Biometric Benchmark based on Handwriting and Hand Geometry Modalities

RanjiRaj Rajendran Nair Study Program INF; DKE Immatriculation no. 226070 ranjiraj.nair@st.ovgu.de Darshit Paresh Shah Study Program INF; DKE Immatriculation no. 225866

darshit.shah@st.ovgu.de

Abstract

Biometrics offers a natural and more promising solution to the identification of humans. Since, human beings can produce more than one modality its more difficult to mimic, share or forge these traits.

This report document encompasses a final presentation on a set of small-scale biometric experiments on the modalities handwriting (handwriting-based user authentication) and hand geometry (hand geometry-based user authentication). The experiments are performed by simple paper and pencil-based work followed by data acquisition, preprocessing and evaluation of the collected data.

1. MOTIVATION

Biometrics is a unique measurement of a person's physical and behavioral characteristics. It is a more secure technique as it is directly linked with the person's biological and physical behavior and because of this misuse by any third party is more difficult.

Some of the modalities in biometrics are as follows:

- Voice: This analyzes the voice of a person as a source of authentication. This can help in speech recognition, emotion detection. It is widely used by users as it is cost-effective to integrate into devices such as home appliances, automobiles, etc.
- Gait: This modality was inspired by the experimental observation where individuals were able to identify other people known to them by looking at the projection of silhouettes of their body movements.
- Keystroke Dynamics: This modality focuses on the timing patterns of a user during typing on a keyboard. This modality is also very useful as all the users have a different style of typing and therefore their timing pattern is also unique to themselves.
- 4. Fingerprint: This is a well-known and used modality. This takes the skin structure of fingertips of a user as a source of authentication which is unique to every human being. Nowadays three levels of structures can be derived from a fingerprint image.

- 5. Iris: Iris is the most visible part of human eyes. Each iris is unique and has more than 250 features (Color and Texture). It remains constant even with aging and under various environments. It is mostly used in military bases for user authentication.
- Face: This modality is accepted by many users. It is fast, requires no user assistance, contactless, fast capturing. It is also used in smartphones, to recognize age, gender, and emotion.
- 7. Hand: Hand biometrics contains palm print, hand geometry, vascular biometrics (Vein Structure). Palm print means individual structures in palm skin with three principal lines. Hand geometry is the geometrical features of the entire hand. It has high accuracy than others. Vascular biometrics is the biometrics of the vein structure in our hands as it has a unique individual structure. This is still in the research area
- 8. Handwriting: This is a process of identifying a person based on his/her handwriting. These biometrics belongs to the behavioral biometrics system as it is based on something that the user has learned to do. This can further be split into two categories:
 - Static: In this, a user writes on a paper and then digitize it by using an optical scanner or a camera. This is also called an off-line method.
 - Dynamic: In this, a user writes on a digitized tablet or a stylus-operated PDA, which analyzes the text in real-time.

2. MODALITY 1: HANDWRITING

For the individual task of data acquisition two different semantic class were chosen:

- a) Based on a pin "77412"
- b) Based on Hometown

Each user has provided with 5 enrollment samples and 5 verification samples as test data. The above samples were collected using a grid template and was instructed to write as naturally as possible. Based on the data provided 7 distinct feature vectors (Table 1) were asked to compute for the samples.

Data Acquisition:

Step 1: Write one single sample of the semantic chosen in the gridded writing area.

Step 2: Construct the Bounding Box (BB) around the handwriting sample, taking the lower-left corner as the origin of ordinates for all feature measurements.

Step 3: Take measurements and calculate the given features (Table 1)

Step 4: Fill in the resulting feature vector and continue with subsequent samples.

Table 1

Feature	Description	fv
Vector Name		component
Aspect Ratio	BB width (mm)*1000 / BB height (mm) *1000	fV ₁
Segment Count	Number of continuous line segments in the sample (including isolated points)	fv ₂
Baseline Angle	(Approximated) Baseline Angle of the sample (in degrees from 0° to perfectly horizontal)	fv ₃
Loop Count	Total number of closed loops in sample	fV ₄
Y-max Count	Total number of vertical local maxima in the y-direction	fv ₅
Y-min Count	Total number of vertical local minima in the y-direction	fv ₆
Intersection Count	Number of intersection (crossings) in sample	fv ₇



