OPEN ACCESS
The Journal of Engineering and Exact Sciences – jCEC, Vol. 10 N. 07 (2024)
journal homepage: https://periodicos.ufv.br/jcec
eISSN: 2527-1075
ISSN: 2446-9416

Advancements in Multi-Modal Face Recognition: A State-of-the-Art Review of Security Biometrics

Article Info:
Article history: Received 2024-10-02 / Accepted 2024-10-20 / Available online 2024-10-27
doi: 10.18540/jcecvl10iss7pp20163

Khaled Merit
ORCID: https://orcid.org/0000-0002-7762-1898
Laboratory of TIT, Department of Electrical Engineering,
Tahri Mohammed University of Bechar, Algeria
E-mail: merit.khaled@univ-bechar.dz
Mohammed Beladgham
ORCID: https://orcid.org/0000-0002-2371-6859
Laboratory of TIT, Department of Electrical Engineering,
Tahri Mohammed University of Bechar, Algeria
E-mail: beladgham.mohammed@univ-bechar.dz

Abstract
This survey is a state-of-the-art review of recent advancements in face recognition, particularly multi-modal face recognition and its associated applications in the domain of security biometrics and identity verification. In this regard, the paper sheds light on the significance of the Sejong Face database and several other leading databases as equally significant tools in studying some of the most challenging tasks in face recognition: occluded and diverse face identification, cross-modality evaluation, and infrared-visible face detection. It also insists on the high challenge of face identification across scenarios and with different forms of disguise. In addition, it emphasizes the need for an effective security framework to be implemented with other major sectors that deal in finances. The purpose is to develop an exhaustive list of methodologies and innovations based on the Sejong Face Database with the intention of finding new methods and algorithms that guarantee accuracy and reliability. Specifically, the application of deep learning methods comprising convolutional neural networks with cross-modality discriminator networks and unit-class loss demonstrates significant advancements in the development of face recognition systems; consequently, these systems exhibit enhanced security and efficacy. The review article is an important source that points out the current status of the research, as well as future possibilities, regarding biometric authentication and its implications in the digital scenario.
Keywords: face recognition. security biometrics. Sejong Face database (SFD). occluded face identification. convolutional neural networks (CNN). biometric authentication.

1. Introduction

In recent years, there has been a significant increase in the advancement of facial recognition technology. This development has resulted in a rise in academic research centered on multi-modal face recognition, particularly regarding its applications in areas such as security, biometrics, and identity verification. Works presented in this study are state-of-the-art contributions to modern research, utilizing the power of the Sejong Face database in addition to other renowned databases. The utilization of broad databases, known for their rich and varied datasets, has thus far enabled researchers to delve deep into intricate aspects of face recognition, including but not limited to occluded and varied pose face recognition, cross-modality assessment, and the fusion of thermal and visible face recognition.

The main focus of these studies is the problems that are there to exactly identify face features under several conditions and disguises. The capability with much importance for security systems and user verification in sectors such as banking and other critically sensitive areas is to recognize faces under various disguises. The Sejong Face database gives the added advantage of providing a dataset that is multi-modal, designed for disguised facial recognition. This collection involves images acquired by different modalities, including visible light, infrared, and thermal, and their combinations. The large amount of data makes possible a more sophisticated approach to face recognition, which successfully overcomes key issues like demonstration attacks and face occlusions.

This survey aims to generate an exhaustive inventory of the methods and developments implemented inside the Sejong Face Database, emphasizing facial recognition techniques' boundaries. This research aims to study new methods and algorithms that improve precision and reliability in this area. The application of convolutional neural networks and deep learning approaches, in conjunction with the examination of multimodal classifier networks and unit-class loss, demonstrated in this study, represents significant progress in advancing facial recognition (FR) techniques that are less risky and more effective. By skillfully navigating the complexities of various face alterations and changing contextual conditions, researchers contribute substantially to the advancing field of biometric verification and its future deployments in our increasingly digitized society.

The paper is structured in the following manner: Section 2 provides an overview of the literature study pertaining to the utilization of the SFD in various research publications. Following this, Section 3 presents an exhaustive compilation of databases that have been employed as authentication or verification systems, in addition to the SFD. Section 4 depicts the methodology used for the systematic review. Section 5 does the analysis of the results and explanation. Concluding remarks and future directions for research are described in section 6.

2. Literature Review

This section provides a review of the literature pertaining to previous research related to applications of face recognition using the Sejong Facial Database. Cheema *et al.* (2021), are credited for novelty in approach when it comes to High Frequency Trading (HFR). Conventionally, HFR has been heavily dependent on approaches that include image preprocessing, feature extraction, and common subspace projection.

However, these methods have been hindered by optimization difficulties and prone to performance error accumulation. The Cross-Modality Discriminator Network has a relational learning capability and implements Unit-Class Loss that are far superior for handling challenging modality differences, such as in visible-to-thermal recognition. This research provides a significant advancement within the domain of security and biometrics by presenting a refined and optimized framework for FR that encompasses various imaging modalities. The importance of this framework cannot be overstated, as it is crucial for the effective execution of facial recognition in areas such as surveillance, access control, and other security-oriented applications. The findings could also be basis Curtain methods adopted for the studies that are likely to influence future research and development studies in cross-modality face recognition systems.

The study of Ahmad *et al.* (2021) basically tries to answer issues that plague face recognition systems as a result of the use of disguises. The authors point to deficiencies in existing algorithms of facial recognition and attribute this to the general lack of large-scale training data, especially for faces accessorized with various disguises like cosmetics, facial hair, or eyewear. The authors proposed a scheme for creating synthetic covert facial images using Generative Adversarial Networks with Cycle-Consistency Loss as a way of solving the problem. It hence involves the creation of a synthetic database in which the facial features are occluded. This methodology thus aims at enhancing the robustness in face recognition algorithms for occluded facial appearances.

The approach followed by the researchers implements CycleGAN-a style transfer technique for generating facial images, incorporating disguises but maintaining the individual's original identity. Furthermore, an automated filtering mechanism is used to remove low-quality images from the synthesized dataset. Experimental results show that training the synthesized database enhances the rank-1 recognition rate in facial recognition systems, hence proving the efficiency of the proposed methodology.

Cheema *et al.* (2023) presented a Deep Neighborhood Difference Relational Network (DNDR) coupled with a Joint Discrimination Loss (JDL) that was able to solve the hidden heterogeneous face recognition problem. Previous approaches have either generated images or pursued domain-invariant features. These normally face serious drawbacks such as the need for large-scale training datasets, misalignment of generated and real images, and loss of distinctive information in generated images. This paper proposes a DNDR network that solves these two problems by making the most of the deep feature relationships and hence bypassing the labor-intensive image preprocessing or gap bridging methods from one domain to another. Unlike other approaches, which are mainly based on the usage of metric loss functions, here a joint discrimination loss mechanism is integrated in the framework, which will further improve identity representation and cross-modality matching. This is confirmed by various analyses as superior to other approaches, particularly on how to handle the well-hidden problems of HFR. The study signifies a milestone in this area of research as it looks for a better and more efficient way of exploiting HFR for challenging applications like masks and different imaging modalities.

