SOAR

STATE-OF-THE-ART REPORT (SOAR)
DECEMBER 2021

HDIAC-BCO-2021-192

ARTIFICIAL INTELLIGENCE (AI) AND MACHINE LEARNING (ML) IN BIOMETRIC DATA FUSION

By Abdul Rahman, Ph.D., Steven R. Knudsen, Ph.D., Deanna C. Milonas, Daniel Fleming, and John Clements
Contract Number: FA8075-21-D-0001
Published By: HDIAC

This Page Intentionally Left Blank

SOAR

STATE-OF-THE-ART REPORT (SOAR)
DECEMBER 2021

ARTIFICIAL INTELLIGENCE (AI) AND MACHINE LEARNING (ML) IN BIOMETRIC DATA FUSION

ABDUL RAHMAN, PH.D., STEVEN R. KNUDSEN, PH.D.,
DEANNA C. MILONAS, DANIEL FLEMING, AND JOHN CLEMENTS

ABOUT HDIAC

The Homeland Defense & Security Information Analysis Center (HDIAC) is a U.S. Department of Defense (DoD) IAC sponsored by the Defense Technical Information Center (DTIC). HDIAC is operated by SURVICE Engineering Company under contract FA8075-21-D-0001 and is one of the three next-generation IACs transforming the DoD IAC program: HDIAC, Defense Systems Information Analysis Center (DSIAC), and Cybersecurity and Information Systems Information Analysis Center (CSIAC).

HDIAC serves as the U.S. national clearinghouse for worldwide scientific and technical information in 8 technical focus areas: alternative energy, biometrics, CBRN defense, critical infrastructure protection, cultural studies, homeland defense & security, medical, and weapons of mass destruction. As such, HDIAC collects, analyzes, synthesizes, and disseminates related technical information and data for each of these focus areas. These efforts facilitate a collaboration between scientists and engineers in the homeland defense and security systems community while promoting improved productivity by fully leveraging this same community's respective knowledge base. HDIAC also uses information obtained to generate scientific and technical products, including databases, technology assessments, training materials, and various technical reports.

State-of-the-Art Reports (SOARs)—one of HDIAC's information products—provide in-depth analysis of current technologies, evaluate and synthesize the latest technical information available, and provide a comprehensive assessment of technologies related to HDIAC's technical focus areas. Specific topic areas are established from collaboration with the greater defense systems community and vetted with DTIC to ensure the value-added contributions to Warfighter needs.

HDIAC's mailing address:

HDIAC
4695 Millennium Drive
Belcamp, MD 21017-1505
Telephone: (443) 360-4600

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE December 2021	**2. REPORT TYPE** State-of-the-Art Report	**3. DATES COVERED**

4. TITLE AND SUBTITLE Artificial Intelligence (AI) and Machine Learning (ML) in Biometric Data Fusion	**5a. CONTRACT NUMBER** FA8075-21-D-0001 **5b. GRANT NUMBER** **5c. PROGRAM ELEMENT NUMBER**
6. AUTHOR(S) Abdul Rahman, Ph.D., Steven R. Knudsen, Ph.D., Deanna C. Milonas, Daniel Fleming, and John Clements	**5d. PROJECT NUMBER** **5e. TASK NUMBER** **5f. WORK UNIT NUMBER**
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) AND ADRESS(ES) Homeland Defense & Security Information Analysis Center (HDIAC) SURVICE Engineering Company 4695 Millennium Drive Belcamp, MD 21017-1505	**8. PERFORMING ORGANIZATION REPORT NUMBER** HDIAC-BCO-2021-192
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Defense Technical Information Center (DTIC) 8725 John J. Kingman Road Fort Belvoir, VA 22060	**10. SPONSOR/MONITOR'S ACRONYM(S)** DTIC **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION/ AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

13. ABSTRACT
This State-of-the-Art Report focuses on recent advances in biometric recognition technologies and the application of artificial intelligence and machine learning to recognition tasks in multimodal identification systems. Multimodal systems use "feature-level" data fusion (e.g., periocular and gait recognition), which provides faster reference set retrieval across identity templates and significantly improves recognition accuracy over a unimodal system. Specifically, the recent field of *biometric data fusion* holds promise to deliver improved biometric data sample capture and analysis to the warfighter regardless of disguised, altered, or occluded facial characteristics. The use of convolutional neural networks, deep neural networks, and recurrent neural networks in biometric data fusion is also explained in this report. In addition, topics in leading-edge biometric recognition research are also presented.

14. SUBJECT TERMS
Biometric system, data fusion, artificial intelligence, AI, machine learning, ML, convolutional neural network, CNN, deep neural network, DNN, facial recognition, voice recognition, human gait recognition, HGR, recurrent neural network, RNN, deep learning, DL

15. SECURITY CLASSIFICATION OF: U			**16. LIMITATION OF ABSTRACT** UU	**17. NUMBER OF PAGES** 40	**18a. NAME OF RESPONSIBLE PERSON** Vincent "Ted" Welsh
a. REPORT UNCLASSIFIED	**b. ABSTRACT** UNCLASSIFIED	**c. THIS PAGE** UNCLASSIFIED			**18b. TELEPHONE NUMBER** *(include area code)* 443-360-4600

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

ON THE COVER:
(Source: 123rf.com)

THE AUTHORS

ABDUL RAHMAN, PH.D.

Dr. Abdul Rahman is a subject matter expert for KeyLogic Systems in the design and implementation of cloud analytics and architectures that support situational awareness tools for Cyber Network Defense and other national security missions. Dr. Rahman holds Ph.D.s in mathematics (algebraic topology) and physics (experimental condensed matter). He has published widely in the fields of physics, mathematics, and information technology.

STEVEN R. KNUDSEN, PH.D.

Steven Knudsen holds a Ph.D. in physics (West Virginia University) and works for KeyLogic Systems in Morgantown, WV as a contractor for the Department of Energy's National Energy Technology Laboratory. As a physicist solving systems engineering problems related to energy, he focuses on the electric grid and its cybersecurity. Dr. Knudsen's Ph.D. work used computational approaches to optimize the dynamics of space elevators for propulsion. His main coding languages are C++ for simulation and MATLAB/Mathematica for analysis.

DEANNA C. MILONAS

Deanna C. Milonas is a Research Analyst for HDIAC. She earned her B.S. degree in chemistry from Florida State University and her M.S. degree in biomedical engineering from the University of Florida. Before pursuing her master's degree, Ms. Milonas worked as an analytical chemist for 3 years performing chemical analysis on a variety of foods to screen for pesticides and antibiotics for the Food and Drug Administration and determining the nutritional content of foods for major businesses. While working on her master's degree, she worked as a biomedical engineering researcher focusing on the synthesis and characterization of polymeric magnetic micro- and nanoparticles and their therapeutic applications for drug delivery.

DANIEL FLEMING

Daniel Fleming is a Research Inquiry Analyst for HDIAC. During his 6 years as a combat engineer in the U.S. Marine Corps, Mr. Fleming applied his engineering skill set domestically for bridging operations, as well as internationally to assist in humanitarian relief efforts. For 4 years at a Berkshire Hathaway pharmaceutical subsidiary, he was the Lead Analytical Technician. He followed strict Good Manufacturing Practices guidelines; executed established methods; and assisted in the development of new, site-specific methods or programs. Mr. Fleming also ensured the accurate publication details in reports generated from tests on thousands of internally or externally developed drug or delivery vehicle suspensions.

