Subscription Fraud Prevention in Telecommunication Using Multimodal Biometric System

Freddie Mathews Kau[1] and Okuthe P. Kogeda[2(✉)]

[1] Tshwane University of Technology, Pretoria 0001, South Africa
[2] University of the Free State, Bloemfontein 9300, South Africa
kogedapo@ufs.ac.za

Abstract. South African telecommunications market has reached a saturation point; as a result, telecommunication companies spend most of their budget on customer acquisition and retention, and very little is spent on fraud prevention or detection systems. This spending pattern has caused an increase in fraud, making it the most significant revenue leakage in telecommunications, where the leading fraud type is subscription fraud. Subscription fraud has a direct negative impact on the company's revenue, bonuses of employees, and customers' credit status. Although the current fraud systems can detect subscription fraud, they cannot identify the fraudster. This enables the fraudster to commit fraud using the same or multiple identity documents during the contract application process without being detected. In trying to change the spending pattern and prevent subscription fraud, we sought to determine the impact of subscription fraud in mobile telecommunication companies. We designed, developed, and implemented a Multimodal Biometrics System (MBS) using Python, SQLite3, and JavaScript to enable telecommunication companies to capture and store customer faces and fingerprints to use them for verification before approving the contract. We used Principal Component Analysis (PCA) algorithm to reduce the dimension of the face and fingerprint images. PCA outperformed Independent Component Analysis and Linear Discriminate Analysis algorithms. To do image matching, we used the PCA-based representation for local features (PCA-SIFT) algorithm, which outperformed Scale Invariant Feature Transform (SIFT) and Oriented FAST and Rotated BRIEF (OBR) algorithms. MBS results gave us biometric matching accuracy of 94.84%. MBS is easy to implement and cost-effective. The system can help identify the fraudster, prevent subscription fraud and reduce revenue leakage.

Keywords: Subscription Fraud · Telecommunication · Fingerprint · Biometrics · Face Biometrics · PCA Algorithm · Multimodal Biometrics System

1 Introduction

South African telecommunication market comprises well-known Information Communication Technology (ICT) companies: Telkom and Neotel, mainly fixed-line or fixed wireless providers, and Mobile Telephone Network, Vodacom, and Cell C, primarily

R. A. Saeed et al. (Eds.): AFRICOMM 2022, LNICST 499, pp. 458–472, 2023.
https://doi.org/10.1007/978-3-031-34896-9_28

mobile telecommunication providers. Fixed-line has over 1.8 million subscribers, and mobile telecommunication has over 103 million subscribers [1]. South African demographics encompass about 59.62 million people [2]. This shows that mobile telecommunication in South Africa (SA) has changed from a fast-growing to a saturation state.

With this saturation and high usage of cellular phones come several challenges like revenue growth and fraud. Casey et al. [3] warned us that as people and organizations become more dependent on mobile phones, computer criminals focus more on how they can victimize individuals and break into corporate networks. The growth and the high usage of technology do not tame fraud. KPMG [4] points out that technological advances are a double-edged sword that provides companies with more-powerful, more defensive tools in defence against fraud and fraudsters to discover areas of vulnerability to attack.

Fraud in mobile telecommunication is defined as illegal access to a mobile operator's network and the use of its services for unlawful interest to the detriment of the network operator and/or its subscribers [5]. Wieland, cited by [6], indicates that fraud is the telecommunication industry's most extensive, most significant revenue leakage area. The survey by [7] shows that telecommunications companies lost over $38 billion due to fraud. The report further lists the top five fraud methods: subscription fraud (SF), private branch exchange (PBX) hacking, dealer fraud, service abuse, and account takeover. SF is the main contributor to revenue leakage in mobile telecommunication. The impact of subscription fraud is enormous [22].

SF is defined as acquiring telecommunications services using fraudulently obtained subscriber documents or false identification [8]. Both prepaid and postpaid subscribers can be victims of SF. The fact that SF uses fake identity makes it the most concerning fraud because mobile telecommunication companies cannot confidently identify subscribers beyond a reasonable doubt. "Mobile devices such as cell phones and smartphones have become an integral part of people's daily lives, and as such, they are prone to facilitating criminal activities or otherwise being involved when crimes occur" [3]; this means that incorrect identification of a subscriber is indirectly defeating the ends of justice.