Kowalski *et al.* (2022) presented a new method for thermal-visible modalities face verification that is particularly helpful under bad conditions such as low light or high movement. Previous studies have generally relied on the visible spectrum data for face recognition and achieved high accuracy but faced problems when applied in a place with low visibility. The thermal-visible method overcomes this weakness, although other methods presented so far have demonstrated limited efficiency. It is a novel method that includes integrating many CNNs to handle various properties of thermal and visual spectra, hence it allows one to achieve much higher accuracy in recognition. What can also be said about it is that it is a very progressive approach, as it combines visible and thermal spectrum information in a way to make it more robust and reliable for the recognition process. Therefore, this proposes a new standard for the research that will be conducted in the future regarding this field.

More specifically, Abdullah *et al.* (2022) proposed a GAN-based approach to synthesize thermal images from visible light images using CycleGAN. This is particularly important given the limitation in the availability of thermal image data with which to train face recognition systems compared to the ample supply of visible light images. This research demonstrates that CycleGAN indeed enables the separation of stylistic features from thermal images, which can then be used on regular visible light images to create realistic thermal representations. The results extracted from the experiment conducted by using the Sejong Face Dataset are quite similar in appearance to actual thermal photography. Numerical proofs for this claim include the Fréchet Inception Distance and the Kernel Inception Distance. This study gives a very promising direction for the betterment of thermal face recognition technologies, mentioning the importance of this technology in security and surveillance sectors where it is indispensable.

Ahmad *et al.* (2022) presented a novel approach to generating fake disguised faces from a legitimate image. The strategy aims to increase the robustness of facial recognition systems against attacks using disguises. This method takes the best parts of both patchwise contrastive loss and cycle-consistency loss and mixes them. The writers show that their process, which only needs one picture of each person to be trained, can make face recognition systems much more accurate when tested on fake faces. Their tests used the Sejong Face Database, which showed promise in raising model precision by 20 to 25 percent based on the mask add-on. The paper adds to the field by discussing that there are few freely available databases of disguised faces and how hard it is to make a dataset that includes all the possible disguises.

Alkadi *et al.* (2023) discussed how Automatic Teller Machines (ATMs) may benefit from a multimodal biometric identification framework founded on machine learning. This method was developed to recognize and identify people even when things like hair and makeup obscured their facial features. The researchers discuss the usability and security issues surrounding automated teller machine access, emphasizing the limits of existing biometric methods, such as presentation attacks (face masks) and facial disguises (makeup).

The research studied the use of multiple modalities to enhance authentication accuracy: traditional visible light, thermal imaging, infrared imaging, and their fusion. More importantly, in training various machine learning models that later inform the proposed methodology, use was made of the Sejong multimodal disguise face dataset. This will go a long way toward reliably identifying users-be they enhanced or not. Hence, following an evaluation between numerous machine learning algorithms, the best and most efficient was selected by the researchers on ResNet-50. Several tests also validate the efficiency of the system in being strong in the detection of people wearing traditional face accessories and different ethnic headgear. Nevertheless, problems were found in identifying people who had undergone dramatic changes to their looks, such as women who had put on a Hijab after being shot without one. Overall, the study contributes greatly to the field of biometric authentication, delivering a complex method that combines security and consumer comfort at ATMs.

3. Databases

This section outlines the datasets utilized with the SFD during this study. The SFD (Cheema *et al.*, 2021) is a comprehensive dataset developed to facilitate the creation of concealed FR solutions appropriate for implementation in business applications such as security inspections, where such solutions are necessitated. The SFD comprises images from 100 individuals divided into two distinct groups. Subgroup-A includes 30 participants (16 male, 14 female), each contributing one image across all modalities. Subgroup-B consists of 70 subjects (44 male, 26 female), each providing five or more photos across all modalities. Subgroup-A contains 1500 images, while Subgroup-B has 23100. Each facial image is captured using four spectra: visible (VIS), visible plus infrared (VIS-IR) (, infrared (IR), and thermal. The dataset incorporates eight facial accouterments (such as false beards, wigs, and spectacles) and seven combinations of these items to create a wide array of masked face images, as shown in Figure 1. The images were captured at two-week intervals to account for temporal changes such as facial hair growth. Gender-specific variations are incorporated, and the dataset encompasses a diverse representation of races and sexes. The subject identification, accessory type, and imaging mode are distinctly labeled on each image. In [8], this structured nomenclature facilitates an investigation into face detection and identification frameworks, particularly in disguised and occlusion scenarios. To demonstrate how challenging the dataset is and how it might help create techniques for masked face identification, we conducted some baseline tests with face detection, classification, and verification. The experiments indicated the need to improve face identification approaches to be more robust against facial add-ons.

The USTC-NVIE dataset of facial expressions contains 215 participants, featuring paired visible and infrared images [(Wang *et al.*, 2010), (Wang *et al.*, 2012)]. This dataset is split into two equal parts: one consisting of images capturing the apex of facial expressions (termed the "posed" dataset) and the other comprising images from the onset and offset of facial expressions (labeled the "spontaneous" dataset). Images are captured under three lighting conditions: left, right, and frontal. As shown in Figure 2, the Emotion Recognition Dataset is utilized by VIS-THE HFR in (Cheema *et al.*, 2021). Out of 215 subjects, 126 were deemed to have adequate data, consisting of sufficient images in both modalities. Scientists employed both sub-datasets to enhance the overall volume of training and test data. The training set comprised 16002 image pairs (visible and thermal) from 100 participants. For the test set, 4162 image pairs were used, sourced from 26 individuals.

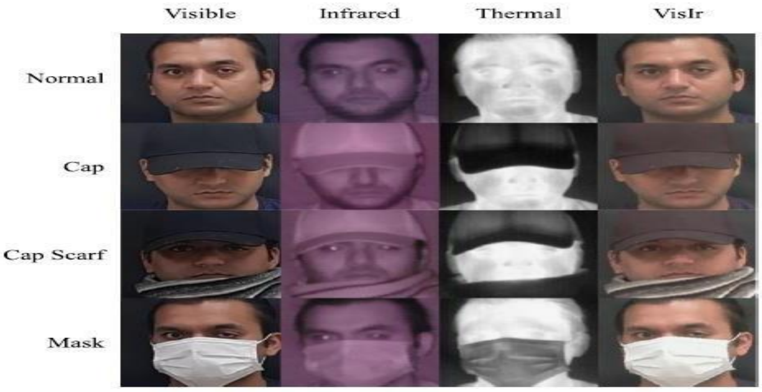

Figure 1 - Examples of pictures for a certain theme in the suggested library.
(Cheema *et al.*, 2021)

Figure 2 - Examples of visual (left) and thermal (right) pictures.
(Cheema *et al.*, 2021)

The TUFTS database (Panetta *et al.*, 2018). encompasses various data types, including two-dimensional visible, thermal, infrared, three-dimensional, three-dimensional LYTRO, sketch, and video. It contains images that showcase differences in pose, facial expressions, and sunglasses. To reduce the complexity associated with pose variations, only frontal images displaying changes in facial expressions and eyewear presence were selected for training and testing purposes. The study in Cheema *et al.* (2021) utilized a combination of thermal and visual images for Heterogeneous Face Recognition (HFR), as shown in Figure 3. The researchers compiled a dataset comprising 4812 pairs of visual and thermal images captured from 74 individuals for training purposes. Cheema *et al.* (2021) employed 2472 image pairs obtained from 38 subjects for testing.