THE AUTHORS, *continued*

JOHN CLEMENTS

John Clements is the Technical Lead for HDIAC. He enlisted in the U.S. Marine Corps Reserve in 2001 and continues to serve as the Regional Contingency Engineering Management Chief for Marine Forces, Pacific. He deployed three times to Iraq in support of Operation Iraqi Freedom. His prior work includes test and evaluation of procedures and systems related to chemical, biological, radiological, and nuclear decontamination; mortuary affairs; cyber insider threats; open-source and social media information; the Common Operational Picture used by Combatant Commands; and the Mounted Computing Environment. He has extensive experience working with Joint, Interagency, and Allied partners at the strategic and tactical levels. Mr. Clements holds an M.A. degree in Homeland Security from the American Military University.

ABSTRACT

This State-of-the-Art Report focuses on recent advances in biometric recognition technologies and the application of artificial intelligence and machine learning to recognition tasks in multi-modal identification systems. Multimodal systems use "feature-level" data fusion (e.g., periocular and gait recognition), which provides faster reference set retrieval across identity templates and significantly improves recognition accuracy over a unimodal system. Specifically, the recent field of *biometric data fusion* holds promise to deliver improved biometric data sample capture and analysis to the warfighter regardless of disguised, altered, or occluded facial characteristics. The use of convolutional neural networks, deep neural networks, and recurrent neural networks in biometric data fusion is also explained in this report. In addition, topics in leading-edge biometric recognition research are also presented.

CONTENTS

CONTENTS, *continued*

FIGURES

SECTION 01 INTRODUCTION

Biometric recognition and identification technologies are used widely across the federal government, primarily by the Department of Defense (DoD), Department of Homeland Security (DHS), Department of Justice (typically but not exclusively by the Federal Bureau of Investigation), and the Department of Commerce. In the Department of Commerce, research by the National Institute of Standards and Technology (NIST) supports the security and interoperability of federal and military biometric programs. Across the national security enterprise, biometric systems are a critical resource for forward-deployed U.S. forces as well as homeland defense operations. Biometrics support the detection and prevention of illegal border entry; help secure access to sensitive DoD and federal government networks and facilities; and aid civilian law enforcement in identifying, tracking, and detaining criminal suspects.

Within the DoD, biometrics is a key enabling activity for what the Joint Force terms "identity activities," which a doctrine note defines as a "collection of functions and actions that appropriately recognize and differentiate one person or persona from another person or persona to support decision making." Biometric recognition systems allow warfighters to fix (and track) the identity of an individual "regardless of disguises, aliases, or falsified documents." Improvements in biometric enrollment, detection, recognition, and verification technologies will improve the ability of the DoD to securely identify friendly forces in theater; protect and secure civilian populations; and track and target persons of interest, insurgents, terrorists, and others who would harm U.S. service members and their allies [1].

Figure 1-1 is the U.S. Army's Operational Viewpoint (OV)-1 graphic of an automated biometric identification system. It clearly depicts the multiple methods of biometric collection as well as the multiple biometric modalities (e.g., iris scan, fingerprints, facial recognition, palm prints), which feed into the system that must process, store, match, and share information.

For roughly a decade, researchers in academia, government, and the private sector have successfully applied artificial intelligence (AI) and machine learning (ML)-enabled algorithms and computational systems to biometric recognition. ML models such as the use of advanced artificial neural networks (ANNs), deep learning (DL) systems, and other "self-learning" algorithms has reduced the processing time and computational load required of traditional biometric systems.

Moreover, the application of AI/ML to biometrics since 2012 has revolutionized the field, enabling significant progress in overcoming the limitations of conducting recognition and identification tasks "in the wild" (i.e., in an uncontrolled, uncooperative, and unconstrained sensing environment). For example, in 2018, researchers applied a DL algorithm to a series of faces wearing eyeglasses, which deteriorate the performance of iris recognition systems. Using a deep neural network (DNN), the

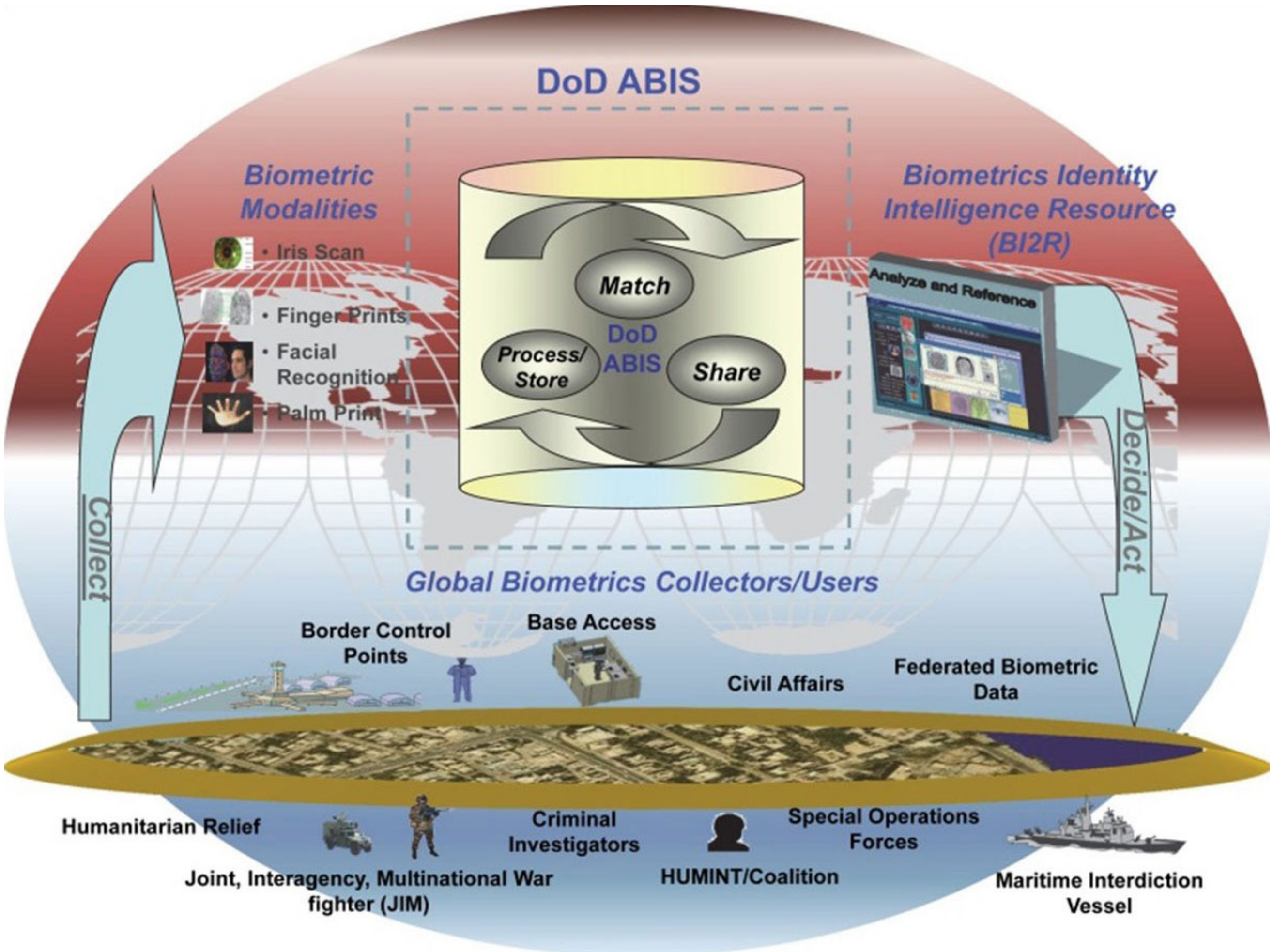

Figure 1-1. Automated Biometric Identification System OV-1 [2].

researchers demonstrated the ability to correct for specular reflection, lens scratches, and other obstructions to improve iris recognition in the near-infrared spectrum [3].

