if the representation learnt during training is not robust enough. This points to a bias associated with the training set. An example of the same is when instances of a gesture in the training set vary widely. This is an indication of an intrinsic encoding of *personality details* in the feature space. Unlike P_{Bias} , $G_{Bias-CU}$ can be inferred from the training set performance itself.

2.2.2. Cross-Session (*CS*) Biases

In a cross-session (*CS*) setting (different sessions for the same user), we characterize biases as (a) *CS* Stability bias (S_{Bias}), and (b) *CS* Gesture Understanding bias ($G_{Bias-CS}$). This also seeds our *CS* decomposition and bias parameter framework of Fig. 3.

(a) ***CS* Stability Bias (S_{Bias})**: In *CS* evaluation, the training is on gesture samples from multiple sessions, while testing is performed over samples from an unseen session. Since a *CS* setting has gestures from a single subject, a performance drop accounts for variability in gestures performed by a subject across sessions. This represents how much variation exists within a gesture class.

(b) ***CS* Gesture Understanding Bias ($G_{Bias-CS}$)**: Similar to $G_{Bias-CU}$, $G_{Bias-CS}$ originates from inadequate gesture understanding at the training level. This incorporates details of variation across sessions for a particular subject.

2.3. Gesture Biometric Quantification Assessment (*GBQA*)

We extend the ideas developed in Section 2.2 to a thorough classification of biases (Fig. 2 and Fig. 3). In this section, we give a mathematical model of the proposed *GBQA* score on the basis of an understanding of different types of biases in cross-subject (*CU*) and cross-session (*CS*) settings.

2.3.1. *CU* Evaluation Decomposition and Biases

We denote the average cross-subject (*CU*) training and testing performance as CU_{Train} and CU_{Test} , respectively. A low CU_{Test} can be explained by two biases (Fig. 2) namely, (i) Personality bias (P_{Bias}) and (ii) *CU* Gesture Understanding bias ($G_{Bias-CU}$) (Section 2.2.1). Moreover, $G_{Bias-CU}$ quantifies the universality of the training representations. Thus, it is reasonable to expect an inverse relationship with the training performance. Further, P_{Bias} estimates biometric uniqueness for an unseen user. Hence, it can be used as a standalone measurement parameter for biometric characteristics. Mathematically,

$$CU_{Test} \propto P_{Bias}^{-1} \times G_{Bias-CU}^{-1} \quad (1)$$

$$CU_{Train} \propto G_{Bias-CU}^{-1} \quad (2)$$

$$GBQA_{CU} \propto P_{Bias} \quad (3)$$

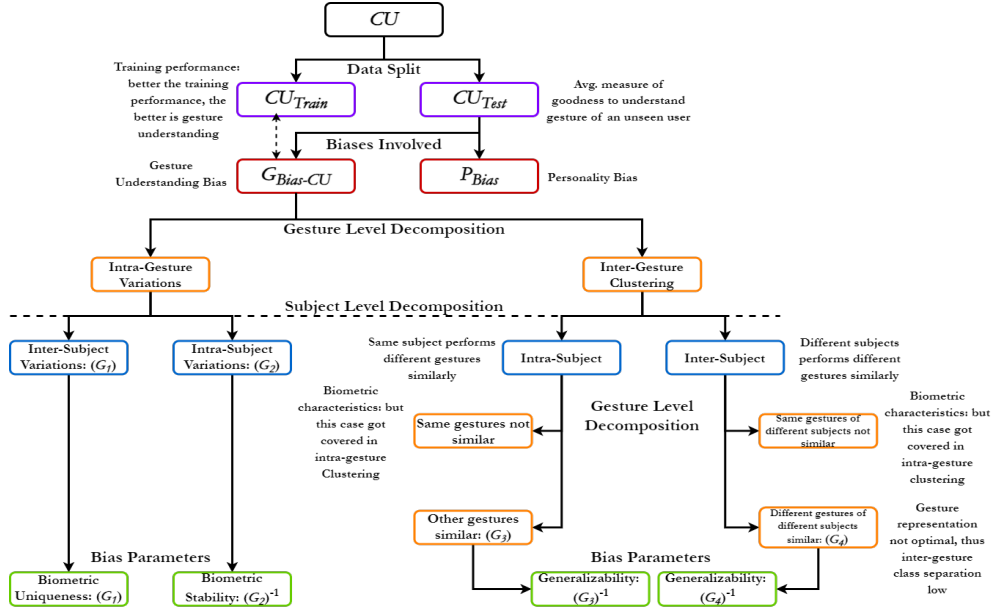


Fig. 2. Derivation of biases (illustrated in last level) that adversely influence generalization in CU settings. It is evident from this decomposition that performance drop in CU evaluation is not only due to variation in gestures of unseen subject, but also due to various pitfalls in gesture understanding. The dotted bidirectional arrow between CU_{Train} and $G_{Bias-CU}$ shows inverse proportionality between them.