Figure 3 - Examples of visual (left) and thermal (right) pictures.
(Cheema *et al.*, 2021)

The UND-X1 dataset [(Flynn *et al.*, 2003), (Chen *et al.*, 2005)] comprises 82 subjects with varying illumination, facial expressions, and time intervals in paired Long Wave Infrared (LWIR) and visible light images. For the training set, images from 50 out of the 82 subjects were selected, as shown in Figure 4. Each individual was represented by 40 distinct image pairs. The training process involved 50 subjects and 10002 image pairs. The evaluation was conducted using 32 subjects and 12802 image pairs.

Figure 4 - Examples of visual (left) and thermal (right) pictures.
(Cheema *et al.*, 2021)

The CASIA NIR-VIS 2.0 (Li *et al.*, 2013) dataset contains 725 subjects, each represented by visible and infrared images. For each subject, there are between 1-22 VIS images and 5-50 IR images. The dataset provides two distinct views for evaluation purposes. The first view is designated for training, while the second is used for testing. Researchers can simulate real-world conditions by employing VIS images as the gallery and NIR images as the probe (Cheema *et al.*, 2021). The gallery set contains only one VIS image per subject. To facilitate direct comparison with other studies, the standard testing protocol of View-2 is implemented, as shown in Figure 5.

The collection D4FLY Thermal and 2D Face includes pictures from 31 individuals (Kowalski *et al.*, 2022). Every individual's face was captured from a distance of 1.5 to 4 meters. Facial pictures were captured when the person approached and paused in front of the camera, with the head in a variety of orientations as demonstrated in Figure 6. Images were taken with a Basler acA2040-90uc camera (2040 x 2046 pixels) and a FLIR A65 camera (640 x 512 pixels).

Figure 5 - Examples of VIS and NIR face images of the same subject that vary in lighting, sharpness, age, and pose.
(Li *et al.*, 2013)

Figure 6 - Examples of face pictures a) Thermal infrared; b) visible range.
(Kowalski *et al.*, 2022)

The IOE_WAT dataset was gathered inside the confines of the Military University of Technology and comprises both visible and thermal infrared pictures of 40 participants (Kowalski *et al.*, 2018). The acquisition of visible pictures was performed using a Microsoft webcam camera and Microsoft Kinect v2, which had resolutions of 1280 × 720 pixels and 1920 x 1080 pixels, respectively as represented in Figure 7. The acquisition of thermal infrared pictures was conducted utilizing two different cameras, namely the Forward-Looking for IR (FLIR) A65 for 16 participants and the FLIR P640 for 24 participants. Both cameras possess a resolution of 640 × 512 pixels and exhibit a Noise Equivalent Temperature Difference (NETD) below 50 mK. Through the process of acquiring data, the participants were positioned in front of the camera at a distance of 1.5 meters. A total of over 100 pictures were acquired for every participant. The dataset comprises a collection of pictures of 12 individuals who are wearing eyeglasses.

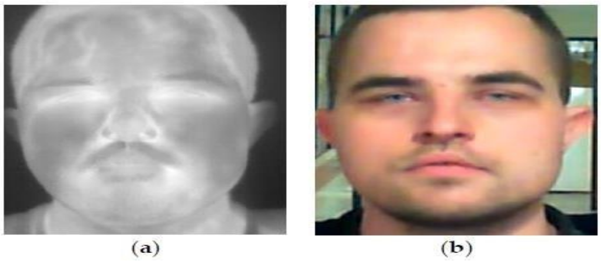

Figure 7 - Examples of face pictures a) Thermal infrared; b) visible range.
(Kowalski *et al.*, 2022)

The Speaking Faces dataset (Abdrakhmanova *et al.*, 2021) is a massive multimodal resource that includes both video and audio recordings of 142 individuals. The same nine camera sites were used for both the visual and thermal data collection sessions. In contrast to the second session, when participants read aloud from a script, participants in this one remained mute throughout. Figure 8 shows how subjects were positioned in front of the camera at a distance of around one meter. Visible pictures were shot with a 1920 x 1080-pixel C920 Pro HD web camera, while thermal images were taken with a 464 348-pixel FLIR T540.

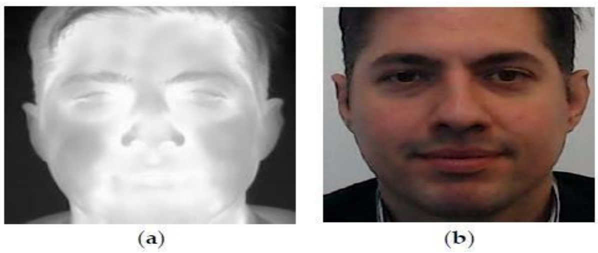

Figure 8 - Examples of face pictures a) Thermal infrared; b) visible range.
(Kowalski *et al.*, 2022)

As outlined in Ng *et al.* (2015), the FaceScrub dataset contains a diverse collection of 106,863 unconstrained visible images featuring 530 individuals. Roughly 200 pictures were collected for each subject, resulting in an extensive compilation of facial photographs. As illustrated in Figure 9, the dataset demonstrates a balanced representation of participants, with an equal number of males and females. Throughout the entire collection, there is noticeable diversity in the characteristics of the images.

Figure 9 - Examples of visible face pictures.
(Kowalski *et al.*, 2022)

The CUFS (CUHK Face Sketch) dataset (Wang *& Tang*, 2008) comprises facial images from three sources: 188 from the Chinese University of Hong Kong's student database, 123 from the AR database (Martinez *et al.*, 1998), and 295 from the XM2VTS database (Messer *et al.*, 2000). For each face, an artist has produced a sketch based on a photograph taken in a frontal position, with a neutral facial expression, and under standardized lighting conditions, as shown in Figure 10.

The IIIT-D Sketch database (Bhatt *et al.*, 2012) encompasses a range of sketch types, including viewable, semi-forensic, and forensic sketches. These can be visually inspected, as illustrated in Figure 11. Researchers assessed the method's effectiveness in (Cheema *et al.*, 2023) using two datasets from the IIIT-D database: the viewed sketch and semi-forensic sketch collections. The viewed sketch subset of IIIT-D comprises 238 pairs of images. The sketch pairs are obtained from three different databases: 67 from FG-Net aging (Fu *et al.*, 2014), 99 from Labelled Faces in the Wild (LFW) (Huang *et al.*, 2008), and 72 from IIIT-D students and staff. Additionally, the IIIT-D Visible-drawing subset contains 140 image pairs. Each pair in this collection consists of a sketch created from memory after observing a digital image.

Figure 10 - Representations of a facial picture and a hand-drawn sketch.
(Wang *& Tang*, 2008)

Figure 11 - Samples of forensic drawings where face attributes are emphasized.
(Bhatt *et al.*, 2012)

The I2BVSD dataset [(Dhamecha *et al.*, 2013), (Dhamecha *et al.*, 2014)] contains frontal facial images of subjects displaying neutral expressions captured under uniform illumination. This collection features various disguise elements, including synthetic facial hair, headgear such as caps, hairpieces, face coverings, and vision aids like glasses, as illustrated in Figure 12a. The dataset comprises 75 South Asian subjects, 60 males and 15 females. It incorporates five distinct disguise categories, including variations in hair, facial hair (beards and mustaches), eyewear, headwear, and face coverings. An additional method involves the simultaneous application of multiple disguise elements. However, the database needs labels for disguises, which creates difficulties in utilizing it to improve disguised face recognition models. The dataset in (Ahmad *et al.*, 2021) contained 681 images for each modality. It's worth noting that every modality includes at least one picture of a frontal face and a variable number of frontal concealed images for each individual, ranging from 5 to 9.

