UDC 004.056.55

An overview of strategic workflow design to develop multimodal biometric recognition systems

Dhiman Karmakar (Department of Computer Science, Surendranath College)

It has been a proven fact that the inclusion of more than one modality makes biometric recognition systems more robust and elevates its recognition accuracy. This paper aims to develop a strategic platform for the upcoming researchers in order to develop multimodal biometric recognition systems. The researchers in this fraternity would be able to design their own procedural strategy using the generalized workflow template depicted in this review. The selection of proper visually interpreted biometric identifiers and modalities, various fusion strategies, choice of performance metrics, formation of training and test sets from databases and the possible challenges in developing the workflow are purposefully portrayed in this article.

Keywords: unimodal, multimodal, recognition, fusion, spoofing, metrics

1. Introduction

Nowadays, Biometrics, the physiological and behavioral traits used to recognize a human being, have become an integral part of human society with the increasing need for security at various levels in forensic, surveillance and commercial fields. A biometric security system should increase the security and secrecy of user data and be capable of authenticating a person’s identity based on his biometric traits like faces, fingerprints, iris etc.

Preference for face and face-like patterns for a child occurs hours after birth and perhaps starts with the face of the mother. According to the Intersensory Redundancy Hypothesis (IRH), during the early development of a child, perception of faces is enhanced in the unimodal visual (i.e., silent dynamic face) rather than bimodal audio-visual (i.e., dynamic face with synchronous speech in the form of lullaby) stimulation. In later days, this capability of a child is developed into the ability of face recognition of an adult by correlating many multimodal traits through expertise gained over the years. However, when the same task is undertaken by machine intelligence, the paradigm becomes altogether challenging and difficult. As human society is developing, increasing needs are felt for security systems that would identify a person from many of his traits and physical forms, such as identification cards, voice, gaits and gesture, and combination of many other parameters under varying conditions and situations.

The broad domain of Multi-Modal Biometrics (MMB) scenario works in either of the following modalities [1].

i. Multi-sensor: Different sensors (e.g. optical and/or electronic) used to capture the same biometric (e.g, fingerprints).

ii. Multi-biometrics: At least two different biometrics (say face and fingerprints) are fused and used for the purpose of recognition.

iii. Multi-unit: Multiple units captured from the same biometric (e.g. fingerprints collected from more than one finger).

iv. Multi-snap: Use of multiple instances of the same biometric (e.g. multiple impressions of the same finger).

v. Multi-matcher: Combining different approaches in feature selection and biometric matching algorithms [25].

A pictorial representation of the MMB scenario is shown in Fig.1. It is important to realize that a multi-biometric system is always multimodal, however the reverse is not true. It is worth noting that a single biometric may possess multiple modalities. For example, the geometric structure of a face may be blended with its behavioural nature to produce facial multimodality [27]. Similarly, finger veins and finger knuckles may be combined to produce multimodality in fingers [35].

Multimodal Biometric Systems (MMBS) have gained popularity among researchers as those provide more variability in information, for processing purposes of an individual rather than a unimodal biometric system [59, 46]. Those are useful to enhance the robustness in many security related areas including passport verification and authentication of persons.

2. Selection of authentication method in workflow

In this context, it is important to distinguish some frequently used terms in literature for the purpose of authentication, determining the identity of an individual in an automated manner. These terms are somewhat used inconsistently and interchangeably in different literature and may create confusion for the readers. Personnel can be authenticated using his (i) personal ID card (what he has), (ii) password (what he knows) and (iii) biometric traits, termed as recognition. Although a general approach to authenticate personnel is password matching or ID card verification, yet, such mechanisms of identity detection can easily be lost, hampered or stolen and thereby undermines the intended security. Incorporating physical and biological traits of human beings in the system has thus become a research issue in modern biometric security.

The process of recognition can further be divided into a couple of categories, namely, verification (“Is this the person who he claims to be?”) and identification (“Is this person in the database?”). It is quite obvious that verification is a 1:1 matching approach as the user’s data is verified only with the claimed person. Similarly, identification is a 1:N matching approach as the user’s identity is compared with all (N) the person’s data present in the database.

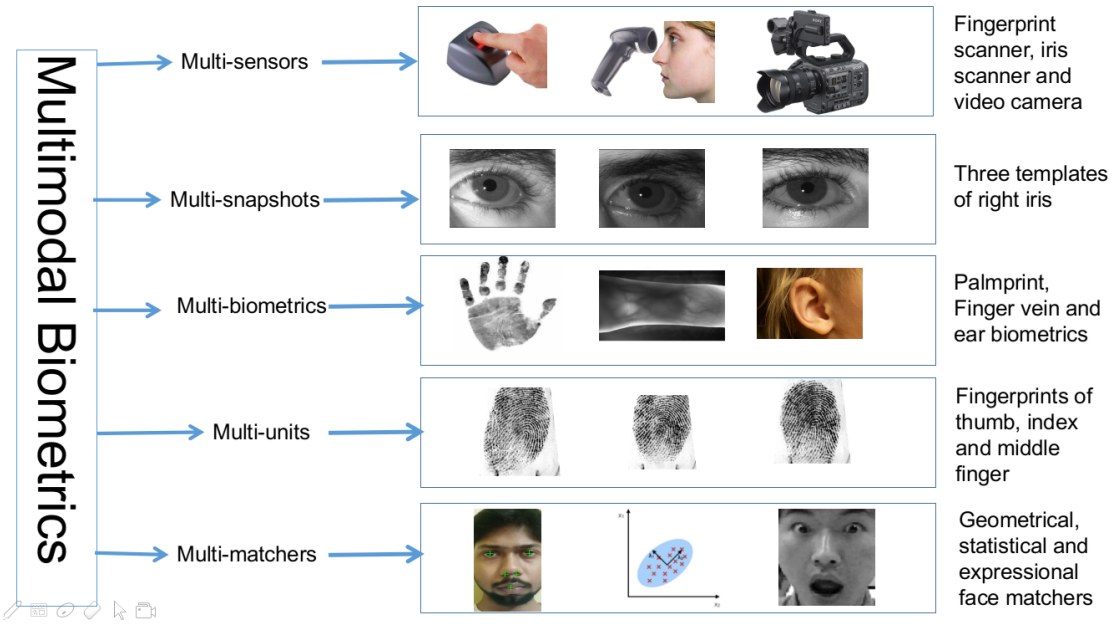

Figure 1: Various scenarios in a multimodal biometric system.