The inverse of a bias parameter (G^{-1}) characterizes the opposite affect of that bias. This encapsulates information on how little a bias G affects the recognition performance. This operation for instance, may be the reciprocal of G , or $(1 - G)$, if $G \in [0, 1]$.

In (2), we represent the intuitive link between failure points in the training and the gesture understanding bias. We identify two major causes of failure in gesture understanding namely, (i) intra-gesture variations and (ii) inter-gesture clustering. The former corresponds to the high variability in gesture samples. The latter accounts for gestures of one class being very similar to those in other classes. A good example of the same is seen for fine-grained gestures such as ‘Pinch-Pinky’ and ‘Pinch-Index’. The ‘Gesture Level Decomposition’ (Fig. 2) classifies failures due to the gesture space characterization of gestures in a gesture set. In order to identify core reasons behind the above-identified causes, we perform the next level of classification with subject-level causes (Fig. 2). For each of intra-gesture variations and inter-gesture clustering, we examine the effect of intra- and inter-subject level decomposition.

(i) Intra-Gesture Variations: Intra-gesture variations could be due to inter-subject variations (disparities in the same gesture performed by different subjects) and intra-subject variations (disparities in the same gesture performed by one subject). *Inter-subject variation* (G_1) is a biometric trait which encodes the uniqueness property of a subject. G_1 accounts for variability in gestures amongst different subjects. Thus, it has a directly proportional relationship with $G_{Bias-CU}$ and $GBQA_{CU}$. The bias term accounting for *intra-subject variations* G_2 provides a measure of the non-repeatability of the motion patterns of a gesture (averaged across subjects). Thus, G_2^{-1} is a measure of biometric stability. Thus, it bears direct proportionality relationship with $GBQA_{CU}$. Since G_2 is a form of bias for gesture recognition, it has a direct proportionality relationship with $G_{Bias-CU}$ (similar

to G_1). Mathematically,

$$GBQA_{CU} \propto G_1 \times G_2^{-1} \quad (4)$$

$$G_{Bias-CU} \propto G_1 \times G_2 \quad (5)$$

(ii) Inter-Gesture Clustering: This broadly accounts for different gestures with an overall similar appearance, such as the ‘Pinch-Pinky’ and ‘Pinch-Index’ gestures. A subject-level decomposition results in two cases: (a) intra-subject inter-gesture clustering and (b) inter-subject inter-gesture clustering (Fig. 2). The first case can be further analyzed at another gesture decomposition level into two key reasons of failure: variability in the same gesture performed by a particular subject, and the same subject performing different gestures in a similar manner. The former encodes cues on biometric stability (and has been covered in intra-gesture variations, above). On the other hand, the latter defeats the notion of generalizability. If different gestures performed by a particular subject are similar, one cannot have much inter-gesture separation. Thus, the corresponding bias parameter G_3 has a direct proportionality relationship with $G_{Bias-CU}$. Furthermore, if generalizability is quantified then inverse of G_3 i.e., G_3^{-1} bears direct proportional relationship with it. Finally, considering the last broad case of inter-subject inter-gesture clustering, we apply the second gesture-level decomposition (Fig. 2). This results in identification of the following two failure causes: dissimilarity in the same gesture performed by different subjects, and similarity in different gestures. The first reason encapsulates information on inter-subject biometric uniqueness (and has been accounted for in the intra-gesture variations, above). The last possible bias case G_4 suggests similarity in different gestures of different subjects which in turn, is not amenable for generalizability. This suggest a direct proportional relationship with $G_{Bias-CU}$, and its inverse (G_4^{-1}) having a direct proportional relationship with generalizability. Mathe-