Kabari et al. [9] state that its complexity makes it hard to detect subscription fraud. According to [4], weak fraud controls are the biggest issue for companies victimized by fraud. They further state that companies are not investing in more robust anti-fraud controls mainly because of economic challenges. Another problem mobile telecommunication faces is that one fraudster can commit multiple SFs. In their report, Estevez et al. [6] discovered that a fraudster committed seven fraud cases within three months without being detected.

Telecommunication fraud is the number one enemy in the industry mainly because, with other revenue losses, companies can still recover money from subscribers. In contrast, a company loses both the cost and expected revenue with fraud. Money lost through fraud can be irrecoverable, especially if the fraud committed is below the fraud insurance access.

This background showed us that there was a need for a system that could prevent SF and proactively stop multiple SFs from the same fraudster. We sought to avoid SF and identify a fraudster beyond a reasonable doubt using multimodal biometric

traits: fingerprints and face. Therefore, our proposed system, connected to a customer relationship management (CRM) system, was used to conduct identity verification using a fingerprint biometrics scanner for Fingerprint and a webcam for the face. Our research was conducted on South African mobile telecommunication focusing on Postpaid SF.

The rest of the paper is organized as follows. In Sect. 2, an overview of subscription fraud is provided. In Sect. 3, biometrics is described. In Sect. 4, an outline of the research methodology is provided. In Sect. 5, we discuss testing results. In Sect. 6, the conclusion and future work is presented.

2 Overview of Subscription Fraud

Subscription fraud is the acquisition of telecom services using fraudulently obtained subscriber documents or false identification [8]. A fraudulently obtained identity document here is defined as identity theft which occurs when a fraudster steals the identity documents of an existing person without their consent to commit fraud [10]. False identification, defined as identity fraud, occurs when a fraudster creates an identity document of a non-existing person [10].

There are three types of SF methods named by [11]:

a) **SF (Identity)** is the utilization of an actual identity without the knowledge of the owner to obtain goods and services with no intention to pay; SF (Identity) contribution to revenue loss was 2 billion USD in 2016 [11].
b) **SF (Application)** is a creation of false details to gain access to goods and services with no intention to pay. SF (Application) contributed 1.9 billion USD to revenue loss in 2016 [11].
c) **SF (Credit Mulling/Proxy)** is the utilization of actual identity details to obtain goods and services with no intention to pay, SF (Credit Mulling/Proxy) contributed 1.8 billion USD to revenue loss in 2016 [11].

3 Biometrics

Biometrics which is made up of two Greek words, *bios* meaning life and *metrics* meaning measure, is defined as the measure of physical or behavioural biological characteristics of a person [12]. For many decades, biometrics characteristics have been used for identification and authentication mainly because they are reliable in determining a person's identity compared to tokens and ID cards, which can be misplaced or shared [13].

There are two characteristics of biometrics [14]:

i. **Behavioral Traits**
   Behavioral traits are biometrics traits that are related to the behaviour of a person. Examples of the commonly used are voice print and signature dynamics [15].
ii. **Physiological Traits**
   Physiological traits are biometrics traits related to the body's shape and differ from person to person; The commonly used physical traits are [15]: face, Fingerprint, iris pattern, and DNA, while the most prevalent behavioural traits are: voice and signature dynamics. In South Africa, the department of home Affairs (DHA) also

uses two biometric traits: face and Fingerprint, to issue ID documents [16]. For this reason, we chose to use Fingerprints and faces in our study because they can be verified with the central database of DHA.

a) **Face**

Face traits are less intrusive and commonly used in biometrics; as humans, we recognize each other by face. Most algorithms have a high failure rate caused by facial expressions, different angles, and slow performance when retrieving images from the DB [17].

b) **Fingerprint**

The Fingerprint is the mother biometric and the most widely used biometric. What makes fingerprints robust is that they remain constant for a person's entire life. The advent of several inkless fingerprint scanning technologies coupled with the exponential increase in processor performance has taken fingerprint recognition out of criminal identification applications to several civilian applications such as access control, time and attendance, and computer login [18].

iii. **Multimodal and Unimodal Biometrics**

Multimodal biometrics is the use of multiple biometrics traits [19]. Unimodal biometrics is defined as the use of one biometric trait. Multimodal systems remove the disadvantages of unimodal biometric systems by combining different biometric traits [20]. According to M. Khan and J. Zhang, as cited by [21], the performance of unimodal biometrics systems has to contend with various problems, such as background noise, signal noise and distortion, and environment or device variations.