However, even advanced AI/ML systems are inherently limited when their analytical power is applied to a single source. Whether used in unimodal (i.e., assessing a single modality, like face, gait, or iris), unisensor, single-sample, or single-instance recognition, such algorithms cannot overcome the constraints imposed on recognition by limited reference sets or degraded environmental conditions. Such limitations include poor illumination, high levels of occlusion, or the shortened duration of biometric data capture (e.g., brief presentation of movement for gait analysis).

This State-of-the-Art Report assesses the recent science of biometric data fusion, which combines two or more biometric data sources, typically aided by an AI/ML algorithm, to compensate for limitations in sample capture and reference set matching to improve a system's predictive power. A major type of biometric data fusion is multimodal (e.g., periocular and gait recognition) "feature-level" data fusion, which provides faster reference set retrieval across identity templates, and significantly improves recognition accuracy over a unimodal system.

The DoD has already identified biometric data fusion as central to improving its capabilities in joint identity operations. In particular, the U.S. Army has solicited information from researchers and vendors to develop "adaptive algorithms" that "improve single modality and fusion accuracy and performance" [4]. In an installation or facility security use case, advances in AI/ML-enabled data fusion may allow for faster, touchless identification of friendly forces to allow entry, freeing up valuable force protection resources. Data fusion may also better resolve potential error-inducing instances like aging, or unanticipated changes to a biometric data template (e.g., a U.S. Soldier who approaches a base displaying an altered gait due to injury).

By improving multimodal data fusion and matching, the DoD also seeks to expand its ability to conduct identity operations in degraded and unconstrained environments, especially via long-range imaging equipment that may enable detection at stand-off or tactical distances [5]. Friendly forces at times only have a limited footprint or must rely heavily on space-based or aerial reconnaissance assets. However, technological breakthroughs in biometric data fusion that allow for the more accurate use of capture samples low in resolution, fidelity, or duration, may significantly bolster U.S. military identity activity capabilities at home and abroad.

This Page Intentionally Left Blank

SECTION 02 CURRENT USE OF AI AND ML IN BIOMETRIC IDENTIFICATION SYSTEMS

2.1 BIOMETRIC SYSTEM MODALITIES

Biometric systems use the observed biological or behavioral traits (modalities) of individuals to create unique identifiers to compare against biometric templates stored within a database. The joint International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC) standardized vocabulary for biometrics defines a *biometric characteristic* as "a biological and behavioral characteristic of an individual from which distinguishing, repeatable biometric features can be extracted for the purpose of biometric recognition" [6]. The most common biometric modalities in use today include the face, fingerprint, DNA, gait, palm print, and iris. Other modalities include traits like voice signature, shape of the periocular region (around the eyes), vascular or vein patterns, cardiac rhythm, and skin texture [7].

For a modality to be considered a usable biometric trait for identification or recognition in a biometric system, it must display seven fundamental qualities, which allow it to yield high accuracy and performance rates: *universality* (qualities per user); *uniqueness* (a subject's distinguishing features); *permanence* (its inability or unlikelihood to change over time); *measurability* (ease of acquisition of the biometric data sample); *performance* (functional and robust properties of the trait); *acceptability* (acceptability by the system user); and *circumvention* (the ability to spoof or deceive the recognition system). Biometric identification systems built around a single characteristic are known as unimodal systems, whereas those that employ two or more characteristics are termed multimodal [8, 9].

2.2 THE BIOMETRIC IDENTIFICATION PROCESS

Enrollment and recognition are the two stages central to most biometric systems (Figure 2-1). The capture and storage of individual biometric characteristics occur in the enrollment stage, where key features of a biometric sample are extracted

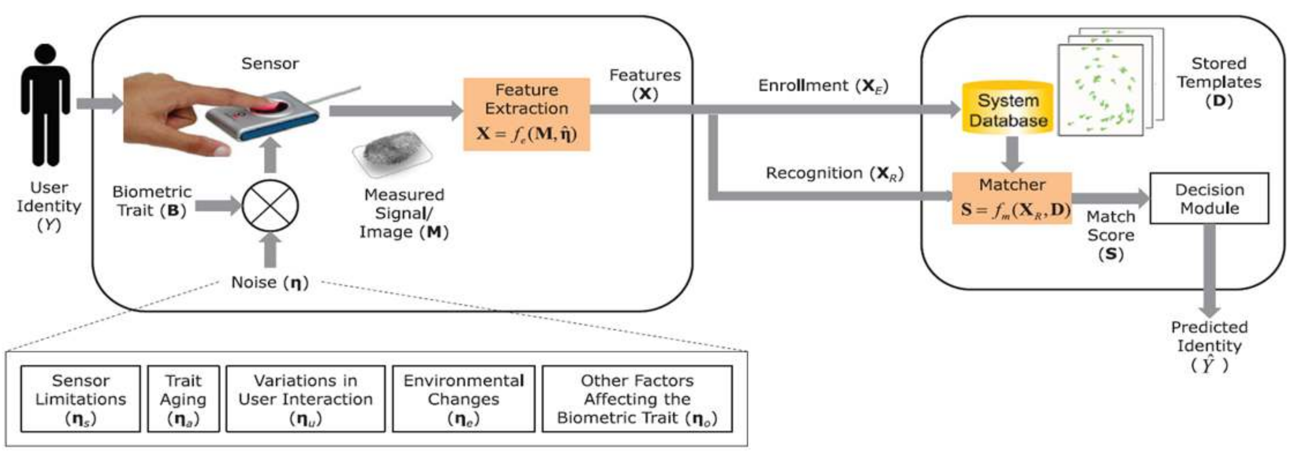

Figure 2-1. Biometric System Operation [7].

and stored in a reference database. Recognition or identification occurs when a new biometric sample is extracted into the template form and compared against the reference set to determine an identity or provide a match. A biometric feature set is a representation of extracted features that are quality checked prior to storage in the template database. These feature sets are processed during enrollment, leading to a usable digital representation of the extracted traits. Recognition is facilitated through the comparison of acquired biometric capture data against previously stored templates to return any matches.

2.2.1 The Enrollment Phase

Figure 2-2 depicts the function of a traditional biometric recognition, sample capture, and identification system. A biometric capture subject's characteristics are scanned by the sensor, after which a second system module conducts preprocessing steps (i.e., treatment or enhancement of the data). Feature extraction is conducted by the next module, which optimizes certain data qualities, and a template (a standardized representation of an extracted modality dataset for an individual) is then produced.