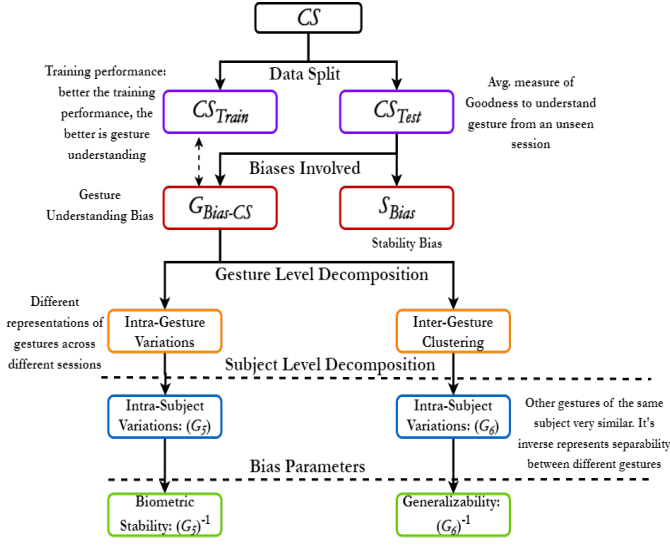


Fig. 3. Decomposition of CS evaluation for derivation of bias parameters.

matically,

$$G_{Bias-CU} \propto G_3 \times G_4 \quad (6)$$

(3) estimates biometric uniqueness for unseen subjects, while (4) formulates biometric uniqueness and stability in a set of previously seen subjects. We club them together in (7) for a more effective formulation:

$$GBQA_{CU} \propto P_{Bias} \times G_1 \times G_2^{-1} \quad (7)$$

For $G_{Bias-CU}$, we use (5) and (6):

$$G_{Bias-CU} \propto G_1 \times G_2 \times G_3 \times G_4 \quad (8)$$

Finally, using (1), (2) and (8), we derive

$$CU_{Test} \propto P_{Bias}^{-1} \times G_1^{-1} \times G_2^{-1} \times G_3^{-1} \times G_4^{-1} \quad (9)$$

(9) suggests that CU evaluation is not entirely dependent upon personality bias/biometrics, but has terms of generalizability (G_3 and G_4) also associated with it (in addition to the inter-subject variation term G_1 and the biometric stability term G_2). Thus, in order to derive $GBQA_{CU}$ from CU evaluation, one needs to decouple the generalizability terms in (9) and then perform the requisite inverse transformations on P_{Bias} and G_1 .

2.3.2. CS Evaluation Decomposition

CS gesture recognition evaluations are instrumental in assessing generalizability in gesture recognition across different sessions. We illustrate the performance decomposition for this evaluation in Fig. 3. Just as for the cross-subject (CU) case (Fig. 2, Section 2.3.1), we derive bias parameters that hinder generalizability, and develop a mathematical expression for $GBQA_{CS}$. Similar to the CU decomposition, we use both training $G_{CS-Train}$ and testing $G_{CS-Test}$ performances. In particular, performance drop during testing can be accredited to either a stability bias (S_{Bias}) or a gesture understanding bias ($G_{Bias-CS}$). Here, S_{Bias} accounts for the average variation in gestures across different sessions, while $G_{Bias-CS}$ (similar to $G_{Bias-CU}$, Section 2.3.1) estimates effectiveness of the learnt representations.

In the same spirit as the CU decomposition, $G_{Bias-CS}$ bears an inverse proportionality relationship with $G_{CS-Train}$. As far as S_{Bias} is concerned, its inverse (S_{Bias}^{-1}) encodes biometric stability for gestures from hitherto unseen sessions. Hence, it has a direct proportionality relationship with $GBQA_{CS}$.

$$CS_{Test} \propto S_{Bias}^{-1} \times G_{Bias-CS}^{-1} \quad (10)$$

$$CS_{Train} \propto G_{Bias-CS}^{-1} \quad (11)$$

$$GBQA_{CS} \propto S_{Bias}^{-1} \quad (12)$$

Exploring the cause of inadequate gesture understanding further, we apply gesture level decomposition over $G_{Bias-CS}$. This results in identification of: (i) intra-gesture variations and (ii) inter-gesture clustering as the preliminary causes. We apply subject-level decomposition over each of these. Since there is a single subject involved in these evaluation schemes, only intra-subject bias terms are relevant. To be specific, from intra-gesture variations, we derive intra-subject variation across different sessions. Denoting this by G_5 , we articulate that G_5 encapsulates inverse of biometric stability. Hence, its bias inverse operation (G_5^{-1}) estimates biometric stability, in turn. $GBQA_{CS}$ is directly proportional to the same. When inter-gesture clustering is considered, its subject level decomposition gives intra-subject inter-gesture clustering. This in turn, suggests that there exist similarities between different gestures, across sessions. We denote this by G_6 , and due to its theoretical nature, we hypothesize that its inverse (G_6^{-1}) measures generalizability. Furthermore, both G_5 and G_6 have direct proportionality relationships with $G_{Bias-CS}$.