The BRSU Spoof Database [(Steiner *et al.*, 2016), (Steiner *et al.*, 2016)] is an extensive collection incorporating various spectral ranges, including visible and infrared. It features image captures at specific wavelengths of 935, 1060, 1200, and 1550 nm, offering a broad spectrum of data for analysis. The database presents various complexities that pose challenges, including variations in facial expressions, makeup, 3D masks, fake beards, glasses, artificial noses, and presentation-based attacks, as illustrated in Figure 12b. These diverse elements contribute to needing help in analyzing and processing the database effectively. However, the database offers a restricted choice of five themes, each comprising a variable number of additional elements ranging from nine to thirty. The BRSU Spoof DB primarily concentrates on the field of multispectral analysis.

The Spectral Disguise Face Database (Raghavendra *et al.*, 2018) contains images of 54 male subjects in both their natural state and with disguises. These images cover a wavelength range from 530 to 1000 nanometers, encompassing both visible light and near-infrared (NIR) regions of the electromagnetic spectrum. The dataset is subsequently divided into two equal parts. The initial half comprises genuine, unobscured images, while the latter includes authentic and concealed versions of identical pictures. However, the disguise option is limited to only two beard lengths: regular and extended, as illustrated in Figure 12c. The authentic dataset comprises 22 subjects sporting beards, while the rest have mustaches. This limited range of facial disguises in the database prevents the effective development of a comprehensive algorithm for detecting various forms of disguise.

The CASIA SURF dataset (Zhang *et al.*, 2020) comprehensively composes various face presentation attack types. It encompasses 1000 Chinese subjects and features three distinct modalities: RGB, Depth, and IR. The system contains six attack categories, mostly involving individuals holding paper cutouts of facial images in six different shapes, as illustrated in Figure 12d. This data is unsuitable for training a public facial recognition application, as these attack methods could be more effective in real-world settings such as airport security checkpoints.

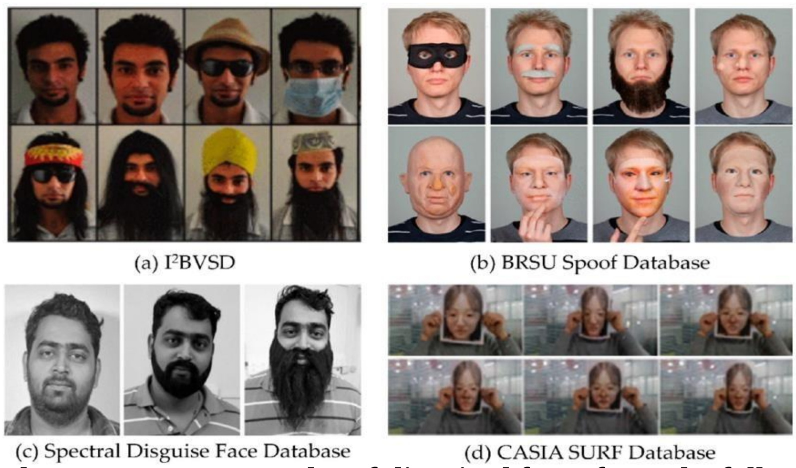

Figure 12 - These are some examples of disguised faces from the following databases: (a) I2BVSD; (b) BRSU Spoof Database; (c) Spectral Disguise Face Database; and (d) CASIA SURF Database.
(Ahmad *et al.*, 2021)

4. Adopted Research Methodologies

The section provides a broad review of the different methodologies applied and adopted in the research and studies mentioned above. The methodology adopted for the research study is the creation of an overall CMDN for HFRs (Cheema *et al.*, 2021). The CMDN has a DRD module, which will help to an idea about the relations among images across different domains. The network further optimizes itself by making use of Unit-Class Loss. Such a loss function has been designed with much stability and precision, beating all other metric-learning loss functions.

The CMDN backbone undergoes initial training using VGGFace2 weights. Subsequently, it is fine-tuned with the IRIS face database to enhance its capability for classifying visual and thermal facial data. The backbone network incorporates the DRD module to enable HFR training. The neural network produces Facial embedding vectors, which the HFR then utilizes for classification purposes. This approach results in a combination of embedding vectors and classification probabilities. The network's performance is evaluated using several datasets, including UND-X1, TUFTS, CASIA NIR-VIS, USTC-NVIE, and the Sejong Face Database.

This is based on the fact that the underlying method relies mainly on the use of complex relationships among deep features to enable effective cross-domain face recognition and, hence, extract modality-agnostic embedding vectors. The Unit-Class Loss function-a part of this method that has been devised to improve feature learning by taking into consideration both the unique properties of individual instances and the holistic distributions of classes-is an important part of this method. Its efficiency is demonstrated in terms of considerable improvements over existing state-of-the-art methodologies in different datasets.

The innovative approach introduced in (Cheema *et al.*, 2023) for tackling the issues related to concealed HFR surpasses the constraints of earlier methods. The DNDR Network, JDL, and a multi-stage training strategy constitute the three critical components of this solution. Notably, the DNDR network's unique design incorporates modality independence, allowing it to process and examine diverse data types across various modalities efficiently. It comprises the training process with interactions between datasets of identical and different modalities. The critical investigation of this research is into deep attribute relations between cross-modality images while ignoring specific traits relevant to real and disguised facial regions during network training. An algorithm proposed in this respect is LJD, which enhances the network to yield high embedding of images that represent identification information optimally in cross-modality matching. A multi-tier training approach here favors the rapid progress of training procedures and enhances the general effectiveness of the network. It outperformed all previous conventional methods on different HFR databases.

The methodology used in (Kowalski *et al.*, 2022) is comprehensive; it starts by integrating all the multiple phases: data collection, selection of algorithms, and establishment of an experimental design. Initially, the research began with preparing and annotating databases in which the datasets were classified into different categories, like glasses or no glasses, since they may have an effect on the thermal-visible face recognition. For the thermal images and the visual ones, two separate face-detection algorithms were developed. Several CNNs were trained for classification tasks to facilitate feature extraction. The training employed a composite dataset comprising both visible and thermal imagery. Pre-training was initially done on the CNNs using the ImageNet database.

Afterward, the study analyzed different methods, such as Siamese-based, Triplet-based, and Verification by Identification schemes. One of the main novelties uses the triple triplet method in which, within each branch in a triplet, three different CNNs are involved (Kowalski *et al.*, 2022). This will compute feature vectors to improve the system's performance in distinguishing between impostors and genuine individuals. The training procedure followed the use of transfer learning techniques with state-of-the-art CNNs. The implementation of this method initially relied on using the FaceScrub database, followed by incorporating a combined dataset comprising D4FLY and IOE_WAT datasets. To evaluate the effectiveness of the developed technique, testing was conducted on subjects wearing and not wearing glasses drawn from various datasets. The work describes an approach differentiated by the broad methodology followed for preprocessing, handling different datasets, and the construction of a robust and efficient thermal-visible FR.

First, the method in (Ahmad *et al.*,2021) consists of some key components. It primarily proposes using a GAN to generate fake faces similar to real ones. The process employs a cycle-consistency loss to maintain distinctive characteristics while incorporating hair, cosmetics, and eyewear to alter facial appearances. Initially, the SFD serves as the primary repository. This database contains numerous disguise enhancements and subject identifiers, which are utilized to expand the fake database with additional facial images.