3. Selection of visually interpreted biometric identifiers

Biometric identifiers are categorized into physiological (say, fingerprint) or behavioral (say, gait) or a blend (say, voice) or some combinations. They can further be classified into three sections, namely hard, soft and hidden. The traditional feature-rich biometrics with self-sufficient ability to detect a human being are called hard biometrics. They are face, fingerprints, iris etc. Their level of accuracy is generally very high.

Soft biometrics, unlike the standalone nature of hard ones, are associated with classical biometrics to assist the overall recognition process. Skin color (related to face), height (related to full body) etc. are examples of soft biometrics. Though accuracy is relatively low, their nature of perceiving physiological traits (say, how tall or dark a person is) with ease, makes them popular. The hidden biometrics cannot be observed by naked eyes and generally they are stored in the form of medical data (for example, DNA report, blood group or X-ray image).

In this article, we confine the discussion to visually interpreted identifiers. They are the traits commonly represented in image or video forms and analyzed using computer vision techniques. For

this reason, DNA samples, odor etc. are not taken into consideration. Some of the traits, like speech data, though represented directly as signals, may be converted to image form for suitable analysis.

Face, fingerprint, palmprint, iris, ear and several others are the different identifiers used for the purpose of automatic human recognition. Each of them carries the pros and cons of their own. Seven characteristics [25], namely universality, distinctiveness, permanence, collectability, performance, acceptability and circumvention should be prioritized as per researcher's goal before the workflow design begins. For example, if permanence is assigned more priority than distinctiveness, then ear should be a preferred choice over face. Similarly, if performance is the key, iris can be an automatic alternative. However, choice of a biometric identifier largely depends upon the researcher's interest, research specific goal and availability of concerned databases.

4. Workflow pattern for a MMBs

MMBS are based on capturing human bodily features and using them for identification and authentication. MMBs function in an abstract manner. Generally, CBIR (Content Based Image Retrieval) system retrieves a query image and primitive features denoting image content, such as colour, textures, and shape, are computed for both stored and query images, and then used to identify stored images, most closely matching the query image. Semantic features such as the type of object present in the image, though difficult to extract, remains an active research area. Similarly, in a MMBs a query image, showing a biometric trait (say, face or fingerprint or iris etc.) is searched from a stored database of the same biometric. An overview of the generic MMBs is illustrated with the help of Fig.2. The entire procedure is divided into two sub-phases, namely training and testing.

4.1. Training phase in workflow

During the training phase, let us consider that the input database consists of multiple biometric modalities and is assumed to be structured and known. The term known indicates that the system and by structured we mean that every instance is properly placed within its class once enrolled. For example, every fingerprint of a person Xj should belong to class j, considering the presence of any person Xi in a database of n persons, where i = 1, 2, .., j, .., n.

The main algorithm of the system used to match a known and an unknown instance is called the matcher. The matcher is used to classify an unknown instance to its proper class. The efficacy of a matcher is adjudged by its score, crudely a number of correct classifications in comparison to wrong ones. The score can be viewed as the rate of recognition of the system. The different input

parameters of a matcher, for example, a global classification threshold value, can be adjusted at the end of each epoch of the algorithm in order to produce a better result. In other words, considering the input parameters and matcher as an encapsulated entity, the matcher is as a whole modified or trained to enhance the level of score.

The dataset used to train the classification algorithm is called the training dataset. Evidently, the term “trained dataset” would be wrong as the dataset is itself not trained and rather used to train the algorithm. Other than the training part the input database is divided into what is called a validation set, the dataset used to validate the matcher. The matcher takes an instance from the validation set and classifies it. Whether it is a mismatch or a correct classification can easily be determined, since the validation set is actually designed from the known database. Based on the result of classification, several input parameters are adjusted to enhance the system outcome. Here human interactions (not automated), for example, manual manipulation of classification threshold value may take place. The training phase concludes once the developer is satisfied with the system outcome. Therefore, the entire training phase helps to train the classification algorithm to a level of significant satisfaction.

4.2. Testing phase in workflow

In the testing phase, the unknown instances are fed to the modified matcher for classification. The database used to test the accuracy of the system is the test database. A robust MMBs should not only classify a test instance (closed set scenario) but also specify if the unknown instance does not fall into any of the existing known classes (open test scenario). In both the cases, the raw data captured is normalized into feature vector form and fusions are carried out at different levels. In the preprocessing phase, one may select the Region of Interest (RoI) of the image instance for further processing. Before feeding the data to the matcher, its dimensionality is reduced for faster processing and the required features (based on which the matcher is designed), generally in vector form, are extracted. Some of the feature extraction and dimensionality reduction schemes commonly used are PCA, KPCA, 2DPCA etc.