4 Research Methodology

i. **Principal Component Analysis**

PCA is one of the most popular and successful techniques for image recognition and compression for feature and data presentation. It achieves this by identifying the maximum variance of data and reducing reconstruction errors. There are five steps involved in the PCA algorithm [22]: Get data, subtract the mean, find the covariance matrix, calculate eigenvectors and eigenvalues, form a feature vector, and retrieve a new dataset.

Step 1: Subtract mean

Once you have identified the data e.g., $Z = [11\ 121\ 30\ 33]$, that you need to perform PCA on, the first thing PCA does is to subtract the mean on the data. Below is the formula to get the mean:

$$\overline{Z} = \frac{\sum_{i=1}^{n} Z_i}{\mathrm{n}} \tag{1}$$

Z is used to refer to the entire data of numbers; to retrieve a certain number, say 33, we will refer to it with Z_4. Also, note that the symbol n refers to the number of elements in dataset Z. $\overline{Z}$ represents the mean for the set of Z as calculated in Eq. (1). Standard deviation Eq. (2) is used to subtract the mean:

$$s = \sqrt{\frac{\sum_{x=1}^{n} \left(Z - \overline{Z}\right)^2}{n-1}} \tag{2}$$

Step 2: Find the covariance matrix

The second step of PCA is to calculate the covariance matrix. The covariance matrix is calculated using Eq. (3):

$$cov(X, Y) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{(n-1)} \tag{3}$$

where $cov(X, Y)$, is the covariance matrix for a 2-dimensional dataset of (X, Y). Since not all the datasets are two-dimensional, when a dataset has more than two dimensions, we use Eq. (4) to calculate covariance and put them in a matric:

$$C^{nxn} = \langle c_{i,j}, c_{i,j} = cov(Dim_i, Dim_j) \rangle \tag{4}$$

where C^{nxn} is the matrix with n rows and n columns and Dim_x is the x^{th} dimension.

Step 3: Calculate eigenvectors and eigenvalues

Equation (5) is used to calculate the eigenvectors and eigenvalues

$$A\vec{v} = \lambda\vec{v} \tag{5}$$

where λ is an eigenvalue of the n * n Matrix A and $\vec{v}$ is an eigenvector of Matrix A.

Step 4: Forming a feature vector

Equation (6) is used to choose these vectors.

$$FeatureVectors = (eigenvector_1 \ldots eigenvector_n) \tag{6}$$

Step 5: Extracting new dataset

The final step of PCA is to form a new dataset using principal components based on the eigenvectors chosen in step 4. To do that, Eq. (7) is used:

$$\text{FinalData} = RowFeatureVector \times RowDataAdjust \tag{7}$$

where *RowFeatureVector* is the transpose columns of an eigenvector matrix, and *RowDataAdjust* is the transposed adjusted-mean variable

ii. **Algorithm**

PCA is one of the most popular and successful techniques used for image recognition and compression for feature and

MBS Algorithm

1. Input (P, ID, F)
   a. P: $Fingerprint$
   b. ID: Identity Number or Passport
   c. F: Face Image
2. C: Customer type (new or existing customer)
3. Output
4. M: Outcome of Verification
5. *PCA Procedure of Fingerprint*
   For each $I_m \in P$ *do*
   fr_m = *PCA_function*(I_m)
   End For

   End PCA
6. M = Biometric_Verification (C, fr_m, R)
7. $R \leftarrow D[\,]$, where R is a Retrieved fingerprint from the database
8. $Contract_{feedback}$ $(status)$
9. Store_Biometrics (ID, F, fr_m)
10. Delete_Biometrics (ID, F, I_m)
11. ENDMBSComputation

a) **Input**

As shown in line 1 of the MBS algorithm, the key inputs of the algorithm are fingerprints (fr_m), ID number (ID), and face image (F) of the customer. The input is used to authenticate or register. Authentication and registration are determined by the input in line 2 C : *CustomerType*.