$$GBQA_{CS} \propto G_5^{-1} \quad (13)$$

$$G_{Bias-CS} \propto G_5 \times G_6 \quad (14)$$

Then, using (12) and (13), we derive final formulation of $GBQA_{CS}$. It is evident that $GBQA_{CS}$ encodes biometric stability across seen as well as unseen sessions.

$$GBQA_{CS} \propto S_{Bias}^{-1} \times G_5^{-1} \quad (15)$$

2.3.3. Derivation of GBQA

We observe that both $GBQA_{CS}$ (Section 2.3.2) and G_2^{-1} from $GBQA_{CU}$ (Section 2.3.1) parameterize biometric stability. The former considers stability in gestures across both hitherto unseen and seen sessions, while the latter is only across the seen sessions. Thus, we consider a proportional relationship between $GBQA_{CS}$ and G_2^{-1} . Using (7) and (15), we derive the final formulation for the $GBQA$:

$$GBQA \propto P_{Bias} \times S_{Bias}^{-1} \times G_1 \times G_5^{-1} \quad (16)$$

Thus, $GBQA$ is dependent on both biometric uniqueness and stability. Furthermore, this formulation of $GBQA$ is as generic as possible, since it incorporates biometric understanding over both seen and unseen sessions (G_5^{-1} and S_{Bias}^{-1}) and subjects (G_1 and P_{Bias}), respectively. It is also essential to highlight that for the sake of simplicity, we used linear relationships between bias parameters and $GBQA$. This can be extended to any non-linear dependence with similar correspondences. **For each gesture class, we compute the $GBQA$ score. Selecting a top-scoring**

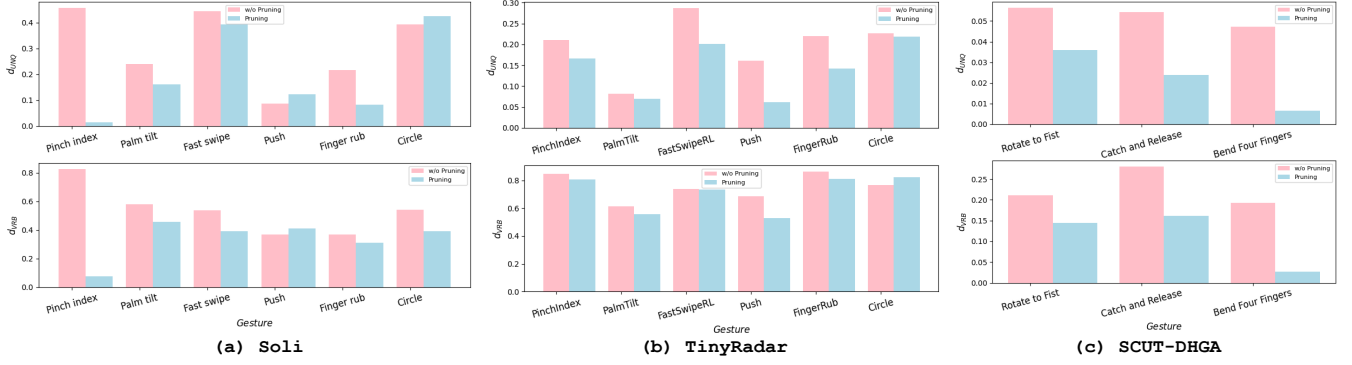


Fig. 4. Comparison between uniqueness and variability with and without gesture pruning. The first row shows uniqueness (d_{UNQ}) and the second row displays variability (d_{VRB}) [14]. d_{UNQ} computes the average inter-subject distance within a gesture cluster. d_{VRB} measures intra-subject variations for a gesture. Gesture pruning causes a significant decrease in uniqueness and variability. This validates the mitigation of $G_{Bias-CU}$ resulting in a reduction of biometric parameters. We also note that in comparison to variability, there is larger reduction in uniqueness.