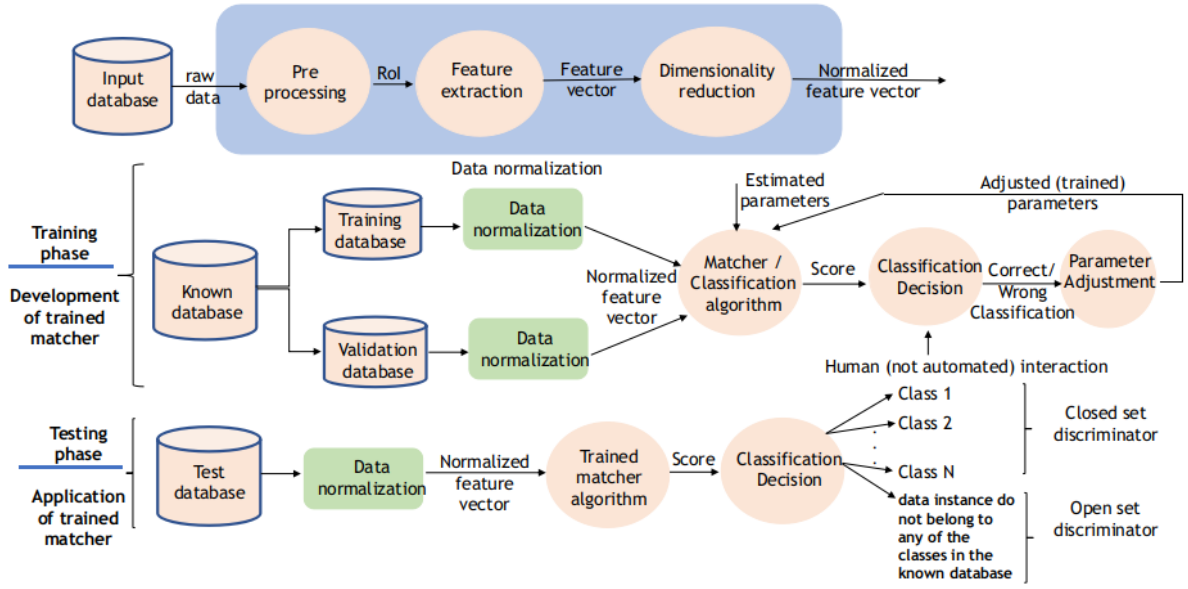

Figure 2: Abstract overview of a biometric recognition system

4.3. Workflow modes

The acquisition and processing in MMBs is carried out in three different modes.

i. Serial mode: Here, a biometric trait is processed before accepting the next one and the decisive classification outcome of the former is forwarded to the latter. It may reduce the overall recognition time as a decision can be drawn without accepting the next modality in sequence.

ii. Parallel mode: Here multiple modalities are processed simultaneously and the conclusion is drawn integrating the outcomes of all modalities.

iii. Hierarchical mode: Here individual matchers are combined in a treelike structure in order to manage several classification approaches.

5. Levels of information fusion in workflow

In MMB scenarios, how to fuse information of two different biometrics of the same person remains a matter of research. These fusion strategies can be performed in raw level, or in feature level, or in score level or even in decision level [24]. In the raw (or sensor) level of fusion, the same characteristics of raw MMB data captured using different sensors (for example, sensing the fingerprint of the same finger of a person in more than one fingerprint scanner) are combined. At times it is found to enhance the recognition accuracy but suffers from the drawback of incompatibility of data from different modalities.

Since the feature set of any biometric trait is supposed to carry the most significant and rich information about raw biometric data, the integration at feature level is expected to provide better recognition performance than the other levels of fusions. Concatenating the feature vectors extracted from the face and fingerprint data is an example of such fusion. Intuitively, the most important features of two biometric traits are expected to provide good performance. Although possessing these advantages, fusion in feature level is relatively understudied in comparison to other fusion levels [46]. The reasons are basically threefold. Firstly, extracted features from different biometric traits may become incompatible to each other; for example, extracted minutiae from a fingerprint and derived eigen coefficients from a face seem to be incompatible. Secondly, when different feature vectors of different biometrics are concatenated together, the resulting feature vector may suffer from the curse of dimensionality [32] and would become very difficult to handle. An effective dimensionality reduction scheme [53] for online authentication of face and signature, however, proves its significance. Thirdly, and most importantly, designing a matcher algorithm for a fused feature vector, containing features of different multimodal traits, is more difficult than generating separate matcher algorithms for different single biometrics. Fusion of information at this level also faces other challenges like large inter-user similarity, small intra-person variability, and unknown relationship between features [24].

A matcher algorithm, as discussed earlier, produces a similarity score based on the proximity of a query feature vector with the template feature vector in the known dataset. In score (or confidence or rank) level fusion, match-scores obtained from various matchers are combined for the final classification decision. The method of weighted average is often used to combine the scores. This level of fusion is attractive due to its simplicity and good performance [24].

Fusion at decision level is least powerful in comparison to other levels [46]. Here a separate decision is taken for each biometric at a very late stage and hence it prohibits enhancing the rate of recognition. A majority voter scheme is often applied to furnish the final decision on classification.

During the last decade several MMB verification and identification algorithms have been proposed including a wide variety of survey papers. Since the majority of the research work emphasizes modality in isolation, judging the efficacy of MMBs depends upon their (modalities) comparative analysis. This, however, is difficult to perform and may not provide a fruitful result.

Technologies are being developed for embedding multiple biometric information in the identity cards. For example, ICAO (International Civil Aviation Organization) encourages American people to use facial image, fingerprints and iris in their travel documents as a verification tool. Indian citizens are using Aadhaar as their identity card, which combines face, iris and fingerprint traits.

In MMB fusion strategy, at least two different biometrics are fused. Fusion at feature level, though difficult and understudied [46], possesses more importance than the other levels, because the extracted feature set from raw data holds most significant and rich information. Several evaluation protocols on closed test set identification have been designed for measuring the performance of different existing algorithms. In open test identification, the challenge is to reject the imposters. However, how to fuse information of two different biometrics of the same person remains a matter of research. The combined classifier approach has been adopted to get better results, especially at score level fusion.

In a brief review of MMB, Ross and Jain [46] have presented the idea of various levels of fusion, various possible scenarios, and different modes of operation, integration strategies and design issues. For homogeneous feature sets (e.g., multiple fingerprint impressions of a user finger), weighted average of the individual feature sets are often used to compute the resultant feature set (e.g., fusion using multiple hand features by Michael et al. [19]). On the other hand, for non-homogeneous feature sets (e.g., feature sets of different biometric modalities like face and hand geometry), we can concatenate them to form a single resultant feature set.

A combination of face and fingerprint authentication approaches using CNN (Convolutional Neural Network) for casting votes in elections through a web portal is explained by Saravanan et al. [3]. A MMB fusion scheme of face and fingerprints (specifically ridge and minutiae) [52] yields a recognition rate of 95%. Riseul et al. [47] designed a survey report towards continuous MMBs. The review has pointed out the deficiency of adequate number of comparisons between biometric types, fusion models and types of machine learning algorithm (supervised or semi supervised) in the published literature in this domain. On the contrary, a multimodal sparse technique of representing MMB data [55] by a scattered linear mixture of training records claims to yield better performance than traditional fusion schemes.