Fig 1. Written Pin on Grid Template

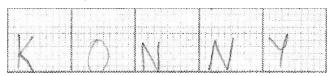


Fig 2. Written Hometown on Grid Template

Applied rules for feature vectors:

Table 2

Table 2						
Feature Vector Name	Customized Measurement Rule					
Aspect Ratio	The horizontal distance (width) of the entire string measured by a ruler and vertical distance (height) from lower to upper corner within the bounding box.					
	CABACIMHA					
Segment Count	Total count of straight lines encountered in each complete string.					
	PHMEDABAD					
Baseline Angle	The angle at which the string is rested upon the lower part of the bounding box.					
	AHMFJARAD					
Loop Count	Detecting where a complete cycle is formed.					
	CHOEDODY)					
Y-max Count	A horizontal line is drawn on the bounding box for the character that touches it in the upper region.					
	PHMEDAGAG					
Y-min Count	A horizontal line is drawn on the bounding box for the character that touches it in the lower region.					
	DAMEDOOVI)					
Intersection Count	Any two lines that crossovers each other, the resultant is considered to be a count.					
	COMED AROD					

Intra-Class Analysis:

In this protocol, we check how strongly does the values of each feature vector for each sample resembles each other. Furthermore, this was carried out as an individual task for each semantic class.

For this experimentation, the data acquisition values were plotted onto Microsoft Excel spreadsheet according to the layout (row-wise enrollment feature vectors to the left; verification feature vectors column-wise on top).

	fv1	fv2	fv3	fv4	fv5	fv6	fv7
e1	10,5	1,9	9,1	13,4	1,15	23,1	13,4
e2	10,5	1,9	9,2	13,2	0,85	24,6	12,8
e3	10,5	2	9,9	13,4	1,1	23,5	13,7
e4	9,5	1,7	9,6	13	1,15	23,5	13,7
e5	10.5	1.9	9.4	13.1	1.1	23.2	13.4

Fig 3. Enrollment sample table

	v1	v2	v3	v4	v5
fv1	10,5	10,5	10,5	10,5	10,5
fv2	1,5	1,7	1,6	1,6	1,7
fv3	9,9	9,4	9,4	9,3	9,4
fv4	12,5	13	12,8	12,5	12,9
fv5	1,15	1,15	1,15	1,1	1,2
fv6	23,6	25,4	24,1	23,7	24,3
fv7	13,2	13,4	13,5	12,7	13,4

Fig 4. Verification sample table

Based on the above work an Intra-Class Distance Scatter matrix is computed. For the calculation, **Minkowski Metrics** were used with r=2 which formulated to **Euclidean Distance**

$$d_{ij} = [\sum_{1}^{k} |x_{ij} - x_{jk}|^r]^{1/r}$$

The implementation of 7-dimensional Euclidean distance was done by the following formula:

=SQRT((\$B16-J\$6)^2+(\$C16-J\$7)^2+(\$D16-J\$8)^2+(\$E16-J\$9)^2+(\$F16-J\$10)^2+(\$G16-J\$11)^2+(\$H16-J\$12)^2)

Where:

- fv1 of e1 located in cell B16
- fv1 of v1 located in cell j6 and
- The upper-left element of the scatter matrix is located in J16

2,254248	2,544502	1	1	3,060612	3,742869	2,803969	3,78504	2,045403	3,510916
2,14021	2,833333	1,759973	1,759973	2,788867	3,270236	2,645751	3,5	3,13717	4,272002
3,156981	2,586503	2,705852	2,705852	1,428286	2,853966	2,115299	2,842821	2,005096	2,842821
5,412099	4,123106	5,30931	5,30931	2,5	2,853966	2,115299	2,842821	2,005096	2,842821
2,812363	2,808914	3,081081	3,081081	2,009975	3,527668	2,5	4	3,894004	4,472136
2,842821	2,603417	3,97995	6,5	3,773592	4,409586	4,409586	3,357734	3,357734	3,357734
3,619048	3,605551	3,761796	5,035982	3,388871	2,872281	3,166667	1,96006	1,96006	1,96006
2,005096	2,533226	3,325719	6,057968	4,164868	3,316625	3,574602	1,737932	1,737932	1,737932
2,005096	2,533226	3,325719	5,718302	3,653235	1,010153	2,533226	2,236068	1	2,236068
2,005096	2,533226	3,325719	6,057968	4,164868	2,236068	4,333333	1,010153	2,240627	1,010153