b) **Biometric Verification**

In line 2, the agent in the store specifies customer type $\boldsymbol{C}$, who can be a new customer or an existing customer; based on this information, the MBS algorithm decides whether to do registration or verification of the customer. MBS does biometric registration when $\boldsymbol{C} = \boldsymbol{New}$ and verification when $\boldsymbol{C} = \boldsymbol{Existing}$.

c) **Retrieve**

In line 7, MBS retrieves biometric information from sets of databases, namely: CRM, Fraud, and DHA, and stores them in $R \leftarrow D[]$. *retrieve*(var) function, which is invoked by Biometric_Verification (C, fr_m, R) procedure only accepts one variable var.

Where: var: Can be the *ID number* or *ALL*.
ID number: South African ID number or Passport number.
IF $var = ALL$, THEN
Retrieve all the fingerprints in the database and store them in $R \leftarrow D[]$
ELSE IF $var = ID\ number$ THEN
Retrieve only fingerprints linked to the *ID number* and store them in $R \leftarrow D[]$
END IF

d) **PCA procedure**

In line 5, the system performs PCA on both biometric information captured by the store agent and biometrics retrieved from the relevant database and store variables in fr_m *and* R respectively.

e) **Matching**

To do the matching, Euclidean and Mahalanobis distances were used. Euclidean distance measures spatial distance, while Mahalanobis distance measures similarity.

1. First, Euclidean distance $T_d(\Gamma_j, \Phi_k)$ between the vectors of the captured biometric information Γ_j and database biometric information Φ_k gets calculated using Eqs. (8) and (9):

$$\Phi = \begin{bmatrix} \Phi_{11} & \Phi_{12} & \Phi_{1M} \\ \Phi_{21} & \Phi_{22} & \Phi_{2M} \\ . & . & . \\ \Phi_{M1} & \Phi_{M2} & \Phi_{MM} \end{bmatrix} \rightarrow \Phi_k := \begin{bmatrix} \Phi_k \\ \Phi_k \\ . \\ \Phi_k \end{bmatrix} \quad and \quad \Gamma_j = \begin{bmatrix} \Gamma_j \\ \Gamma_j \\ . \\ \Gamma_j \end{bmatrix} \tag{8}$$

$$T_d(\Gamma_j, \Phi_k) := \sqrt{\sum_{b=1}^{n} (\Gamma_j - \Phi_k)^2} \tag{9}$$

2. Secondly, the Mahalanobis distance $T_e(F_{xv}, \Phi_k)$ between vectors of captured biometric information PF_{xv} and the biometric information stored in the database $\Phi_{1,2\ldots M}$. Mahalanobis is recalculated and rewritten as indicated in Eqs. (10) and (11):

$$T_e(F_{xv}, \Phi_k) = \sqrt{(\Phi_k - PF_{xv})^T * [\lambda_{kk}]^{-1} * (\Phi_k - PF_{xv})} \tag{10}$$

where:

$$\Phi_k := \begin{bmatrix} \Phi_k \\ \Phi_k \\ . \\ \Phi_k \end{bmatrix} \quad and \quad PF_{xv} = \begin{bmatrix} PF_{11} \\ PF_{21} \\ PF_{M1} \end{bmatrix} \tag{11}$$

λ_{kk} are the eigenvalues of the matrix which correspond to the eigenvectors as described in Eq. (5).

3. Lastly, the system does verification to check if F_x matches the biometric information in our databases; because it is not always necessary to make a full database search, the maximum search value is set as given by Eq. (12):

$$W_L := \left(\frac{1}{2}\right) max\left[T_d(\Gamma_j, \Phi_k)\right] \tag{12}$$