5.1. Feature level fusion

After formation of the MMB class of each person of the dataset, the feature level fusion is carried out on the biometrics. Ross and Jain [46], in a MMB review article elaborate different levels of fusion, integration strategies and design issues. Experimental results claim that MMB fusion improves both throughput and performance of the system. Rattani et al. [45] apply a feature level fusion of face and iris. Their algorithm computes the SIFT (Scale Invariant Feature Transform) features from the biometric sources. However, the method requires segmentation of the captured

images. Fortunately SIFT method can be experimentally combined with color image segmentation strategy for improved processing speed and better performance. Face and iris features may be extracted separately and fed into a wavelet probabilistic neural network classifier. The results are then calculated on the decision layer of the neural network. Nasrabadi et al. [56] present a deep learning approach to feature level fusion of face and iris. Their integration of deep hashing, a binarization technique, into the fusion architecture, generates robust MMBs. Utilization of modified gravitational search algorithm (GSA) in feature level fusion yields satisfactory results [23].

Finger based MMBs, being highly secured and stable, attract the attention of the research fraternity. Shuyi et al. [35] propose a MM fusion technique by combining finger veins with finger knuckle point patterns. This feature learning algorithm, maximizes the correlation between inter-modality samples. The problem has also been handled using CNN based approach by further inclusion of fingerprints with the aforementioned bimodal traits of fingers [36]. Sarangi et al. [50] develop a feature level fusion based MMB recognition of ear and profile face. The inclusion of face stabilizes the system in terms of its recognition rate, which may otherwise be hindered by the uncontrolled environment during ear enrollment. Construction of a unique template for each person, fusing his face, ear and palmprint data at feature level, and yet maintaining a relatively low-dimensional feature vector is discussed by Bokade et al. [5]. A multi-level fusion of unimodal methods with trimodal feature level fusion of face, iris and ear is proposed by Purohit et al. [43]. An accuracy of 95% in the trimodal case is found better than the highest achieved unimodal (iris) accuracy of 94%.

5.2. Fusion other than feature level

Some of the important work regarding multimodal dataset fusion strategies other than feature level are briefly summarized. Byeon et al. [6] proposed a deep learning model for multi modal biometric fusion, where they firstly used some fusion strategies at pixel level to optimize the process. Secondly, back propagation was used at feature level to establish relationships among the modalities and lastly, some intelligent fusion techniques were used at the score level. Claims are also made that MMBS security can be enhanced by subsequent use of early fusion, late fusion, and score-level fusion [2]. Performance of different fusion approaches, including image-level fusion, feature-level fusion, and two score-level fusion methods are also explored with the help of deep learning approach [22].

Choudhury et al. [12] apply a new framework of person authentication through adaptive rank level fusion. The proposed approach builds up a meta-heuristic design using ant colony optimization techniques on fingernail plates. A less explored combination of iris with palmprint

seems to be useful [57] in order to authenticate personnel by score level fusion. The article shows an interesting approach of bit transition code in Gabor filtered images. Inclusion of iris with face in uncontrolled condition may, however, worsen the system performance. A face-iris quality assessment network is proposed by Luo et al. [37] in order to decrease the effect of poor quality samples. Here the adaptive weights are also assigned to face and iris, based on their relative quality score and thereafter a score level fusion is applied to obtain satisfactory results. Being unobtrusive, soft biometrics, like gestures and postures are interesting choices in MMBs. Cherifi et al. [9] obtain an EER of 5.15% in score level fusion, where the user of a mobile phone is verified by his single action arm gesture while answering the phone. Ear shape structures are also extracted using the local phase quantization method. A normalized score level fusion strategy on iris and finger traits using hybrid genetic algorithm and Particle Swarm Optimization are applied by Sujatha et al. [54] to reduce FAR and FRR of the system. A score level fusion of fingerprints, finger-knuckle point and palmprint in a MMBs is illustrated by Kant et al. [28]. The suitability of cryptographic MMB authentication in medical application has been discussed by Mohsen et al. [14]. In this work, face and voice are used in both feature as well as score-level fusion.

5.3. Fusion of classifiers

The use of Artificial Neural Network, Artificial Intelligence, Genetic Algorithm etc. as tools in developing the classifiers or matchers has been attempted nowadays. Mikel et al. [34] clarifies the importance of MMB authentication for online student evaluation in COVID-19 pandemic scenario, by the development of an AI based procedure and thereafter testing it in a large scale system. ANN was used as a tool for MMB extensively. For example, Gokulkumari [21] traines the ANN using a modified dragonfly algorithm by selecting optimal weight in order to achieve classification accuracy. Rahman et al. [15] applied CNN successfully on ECG signal and fingerprint in both parallel and sequential models. CNN was also used to train the workflow model while integrating traits like Photoplethysmography and Electrocardiogram signals [13]. Rajasekar et al. [44] proposed a deep learning approach using CNN to integrate face, fingerprint and iris in MMBs.

A MMB recognition approach using the firefly algorithm of ONN (Optimal Neural Network) on fingerprint and ear is used by Chanukya et al. [8]. A median filter approach in preprocessing to identify the RoI of the given traits for further cropping is also applied. Alshardan et al. [2] embraced three renowned CNN architectures viz. ResNet, VGGNet, and DenseNet, to extract features from finger vein and fingerprint images.

In recent times, the combined classifier approach has been adopted to get better results, especially at score level fusion. That is, we may have different feature sets, different training sets, different classification methods or different training sessions, all resulting in a set of classifiers whose outputs may be combined, with the hope of improving the overall classification accuracy. A classifier combination is especially useful if the individual classifiers are largely independent. Various re-sampling techniques like bootstrapping may be used. Examples are stacking, bagging and boosting (or ARCing).

6. Performance metrics: The workflow verdict

The proper feature exaction and computationally efficient matching, clustering and classification algorithms make the MMB recognition systems reliable and robust. The efficiency of a MMBs is measured by its performance metrics which deal with the rates of success and failure (error) of the system. These metrics are expressed in terms of ratio or percentage or frequency. During the enrollment phase, the system may sometimes be unable to sense the biometric modalities (say, due to the lack of ridges in the fingerprint) of some users or fails to capture user's biometric data (say, due to the technological fault of the sensors). The former is known as *FTE* (Failure to Enroll) error while the latter *FTAR* (Failure to Acquire).