Fig 5(a) Scatter matrix for Pin

10,09224	11,71537	11,95366	12,05197	11,3901	22,35369	22,84927	22,8147	22,14747	22,97145
5,616593	10,87014	10,53755	10,96175	7,729884	23,22951	23,55844	23,5034	22,98717	23,57138
6,819933	10,55474	10,15197	10,45957	9,064966	23,90418	24,3072	24,25469	23,67045	24,31892
6,037299	12,20819	11,40175	11,70641	8,931841	24,28189	24,63737	24,58475	24,05016	24,69028
5,877595	10,91604	10,58489	10,82405	7,533333	22,39913	22,92781	22,88379	22,2636	23,01567
23,75479	20,60679	23,44275	22,75961	20,53387	7,321885	6,588205	6,9	7,012365	7,012365
22,83311	19,85282	22,63862	22,24922	19,77712	9,013323	8,41434	9,013323	8,987474	8,987474
23,79685	20,31354	23,48537	22,80351	20,14051	8,993887	8,379605	8,993887	8,966388	8,966388
23,29247	19,88547	22,93027	22,24263	19,75935	7,930322	7,226187	7,803204	7,771494	7,771494
23,29247	19,88547	22,93027	22,24263	19,75935	9,033825	8,210021	8,695401	8,784832	8,784832

Fig 5(b) Scatter matrix for Hometown

Inter-Class Analysis:

In this protocol, the similarity of values in the feature vector of both users is compared which gives a resultant Inter-Class Scatter matrix.

On the creation of the scatter matrix some observations are drawn:

- Size [10 x 10]
- A subrange of Intra-class distance values [5.61, 12.20]
- A subrange of Inter-Class distance values [19.77, 24.69]
- Redundancy in the matrix Repeated values due to measurement
- Type of redundancy Data duplication

Statistical Analysis (FAR/FRR/EER error rate diagram):

<u>False Acceptance Rate (FAR)</u>: Ratio between the number of false acceptances of non-authentic persons and the total number of tests.

<u>False Rejection Rate (FRR)</u>: Ratio between the number of false rejections of authentic persons and the total number of tests.

Equal Error Rate (EER): FAR and FRR are identical.

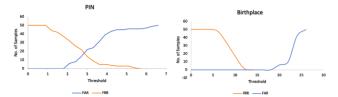


Fig 6(a) Error plot: Pin Hometown

Fig 6(b) Error plot:

Conclusions:

- For Pin: For 50 samples the observed EER is 14
- For Hometown: For 50 samples the observed EER is 0

Forgery Attempts:

Forgery experiments are also carried out as part of this task to check how each user can well spoof or fake each other's modality. To check this the semantic class of hometown was taken into account since it gives a better basis for statistical evaluation. Additionally, three different forgery cases were performed:

- a) With no knowledge The user does not mention their hometown.
- b) With partial knowledge User verbally mentions about their hometown.
- with visual appearance knowledge User discloses their hometown.

3. MODALITY 2: HAND GEOMETRY

In this task, we determined our hand geometry code. These codes can be used for identification in a variety of situations.

For the task, we collected 5 enrollment samples and 5 verification samples for each. The samples were collected on an A4 sheet and we were provided with the rules for extraction of code. Here 5 samples were generated by self and the other 5 samples were traced by my teammate and the same procedure was carried out for my teammate's sample. The rules are as follows:

Table 3

Feature Vector Name	Pre-set Measurement Rule
А	Distance from index fingertip to bottom knuckle
В	Width of the ring finger, measured across the top knuckle.
С	Width of palm across 4 bottom knuckles
D	Width of palm from the middle knuckle of thumb across the hand

- There were also some basic rules which we followed while doing this task like hand should be placed properly on the paper and fingers should be well spread across the paper, we chose to trace the right hand and the tracing should be done by keeping the pencil as close to the skin as possible. Then the measurements were taken using a ruler in centimeters.
- We performed the task with the above rules and then generated the personal hand geometry code by placing the values of the measurement in the sequence of A B C D. For example if the values are as follows: A = 8.5, B = 1.7, C = 9.5, D = 12.3 then the biometric code will be 8.7 1.7 9.5 12.3