There are three scenarios where the match is confirmed, as indicated in Eq. (13):

$$\begin{aligned}
&\text{a. If} \quad \min[\mathrm{T}_e(\mathrm{F}_{xv}, \Phi)] \leq W_L \text{ then} \\
&\qquad \mathrm{M} := \mathit{Match} \\
&\qquad \text{Return M} \\
&\text{b. If} \quad \min[\mathrm{T}_M(\mathrm{F}_{xv}, \Phi)] \leq a\lambda_{11} \text{ then} \\
&\qquad \mathrm{M} := \mathit{Match} \\
&\qquad \text{Return M} \\
&\text{c. If} \quad \min[\mathrm{T}_e(\mathrm{F}_{xv}, \Phi)] \leq \beta\lambda_{11} \text{ then} \\
&\qquad \mathrm{M} := \mathit{Match} \\
&\qquad \text{Return M}
\end{aligned} \tag{13}$$

If the above three scenarios are not met, then $\mathrm{Biometric}_{\mathrm{Verification}}(C, fr_m, R)$ Return M := *MisMatch*.

Lines 3, 4, 7, 8, 9, and 10 are triggered by Biometric_Verification (C, fr_m, R) in line 6.

In line 8 function $Contract_{feedback}$ $(status)$returns $Approved$when M := *Match* in line 7 and returns $(Decline)$when is M := *MisMatch*.

Line 9 is function *store_biometric*(ID, F, fr_m) which is used to store the ID number, face image, and Fingerprint after verification or authentication. If the Fingerprint is successfully authenticated, info is stored in CRM DB; otherwise in Fraud DB.

Function *delete_biometric*(ID, F, fr_m) in line 10 is used to delete the ID number, face image, and Fingerprint, which were temporarily stored in the CRM Biometric DB.

iii. **Data Collection and Description**

a) **Face Images Collection -** The face images were obtained from the AT&T dataset, which has four hundred (400) images made up of ten (10) images per person from forty (40) individuals.

b) **Fingerprints images Collection -** The fingerprints images were obtained from FVC2004 (Fingerprints Verification Competition) dataset. The dataset consists of eight (8) fingerprints per person collected from thirty (30) random people.

c) **Telecommunication CRM Data Collection -** The fraud data was obtained from a South African telecommunication mobile company. The dataset is for postpaid customers from January to April 2017, having 1510 churn cases with 38 fields.

iv. **Biometric Live Data Capture**

a) **Capture Face Image -** An internal webcam was used to capture the face image and store it locally in the working directory before saving it in the database.

b) **Capture Index and Middle Fingerprint -** Both fingerprints were captured using the '*BioMini Slim 2'* fingerprint scanner that is FBI PIV and FBI Mobile ID FAP20 certified; the scanner also provides advanced live finger detection technology, which detects various materials including film, paper, glue, silicon, rubber, clay and many more.

v. **Face Image Lifecycle**

Figure 1 shows the lifecycle that a captured face image goes through before matching is performed. *Step 1* is the first step where an image gets captured. *Step 2* is the second step that checks whether the captured image is a face; the check is to determine whether the image contains a face and also if there are two eyes. *Step 3* is the third step, where only a face gets extracted from the image, and cut out the unnecessary portion of the image; this step is necessary because in cases where we have a full body image, this step ensures that only a face is cut out and used in the model. *Step 4* is the last step, where PCA is applied to a captured image; this is the image the system uses to do recognition tests. In this study, we used and recommended the image size 200 × 200 to be used for testing.

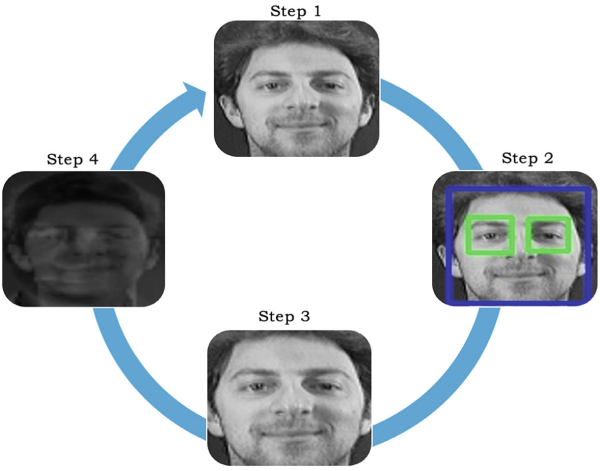
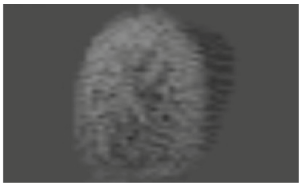


Fig. 1. Face image lifecycle

vi. **Face Image Lifecycle**

Table 1 shows the lifecycle the image goes through before verifying the captured image and any image stored in the database. *Step 1* is the first step, where a fingerprint image is captured. *Step 2* is the second step, where PCA is performed on the captured image, and this is the image our system uses to do recognition tests.