Table 1. Results of gesture pruning based $GBQA$ scores. The effect of dataset pruning on CU_{Test} and $GBQA$ (EER and CU_{Test_i} are in %). The gestures with a high $GBQA$ score after pruning also correspond to a relatively lower EER. Overall, gesture set pruning reduces biases due to gesture understanding, leading to better extraction of biometric characteristics within a gesture.

Gesture	EER	Without Pruning		With Pruning	
		CU_{Test_i}	$GBQA_i$	CU_{Test_i}	$GBQA_i$
Soli [15]					
Pinch index	19.23	43.20	56.80	100.00	0.00
Palm tilt	19.20	90.40	9.60	98.40	1.20
Fast swipe	7.10	80.00	20.00	85.60	14.40
Push	14.73	96.80	0.80	96.80	3.40
Finger rub	17.60	89.60	10.40	99.20	0.80
Circle	9.10	70.40	29.60	68.00	32.00
TinyRadar [16]					
Pinch index	18.64	45.00	55.00	84.00	16.00
Palm tilt	24.21	98.00	2.00	98.00	2.00
Fast swipe	12.95	70.00	30.00	91.00	9.00
Push	17.08	90.00	10.00	99.00	1.00
Finger rub	21.31	72.00	28.00	90.00	10.00
Circle	10.33	89.00	11.00	90.00	10.00
SCUT-DHGA [5]					
Rotate to fist	7.60	99.30	0.70	100.00	0.00
Catch and release	5.19	99.40	0.60	99.80	0.20
Bend four fingers	5.13	99.80	0.20	100.00	0.00

gesture can be expected to work well for authentication purposes. Thus, this framework can be employed as an off-the-shelf component (a preliminary step) to decide which gestures are suitable for generalization and/or personalization. A gesture with low $GBQA$ score is well-suited for gesture recognition, as it would have limited biometric uniqueness (refer (16)). On the other hand, a gesture with an average $GBQA$ score would be more suitable for personalized device control: the gesture will have both generalization and biometric traits. Thus, based on the application, different thresholds for $GBQA$ score can be considered.

3. $GBQA$ Measurement

In this section we discuss strategies to measure $GBQA$. We classify the methods broadly under two heads: (i) direct measurements and (ii) relative measurements.

3.1. Direct Measurement

These techniques directly infer P_{Bias} from CU accuracy, by eliminating or minimizing $G_{Bias-CU}$ (refer (8) and (9)). In these methods, the drop in performance can be mainly attributed to P_{Bias} , but there exists some trace elements of G_1 (uniqueness) and G_2 (intra-subject variability). One possibility is to improve the representation within gestures. This could be achieved by incorporating more domain-specific representation backbones. The system of [17] introduces attention-based graph convolutions and recurrent layers for hand gesture recognition. An alternative is a robust feature representation. Metric learning-based approaches [18] typically obtain tight clusters with wide inter-class separation. This leads to elimination of bias parameters which hinder generalizability. The next section considers another approach: pruning out low-performing gestures.

3.1.1. Pruning Gestures from a Dataset

By removing low-performing gestures during training, we mitigate $G_{Bias-CU}$. Specifically, we prohibit inter-gesture clustering, thereby reducing G_3 and G_4 . Thus, when we measure CU accuracy, we measure primarily P_{Bias}^{-1} and some trace amounts of G_1^{-1} (inverse of biometric uniqueness) and G_2^{-1} (biometric stability) (refer (8) and (9)). Using the $(1-G)$ representation for the inverse of bias parameter G (Section 2.3.1), $GBQA_i$ (for gesture i) becomes

$$GBQA_i = \left(1 - \frac{CU_{Test_i}}{100}\right) \times 100 \quad (17)$$

Thus, (17) primarily estimates personality bias and remaining biometric uniqueness. To validate the proposed method, we perform experiments on three datasets namely, Soli [15], TinyRadar [16], and SCUT-DHGA [5]. For more details on dataset description and experimentation protocol, please refer Section 2 of the supplementary material.