This biometric recognition process is typically referred to as the *enrollment phase*, in which the biometric data of a subject are obtained (whether cooperatively [voluntarily] or noncooperatively [nonvoluntarily]) and stored for later matching attempts. Figure 2-3 shows a cooperative enrollment (far left photograph) versus a noncooperative enrollment (right three photographs). Cooperative enrollments are voluntary and are accompanied by a verified name and identity, which typically results in near-perfect reference templates (such as a driver's license photograph) [6]. Noncooperative enrollments may establish a template for an unknown subject for use in tracking, but not-yet-positive identification tasks.

2.2.2 The Recognition Phase

The recognition phase occurs when a modality is presented to the biometric system and is compared against the database of enrolled images. This phase involves a large amount of computation power, growing larger as the enrollment database is filled and as more biometric modalities can be measured.

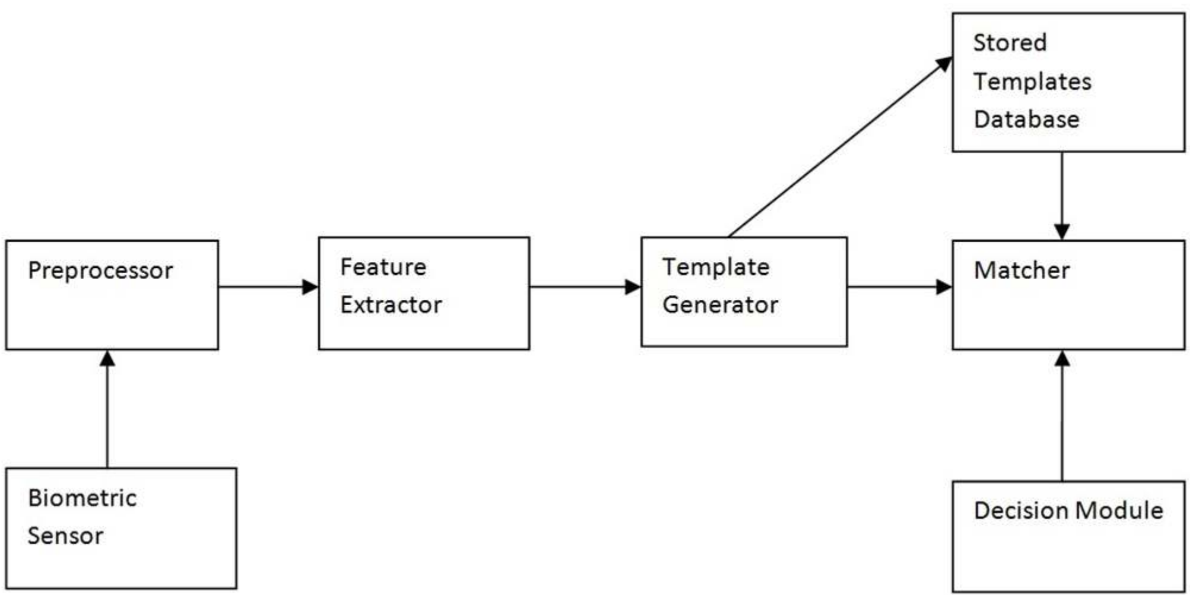

Figure 2-2. Simple Diagram of a Biometric Recognition System, Including Matching to Database of Template References [10].

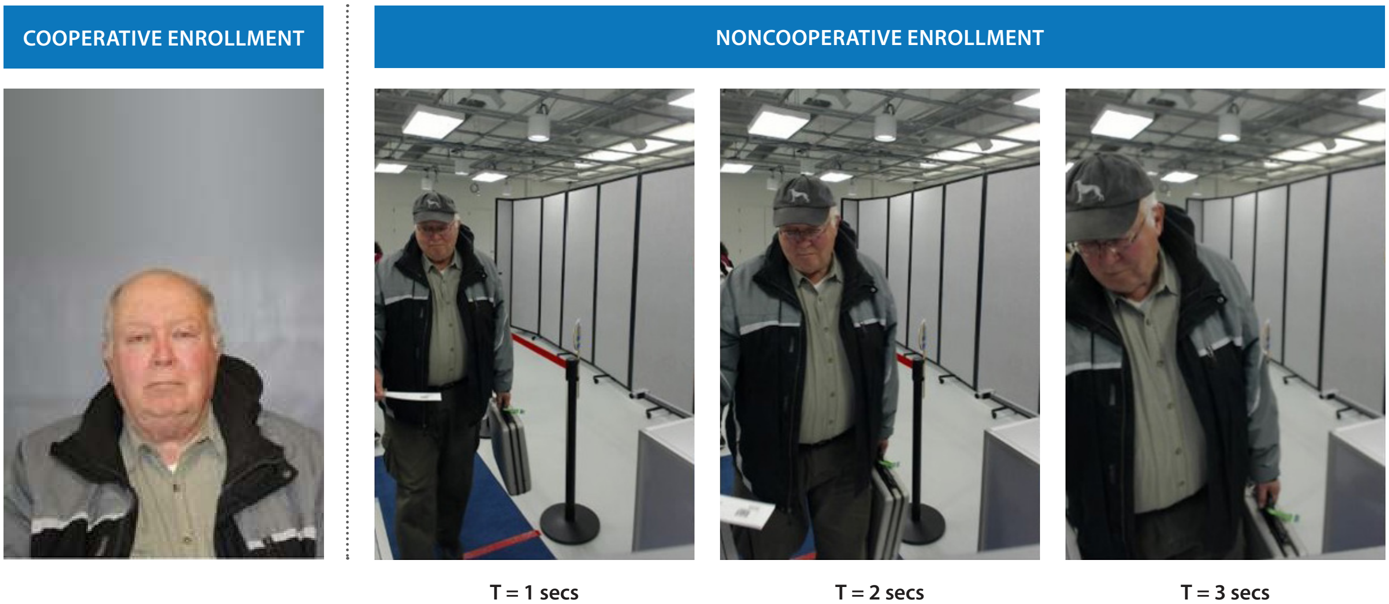

Figure 2-3. Cooperative Versus Noncooperative Enrollment [11].

2.3 ML METHODS

There are several models of ML that may be implemented to increase the efficiency of the recognition phase. This section will discuss several of those models.

2.3.1 DNNs

DL-based biometric models [12] provide an end-to-end learning framework, which jointly learns the feature representation process while performing classification/regression. Learning and classification are achieved via the application of multilayer neural networks called DNNs to learn multiple levels of representations that correspond to different levels of abstraction, which is better suited to uncover underlying patterns within the data.

Prior to the use of AI/ML in face recognition, for example, discrete facial points were applied to a processed facial image to determine and classify the face and identify the person [11]. Early methods explicitly modeled the geometric shape or texture of faces, marking coordinates of facial landmarks such as the eyes, ears, nose, and chin. This process is relatively slow, computationally resource-intensive, and its utility is limited when variables such as pose, illumination, facial expression, image orientation, and imaging conditions vary from the strict standards used in a traditional enrollment reference template.

To address these challenges, researchers began applying AI and ML algorithms and systems to a variety of approaches well known to the biometrics field, including Eigenfaces, Distribution-Based Methods, Support Vector Machines (SVMs), Sparse Network of Winnows (SNoW), Naïve Bayes Classifier, Hidden Markov Models (HMMs), Information Theoretic Approaches, and Inductive Learning. When used in the absence of AI/ML, such approaches require extensive tuning and remain largely statistical, limiting their ultimate usefulness in advanced recognition tasks (e.g., principal component analysis [PCA]) [13]. These methods, while beneficial, lack the analytical strength necessary to produce accurate results in unstable or unpredictable capture environments.