Table 1. Fingerprint image lifecycle

Step 1	Step 2
	

5 Testing Results and Discussions

i. **Face Accuracy Test**

Face accuracy evaluation consist of recognition before applying PCA, PCA face recognition, and a full database face accuracy test.

In Fig. 2, the recognition/match test is done before applying PCA to the image. Images for both captured on the left and the database image on the right were used to do the verification. Both images had 91 key points, and all the key points were matched by our system.

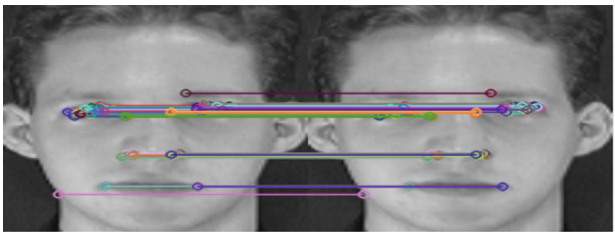


Fig. 2. Face recognition before PCA

Figure 3 shows the final face test conducted, showing the PCA image recognition. PCA was applied in Fig. 3. PCA reduced the number of key points from 91 to 65. When the match was done, all 65 key points were matched successfully.

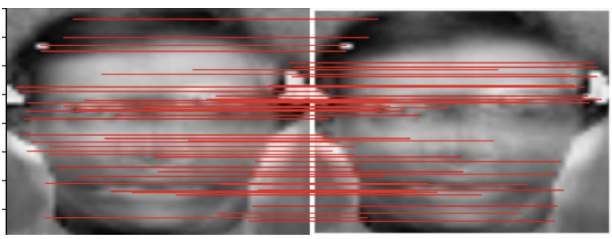


Fig. 3. Face ricognition PCA image

ii. **Full database Face Accuracy Test**

For the *Full database test,* all faces from the *database* were selected to do a recognition against a captured image; the model was 93% accurate. The initial tests of the system were not always accurate due to different hairstyles introducing variance. The introduction of *Step 2* in Fig. 1*: Face Image Lifecycle* improves the system's performance and is 93% accurate, which is better than the earlier 80%. Table 2 shows the results of a full database test for the face image.

Table 2. Face CRM images compared with DHABIO database images results table

Row ID	Results description	Items
1	*Correct*	*71*
2	*Wrong*	*5*
3	*Total Test Images*	*76*
4	*Accuracy Percent*	*93.42*
5	*Total Person*	*38*
6	*Total Train Images*	*76*
7	*Total Time Taken for recognition:*	*0.015625* s
8	*Time Taken for one recognition:*	*0.000205* s
9	*Training Time*	*7.46875* s

iii. **Fingerprint Match Testing**

Fingerprint accuracy evaluation consists of fingerprint recognition before applying PCA, PCA fingerprint recognition, and a full database fingerprint accuracy test.

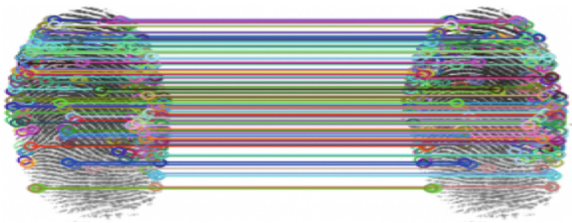


Fig. 4. Fingerprint recognition before PCA

Figure 4 shows the accuracy test done before applying PCA. Images of the same finger were used for both captured image on the left and the database image on the right. Both images had 364 key points, and our system successfully matched all the key points.

Fig. 5. PCA Fingerprint recognition

Figure 5 shows the PCA testing done on the fingers used in Fig. 4. PCA reduced the number of key points from 364 to 283. The system successfully matched all 283 key points.

iv. **Full database Fingerprint Accuracy Test**

For the *Full database test,* all fingerprints from the *DHABIO* database were selected and used to do recognition against a captured image. The system was 95% accurate. Table 3 shows the results of a full database test for fingerprint images.