Results of gesture pruning: In Table 1, we compare the $GBQA$ scores (refer (17)) against EER. Gestures with high $GBQA$ scores (implying higher biometric characteristics) achieve relatively low EER values on a pruned dataset. For instance, ‘Fast swipe’ and ‘Circle’ have the highest $GBQA$ score for the Soli

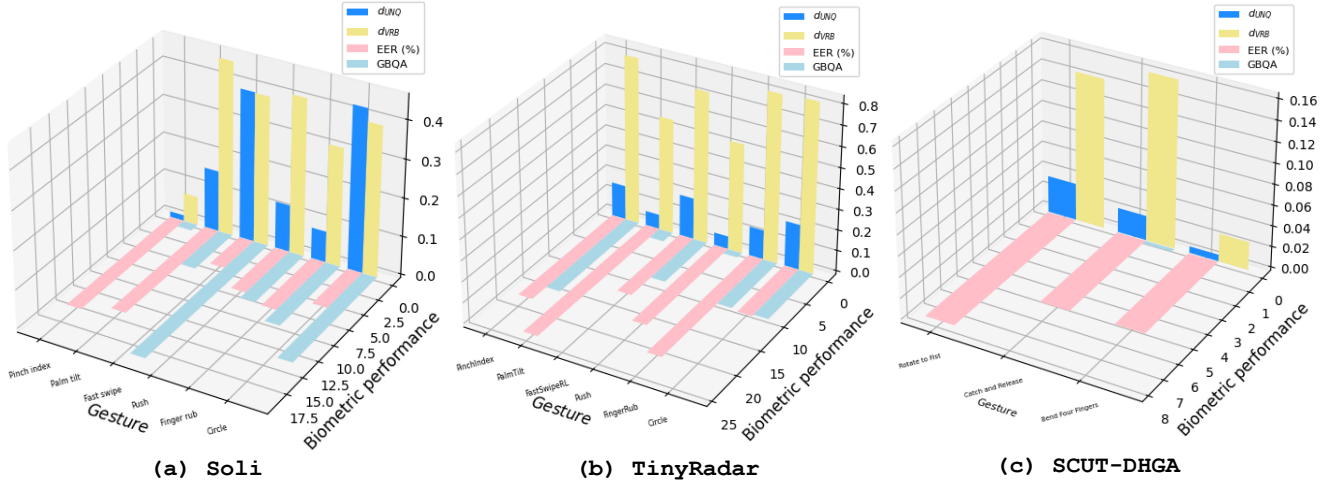


Fig. 5. Comparison between biometric uniqueness and variability. Our parameters d_{UNQ} and d_{VRB} [14] quantify uniqueness and variability, respectively. The plot reveals that for gestures having higher EER, variability dominates uniqueness. If a gesture is having higher uniqueness it tends to achieve high GBQA score.

dataset, and achieve the lowest EER. Similarly, ‘Fast swipe’ and ‘Circle’ in TinyRadar, and ‘Catch and release’ in SCUT-DHGA with lowest EER values have higher GBQA scores in comparison to other gestures. Additionally, gestures with low GBQA scores, have higher EER values. This provides an empirical justification for the theoretically decomposed structure of the GBQA score. However, for the dataset without pruning, GBQA scores are not aligned with corresponding EER. This suggests that gesture pruning is effective in mitigating generalization biases.

Impact of pruning: To analyze the significance of gesture pruning, we firstly compare the tSNE plots with and without pruning (refer Fig. S1 in the supplementary material). It can be observed that gesture pruning leads to reduction in inter-gesture clustering. This indicates that it is instrumental in mitigating $G_{Bias-CU}$. To further analyze the impact, we compare the reduction in d_{UNQ} (biometric uniqueness: average inter-subject distance) and d_{VRB} (biometric variability: average intra-subject distance) parameters [14] after pruning. A bar plot-based representation of these results is shown in Fig. 4. We note that these parameters require identity labels, whereas our formulation operates without explicit identity information (labels). We use them only for analysis purpose as they serve as proxy for G_1 and G_2 . In Fig. 4, we observe a reduction in uniqueness and variability after pruning. This is seen for all the three datasets. This highlights that with a decrease in $G_{Bias-CU}$, biometric uniqueness (G_1) and variability (G_2) are also reduced. This further validates the proposed theoretical decomposition and approach.