Table 4

Feature Vector Name	Customized Measurement Rule
Е	Thumb Radius
F	Distance between fingers from fingertip
G	Distance between bottom knuckles

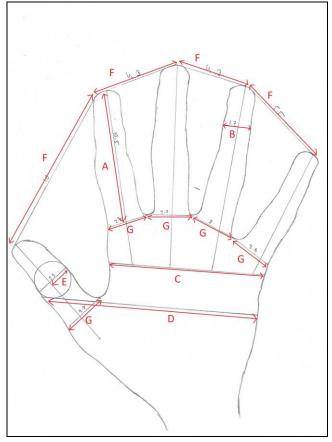


Fig 7. Hand Feature Extraction

1,26885775	0,53851648	0,73484692	0,96953597	0,6164414	3,9179076	4,38406204	5,13711982	4,94873721	4,99199359
1,06770783	0,34641016	0,53851648	0,76811457	0,41231056	3,76563408	4,22492603	4,98397432	4,80624594	4,84664833
1,02956301	0,70710678	0,87749644	1,15325626	0,76811457	4,18090899	4,62493243	5,41664103	5,23641098	5,2848841
1,17473401	1,0198039	1,04403065	1,161895	1,02469508	3,43802269	3,9	4,72863617	4,45196586	4,5177428
0,87749644	0,2236068	0,42426407	0,678233	0,28284271	3,73764632	4,18807832	4,95883051	4,78852796	4,82700735
3,15753068	3,39263909	3,21869539	2,94278779	3,30756708	0,41231056	0,76157731	1,61554944	1,3892444	1,40712473
2,45356883	2,65706605	2,49599679	2,25166605	2,57875939	1,18321596	1,61554944	2,46170673	2,17715411	2,22485955
2,79642629	2,97657521	2,81247222	2,55147016	2,89654967	0,78740079	1,23693169	2,06397674	1,78325545	1,82482876
2,81247222	3,17962262	3,00333148	2,74590604	3,09192497	1,14455231	1,36381817	2,19317122	2,00748599	2,04450483
2.80713377	3.06267857	2.89309523	2.62868789	2.97825452	0.78740079	1.17046999	2.02484567	1.77763888	1.83030052

Fig 8. Scatter Matrix for Hand Geometry

Inter-Class Analysis:

On the creation of the scatter matrix some observations are drawn:

- Size [10 x 10]
- A subrange of Inter-Class distance values [2.2, 5.4]
- Redundancy in the matrix Repeated values due to measurement
- Type of Redundancy Data duplication

Intra-Class Analysis:

On the creation of the scatter matrix some observations are drawn:

- Size [10 x 10]
- A subrange of Intra-class distance values [0.2, 2.4]
- Redundancy in the matrix Repeated values due to measurement
- Type of Redundancy Data duplication

Conclusion:

For 50 samples the observed EER is 1.

After adding additional features following are the observations:



Fig 9. Scatter Matrix after adding new features

Inter-Class Analysis:

On the creation of the scatter matrix some observations are drawn:

- Size [10 x 10]
- A subrange of Inter-Class distance values [2.4, 6]
- Redundancy in the matrix Repeated values due to measurement
- Type of Redundancy Data duplication

Intra-Class Analysis:

On the creation of the scatter matrix some observations are drawn:

- Size [10 x 10]
- A subrange of Intra-class distance values [0.8, 3.1]

- Redundancy in the matrix Repeated values due to measurement
- Type of Redundancy Data duplication

Conclusion:

With the addition of new feature vectors for 50 samples, the observed EER was increased to 2.

4. SUMMARY OF THE RESULTS AND COMPARISON OF BOTH MODALITIES

From the graphs obtained from FAR/FRR/EER plots, the EER observed in handwriting (EER = 0) was less as compared to hand geometry (EER = 1). After adding new features to the hand geometry, the EER was increased to 2.

From the graphs obtained by FAR/FRR/EER plots, the EER observed for the case of attempting forgery without any knowledge yielded 5, with full knowledge the ERR narrowed to 3 and with little knowledge EER dropped to 0. From this, we can infer that the forgery attempt was moderate.

Based upon the results achieved the results were projected onto the Doddington's Zoo concept which assigns the resulting classification of users into several categories labeled as:

<u>Sheep</u>: Well behaved users exhibiting low FRR and features well-separated from other users.