It also has a process for filtering through the generated images. The tool is able to sort through the generated images, removing those not of quality, hence allowing only the best and most realistic photos. Such filtering is important in ensuring the quality of the data used for training and experimenting with face recognition is guaranteed.

A crucial research component involves conducting numerous experiments to evaluate the effectiveness of synthetic facial images. These tests aim to assess how well artificial data performs compared to genuine data when training facial recognition systems. The findings indicate that facial recognition systems perform more effectively when trained using the proposed synthetic database, demonstrating the technique's utility for generating high-quality, artificial facial images.

GANs create novel infrared and thermal facial images (Abdullah *et al.*, 2022). The CycleGAN, a model designed for unpaired image-to-image translation, transfers stylistic features from one data distribution to target images. This approach was selected due to the limited availability of training data for thermal face recognition. For validation purposes, the researchers employed a subset of the SFD consisting of 400 pairs of visual and thermal images from 70 patients. The separation of training and test data was accomplished using subject IDs. Subsequently, CycleGAN generated thermal versions based on the visible images. The approach was evaluated using Nvidia GTX 3080s and PyTorch. The study utilized Fréchet Inception Distance (FID) and Kernel Inception Distance (KID) as evaluation metrics. Results indicated that the generated images were high quality and appeared natural, resembling the original data. The approach's effectiveness was demonstrated through the obtained FID and KID scores.

The method outlined in Ahmad *et al.* (2021) seeks to create synthetically disguised facial images utilizing a single reference photo for each person. This approach was developed to bolster the ability of face recognition systems to withstand various deceptive tactics. The approach employs Contrastive Unpaired Training (CUT) and a cycle consistency loss-based technique to generate masked face images. CUT streamlines the training procedure using a single image and an integrated network that combines discriminator and generator functions. This study integrates CycleGAN’s cycle consistency loss with CUT's patchwise contrastive loss, employing a framework with dual generators and discriminators to overcome its shortcomings. Using StyleGAN2 to convert individual images from exposed to concealed facial features, the integrated method is subsequently enhanced. Most pictures and disguise augmentations used in the training process are sourced from the expanded subset B of the SFD. This approach aims to generate new facial databases or incorporate this technology into existing facial recognition training platforms to improve the accuracy of identifying disguised individuals.

Alkadi *et al.* (2023) introduce a machine learning-based multi-modal biometric authentication system for ATMs. This system employs various imaging techniques (VIS, IR, and thermal) to capture users' facial images. It is then trained to identify these faces even when they have been modified by elements such as hairstyles and facial accessories.

Such diversity helps in preventing different kinds of threats, including false identity theft and presentation attacks. Further enhancements in this regard include that the Sejong multi-modal disguised face dataset contains images in the visual, IR, thermal, and a combination of IR and visual spectra. The data is then preprocessed and divided into three clear categories: training, testing, and validation. Besides, we have used ML-based face recognition algorithms, namely, VGG-16, InceptionV3, and ResNet-50. The prepared dataset trains the models. The different scenarios involve testing machine learning models on a wide range of demographics and with faces having modifications in place. Moreover, because of their incorporation into an ATM system for user authentication, they maintain robustness for various disguises and presentation attacks under varying conditions without compromising on accuracy or usability. Therefore, the aim of this technology will be to identify one person from another continuously in a changing environment and also in changing facial alterations while enhancing the security without degrading the usability.

5. Results and Discussion

This segment allows the presentation of all the experimental results. It is portrayed from Cheema *et al.* (2021) that the proposed approach outperforms existing state-of-the-art methodologies in HFR considerably. Such an improvement in network efficacy is demonstrated across several datasets. The TUFTS Face Database, UND-X1, CASIA NIR-VIS, USTC-NVIE, and the Sejong Face Database are among these. The three critical contributions made by the researchers are the efficient CMDN for HFR, a Deep Relational Discriminator module that does away with the requirement for handmade loss functions, and the innovative Unit-Class Loss for optimizing the CMDN. This novel method improves the network's feature learning by considering individual samples and the entire class distributions, making it more resistant to limited training data, noisy samples, and huge modality disparities (Cheema *et al.*, 2021).

Given the successful application of VIS-THE, VIS-NIR, and other challenging datasets, researchers and developers should consider applying these techniques in diverse modalities and challenging scenarios. The techniques used in this research could be explored further for other applications of computer vision and deep learning, particularly those involving challenging cross-domain image processing tasks.

In the investigation that follow Cheema *et al.* (2023), this DNDR methodology showed marked improvements in Rank-1 identification and classification rates compared with many state-of-the-art algorithms for a range of test conditions. Results obtained on the SFD showed an improvement of 1.76% in VIS-NIR tests against the CMDC methodology and an improvement of 3.68% in VIS-THE trials over the ADCANS approach. Results obtained using the Tufts database proved that DNDR performed higher accuracy in the pairing of VIS-NIR, VIS-THE, and VIS-CSKE images in comparison with other methodologies. On CASIA NIR-VIS 2.0, the highest Rank-1 accuracy has been attained with a proposed DNDR methodology that outperformed other methods, DVG-Face, MAELR, DMBN, RGM, SSN, ADCANS, VSANet, DFAL, SSN, and CMDC. These abovementioned results indicate the efficiency of the DNDR network in solving HFR occlusion problems in the hidden challenging modalities.

The DNDR represents a fresh approach to face recognition, which has the excellent capability of recognizing faces across different modalities and disguises. Therefore, it would be ideal for applications in security and surveillance. Researchers into computer vision and biometrics should study this approach, since it can handle disguised HFR complexities. It could also be beneficial in other domains, such as access control, fraud detection, and smart city applications. Ongoing experimentation and adjustment to new challenges are crucial, while collaboration with industry partners may provide opportunities for the practical application of such sophisticated methods.

Kowalski *et al.* (2022) have used CNN architectures and a new approach called the triple triplet. The triple triplet strategy has three CNNs per triplet branch and gave the best results outperforming the earlier methods improving the TAR at FAR 1% to as high as 90.61%. Public datasets are used to study the design of Siamese, Triplet, and Verification Through Identification algorithms. Three triplets separated positives and impostors better, showing more similarity among photos of the same topic, and greater dissimilarity between subjects. Datasets used were IOE_WAT, D4FLY, and Speaking Faces to separate tests among those with and without spectacles. Biometric rates were quantified by TAR and FAR.

It is an advanced technology that detects and identifies faces over a range of lighting conditions and with the movement of subjects in real-world environments. This makes the triple triplet technique particularly suited to security-related applications. It could also be used as an enhancement for present facial recognition systems where improved lighting conditions or moving subjects degrade their effectiveness. Therefore, infrared and visible imaging techniques should be combined in biometric recognition systems as a method of increasing their robustness and accuracy. Since the proposed method focuses on execution time, it will be most applicable in fields such as real-time implementations: public transportation, event venues, and identity verification contexts. Stakeholders in industries and collaboration with government agencies will play an important role in successfully implementing this technology.

The research in Ahmad *et al.* (2021) introduces a "Synthetic Disguised Face Database" designed to evaluate and improve face recognition algorithms' performance when dealing with disguises. This database encompasses 13 different synthetic disguise types. Additionally, the study employs CycleGAN, which utilizes cycle consistency loss to learn mappings between domains, ensuring that the generated images are consistent with the original dataset. A sophisticated filtering system automatically removes subpar images to maintain training quality. This method utilizes a facial recognition (FR) model based on SqueezeNet, which is trained using genuine and manipulated images. The process eliminates the need for manual intervention and prevents overtraining. The outcomes demonstrate significant improvements in facial recognition technology.