Failure cases: ‘Pinch index’ and ‘Finger rub’ in the TinyRadar dataset have the highest EER but surprisingly achieve high GBQA scores. To better understand this, we explore the trade-off between biometric uniqueness and variability. For each gesture, we compare d_{UNQ} (uniqueness) and d_{VRB} (variability). The corresponding results are shown in Fig. 5. It is seen from these results that these gestures encompass high variability while their uniqueness is limited. In contrast, the gestures

which perform well have significantly higher uniqueness, e.g., ‘Circle’ and ‘Fast swipe’. Thus, the major misclassifications for ‘Pinch index’ and ‘Finger rub’ can be attributed to variability which is an important parameter for generalization as well. In this case, the biometric uniqueness is a minor contributor in misclassification. Hence, the higher GBQA score can be majorly attributed to generalization characteristics. This highlights the entanglement between generalization and biometric characteristics. Another important observation is that results in SCUT-DHGA dataset are suboptimal. This is because of several reasons: (i) **limited P_{Bias} :** as the number of subjects in SCUT-DHGA dataset are significantly higher, biometric uniqueness is lower (can be observed Fig. 5 and Fig. S1 (supplementary material)). As a result, data of training subjects provides a good coverage for testing subjects. (ii) **Impact of visual cues:** In SCUT-DHGA dataset, visual cues are available. This allows the feature extractor to extract even richer features, resulting in tighter clustering (refer to Fig. S1, supplementary material). Consequently, we observe reduction in biometric uniqueness and P_{Bias} . This reduction causes the GBQA score to be suboptimal, as it primarily depends on these two parameters. Finally, it is seen from Fig. 5 that if uniqueness is higher and variability is lower then GBQA score is higher, irrespective of the dataset. This aligns with our understanding of biometric characterization.

Impact of gesture complexity in biometric characterization: From Fig. 5, we note that gestures with high complexity have high uniqueness but at the same time significantly higher variability. For e.g., ‘Pinch index’ and ‘Finger rub’ in Soli and TinyRadar dataset. Owing to the same they have lower biometric characteristics. However, because of higher variability they are susceptible to misclassification. As a result, there is a possibility that they might attain delusive but high GBQA scores.

Trade-off between biometric parameters and $G_{Bias-CU}$: Our bias decomposition postulates that with decrease in $G_{Bias-CU}$ we have reduction in biometric uniqueness (G_1) and variability

Table 2. Comparison of the proposed *GBQA* score with SOTA in terms of rank deviation. Lower rank deviation is desired. The proposed method outperforms other methods in Soli dataset, while achieves comparable results on the other datasets. Unlike other methods, it does not use identity labels.

Method	Soli	TinyRadar	SCUT-DHGA
Quality [11]	0.67	2.00	1.34
Distinctiveness [10]	1.0	1.34	1.24
Repeatability [10]	2.66	2.34	0.00
Delta distance [7]	2.0	2.34	0.67
MasterFace [12]	1.0	1.0	1.0
Generative capacity [13]	1.33	2.0	0.67
DGBQA [14]	1.0	1.34	1.34
GBQA (Proposed)	0.33	2.16	1.0

ity (G_2). The first validation of this claim can be observed in Fig. 4. With gesture pruning (effective in better gesture understanding), we find decrease in both d_{UNQ} and d_{VRB} . To further validate this, we analyze this tradeoff using: (i) better gesture recognition models and (ii) metric-learning. For a detailed experimental analysis, please refer to Section 3 of the supplementary material. The obtained results justify the theoretical basis behind decomposition.

Comparison with state-of-the-art: We compare the proposed *GBQA* with existing biometric characterization methods (Table 2). We use our rank deviation [14] as the evaluation measure. *GBQA outperforms other methods on Soli dataset, while attaining comparable results on the other datasets.* This establishes the significance of the proposed method as unlike other methods [10, 7, 12, 13, 14] it does not utilize explicit identity information (labels). For SCUT-DHGA and TinyRadar datasets, methods that utilize intra-subject distance (inverse of biometric stability) [10, 7] and inter-subject distance (uniqueness) [10, 12, 14] respectively obtain better results. This is because the former has significantly lower uniqueness, while there is higher variability in the latter (refer Fig. 5). *GBQA* score on the other hand majorly bases its estimation on P_{Bias} and biometric uniqueness (refer (17)). This highlights two challenges for *GBQA* score: (i) performing well when the number of subjects is high, because then P_{Bias} will be low. (ii) Performing in the presence of outliers which can in turn cause misclassifications, and hence lead towards delusively high *GBQA* scores.

3.2. Relative Measurement

This class of methods estimate P_{Bias} by comparative measurements i.e., they first compute the performance of the model under probe, and then mitigate or intensify the personality bias for the same model. Final results are computed and compared against the original results. This helps in estimating the associated P_{Bias} . We categorize these methods further as: bias intensification and bias mitigation methods. For further discussion on this, please refer the supplementary material.