Table 3. CRM Fingerprint compared with DHABIO database images results table

Row ID	Results description	Items
1	*Correct*	*72*
2	*Wrong*	4
3	*Total Test Images*	*76*
4	*Accuracy Percent*	*94.84%*
5	*Total Person*	*38*
6	*Total Train Images*	*76*
7	*Total Time Taken for recognition:*	*0.015625* s
8	*Time Taken for one recognition:*	*0.0002067* s
9	*Training Time*	*7.46875* s

v. **Performance of PCA**

We compared the performance of the proposed algorithm with other algorithms based on accuracy. The algorithms are Independent Component Analysis Algorithm (ICA), Linear Discriminate Analysis Algorithm (LDA), and Principal Component Analysis PCA. Table 4 shows the comparison between PCA, ICA, and LDA based on accuracy, key points recovery, and execution time.

Table 4. Comparison results between PCA, ICA, and LDA

Evaluation	PCA	ICA	LDA
Accuracy	94.84%	93.2%	94%
Keypoints Recovery	47.40%	44.47%	45.75%
Execution Time	7.46875 s	10.26437 s	7.59831 s

a) **PCA** – the PCA reduced the number of key points of the image and still retained 47.4% of the key points with the most variance. Even with reduced key points, we could still get the desired accuracy of 94.84% at the execution time of 7.46875 s for a full database evaluation, as illustrated in Table 5. This makes the deployment of the algorithm cost-effective by saving space and the number of required CPUs.

The principal Component Analysis Algorithm was chosen to solve subscription fraud based on the best accuracy, key points recovery, and fast execution time.

b) **ICA** – we compared the proposed algorithm with ICA. The idea behind ICA is to minimize statistical dependence between components. We compared our algorithm with ICA because both reduce dimension, and both are widely used for recognition; however, ICA first executes PCA and then selects the best components of our data. This makes the algorithm slower than PCA, needs more processing power and makes the system expensive to deploy. PCA outperforms ICA with accuracy as well. ICA is 93.2% accurate, recovered 44.7% of the key points, and is 3.1 s slower than PCA.

c) **LDA** – The LDA is similar to PCA since both models focus on linear combinations of variables that best describe data. The goal behind LDA is that it explicitly shows the difference between the data. However, LDA is 94% accurate and managed to recover 45.75% of the key points. PCA outperformed LDA. On the other hand, LDA was more accurate and faster than ICA.

Table 5 shows the comparison between the matching algorithms: Scale Invariant Feature Transform (SIFT), Oriented FAST and Rotated BRIEF (ORB), and PCA-based representation for local features (PCA-SIFT). False Acceptance Rate (FAR) and False Rejection Rate (FRR).

Table 5. Comparison between PCA-SIFT, SIFT, and OBR

Evaluation	PCA-SIFT	SIFT	OBR
Accuracy	94.84%	93%	0%
FAR	0.2791	0.329	N/A
FRR	0.0269	0.0279	N/A
Execution Time	7.46875 s	12.67435 s	N/A

a) **PCA-SIFT** was chosen to do matching for our system. PCA-SIFT uses PCA to replace the gradient histogram method in SIFT. It produces vectors that are smaller than vectors produced by SIFT. PCA-SIFT is 94.84% accurate with FAR of 0.2791, FRR of 0.0269, and execution time of 7.46875 s, as illustrated in Table 6.

b) **SIFT** - we compared the efficiency and accuracy of our matching algorithm with SIFT. SIFT is a feature detector proven to be very efficient in recognition applications. SIFT identifies the key points on the image and uses them to do recognition. As illustrated in Table 6, results show that SIFT is 93% accurate with FAR of 0.329, FRR of 0.0279, and execution time of 12.67435 s. PCA-SIFT is more accurate and faster than SIFT. SIFT is CPU intensive and costly to deploy.

c) **ORB** - uses FAST to detect the key points of the image and BRIEF descriptor. ORB is cost-effective since it is free to use for everyone. It is very fast in matching and identifying the key points. This algorithm is unable to identify the key points on a PCA image.