The application of DNNs to the field of biometrics began in 2012, with promising results attained first in face recognition tasks [14]. A major milestone was achieved in 2014, when Taigman and colleagues proposed one of the earliest DL-based tasks for face recognition in a paper titled "DeepFace" [15], achieving what was then a state-of-the-art level of accuracy when applied to the "Labeled Faces in the Wild" (LFW) database benchmark (a public benchmark for face verification). The accuracy of Taigman's algorithm approached that of humans for the first time (DeepFace: 97.35% vs. human: 97.53%). This work set a milestone in face recognition, and researchers began applying DL-based approaches to other components of the face recognition task and soon to other modalities as well. The occurrence of this effort just 7 years ago indicates the recency of DL within biometrics.

Hardware-based advances have also played a role in the successful use of DL. These advances have included improved Graphics Processing Units (GPUs) (computer video cards) and the development of General-Purpose GPUs (GPGPUs). These improvements in turn have enabled the development of new techniques for training neural networks (such as the Dropout regularization method), which is added to each layer of the network [16]. These techniques result in a lower chance of over-fitting, enabling researchers to train very deep neural networks much faster. Relevant types of DL-based biometric recognition algorithmic approaches include convolutional neural networks (CNNs); recurrent neural networks (RNNs); autoencoders; and generative adversarial networks (GANs). A brief assessment of each algorithm type is provided in the following subsections.

2.3.2 CNNs

A CNN is a type of DL algorithm primed to receive input images and assign importance (i.e., learnable weights and biases) to various aspects and/or objects in the image. CNNs also differentiate among features between images. A key benefit of employing a CNN is the smaller preprocessing load it requires compared to other classifiers. The CNN architecture is analogous to that of the connectivity pattern of neurons found in the human brain. Indeed, the design of CNNs is comparable to the organization of the visual cortex, in which individual neurons respond to stimuli only in a restricted region of the visual field, known as the "receptive field." A collection of such fields in the artificial neural map overlaps in a network to cover the entire visual area.

2.3.3 RNNs

RNNs are a class of ANNs where connections between nodes form a directed graph along a temporal sequence [17]. This graphing allows the RNN to exhibit temporal, dynamic behavior. Temporal sequences are not more difficult to analyze simply because they address a sequence of images in time; rather, together with spatially separated measurements, they are "spatio-temporal" and are extremely complex. The directed graphs within RNNs use arrows to imply causality, which is necessary for coding.

2.3.4 Self-Autoencoders (Self-Organizing Maps)

An autoencoder is a type of ANN used to learn efficient codings of unlabeled data (a process known as *unsupervised learning*). The encoding is validated and refined by attempting to regenerate the input from the encoding [12]. The autoencoder learns a representation (or encoding) for a set of data, typically for dimensionality reduction, by training the neural network to ignore insignificant data, or "noise."

2.3.5 GANs

GANs create new data instances that resemble training data. For example, GANs can create images that look like photographs of human faces, even though the faces are fully artificial and do not correspond to any real person. GANs achieve this level of realism by pairing a generator (which learns to produce the target output), with a discriminator (which learns to distinguish true data from the output of the generator). In a sense, the generator tries to "fool" the discriminator, and the discriminator tries to keep from "being fooled." Although it may seem like this capability of GANs makes them unsuitable for face recognition of real individuals, GANs provide an expanded data set for training, and aid in deciding whether two images of a face belong to the same individual. GANs can resolve issues like pose or illumination in face images, making them uniquely suited to application in constrained or degraded sample capture conditions.

While research and development on the use of AI/ML in biometrics remain popular within the domain of face recognition in large part because its reference database is the largest, some lines of inquiry suggest that gait recognition may be more accurate and powerful when conducting the match function [7]. DL-enabled systems are likely to decrease the number of pixels and duration needed on a subject to capture a gait sample. For the challenge of recognizing and identifying faces that are disguised, occluded, at turned-pose, or captured in low-illumination environments, DNNs have already significantly improved the ability of biometric systems to extract a standardized and "comparable" face image to match against a reference set.

This Page Intentionally Left Blank

SECTION 03 AI AND ML FOR MULTISENSOR AND MULTIMODAL DATA FUSION

The prospect of data fusion is attractive to biometrics researchers because early AI/ML-enabled recognition systems, while powerful, were trained on massive, web-based datasets that were generally invariant to effects like pose, illumination, resolution, and facial occlusion [6]. This section explores recent research contributions towards the efficient fusion of multimodal biometrics to improve identification and recognition tasks across multiple modalities and sensor types. In particular, the application of DL methods to biometric data fusion holds promise to overcome the limitations of traditional, unimodal methods [18].

Prior DL approaches have relied on weight combination, or feature concatenation, which constructs representation layers for recognition stages [19]. Due to the difficulty of fusing multiple modalities for recognition based on inherent modality inconsistencies and technical fusion obstacles [7, 20, 21], these early DL methods are generally understood to be inefficient. The use of CNNs can overcome these challenges in that they enable highly improved recognition via multifusion network layers that drive robust and informative model learning [19, 21, 22, 23]. Furthermore, the broad applicability of these CNNs to modality fusion supports more compact and discriminative feature representations. This improvement then increases the predictive power in multibiometric systems [21, 23].

3.1 BIOMETRIC SYSTEMS CHALLENGES

An acquired biometric signal from a capture subject may exhibit variations in quality due to "noise" (Figure 3-1). Stability across acquired trait measurements, that is, across intrasubject variations, may require optimization to reduce noise across five areas according to Jain, Nandakumar, and Ross [7]: (1) sensor limitations, (2) intrinsic aging of the biometric trait, (3) variations in user interaction, (4) changes in the acquisition environment, and (5) all other factors affecting the biometric trait [18]. Figure 3-1 demonstrates the aging in the biometric trait, changes in the acquisition environment

Figure 3-1. Noise Areas in Biometric Capture [13].

(lighting, background), and variations in user interaction (eyes closed, open).

3.2 THE NEED FOR DATA FUSION IN BIOMETRICS

The potential drawbacks of a unimodal biometric system pertain to its data quality, information resilience, identity overlap, and limited discriminability [7, 19]. Therefore, other modalities need to be integrated into a concurrent system to increase accuracy in recognition. This need highlights the critical importance of techniques for the fusion of biometric modalities in a multibiometric system [24].

The development of a multibiometric system focuses on three fundamental questions: (1) what to fuse, (2) when to fuse, and (3) how to fuse [25]. According to Singh, Singh, and Ross,

> What to fuse involves selecting the different sources of information to be combined, such as multiple algorithms or multiple modalities. When to fuse is answered by analyzing the different levels of fusion, that is, the various stages in the biometric recognition pipeline at which information can be fused. How to fuse refers to the fusion method that is used to consolidate the multiple sources of information [25].