<u>Goats</u>: Difficult to identify and thus leads to extreme degradation in performance by increasing FRR.

<u>Lambs</u>: Features that overlap with the rest of the users attaining high FAR.

<u>Wolves</u>: Capability to mimic the biometric characteristics of other users.

Our results yielded a classification to Lamb and Wolf and the behavior was seen to be overlapping because a very small set of samples were taken for experimentation.

5. SHARES

Chapters	Written By					
Abstract	RanjiRaj Rajendran Nair					
1	Darshit Paresh Shah					
2	RanjiRaj Rajendran Nair					
3	Darshit Paresh Shah					
4	RanjiRaj Rajendran Nair, Darshit Paresh Shah					
5	Darshit Paresh Shah					
6	RanjiRaj Rajendran Nair, Darshit Paresh Shah					
Appendix A	RanjiRaj Rajendran Nair					
Appendix B	RanjiRaj Rajendran Nair					
Appendix C	Slide 1-9: Darshit Paresh Shah,					
Appendix o	Slide 9-23: RanjiRaj Rajendran Nair					
Appendix D	RanjiRaj Rajendran Nair, Darshit Paresh Shah					

6. LITERATURE & FURTHER SOURCES

- [1] Kraetzer, C., and Dittmann, J. 2018. Student task description. Course "Biometrics and Security" Winter term '18 / '19. Otto-von-Guericke University of Magdeburg, Dept. of Computer Science, 2018.
- [2] Abir Mhenni, Estelle Cherrier, Christophe Rosenberger, Najoua Essoukri Ben Amara. Analysis of Doddington Zoo Classification for User Dependent Template Update: Application to Keystroke Dynamics Recognition. Future Generation Computer Systems, Elsevier, 2019.hal-02050173
- [3] Mohammad Nayeem Teli, J. Ross Beveridge, P. Jonathon Phillips, geof H. Givens, David S. Bolme, Bruce A. Draper. Biometrics Zoos: Theory and Experimental Evidence*

- [4] Chapran, J. (2006). "Biometric Writer Identification: Feature Analysis and Classification". International Journal of Pattern Recognition & Artificial Intelligence.
- [5] Schomaker, L. (2007). "Advances in Writer Identification and Verification" (PDF). Ninth International Conference on Document Analysis and Recognition.
- [6] Said, H. E. S.; TN Tan; KD Baker (2000). "Personal identification based on handwriting". Pattern Recognition.
- [7] Schlapbach, A.; M Liwicki; H Bunke (2008). "A writer identification system for on-line whiteboard data". Pattern Recognition.
- [8] Sesa-Nogueras, Enric; Marcos Faundez-Zanuy (2012). "Biometric recognition using online uppercase handwritten text". Pattern Recognition.
- [9] From, https://www.aware.com/voice-authentication/, website request 27.01.2020

APPENDIX A – DOCUMENTATION OF THE DATA GENERATED / USED

1. Sensor/Source description:

Non-sensor techniques were used to collect the data for the experimentation procedures.

Sources used for Handwriting task:

A sheet of grid template provided by the instructor for writing the data and material used for writing is a normal pencil by which the sample can be erased off easily in doubt of quality. For the measurement of the feature vectors, a ruler of 15 cm was used.

Sources used for Hand Geometry:

A sheet of white A4 paper with GSM: 80 g/m², ISO: 70 and dimensions: 210 x 297 mm. The tracing of hand was done by a normal pencil. For the measurement of the non-linear surfaces of the hand, a ruler of 15 cm was used. Also, for certain additional features (like the measure of thumb) a compass was used.

2. Acquisition (capturing and sampling):

Measurement techniques for Handwriting task:

For data capturing the customized measurement rule as stated in Table 2 was followed.

For each of the semantic class (pin and hometown), 5 enrollment and 5 verification samples were collected which was further fed into Microsoft Excel for the basis for matrix calculation.

Measurement techniques for Hand Geometry task:

For data capturing the customized measurement rule as stated in Table 4 was followed.

For each pattern of hand geometry, 5 enrollment (traced by the user itself) and 5 verification (traced by another user) samples were collected which was further fed into Microsoft Excel for the basis for matrix calculation.