6 Conclusion and Future Work

In this study, we designed and built MBS that prevents subscription fraud, and we also managed to identify a fraudster and save their details to avoid the future reoccurrence of subscription fraud. We demonstrated successfully how telecommunication companies could capture biometrics information of potential customers and verify it against biometrics information stored in a central database of the DHA before approving a contract. We compared the performance of PCA with LDA and ICA. We compared our matching algorithms and found that PCA-SIFT is better than SIFT and ORB. In the future, the integration of MBS into the police system can be investigated.

References

1. ICASA: The state of the ICT sector in South Africa. Independent Communications Authority of South Africa, p. 115, March 2021
2. StatsSA: Statistical release P0302: mid-year population estimates 2020, pp. 1–22. Stats SA, July 2020
3. Casey, E., Turnbull, B.: Digital evidence on mobile devices. Digit. Evid. Comput. Crime **3**, 1–44 (2011)
4. KPMG: Global profiles of the fraudster, pp. 1–27. KPMG, May 2016
5. Ogundile, O.O.: Fraud analysis in Nigeria' s mobile telecommunication industry. Int. J. Sci. Res. Publ. **3**(2), 1–4 (2013)
6. Estévez, P.A., Held, C.M., Perez, C.A.: Subscription fraud prevention in telecommunications using fuzzy rules and neural networks. Expert Syst. Appl. **31**(2), 337–344 (2006)
7. CFCA: 2015 global fraud loss survey. Report, p. 26 (2015)
8. Van Heerden, J.H.: Detecting fraud in cellular telephone networks, December 2005
9. Kabari, L.G., Ajuru, I., Harcourt, P.: Telecommunications subscription fraud detection using Naïve Bayesian network. Int. J. Comput. Sci. Math. Theory **2**(2), 1–10 (2016)
10. Koops, B.-J., Leenes, R.: Identity theft, identity fraud and/or identity-related crime: definitions matter. Datenschutz und Datensicherheit - DuD **30**(9), 553–556 (2006). https://doi.org/10.1007/s11623-006-0141-2
11. CFCA: 2017 global fraud loss survey. Report, p. 26 (2018)
12. Adeoye, O.: Multi-mode biometric solution for examination malpractices in Nigerian schools. Int. J. Comput. Appl. **4**(7), 20–27 (2010)
13. Shah, D., Haradi, V.: IoT based biometrics implementation on Raspberry Pi. Procedia Comput. Sci. **79**, 328–336 (2016)
14. Ugale, A., Ingole, A.: Bimodal biometric recognition using PCA. Int. J. Innov. Res. Comput. Commun. Eng. **4**(6), 2257–2263 (2016)
15. Bhattacharyya, D., Ranjan, R., Alisherov, F., Choi, M.: Biometric authentication: a review. Int. J. Serv. Sci. Technol. **2**(3), 13–28 (2009)
16. Republic of South Africa: Department of Home Affairs - Identity Documents. http://www.dha.gov.za/index.php/civic-services/identity-documents. Accessed 23 June 2018
17. JKCS: Biometrics white paper, pp. 1–46 (2012)
18. Ratha, N.K., Senior, A., Bolle, R.M.: Automated biometrics. In: Singh, S., Murshed, N., Kropatsch, W. (eds.) ICAPR 2001. LNCS, vol. 2013, pp. 447–455. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-44732-6_46
19. Kale, P.G., Khandelwal, C.S.: IRIS & finger print recognition using PCA for multi modal biometric system, pp. 78–81 (2016)

20. Jain, A.K., Hong, L., Kulkarni, Y.: A multimodal biometric system using fingerprint, face, and speech. In: International Journal of Computer Science and Mathematical Theory, vol. 1, no. 4, pp. 182–187 (1999)
21. Wang, Z., Wang, E., Wang, S., Ding, Q.: Multimodal biometric system using face-iris fusion feature. J. Comput. **6**(5), 931–938 (2011)
22. Kau, F.M., Kogeda, O.P.: Impact of subscription fraud in mobile telecommunication companies. In: IEEE 2019 Open Innovations Conference, Cape Town, South Africa, 2–4 October 2019, pp. 42–47. Cape Peninsula University of Technology (2019). ISBN: 978-1-7281-3464-2. https://doi.org/10.1109/OI.2019.8908261