The Synthetic Disguised Face Database can significantly improve the resilience of existing facial recognition (FR) systems against presentation and disguise attacks. Moreover, the proposed GAN and cycle-consistency loss method enables researchers to expand facial datasets by incorporating new disguise variations, thereby advancing FR technology. Given its effectiveness in handling disguises, this approach could be particularly advantageous for facial recognition (FR) systems in security applications that need to withstand spoofing and presentation attacks. The synthetic database and methodology can be utilized to train and evaluate FR algorithms, especially in scenarios where disguises are common. Moreover, the principles of automated filtering and synthetic image generation can be applied to other areas of computer vision and machine learning that emphasize data augmentation and training data quality control. Finally, valuable insights could be gained for refining and implementing these techniques through partnerships with industry stakeholders to test them in practical, real-world scenarios.

Quantitative analysis in Abdullah *et al.* (2022), verified through FID and KID metrics, demonstrated that the thermal images generated by CycleGAN closely matched the original dataset. The visual assessment also confirmed the strong resemblance between these two sets of images. Research indicates that Generative Adversarial Networks (GANs), particularly CycleGAN, effectively convert images from visible light to thermal spectrum. This capability has potential applications in areas requiring thermal image data, such as facial recognition systems.

GANs can be employed in surveillance and monitoring systems to improve the effectiveness of facial recognition in poorly lit conditions. This technique addresses the lack of infrared and thermal imaging data by creating artificial thermal images, which can augment training databases for facial recognition algorithms. This technology has potential applications for police forces and emergency responders. Additional studies are necessary to explore the conversion of images between various modes. For example, visual information could be transformed into formats similar to radar or sonar displays. Partnerships with government agencies and private companies may lead to practical implementations and further developments in this field.

Ahmad *et al.* (2021) study found that patchwise contrastive loss lets you train with as few as one sample per picture, and cycle consistency is used to further shape the networks for better results. Demonstrating its usefulness, the method can raise the accuracy of face recognition by 20 to 25 percent, relying on the disguise add-on, when taught on fake disguised data and tested on real disguised face pictures. There has been a big step forward in face recognition technology with this discovery, especially since the models were trained with just one picture per subject per add-on.

Security systems might be able to recognize people better even when they are wearing masks if they use a way to make fake faces that look like people. This way can make face recognition systems more reliable, especially in situations where people often wear masks. It can also add more training data for face recognition models by giving them a wider range of pictures to use as training data. In the real world, working together with law enforcement could lead to useful apps that improve safety and security for everyone.

Alkadi *et al.* (2023) study, a multi-modal facial recognition system for ATM biometric identification is examined. Machine learning algorithms trained on a multi-modal dataset identified people wearing masks or cosmetics. VGG-16, InceptionV3, and ResNet-50 were examined, and ResNet-50 had the greatest accuracy-to-resource efficiency ratio.

The various experiments done on the system showed high accuracies of identification in visible light, thermal imaging, infrared, and their combination. The performance of ResNet-50 in visual and thermal infrared modalities was at 99% with an average of 95% for general accuracy (Alkadi *et al.*, 2023). This technology fared exceptionally well in identifying people with normal facial accessories and distinguishing between a real user and an imposter using masks or photos.

The methodology was additionally evaluated with participants wearing Hijabs and Ghutras. The system demonstrated high levels of accuracy across various scenarios; however, it struggled to identify female participants who were initially not wearing a Hijab and later put one on during the evaluation, suggesting limitations in training and recognition (Alkadi *et al.*, 2023). Moreover, the system exhibited remarkable accuracy and resilience when confronted with challenges related to facial recognition, thereby positioning it as a promising alternative for biometric identification in ATMs.

A multi-modal face recognition system can make ATMs safer by going beyond standard ways of logging in, like PINs. You can use this method with other types of security systems, like entry control systems in safe buildings. To train machine learning models, you need a varied collection, with different masks and add-ons (Alkadi *et al.*, 2023). Developing such systems opens up new dimensions of research in face recognition technologies, which can solve problems regarding presentation attacks and disguises. They need to be tested and implemented for real applications with the collaboration of banks and other financial institutions. For added security, a multi-model approach can be combined with other fingerprint systems.

Table 1 provides an extensive overview of various scholarly articles that have utilized the Sejong Face Database in conjunction with other databases related to facial recognition technology. The research examines multi-modal facial recognition, emphasizing its security, biometrics, and identity verification applications. A table summarizes each paper, detailing the databases employed, methodologies implemented, outcomes achieved, and challenges encountered.

The various research efforts in facial recognition technology collectively advance the field by tackling key issues such as expensive imaging equipment, limited data availability, and the necessity for sophisticated algorithms that can handle disguised faces and multi-modal recognition. While these studies progress significantly, they underscore persistent challenges and the ongoing need for additional research in this rapidly developing area.

6. CONCLUSION

This review thoroughly examines the latest developments in face recognition systems, with particular emphasis on multi-modal techniques and their implementation in security-related biometric applications and identity confirmation processes. The text emphasizes the application of the Sejong Face database in investigating various facets of facial recognition technology. These aspects include concealed and diverse face recognition, analysis across different modalities, and the recognition of faces using thermal and visible imaging techniques. The study examines the difficulties in accurately recognizing faces across different circumstances and disguises, emphasizing its importance for security systems and critical industries such as banking.

However, the survey also reveals several knowledge gaps in the current state of facial recognition technology. These gaps include the high cost of thermal imaging devices, the lack of large-scale visible-thermal face datasets, the significant difference between visible and thermal/infrared modalities, dependence on data quality and availability, generalization and applicability concerns, the resource-intensive nature of network implementation and training, challenges with subjects wearing glasses, limited adaptability to rapid technological changes, and complexity in implementation.

The identified gaps in understanding highlight the importance of ongoing research and innovation in facial recognition systems. Future efforts should concentrate on tackling these issues to improve such technologies' precision, dependability, and real-world effectiveness. The study seeks to create a thorough catalog of methods and innovations within the Sejong Face Database, exploring novel techniques and computational processes to enhance precision and dependability.

The future works of this research direction of Multi-Modal Facial Recognition can be summarized in the following areas: (i) Apply VGG16 as a feature extraction for each type of database such as VIS, IR, VIS-IR, and Thermal images. (ii) Apply PCA to VGG16-extracted features to reduce their dimensionality. (iii) Combine those two models with the NN to improve image classification performance compared to using VGG16 features alone or PCA alone. (iv) Compare our results with other papers like Kowalski *et al.* (2022).