Biometric verification scenario may be viewed as a binary classification problem and can be analyzed with the help of a confusion matrix [33]. Two types of system errors are encountered in the verification process.

i. Type-I error or *FRR* (False Rejection Rate), which indicates the proportion of genuine users falsely predicted as impostors and hence rejected to grant the system access. It is expressed as below.

$$FRR = \frac{Total\ false\ rejection}{Total\ true\ attempts}$$

This error is also commonly termed as False Non-Match Rate or False Negative Rate.

ii. Type-II error or *FAR* (False Acceptance Rate), which signifies the proportion of impostors falsely predicted as genuine users and being accepted to grant the system access. Its expression is shown below.

$$FAR = \frac{Total\ false\ acceptance}{Total\ false\ attempts}$$

This error is also commonly termed as False Match Rate or False Positive Rate.

iii. Analogously, True Rejection Rate (*TRR*) and True Acceptance Rate (*TAR*) are used as two success rates of the system. In an ideal MMB verification scenario, FAR = 0, FRR = 0, TAR = 1 and TRR = 1.

$$TRR = \frac{Total\ true\ rejection}{Total\ false\ attempts}$$

$$TAR = \frac{Total\ true\ acceptance}{Total\ true\ attempts}$$

iv. The thresholds set in a MMBs trivially estimating the training set result and test set result are termed as a *priori* and a *posteriori* respectively. A posteriori threshold can be considered as a finally adjusted parameter in the training phase of Fig.2. Other than soft and hard, adaptive threshold [26] has its own significance in biometric recognition. Ideally, if a MMBs matcher maintains a very low threshold, there will be maximum acceptance (or minimum rejection) irrespective of true or false attempts. Similarly, a maximum rejection (or a minimum acceptance) will occur for a very high threshold value. Since FAR and FRR are inversely related, no adjustment of threshold can decrease them simultaneously. The error percentage (in terms of FAR and FRR) against threshold is plotted. To maintain a trade-off between such a high and low threshold value, the intersecting point of FAR and FRR, where both the error rates are equal, is estimated as the threshold value. This point is termed as Equal Error Rate (*EER*) or Crossover Error Rate (*CER*).

v. The Receiver Operating Characteristic (*ROC*) curve (uses the normal deviate scale) or Detection Error Trade-off *(DET*) curve (uses linear, semi-logarithmic or logarithmic scale) and designed to measure the performance of a classification model. ROC serves as another ploy to detect the EER. Here, FAR vs FRR is plotted and the point on the curve, where FAR equals FRR signifies EER. Obviously, lower the EER, better is the system.

vi. Some of the MMBs prefer to use Half Total Error Rate (*HTER*), the averaged value of FAR and FRR as their performance measure, as indicated below.

$$\text{HTER} = \frac{FRR + FAR}{2}$$

However, *HTER* can only impose a gross effect on system efficacy and is not an issue for a MMBS of specific purpose. For example, a MMB whose security is the main concern, given two systems of equal *HTER* should opt for the one with lesser FAR.

vii. The system performance in the identification scenario is measured in terms of its accuracy or Recognition Rate (RR), as stated below.

$$RR = \frac{total\ correctly\ identified\ sample}{total\ test\ sample}$$

The correct or wrong identification of a test sample is decided as follows. A test sample x_i out of total n test samples x_1 , x_2 , ..., x_n is fed to a matcher $m(x_i , r)$ of rank r. The matcher returns r most closely matched samples of xi from the training dataset, out of which k are truly and $(r - k)$ are falsely matched. The rank value, $rank_r = \frac{100k}{r}\%$, more than a threshold percentage $\theta\%$, portrays true identification. For example, given $r = 10$ and $\theta = 80$, say the matcher returns $k = 9$ true matches resulting $rank = 90\%$ and since $90 > \theta$, the match (identification) is concluded to be correct. A plot of $rank_r$ vs $r = 1, 2, ...n$, known as *CMC* (Cumulative Match Characteristic) curve yields a summarization of the identification effectiveness. The proportion of test samples misclassified to a wrong bin (any class other than where it originally belongs to) termed as *bin error rate* or misclassification rate signifies the failure rate in identification.

viii. The duration of the matching process from the end of the enrollment phase until the classification decision, is called Time to Match (*TTM*) rate and is often used as a final conclusive way of estimating the performance of a MMBs.

Spoofing attacks are often responsible for downgrading or even compromising MMBs performance, which emerges from the necessity of anti-spoofing systems. Safavipour et al. [48] claimed that multimodal templates obtained from their deep learning strategy of feature spaces are extremely secure against spoof attacks. A cancellable MMBs capable of protecting the actual biometric features from the intruders was developed by Umer et al.[51]. The security of online and IoT-enabled authentication was also maintained through a method of encryption-decryption. Liveness detection is a modern anti spoofing technique, where the system diagnoses whether the impersonation is caused by representing a fake biometric sample instead of the actual live human being. Dhiman et al. developed two unique liveness detection approaches, namely, multivariate gradient descriptor and multi dimensional Fourier transform applied on facial micro-expression regions [29, 31].

A self-explanatory tabular form summarizes some state-of-the-art MMBs methods for researchers' benefit (see Table 1). The identifiers fused, their levels of fusion and the keynotes in the proposed workflow model are depicted in separate columns.

Table 1: State-of-the-art MMBs methods in a nutshell

Ref.	**Biometrics**	**Fusion level**	**Key points in workflow design**

[2]	Fingerprints finger and veins	Early, late and score	Claims were made that subsequent levels of fusion could make a MMB more robust.
[3]	Face and fingerprints	Decision	CNN based voting system in web portal.
[5]	Face, ear and palmprint	Feature	Feature vector was restructured even after considering three biometric traits, to lesser dimension.
[6]	Face, iris and fingerprints	Pixel, feature and score	Used intelligent fusion techniques through deep learning.
[8]	Fingerprint and ear	Feature	Firstly a median filter was used to extract the RoI and thereafter firefly algorithm of ONN was applied for recognition.
[9]	Ear and arm	Score	User of a mobile phone was verified by his arm gesture and ear portion.
[12]	Index, middle and ring fingernails plates	Rank	The work contributed towards optimal performance accuracy using ant colony optimization and deep learning.
[13]	PPG and ECG signals	Feature and score	Federated learning, an efficient machine learning approach, was used to collaborate decentralized modules without effective data exchange, for security management.
[14]	Face and voice	Feature and score	Established the importance of cryptographic MMB authentication in medical imagery.
[15]	ECG and fingerprints	Decision and score	Deep learning and traditional classification module were used to evaluate the proposed system.
[22]	Face and iris	Pixel, feature and score	An effective deep learning approach was used in the extracted region of interest of the images.
[23]	Face, iris and fingerprints	Feature	Used a threshold specific optimization technique in gravitational search algorithm to outperform the traditional methods.
[28]	Fingerprints, finger knuckle and palmprint	Score	The resulting match score was used to detect authenticity by comparing different performance metrics.
[34]	Face, audio and	Decision	An AI driven monitoring system for online