3.2.1 What to Fuse: Sources of Information for Fusion

In many instances, the addition of contextual metadata and subject demographics into a recognition system may greatly increase its accuracy. All relevant data can be captured from an area of interest, such as a corridor or battlefield, whether it is from one sensor used repeatedly; one sensor with two different views of the subject; or multiple sensors on one subject. Effective functioning of the algorithm only requires optimal data vectors at the preclassification level and/or more than one classification result for merging. AI/ML DL relies less on the meaning of a particular piece of data and more on its ability to compute using it. Figure 3-2 depicts the five sources of information for fusion: (1) multisensor, (2) multi-algorithm, (3) multi-instance, (4) multisample, or (5) multimodal [25].

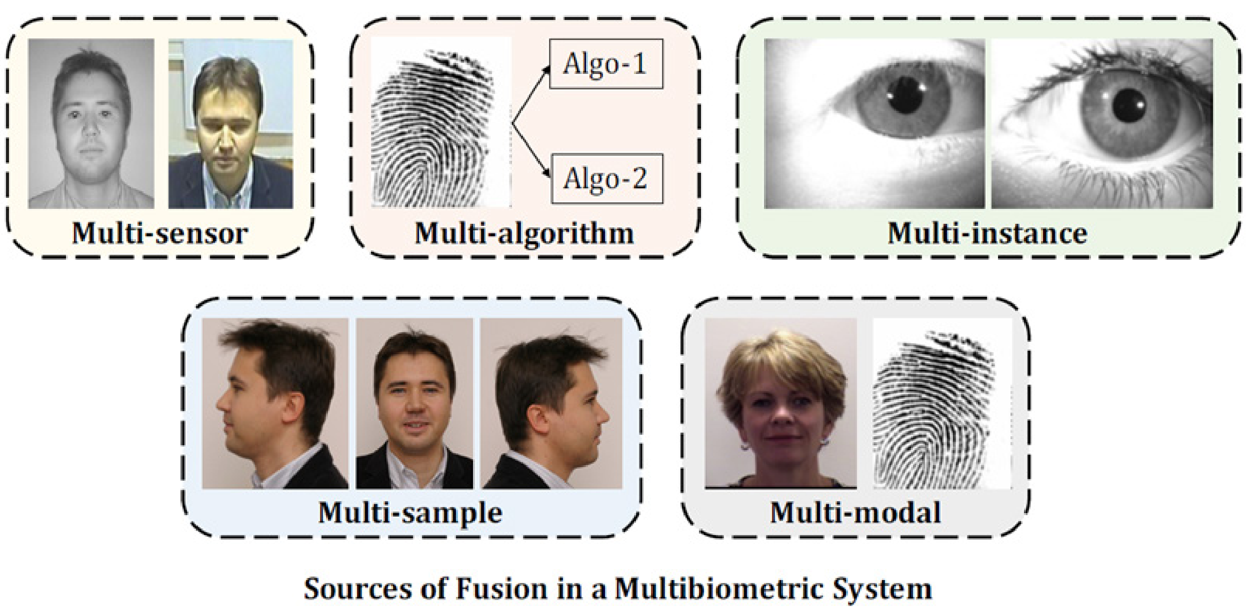

Figure 3-2. Five Sources of Information Fusion in a Multibiometric System [25].

Multisensor. Multisensor systems combine raw data of the same biometric modality captured through multiple sensors [25]. Facial data, fingerprints, the iris, and the gait can be sensed, but if more than one modality is used, the approach qualifies as multimodal data fusion.

Multi-algorithm. Multi-algorithm systems use different algorithms to process the input data of a single biometric modality [25]. This processing occurs during the feature extraction and matching steps [26].

Multi-instance. Multi-instance systems capture multiple instances of the same biometric modality [26]. RNNs are ideal for this type of time series data. Note that each image in the time series may be strongly dependent on the previously captured image. Long-short-term memory (LSTM) is an RNN architecture including memory to model temporal dependencies in time series problems. As Ordonez and Roggen explain,

> [T]he combination of CNNs and LSTMs in a unified framework has already offered state-of-the-art results in the speech recognition domain, where modelling temporal information is required. This kind of architecture is able to capture time dependencies on features extracted by convolutional operations [27].

LSTM RNNs take advantage of the correlation between images at different times, assuming that the sensor is not undersampling significantly.

Multisample. Multisample systems collect a variety of samples for the same biometric modality [25]. The samples are not necessarily independent. For instance, two images of the same face from different angles could display similar results; if the eye is shown in each, the periocular region would be present in each. The fusion step, therefore, might involve a classic 3D transformation before AI/ML techniques commence. This step distinguishes multisample from multimodal data capture in which the streams of data are independent.

Multimodal. Multimodal systems combine multiple biometric characteristics [25]. This is the most relevant sort of data fusion to next-generation sensing and control systems, as it can leverage two or more biometric modalities such as the face, finger, hand, legs, iris, voice, etc., yielding face recognition, fingerprints, hand geometry, gait, iris scan, and voice recognition [28].

A preferred AI/ML method for multimodal fusion is the CNN. A notable example in the literature is a dual stream (or two-mode) approach described by Tiong, Lee, and Teoh, which "accepts [Red Green Blue] RGB ocular image and a novel color-based texture descriptor, known as Orthogonal Combination-Local Binary Coded Pattern (OCLBCP)" [29]. This periocular recognition in the wild exploits the color information in the periocular texture to better represent the periocular features for recognition in the wild. Fusion occurs late in the process, in the last convolutional layer before the fully connected layer; the networks would otherwise share the same parameters.

An interesting application of multimodal data fusion is to avoid spoofing/jamming of signals. If the fusion is based on one signal, perhaps one mode is less susceptible to noise than another [30]. If two signals are used, one signal may be lost to jamming without jeopardizing the classification. This concept can be extended to suppressing less-desirable modes by amplifying the desirable mode's influence in the classification. One underrepresented mode in the literature is voice [31]. The voice mode may be particularly useful for military applications because of the ability to identify an individual using stand-off or cyberspace-based sensors. The approach described by Kusmierczyk et al. [31] draws features from the iris and combines them with voice recognition, both of which could be valuable for military applications.

Hybrid Data Fusion. Hybrid biometric systems combine more than one source of fusion; for example, a single system could combine multisample periocular sample captures with gait recognition [26].

3.2.2 When to Fuse: Data Fusion Levels or Stages

Postmapping fusion using a matching-score approach requires combining scores from biometrics of varying characteristics. This approach largely depends on feature vector proximity, as depicted in Figure 3-3.

Processed scores are sent to adjacent decision modules, where simplicity contributes to good performance. Score-level fusion (also referred to as measurement level or confidence level) is used when the output data includes a set of matches along with the quality of each match [7]. There are two different types of postmapping fusion:

- Score-level: In score-level fusion, the classifier (matcher) compares each enrolled biometric trait against all the identities stored in the database.
- Decision-level: Decision-level fusion (also referred to as the abstract level) fuses together information taken from different sources after each has been classified individually; it then makes the final decision based on methods such as the "AND" and "OR" rules, using weighted voting. This is a later-stage approach for fusing data for each biometric and is the least powerful [33].

Premapping fusion involves using feature vectors originating from multiple biometric sources or feature vectors from the same source using multiple feature extraction. Feature vectors are fused to create a new, single feature that represents an individual [25]. This fusion is accomplished via appropriate feature normalization, transformation, and reduction schemes [34]. Feature-level fusion occurs prior to matching or classification [26].