References

Abdrakhmanova, M., Kuzdeuov, A., Jarju, S., Khassanov, Y., Lewis, M., & Varol, H. A. (2021). SpeakingFaces: A large-scale multimodal dataset of voice commands with visual and thermal video streams. Sensors, 21(10), 3465. doi: https://doi.org/10.3390/s21103465

Abdullah, M., Lee, A., & Han, D. (2022, October). GAN based Visible to Thermal Image Translation. In 2022 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia) (pp. 1-3). IEEE. doi: https://doi.org/10.1109/icce-asia57006.2022.9954654

Ahmad, M., Abdullah, M., & Han, D. (2022, July). One-Shot Synthetic Disguised Face Generation to improve Robustness against Disguise Attacks. In 2022 37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC) (pp. 1038-1041). IEEE. doi:https://doi.org/10.1109/itc-cscc55581.2022.9895033

Ahmad, M., Cheema, U., Abdullah, M., Moon, S., & Han, D. (2021). Generating synthetic disguised faces with cycle-consistency loss and an automated filtering algorithm. Mathematics, 10(1), 4. doi: https://doi.org/10.3390/math10010004

Alkadi, A. M., AlMahdawi, R. A., Anwahi, S. M., & Soudan, B. (2023, February). Biometric Authentication Based on Multi-Modal Facial Recognition Using Machine Learning. In 2023 Advances in Science and Engineering Technology International Conferences (ASET) (pp. 1-6). IEEE. doi:https://doi.org/10.1109/aset56582.2023.10180532

Bhatt, H. S., Bharadwaj, S., Singh, R., & Vatsa, M. (2012). Memetically optimized MCWLD for matching sketches with digital face images. IEEE Transactions on Information Forensics and Security, 7(5), 1522-1535. doi:https://doi.org/10.1109/tifs.2012.2204252

Cheema, U., & Moon, S. (2021). Sejong face database: A multi-modal disguise face database. Computer Vision and Image Understanding, 208, 103218. doi: https://doi.org/10.1016/j.cviu.2021.103218

Cheema, U., & Moon, S. (2023). Disguised heterogeneous face recognition using deep neighborhood difference relational network. Neurocomputing, 519, 44-56. doi: https://doi.org/10.1016/j.neucom.2022.11.058

Cheema, U., Ahmad, M., Han, D., & Moon, S. (2021). Heterogeneous visible-thermal and visible-infrared face recognition using unit-class loss and cross-modality discriminator. doi: https://doi.org/10.1155/2022/4623368

Chen, X., Flynn, P. J., & Bowyer, K. W. (2005). IR and visible light face recognition. Computer Vision and Image Understanding, 99(3), 332-358. doi:https://doi.org/10.1016/j.cviu.2005.03.001

Dhamecha, T. I., Nigam, A., Singh, R., & Vatsa, M. (2013, June). Disguise detection and face recognition in visible and thermal spectrums. In 2013 International Conference on Biometrics (ICB) (pp. 1-8). IEEE. doi:https://doi.org/10.1109/icb.2013.6613019

Dhamecha, T. I., Singh, R., Vatsa, M., & Kumar, A. (2014). Recognizing disguised faces: Human and machine evaluation. PLOS One, 9(7), e99212. doi: https://doi.org/10.1371/journal.pone.0099212

Flynn, P. J., Bowyer, K. W., & Phillips, P. J. (2003). Assessment of time dependency in face recognition: An initial study. In Audio-and Video-Based Biometric Person Authentication: 4th International Conference, AVBPA 2003 Guildford, UK, June 9–11, 2003 Proceedings (pp. 44-51). Springer Berlin Heidelberg. doi:https://doi.org/10.1007/3-540-44887-x_6

Fu, Y., Hospedales, T. M., Xiang, T., Gong, S., & Yao, Y. (2014, September). Interestingness prediction by robust learning to rank. In European Conference on Computer Vision (pp. 488-503). Cham: Springer International Publishing. doi:https://doi.org/10.1007/978-3-319-10605-2_32

Huang, G. B., Mattar, M., Berg, T., & Learned-Miller, E. (2008). Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition. doi:https://inria.hal.science/inria-00321923v1

Kowalski, M. Ł., & Grudzień, A. (2018). High-resolution thermal face dataset for face and expression recognition. Metrology and Measurement Systems, 25(2). doi:https://doi.org/10.24425/119566

Kowalski, M., Grudzień, A., & Mierzejewski, K. (2022). Thermal–visible face recognition based on CNN features and triple triplet configuration for on-the-move identity verification. Sensors, 22(13), 5012. doi: https://doi.org/10.3390/s22135012

Li, S., Yi, D., Lei, Z., & Liao, S. (2013). The CASIA NIR-VIS 2.0 face database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 348-353). doi:https://doi.org/10.1109/cvprw.2013.59

Martinez, A., & Benavente, R. (1998). The ar face database: cvc technical report, 24.

Messer, K., Matas, J., Kittler, J., Luettin, J., & Maitre, G. (1999, March). XM2VTSDB: The extended M2VTS database. In Second International Conference on Audio and Video-Based Biometric Person Authentication (Vol. 964, pp. 965-966).

Ng, H. W., & Winkler, S. (2014, October). A data-driven approach to cleaning large face datasets. In 2014 IEEE International Conference on Image Processing (ICIP) (pp. 343-347). IEEE. doi:https://doi.org/10.1109/icip.2014.7025068

Panetta, K., Wan, Q., Agaian, S., Rajeev, S., Kamath, S., Rajendran, R., & Yuan, X. (2018). A comprehensive database for benchmarking imaging systems. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(3), 509-520. doi:https://doi.org/10.1109/tpami.2018.2884458

Raghavendra, R., Vetrekar, N., Raja, K. B., Gad, R. S., & Busch, C. (2018, August). Detecting disguise attacks on multi-spectral face recognition through spectral signatures. In 2018 24th International Conference on Pattern Recognition (ICPR) (pp. 3371-3377). IEEE. doi: https://doi.org/10.1109/icpr.2018.8545076

Steiner, H., Kolb, A., & Jung, N. (2016, June). Reliable face anti-spoofing using multispectral SWIR imaging. In 2016 International Conference on Biometrics (ICB) (pp. 1-8). IEEE. doi:https://doi.org/10.1109/icb.2016.7550052

Steiner, H., Sporrer, S., Kolb, A., & Jung, N. (2016). Design of an active multispectral SWIR camera system for skin detection and face verification. Journal of Sensors, 2016(1), 9682453. doi:https://doi.org/10.1155/2016/9682453

Wang, S., Liu, Z., Lv, S., Lv, Y., Wu, G., Peng, P., & Wang, X. (2010). A natural visible and infrared facial expression database for expression recognition and emotion inference. IEEE Transactions on Multimedia, 12(7), 682-691. doi:https://doi.org/10.1109/tmm.2010.2060716

Wang, S., Liu, Z., Wang, Z., Wu, G., Shen, P., He, S., & Wang, X. (2012). Analyses of a multimodal spontaneous facial expression database. IEEE Transactions on Affective Computing, 4(1), 34-46. doi:https://doi.org/10.1109/t-affc.2012.32

Wang, X., & Tang, X. (2008). Face photo-sketch synthesis and recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(11), 1955-1967. doi:https://doi.org/10.2139/ssrn.3998904

Zhang, S., Liu, A., Wan, J., Liang, Y., Guo, G., Escalera, S., & Li, S. Z. (2020). Casia-surf: A large-scale multi-modal benchmark for face anti-spoofing. IEEE Transactions on Biometrics, Behavior, and Identity Science, 2(2), 182-193. doi:https://doi.org/10.1109/tbiom.2020.2973001

JCEC
THE JOURNAL OF ENGINEERING AND EXACT SCIENCES

OPEN ACCESS
The Journal of Engineering and Exact Sciences – jCEC, Vol. 10 N. 07 (2024)
journal homepage: https://periodicos.ufv.br/jcec
eISSN: 2527-1075
ISSN: 2446-9416

Table 1: Summary of all articles that have used Sejong Face Datasets and other datasets with their methods and results.