	keystroke dynamics		student evaluation.
[35]	Finger veins and finger knuckle	Feature	Portrayed multimodality in unibiometrics (finger) and maximized the correlation between inter-modality samples.
[36]	Fingerprints, finger veins and finger knuckle	Feature	Trimodal representation of uni biometric (finger).
[37]	Face and iris	Feature	Used generalized divisive normalization and assigned adaptive weights in image samples to reduce poor image acquisition quality.
[43]	Face, iris and ear	Feature	Recognition rate for unimodal and multimodal data were compared and contrasted in detail.
[44]	Face, fingerprints and iris	Feature and score	CNN was used in conjunction with softmax classifier for identification purposes.
[48]	Face, iris and fingerprints	Feature	Feature vectors were mapped from the feature space into the reproducing kernel and deep learning combined them in fully connected in-depth layers.
[50]	Ear and face	Feature	Uncontrolled enrollment of ear was stabilized by the inclusion of face traits.
[51]	Face, iris and palm prints	Score	A cancelable biometric system (CBS) was introduced to preserve the original traits from possible external misuse.
[52]	Face and fingerprints	Feature	Used variable crossing number(CN) which determines minutiae and ridge ending or bifurcation.
[53]	Face and signature	Feature	An online authentication system with novel dimension reduction approach.
[54]	Iris and finger	Score	Hybrid genetic algorithm and Swarm optimization were used to enhance the recognition rate.
[56]	Face and iris	Feature	Integrated the learning approach of deep hashing in fusion strategy.
[57]	Iris and palmprints	Score	Used bit transition codes in Gabor filter images.

7. Database selection in workflow

It has been observed in the recent past that inclusion of more than one modality in the image dataset, though increases the data handling and manipulation effort, drastically improves the efficiency of the system in terms of its rate of recognition. In this context, the terms database and dataset are interchangeably used. However, specifically, a database is an organized collection of data stored as multiple datasets which are in turn storage of structured data for a specific purpose.

For a given query, which can be an image in case of iris, face, fingerprint and palmprint recognition problems or can also be a signal in case of speech data, the system should be able to conclude whether it matches with any in its database. The system should have the option of deciding that the query does not belong to the given data set, and hence classify it as an imposter.

Some popular databases for faces are FIA [20], YALE [18], ORL [49], FERET [41] etc. and for iris are MMU [39], CASIA [10], DOBES [38] etc. FIA and YALE datasets consist of frontal faces with varying facial expressions while FERET contains faces with different angular directions. A collection of A-Z face database repositories is collectively uploaded in Princeton University's webpage [42]. NIST [16] and CASIA [10] maintain databases containing various forms of fingerprint data for research purposes. SVC 2004 [58] is one among the very few online signature databases publicly available to the research community. Including the well-known corpora [7] many other databases are freely available in case of speaker recognition.

Multimodal databases are not as widely obtainable as unimodals. To handle such scarcity, researchers either create their own database, which is typically effort-heavy, or may opt for the following interesting approach. If two different traits are uncorrelated (features of one trait does not depend on another, for the same person), like face and iris, identifiers obtained from different sources may be assumed as the same person's data [30]. However, for correlated traits, like ear and skin-color, such assumption is inane. In multimodal scenarios, the databases available are BIOMET [17] (face, hand, fingerprint, voice and signature), BANCA [4] (face and voice), MYCT [40] (fingerprint and signature) etc.

8. Conclusion in workflow design: challenges and benefits

Some of the major problems commonly come across in MMB research scenarios are enlisted below.

i. Class-variability: In vision-based MMB recognition, the presence of a high degree of variability in human biometric information has been a major concern. In other words, there exists an

infinite number of possible classes in a biometric database. There can be potentially very large intra-class variations and also rather small inter-class variations (due to the similarity of individual appearances).

ii. Noisy data: The presence of noisy and distorted data in the captured sample and thereby production of degraded quality samples is causing database enrollment error. This may further lead to a failure in identification algorithm and thereby arrival at a faulty decision.

iii. Spoofing attack: In order to get unauthorized access in MMBs, biometric spoofing has often been applied to compromise the system. To combat such attacks, researchers are developing various anti-spoofing technologies.

iv. Population coverage: The lack of population coverage is another challenge faced by the system designers to construct proper databases. However, construction of large sample sizes manifests big data problems. Naturally, the sensed data is stored in compressed encrypted form for effective space utilization and speedy probing. However, such a storing scheme makes it difficult to retrieve the data in originally captured lossless form.

v. Social acceptance: Social issues pose barriers in the acquisition level of a MMBS. For example, face may be considered more user friendly than iris as the former requires less human interaction. On the contrary, due to the same reason the former can be captured without human intervention and causes privacy threats. A common threat in social media, function creep, where the captured traits are not utilized for the legitimate purpose, is difficult to restrain.

This review article offers several benefits to the emergent researchers in the field of MMBS, enlisted below.

i. Substantive amount of referred research articles to foster the circumstantial workflow designing.

ii. Choice of biometric identifiers as per researcher’s need.

iii. Selection of multimodal databases or formation of multimodal databases using different unimodal sources.

iv. Detailed comparison of fusion strategies to design the workflow.

v. Selection of a particular classification algorithm or developing the classifier fusion strategy in workflow.

vi. The explicit choice of authentication method and subsequently arriving at final verdict using proper performance metrics.