3.3 SOURCES OF DATA FUSION ERRORS

Data fusion errors may arise based on varying conditions when different modalities are fused to support identification. It is important to note that these errors describe aspects of future systems and how their architectures may be designed to address them. In general, four types of errors may be admitted in a biometric recognition system [32, 35]:

- False Acceptance (FA): System can accept an impostor.

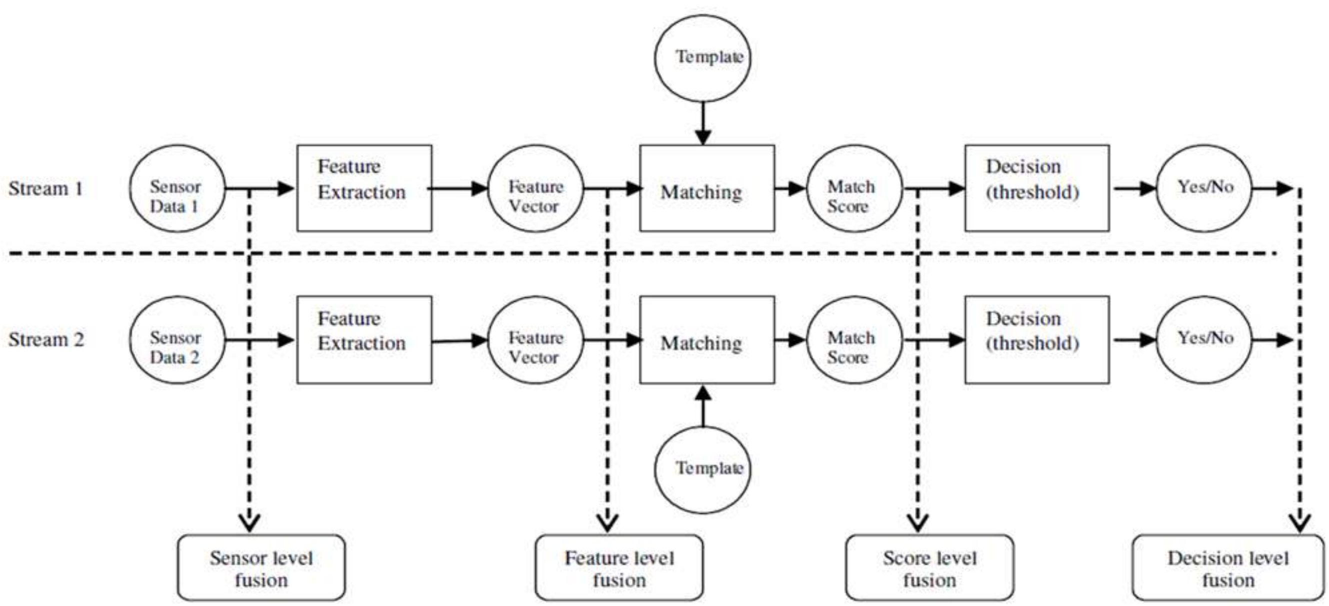

Figure 3-3. Sensor, Feature, Score, and Decision Level Biometric Data Fusion Levels [32].

- False Rejection (FR): System rejects a client.
- Failure to Capture (FTC): Inability to capture biometric modality.
- Failure to Enroll (FTE): Inability to enroll a user.

There are some unique characteristics of error within biometric systems pertaining to data availability or performance issues:

- When a signal of a biometric is not able to be located due to weakness in the signal (like a weak fingerprint), an FTC error occurs.
- When a user cannot be enrolled into the recognition system, an FTE error occurs.
- The Equal Error Rate (ERR) provides a measure of system performance in verifying identities [32].
- Open Set Identification (OSI) involves identification and verification processes.
- In the verification process for OSI, the performance of the system can be measured by OSI ERRs under certain conditions.
- The performance of the identification process is measured by the Identification Error Rate (IER), which occurs when a person is misidentified in a database.

3.4 SOURCES OF DATA VARIATION

Biometric system performance is often metered by variations in data. There are two types of data variation: 1) data arising from uncontrolled operating conditions and 2) data arising from data degradation [32]. Some examples of uncontrolled operating conditions include poor resolution of the face due to improper illumination, or substantial background noise that contributes to the inability to perform proper voice recognition [35]. Sources of data degradation are usually attributed to poor performance of the biometric capture device or malfunctions in the sensor(s). In this way, data variation due to an uncontrolled environment can be attributed to sources of degradation within the environment or close to the user (e.g., white-noise generation) [32].

This Page Intentionally Left Blank

SECTION 04

RECENT RESEARCH IN DL FOR BIOMETRICS

4.1 CNNS USING TEXTURE DESCRIPTORS AND RGB DATA

Image processing and computer vision have benefitted from tremendous strides made in DL in recent years [23]. The architecture of these neural networks uses an image's RGB raw data pixels for the input layers to facilitate (model) learning inside these network layers (typically a very difficult task [19]). Illuminations and occlusions are the primary products of filtering by DL models and are usually viewed as the key confounding factors.

Hence, persona recognition remains a difficult task driven by challenges from occlusions, poses, illuminations, and loss of distinctiveness. The use of a CNN minimizes these drawbacks by using more than a single modality within the DL architecture. Two stream inputs are used: the texture descriptor and RGB data. Texture descriptors support discriminatory feature extraction to distinguish against any variation in illumination. Assembly of the model using these RGB and texture descriptor data into a two-stream CNN supports compensation of any hidden information, thereby making it more robust than other approaches. The inputs explicitly describe the "hidden layers" of data, enabling easier recognition (which differs from alternative approaches centered on biometric feature estimation). Removal of single-source dependencies through this approach to CNNs improves upon single-source feature modalities, thereby making them better candidates for facial recognition.

Recent research by Tiong, Kim, and Ro (2020) addressed the efficacy of using different texture descriptors versus raw RGB data for facial recognition [22]. Their investigations involved fusion-layer design approaches to support a robust CNN using dual descriptors to determine if analysis of textures within a CNN leads to substantial improvement in recognition. These researchers used a DL approach for texture descriptor analysis, formalizing another network layer for processing texture descriptors. This layer served as the second multifeature DL network (MDLN) stream for additional explanatory factor extraction. They determined uniquely that there is an invariance in using texture descriptors for complex input transformations.

Their work also emphasized a fusion strategy based on features within the two-stream input, where the convolutional layers are considered a filter bank. This strategy culminates in feature representations that are new and helps solve for any differences in information that could be hidden.

Finally, the authors describe a fusion layer based on scoring using a similarity measurement vector using a joint feature approach. Such a layer enhances the model to support improved recognition by formulating measurement functions (periocular and facial) after sampling the measurement of highest weight. The model showed greater accuracy in recognition than unimodal DL methods.

In another work, Tiong, Kim, and Ro (2019) used the MDLN approach to report substantial increase in accuracy against the ethnic facial dataset [21]. The authors developed fusion layers for multiple features that support independent, dual-stream CNNs using texture descriptors and RGB data. Their work was centered on a strategy for rank weighting to support multimodal biometric features from two networks that incorporate rank-K scores for improved decisions. Using the ethnic facial data set, the authors were able to train and test their model against variations like location, uncontrolled distances of camera, appearances, and different ethnicities.