Article Name	Used Database	Methods	Results	Limitations	Ref
Heterogeneous Visible-Thermal and Visible- Infrared Face Recognition Using Cross- Modality Discriminator Network and Unit-Class Loss	USTC-NVIE TUFTS UND-X1 Sejong Face CASIA NIR- VIS 2.0	A unified end- to-end Cross Modality Discriminator Network (CMDN) for HFR.	• The proposed fus method in the **USTC-NVIE** dataset achieves a remarkable 99.7% Rank-1 recognition rate, outperforming G-HFR. • In **TUFTS**, hfr and fus show significant improvements of 21.3% and 22.8%, respectively, over current methods. • In **UND-X1**, the fus output attains an impressive 95.21% Rank-1 accuracy, surpassing DPM and CpGAN. • **Sejong Face** sees a substantial improvement with the proposed method, achieving a 92.4% Rank-1 recognition rate for cross- domain matching. • In **CASIA NIR-VIS**	• The high cost of thermal imaging devices makes it harder to build and use systems that can recognize faces using heat. • Lack of Large-Scale Visible- Thermal Face Datasets: This makes it harder to train and test models thoroughly. • There is a big difference between what can be seen and what can be felt, which makes it hard to recognize faces correctly. • Focusing on the Cross-Modality Discriminator Network may make it harder to look into other methods. • Limited Generalization Due to Specific Datasets Used: The results can't be used with other datasets or in real life. • Reliance on Preprocessing and Data Augmentation: Could make it less useful with raw, unstructured data. • Networking implementation and training possess some resource-intensive characteristics that demand much time and computation.	**(Cheema *et al.*, 2021)**

			2.0, the proposed method excels with a 99.5% Rank-1 recognition rate.		

Thermal–Visible Face Recognition Based on CNN Features and Triple Triplet Configuration for On-the-Move Identity Verification	D4FLY Thermal and 2D Face IOE_WAT Speaking Faces Sejong Face FaceScrub	Triplet-Triplet method to compare faces in thermal and visible images.	• The triple triplet method outperforms other reference methods and achieves TAR @FAR 1% values up to 90.61%.	Limitations of this study include the following: • Data Bias: The datasets used in this study are biased towards gender; most of the samples were males. • Lack of Racial Diversity in Data: Geographical annotations are not present in the data collections, and hence, the racial diversity of participants cannot be assessed. • This would then limit the generalization of the results across various racial demographics. • Challenges with Subjects Wearing Glasses: the success of the face recognition system is greatly affected when people are wearing glasses. • This might make the problem challenging in computational resources and/or processing time, particularly in dynamic scenarios where speed plays a big role. • While this is a common practice, it means that the findings are contingent on the characteristics and limitations of these databases and models. • Real-world applications, especially in on-the-move scenarios, may present additional challenges not accounted for in this study. • Modality Gap: There is a significant modality gap between visible and thermal infrared images, which poses challenges for the face detection algorithms. • Performance Variation with Different CNN Models: The study indicates that the performance of the face recognition system varies significantly with different CNN models.	**(Kowalski *et al.*, 2022)**

Generating Synthetic Disguised Faces with Cycle- Consistency Loss and an Automated Filtering Algorithm	Database with Disguised Facial Images contains: I2BVSD BRSU Spoof Spectral Disguise Face CASIA SURF Sejong Face Synthetic Face Database which contains: Specs on Faces Virtual Makeup database	A methodology for generating synthetic facial images for the desired disguise add- ons. Cycle-consistency loss is used to generate facial images with disguises, e.g., fake beards, makeup, and glasses, from normal face images. Additionally, an automated filtering scheme is presented, for automated data filtering from the synthesized faces.	• Training on the proposed database achieves an improvement in the rank- 1 recognition rate (68.3%), over a model trained on the original non-disguised face images.	• Disguising Add-Ons Have a Gender Bias Beards and cosmetics are traditionally seen as feminine and masculine accessories, respectively. • Problems with Identifying Someone: It's hard to tell who someone is when they're covering up their face with a mask or a hood, thus combos like "scarf and cap" might be problematic. • Low-Quality Generated Images: Often require further filtering (both human and machine) to remove artefacts like blur and pixelation. • As a result of human mistake and prejudice, manual filtering can be inconsistent and add subjectivity. • High computational complexity in the approaches; this is mitigated by the use of a single generator for inference. • Possible but questionable success in a wide variety of real-world contexts; more confirmation of practical application is required.	**(Ahmad *et al.*, 2021)**

GAN based Visible to Thermal Image Translation	Sejong Face	A Generative Adversarial Network (GAN) based solution to solve the problem of data scarcity. The closed loop GAN system CycleGAN is used to generate novel thermal images from visible light images.	• The quality of generated images with very good and close to natural.	• The training procedure for the deep learning algorithm is said to be time-consuming and resource-intensive, which might be a problem in cases when such resources are few. • Generalizability Concerns: The paper does not clearly describe the method's generalizability to more varied or real-world datasets, which could be a problem in many practical applications. • Thirdly, the network's sophisticated architecture may make it difficult to apply and fine-tune, especially for those who aren't deep learning specialists. This is because the network is designed with several modules and complex training phases. • Dependence on High-Quality Data: The network's performance relies strongly on the quality and variety of training data, which might reduce its efficacy in situations with poor-quality or limited datasets. • Adaptability to Rapid Technological Changes: It is not made clear whether or not the approach can integrate or adapt with quickly expanding facial recognition technology and methodologies, which might be a hindrance in a rapidly growing area.	**(Abdullah *et al.*, 2022)**

One-Shot Synthetic Disguised Face Generation to improve Robustness against Disguise Attacks	Sejong Face	The proposed method utilizes patchwise contrastive loss and cycle- consistency loss to take benefit of both approaches.	• The results are satisfactory given that the models were trained with only one image per subject per add-on.	• Not required to use large number of images.	**(Ahmad *et al.*, 2022)**

Biometric Authentication Based on Multi- Modal Facial Recognition Using Machine Learning	2BVSD BRSU Spoof Spectral Disguise Face Sejong Face (Used as a reference result)	Three Machine Learning Algorithms: ResNet-50, VGG-16, InceptionV3 5 Different experiments: Sejong Face as a reference without Add- ons, Sejong Face as a reference with Add-Ons, Expand the datasets by adding 30 images with Hijab. In images, all modalities of a female without a Hijab were added to the dataset (without add- ons), Some images of male subjects were added to the dataset without the head cover. The model was then trained using the Sejong with add-ons and the male subject without Ghutra & add- ons.	• The system was able to achieve above 99% recognition accuracy on the Sejong dataset, even with the addition of common facial accessories such as sunglasses, face masks, caps, and scarves. Additionally, the system was able to correctly authenticate a Hijab-wearing female subject with different permutations of add- ons, except for the cap. The system was also able to correctly authenticate a male subject who was originally photographed with a bare head, then added a male head cover (Ghutra) and add-ons for the authentication test. The system authenticated the subject correctly regardless of the add-ons present. On the other hand, the system was not able to correctly authenticate a female subject who was originally photographed without a Hijab, and then donned the Hijab for the authentication test.	• The limitation is is its inability to correctly authenticate a female subject who was initially photographed without a Hijab and later donned the Hijab for the authentication test. The system struggled with this scenario because the Hijab covered significant facial features that the system was trained to recognize.	**(Alkadi *et al.*, 2023)**

Source: Authors.