References

1. Afaneh A., Alqaralleh E., Toygar O. Person identification using multimodal biometrics under different challenges // In: Person Identification Using Multimodal Biometrics Under Different Challenges. 2018. Chapter 5. P. 81–100.
2. Alshardan A., Kumar A., Alghamdi M., Maashi M., Alahmari S., Alharbi A. A. K., Almukadi W., Alzahrani Y. Multimodal biometric identification: leveraging convolutional neural network (CNN) architectures and fusion techniques with fingerprint and finger vein data // PeerJ Computer Science. 2024. Vol. 10. Article e2440.
3. Arputhamoni S., Jireh J., Sarvanan A. G. Online smart voting system using biometrics based facial and fingerprint detection on image processing and CNN // Proc. of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV). IEEE, 2021.
4. Bailly-Bailliere E., Bengio S., Bimbot F., Hamouz M., Kittler J., Mariethoz J., Matas J., Messer K., Popovici V., Poree F., Ruiz B., Thiran J. P. The BANCA database and evaluation protocol // Proceedings of the 4th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA). LNCS 2688. 2003. P. 625–638.
5. Bokade G. U., Kanphade R. D. A novel approach for secured multimodal biometric authentication based on data fusion technique // International Journal of Computational Vision and Robotics. 2021. Vol. 11, no. 2. P. 214–243.
6. Byeon H., Raina V., Sandhu M., Shabaz M., Keshta I., Soni M., Matrouk K., Singh P. P., Lakshmi T. R. V. Artificial intelligence-enabled deep learning model for multimodal biometric fusion // Multimedia Tools and Applications. 2024. P. 1–24.
7. Campbell J., Reynolds D. Corpora for the evaluation of speaker recognition systems // Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP’99). 1999. Vol. 2. P. 829–832.
8. Chanukya P. S. V. V. N., Thivakaran T. K. Multimodal biometric cryptosystem for human authentication using fingerprint and ear // Multimedia Tools and Applications. 2020. Vol. 79, no. 1. P. 659–673.
9. Cherifi F., Amroun K., Omar M. Robust multimodal biometric authentication on IoT device through ear shape and arm gesture // Multimedia Tools and Applications. 2021. Vol. 80, no. 10. P. 14807–14827.

10. Chinese Academy of Sciences' Institute of Automation (CASIA) fingerprint dataset. Available at: http://biometrics.idealtest.org/findTotalDbByMode.do?mode=Fingerprint#/ (accessed June 05, 2025).
11. Chinese Academy of Sciences' Institute of Automation (CASIA) iris dataset. Available at: http://biometrics.idealtest.org/ (accessed June 05, 2025).
12. Choudhury S. H., Kumar A., Laskar S. H. Adaptive management of multimodal biometrics—a deep learning and metaheuristic approach // Applied Soft Computing. 2021. Vol. 106. Article 107344.
13. Coelho K. K., Tristão E. T., Nogueira M., Vieira A. B., Nacif J. A. M. Multimodal biometric authentication method by federated learning // Biomedical Signal Processing and Control. 2023. Vol. 85. Article 105022.
14. El-Bendary A. M., Mohsen H., Kasban H., Haggag A. Investigating of nodes and personal authentications utilizing multimodal biometrics for medical application of WBANs security // Multimedia Tools and Applications. 2020. Vol. 79, no. 33. P. 24507–24535.
15. El Rahman S. A., Alluhaidan A. S. Enhanced multimodal biometric recognition systems based abstract: on deep learning and traditional methods in smart environments // PLOS ONE. 2024. Vol. 19, no. 2. Article e0291084.
16. Fiumara G., Flanagan P., Grantham J., Bandini B., Ko K., Libert J. NIST Special Database 300 uncompressed plain and rolled images from fingerprint cards. National Institute of Standards and Technology. DOI: 10.18434/T4/1502472. Available at: https://doi.org/10.18434/T4/1502472 (accessed December 17, 2019).
17. Garcia-Salicetti S., Beumier C., Chollet G., Dorizzi B., Leroux les Jardins J., Lunter J., Ni Y., Petrovska-Delacretaz D. BIOMET: A multimodal person authentication database including face, voice, fingerprint, hand and signature modalities // Proceedings of the 4th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA). LNCS 2688. 2003. P. 845–853.
18. Georghiades A. S., Belhumeur P. N., Kriegman D. J. From few to many: Illumination cone models for face recognition under variable lighting and pose // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001. Vol. 23, no. 6. P. 643–660.
19. Goh K. O. M., Tee C., Teoh A. B. J. A contactless biometric system using multiple hand features // Journal of Visual Communication and Image Representation. 2012. Vol. 23, no. 7. P. 1068–1084.

20. Goh R., Liu L., Liu X., Chen T. The CMU face in action (FIA) database // In: Zhao W., Gong S., Tang X., editors. Analysis and Modelling of Faces and Gestures. Springer Berlin Heidelberg, 2005.
21. Gokul Kumari G. Metaheuristic-enabled artificial neural network framework for multimodal biometric recognition with local fusion visual features // The Computer Journal. 2021.
22. Hattab A., Behloul A. Face-iris multimodal biometric recognition system based on deep learning // Multimedia Tools and Applications. 2024. Vol. 83, no. 14. P. 43349–43376.
23. Ipeayeda F. W., Oyediran M. O., Ajagbe S. A., Jooda J. O., Adigun M. O. Optimized gravitational search algorithm for feature fusion in a multimodal biometric system // Results in Engineering. 2023. Vol. 20. Article ID 101572.
24. Jain A. K., Nandakumar K., Ross A. Score normalization in multimodal biometric systems // Pattern Recognition. 2005. Vol. 38. P. 2270–2285.
25. Jain A. K., Ross A., Prabhakar S. An introduction to biometric recognition // IEEE Transactions on Circuits and Systems for Video Technology. 2004. Vol. 14, no. 1. P. 4–20.
26. Joshi V. B., Raval M. S. Adaptive threshold for fingerprint recognition system based on threat level and system load // Procedia Computer Science. 2020. Vol. 171. P. 498–507.
27. Kakadiaris I. A., Passalis G., Theoharis T., Toderici G., Konstantinidis I., Murtuza N. Multimodal face recognition: combination of geometry with physiological information // Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). 2005. Vol. 2. P. 1022–1029.
28. Kant C., Chaudhary S. A multimodal biometric system based on finger knuckle print, fingerprint, and palmprint traits // In: Innovations in Computational Intelligence and Computer Vision. Springer, Singapore, 2021. P. 182–192.
29. Karmakar D., Mukherjee P., Datta M. Spoofed facial presentation attack detection by multivariate gradient descriptor in micro-expression region // Pattern Recognition and Image Analysis. 2020. Vol. 31, no. 2. P. 285–294.
30. Karmakar D., Murthy C. A. Generation of new points for training set and feature-level fusion in multimodal biometric identification // Machine Vision and Applications. Springer. 2013. Vol. 24, no. 6. July.
31. Karmakar D., Sarkar R., Datta M. Spoofed replay attack detection by multidimensional Fourier transform on facial micro-expression regions // Signal Processing: Image Communication. Elsevier. 2021. Vol. 93.