4. Conclusion

This work introduces *GBQA*, a generic framework to quantify biometric characteristics within a hand gesture. The *GBQA* framework develops an understanding and classification of biases from cross-subject/-session gesture recognition. We quantify the entangled relationship between biometric characteristics and generalization, in a gesture recognition framework.

Our experiments give an empirical validation of the proposed models. We show results of dataset gesture pruning and its effects on bias mitigation. The proposed method outperformed several biometric characterization techniques from different domains such as online signature verification and face recognition. This validates its generalizability. This also highlighted its application in other behavioral biometrics such as gait and keystroke dynamics. Future work in the area includes quantifying each term in the *GBQA* formulation, and developing suitable linear/non-linear models.

References

- [1] W. Song, et al., TDS-Net: Towards Fast Dynamic Random Hand Gesture Authentication via Temporal Difference Symbiotic Neural Network, in: Proc. IEEE Intl. Joint Conf. on Biometrics (IJCB), 2021, pp. 1–8.
- [2] D. Thapar, et al., Recognizing Camera Wearer from Hand Gestures in Egocentric Videos, in: Proc. ACM Intl. Conf. on Multimedia, 2020, pp. 2095–2103.
- [3] O. Mendels, et al., User Identification for Home Entertainment based on Free-Air Hand Motion Signatures, IEEE Transactions on Systems, Man, and Cybernetics: Systems 44 (2014) 1461–1473.
- [4] M. Shahzad, et al., Behavior based Human Authentication on Touch Screen Devices using Gestures and Signatures, IEEE Transactions on Mobile Computing 16 (2016) 2726–2741.
- [5] C. Liu, et al., Dynamic-Hand-Gesture Authentication Dataset and Benchmark, IEEE Transactions on Information Forensics and Security 16 (2020) 1550–1562.
- [6] A. K. Jain, et al., Biometrics: Trust, but Verify, IEEE Transactions on Biometrics, Behavior, and Identity Science 4 (2021) 303–323.
- [7] A. Zunino, et al., Revisiting Human Action Recognition: Personalization vs. Generalization, in: Proc. International Conference on Image Analysis and Processing, 2017, pp. 469–480.
- [8] S. Hazra, et al., Cross-Modal Learning of Graph Representations using Radar Point Cloud for Long-Range Gesture Recognition, in: Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM), 2022, pp. 350–354.
- [9] J. Wu, et al., Leveraging Shape and Depth in User Authentication from In-Air Hand Gestures, in: Proc. IEEE Intl. Conf. on Image Processing (ICIP), 2015, pp. 3195–3199.
- [10] N. Sae-Bae, et al., Distinctiveness, Complexity, and Repeatability of On-line Signature Templates, Pattern Recognition 84 (2018) 332–344.
- [11] M. Santopietro, et al., Assessing the Quality of Swipe Interactions for Mobile Biometric Systems, in: Proc. IEEE Intl. Joint Conf. on Biometrics (IJCB), 2020, pp. 1–8.
- [12] P. Terhörst, et al., On the (limited) Generalization of Masterface Attacks and its relation to the Capacity of Face Representations, in: Proc. IEEE Intl. Joint Conf. on Biometrics (IJCB), 2022, pp. 1–9.
- [13] V. N. Boddeti, et al., On the Biometric Capacity of Generative Face Models, in: Proc. IEEE Intl. Joint Conf. on Biometrics (IJCB), 2023.
- [14] A. Verma, et al., Quantifying Biometric Characteristics of Hand Gestures Through Feature Space Probing and Identity-Level Cross-Gesture Disentanglement, in: Proc. IEEE Intl. Conf. on Automatic Face and Gesture Recognition (FG), 2024, pp. 1–9.
- [15] S. Wang, et al., Interacting with Soli: Exploring Fine-Grained Dynamic Gesture Recognition in the Radio-Frequency Spectrum, in: Proc. Symposium on User Interface Software and Technology, 2016, pp. 851–860.
- [16] M. Scherer, et al., TinyRadarNN: Combining Spatial and Temporal Convolutional Neural Networks for Embedded Gesture Recognition with Short Range Radars, IEEE Internet Things J. 8 (2021) 10336–10346.
- [17] J.-H. Song, et al., Dynamic Hand Gesture Recognition using Improved Spatio-Temporal Graph Convolutional Network, IEEE Transactions on Circuits and Systems for Video Technology (2022).
- [18] M. Kim, et al., Adaface: Quality Adaptive Margin for Face Recognition, in: Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 2022, pp. 18750–18759.