3. Acquisition Protocol and pre-processing steps performed:

The general evaluation protocol followed in the course description was:

- Data Acquisition (individual task)
- Intra-Class Analysis (individual task)
- 3. Inter-Class Analysis and Forgeries (team task)
- 4. Result comparison (course task)

Since the value ranges were very much wider, **binning** the value into a pre-selected threshold was used as a pre-processing step which provides the basis for FAR/FRR/EER plots.

The values from the scatter matrix for Intra-Class distances and Inter-Class distances were sorted in an ascending manner. Then identifying the value sub-ranges from lower to higher was made. Then depending upon the value ranges the optimal threshold was determined (e.g. 0.3 or 0.5) which gave the count of values lying in that range. This set of new bin values gave the basis for FAR and FRR values which later was used for FAR/FRR/EER plots.

APPENDIX B – PROBLEMS ENCOUNTERED AND OPEN ISSUES

- Data acquisition was carried out using a normal paper and pen/pencil there are some inconsistencies in the values.
- As there were no pre-standard measurements for the computation of feature vectors the notion of defining features differs.
- For a team size of 2, only fewer data samples were taken for result analysis which would not draw any significant conclusions for classification and visualizations.
- There was a large variation in the value ranges for calculation in scatter matrix which was caused by the ethical and cultural background of the users which required pre-processing into multiple bins to generate graphs for interpretations.

APPENDIX C - PRESENTATION SLIDES

Shared as a separate file.

APPENDIX D - STATEMENTS ON PANEL SESSION PARTICIPATION

Active Panel session statement

RanjiRaj Rajendran Nair:

(Panel 4: Keystroke Dynamics on December 4, 2019; Role: Data Protection Officer)

Statement: Depending upon the region or country the usage of keylogging software requires legal advice/consent before implementing, collecting or experimenting from the residents.

(Panel 9: Fusion of Active and Passive Modality for user authentication on January 15, 2020; Role: Contra Modality)

Statement: (Active modality: Voice; Passive modality: Face) The active modality can be mimicked by some voice modulation techniques which can fail at the primary level of the multi-biometric system. However, the introduction of a second passive modality can provide a stronger guard against any threat or attack. If this authentication is carried out in an uncontrolled environment the impact of weather and poor illumination can compromise.

Darshit Paresh Shah:

(Panel 4: Keystroke Dynamics on December 4, 2019; Role: Pro Modality)

Statement: Yes, it can be implemented in the ATMs keypad. So, while withdrawing money from the ATM the machine asks for the pin and then we can identify the user through the keystroke dynamics.

(Panel 7: Face image-based user authentication on January 15, 2020; Role: ABC Gates)

Statement: This modality has the potential to be used in ABC Gates as the person with the biometric passport will have their face data in it so while entering the gates the camera will scan the face and authenticate it with the face data present in the biometric passport.

Passive Panel session statement

RanjiRaj Rajendran Nair:

(Panel 7: Face image-based user authentication on January 15, 2020)

Statement: Face images remain unique and distinct for two individuals even there will a slight variation in the feature vectors for conjoint twins. Images can be captured from long stand-off distances with biometric cameras using non-contact sensors which can also portray user emotions ethnicity and biographic information.

(Panel 8: Fusion of two Active modalities for user authentication on January 15, 2020)

Statement: (Modalities: Voice and Keystroke Dynamics) A multi-biometric system with a two active layer modality can be a more powerful mechanism for authentication. The chosen modalities can be applied over a telephonic verification system where the user is asked to reveal some pin, as well as simultaneously, would type the same pin to ensure that data is secured instead of some robotic voice.

(Panel 6: Iris image-based for user authentication on January 8, 2020)

Statement: By giving away the iris data of a user it can disclose information about the individual's health condition. Moreover, just by looking at the image of some random iris the gender, age, and other such information are hard to derive.

Darshit Paresh Shah:

(Panel 2: Handwriting-based user authentication on November 13, 2019)

Statement: Handwriting behaviour can change due to the skill set of a person, also by age and more by weariness which can severely affect the data acquisition.

(Panel 3: Gait-based user authentication on November 27, 2019)

Statement: Walking manner of a person can change due to various health situations. Also, if there is any knee surgery done for that person then gait changes.

(Panel 5: Fingerprint-based user authentication on December 11, 2019)

Statement: People doing rigorous work can have degraded finger which makes unsuitable for performing fingerprint-based authentication.