32. Keogh E., Mueen A. Curse of dimensionality // Encyclopedia of Machine Learning and Data Mining. 2017. P. 314–315.
33. Kulkarni A., Chong D., Batarseh F. A. Foundations of data imbalance and solutions for a data democracy // Data Democracy. Academic Press, 2020. P. 83–106.
34. Labayen M., Vea R., Flórez J., Aginako N., Sierra B. Online student authentication and proctoring system based on multimodal biometrics technology // IEEE Access. 2021. Vol. 9. P. 72398–72411.
35. Li S., Zhang B., Fei L., Zhao S. Joint discriminative feature learning for multimodal finger recognition // Pattern Recognition. 2021. Vol. 111. Article ID 107704.
36. Li S., Zhang B., Zhao S., Yang J. Local discriminant coding based convolutional feature representation for multimodal finger recognition // Information Sciences. 2021. Vol. 547. P. 1170–1181.
37. Luo Z., Gu Q., Su G., Zhu Y., Bai Z. An adaptive face-iris multimodal identification system based on quality assessment network // In: International Conference on Multimedia Modeling. Springer, Cham, 2021. P. 87–98.
38. Machala L., Dobes M. Upol iris image database, 2008. Available at: http://phoenix.inf.upol.cz/iris/ (accessed June 05, 2025).
39. Multimedia University. MMU1 and MMU2 iris image databases, 2008. Available at: https://www.kaggle.com/datasets/naureenmohammad/mmu-iris-dataset (accessed June 05, 2025).
40. Ortega-Garcia J., Fierrez-Aguilar J., Simon D., Gonzalez J., Faundez Zanuy M., Espinosa V., Satue A., Hernaez I., Igarza J. J., Vivaracho C., Escudero D., Moro Q. I. MCYT baseline corpus: A bimodal biometric database // IEE Proceedings - Vision, Image and Signal Processing. 2003. Vol. 150, no. 6. P. 395–401.
41. Phillips P. J., Wechsler H., Huang J., Rauss P. J. The FERET database and evaluation procedure for face-recognition algorithms // Image and Vision Computing. 1998. Vol. 16, no. 5. P. 295–306.
42. Princeton University Library. Face image databases. Available at: https://libguides.princeton.edu/facedatabases#s-lg-box-27701497 (accessed June 05, 2025).
43. Purohit H., Ajmera P. K. Multimodal multilevel fusion of face ear iris with multiple classifications // In: International Conference on Modelling, Simulation and Intelligent Computing. Springer, Singapore, 2020.
44. Rajasekar V., Saracevic M., Hassaballah M., Karabasevic D., Stanujkic D., Zajmovic M., Tariq U., Jayapaul P. Efficient multimodal biometric recognition for secure authentication

based on deep learning approach // International Journal on Artificial Intelligence Tools. 2023. Vol. 32, no. 03. Article 2340017.

45. Rattani A., Tistarelli M. Robust multi-modal and multi-unit feature level fusion of face and iris biometrics // University of Sassari - Computer Vision Laboratory. (CVL 2008/003). Technical Report, 16 Dec. 2008. P. 15.
46. Ross A., Jain A. K. Multimodal biometrics: An overview // Proc. of 12th European Signal Processing Conference (EUSIPCO). Vienna, Austria, Sept. 2004. P. 1221–1224.
47. Ryu R., Yeom S., Kim S. H., Herbert D. Continuous multimodal biometric authentication schemes: A systematic review // IEEE Access. 2021. Vol. 9. P. 34541–34557.
48. Safavipour M. H., Doostari M. A., Sadjedi H. Deep hybrid multimodal biometric recognition system based on features-level deep fusion of five biometric traits // Computational Intelligence and Neuroscience. 2023. Vol. 2023, no. 1. Article 6443786.
49. Samaria F. S. ORL database of faces. Face recognition using hidden Markov models, Doctoral dissertation. University of Cambridge, 1994.
50. Sarangi P. P., Nayak D. R., Panda M., Majhi B. A feature-level fusion based improved multimodal biometric recognition system using ear and profile face // Journal of Ambient Intelligence and Humanized Computing. 2021. P. 1–32.
51. Saiyed U., Sardar A., Rout R. K., Tanveer M., Razzak I. IoT-enabled multimodal biometric recognition system in secure environment // IEEE Internet of Things Journal. 2023.
52. Singh L. K., Khanna M., Garg H. Multimodal biometric based on fusion of ridge features with minutiae features and face features // International Journal of Information System Modeling and Design (IJISMD). 2020. Vol. 11, no. 1. P. 37–57.
53. Singhal M., Shinghal K. Secure deep multimodal biometric authentication using online signature and face features fusion // Multimedia Tools and Applications. 2024. Vol. 83, no. 10. P. 30981–31000.
54. Sujatha E., Sundar J. S. J., Deivendran P., Indumathi G. Multimodal biometric algorithm using iris, finger vein, fingerprint with hybrid GA, PSO for authentication // In: Data Analytics and Management. Springer, Singapore, 2021. P. 267–283.
55. Suntharam V. S., Chandu R., Rajan D. P. Robust multimodal biometric recognition based on joint sparse representation // ICCCE. Springer, Singapore, 2021. P. 265–274.
56. Talreja V., Valenti M. C., Nasrabadi N. M. Deep hashing for secure multimodal biometrics // IEEE Transactions on Information Forensics and Security. 2020. Vol. 16. P. 1306–1321.

57. Vyas R., Kanumuri T., Sheoran G., Dubey P. Accurate feature extraction for multimodal biometrics combining iris and palmprint // Journal of Ambient Intelligence and Humanized Computing. 2021. P. 1–9.
58. Yeung D. Y., Chang H., Xiong Y., George S., Kashi R., Matsumoto T., Rigoll G. SVC2004: First international signature verification competition // Proceedings of the 2004 Biometric Authentication: First International Conference (ICBA 2004), Hong Kong, China. 2004. P. 16–22.
59. Waleed D., Fadewar H. S. Multimodal biometric system: A review // International Journal of Research in Advanced Engineering and Technology. 2018. Vol. 4. P. 25–31.