Other research employed this dual RGB/texture descriptor CNN approach to support periocular biometric recognition from images in the wild [36]. The validity of this approach is a substantial contribution to facial recognition and proves the strength of using both texture descriptors and raw RGB data in a dual-layer network as an alternative to older, unimodal biometric methods for facial recognition.

4.2 IMAGE PARTITIONING

Additional data management approaches for facial recognition involve partitioning images for processing [36]. Partitioning an image improves facial recognition system functionality and performance because it enables larger images to be efficiently stored and indexed for search and reference. Dividing images into smaller and more manageable units (e.g., fixed-size sub-blocks) greatly simplifies compression, storage, access, and retrieval [36]. There is an inherent tradeoff between larger or smaller image block sizes: smaller sizes contain more details but require more computation, while larger block sizes reduce time for computation but generally lack finer image detail.

Segmentation involves ignoring occluded or hidden parts of an image. It is accomplished by the fuzzy c-means (FCM) algorithmic rule to support performance improvements. Using this occlusion segmentation approach, Jyothi and Ramanjaneyulu (2021) demonstrated accurate identification of faces with sunglasses, scarfs, or masks [36].

4.3 BIMODAL DATA FUSION

While CNNs have shown tremendous image-recognition capabilities, their utility does not immediately extend to some other modalities. (Recent work by Luo, Li, and Zhu in 2021 may be an exception [37].) In the case of poor lighting and irregular facial angles, recognition and identification may be near impossible using facial recognition alone. In this case, other modalities could be employed as described in the recent work by Cuzzocrea and Mumolo (2021), in which fingerprints and voice data were fused using a weighted sum and fuzzy system [38]. Wang et al. (2004) also combined voice and fingerprint data [39]. Cuzzocrea and Mumolo employed a universal background model to address background noise contributions along with a Gaussian Mixture Model to represent acoustic features associated with voice recognition. The novelty in their approach involves the application of the Dempster-Shafer algorithm that facilitates verification of decisions sourced from many data fusion algorithms (Figure 4-1) [38].

Fingerprint and palm print modalities were fused by Soviany, Pușcoci, and Săndulescu (2021) using a kernel SVM model and a multiclass extension approach that supports identification rather than verification [40]. Their very recent approach fusing these two modalities using SVM (without concatenation) offers additional alternatives when voice or facial imagery is difficult to acquire.

4.4 DEEP-FEATURE FUSION AND DL

4.4.1 Use of Periocular Regions

Fixed-fusion schemes like element-wise feature sum, element-wise feature product, and simple feature concatenation lack adaptation and flexibility. The inability to use multimodal features for optimizing fused features is the primary drawback of fixed-fusion schemes. Recent work by Luo, Li, and Zhu (2021) involves using CNNs with deep-feature fusion to bridge these gaps for joint iris and periocular biometric recognition [37]. The benefits

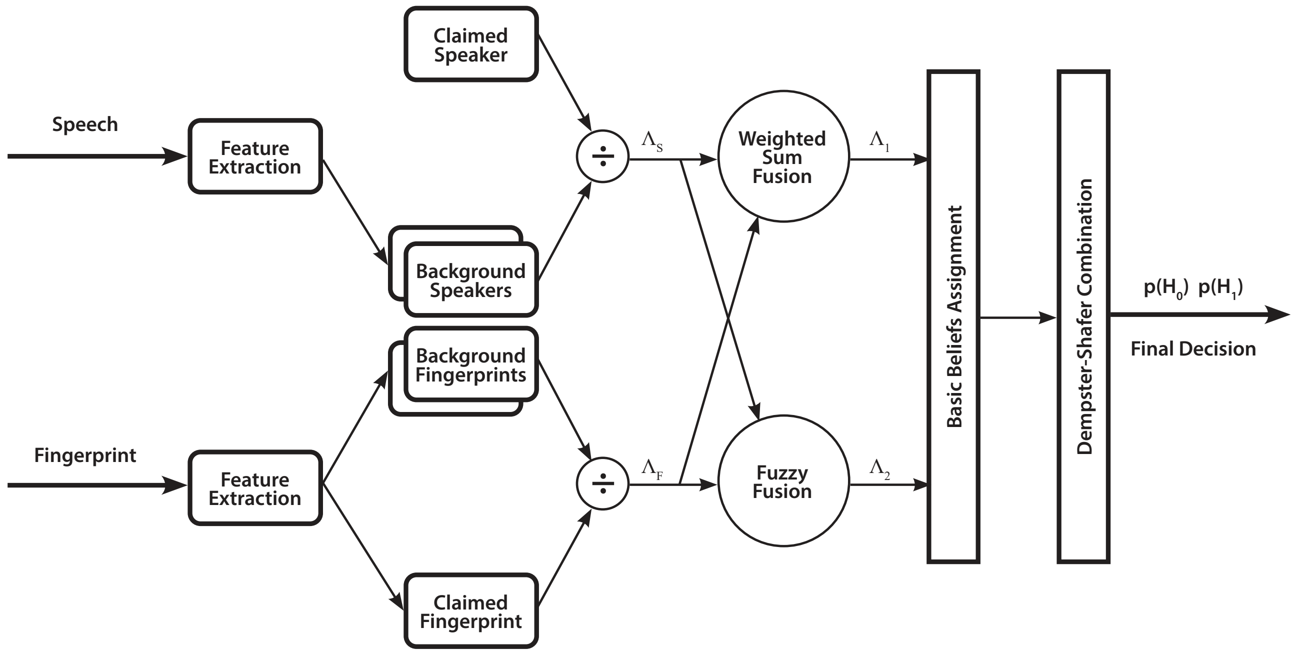

Figure 4-1. Bimodal Data Fusion Combination Algorithm [38].

of using periocular regions [41] over iris, retina, conjunctiva, and sclera traits provide a larger field depth and require less subject cooperation in their acquisition. Additionally, their utility at a wide range of distances has advantages over using the full or partial face for recognition [41]. COVID-19 mask mandates have underscored this need for exploring multimodal biometric identification and recognition by emphasizing periocular features with other ocular or physical traits when faces are covered (Figure 4-2) [42, 43].

4.4.2 Feature Descriptors

Periocular image matching was initially performed through development of hand-crafted features which were divided into global and local feature descriptors. According to Kumari and Seeja (2021),

> Global feature descriptors consider [the] image as a whole and create a single feature vector for the whole image, whereas local feature descriptor[s] [divide] the image into patches (group of pixels), create feature vector[s] for every patch and finally combine them to create a single feature vector [34].

Additional classification of global feature descriptors requires viewing them in three groups based on texture, color, and shape. Zhao and Kumar (2018) emphasized the attention model concept based on a DL architecture using certain periocular traits, like the shape of the eye and the eyebrow. The model's recognition performance improved, which indicates that highly feature-rich traits like the eyebrow and eye shape can improve a model's discriminating power [44].

Feature-based descriptors based on texture have received a lot of attention in the field, since they are easy to implement due to their low-to-manageable computational cost. However, their primary drawbacks are influences from image rotation, blurring effects, and noise (e.g., blinking and off-angle poses). Works by Tiong, Kim, and Ro (2019, 2020) demonstrate the value of employing texture along with effective use of both handcrafted and non-handcrafted feature creation for improved recognition [21, 22].

Examples of multimodal approaches including periocular modalities can be found in previous work fusing iris-periocular elements that involve

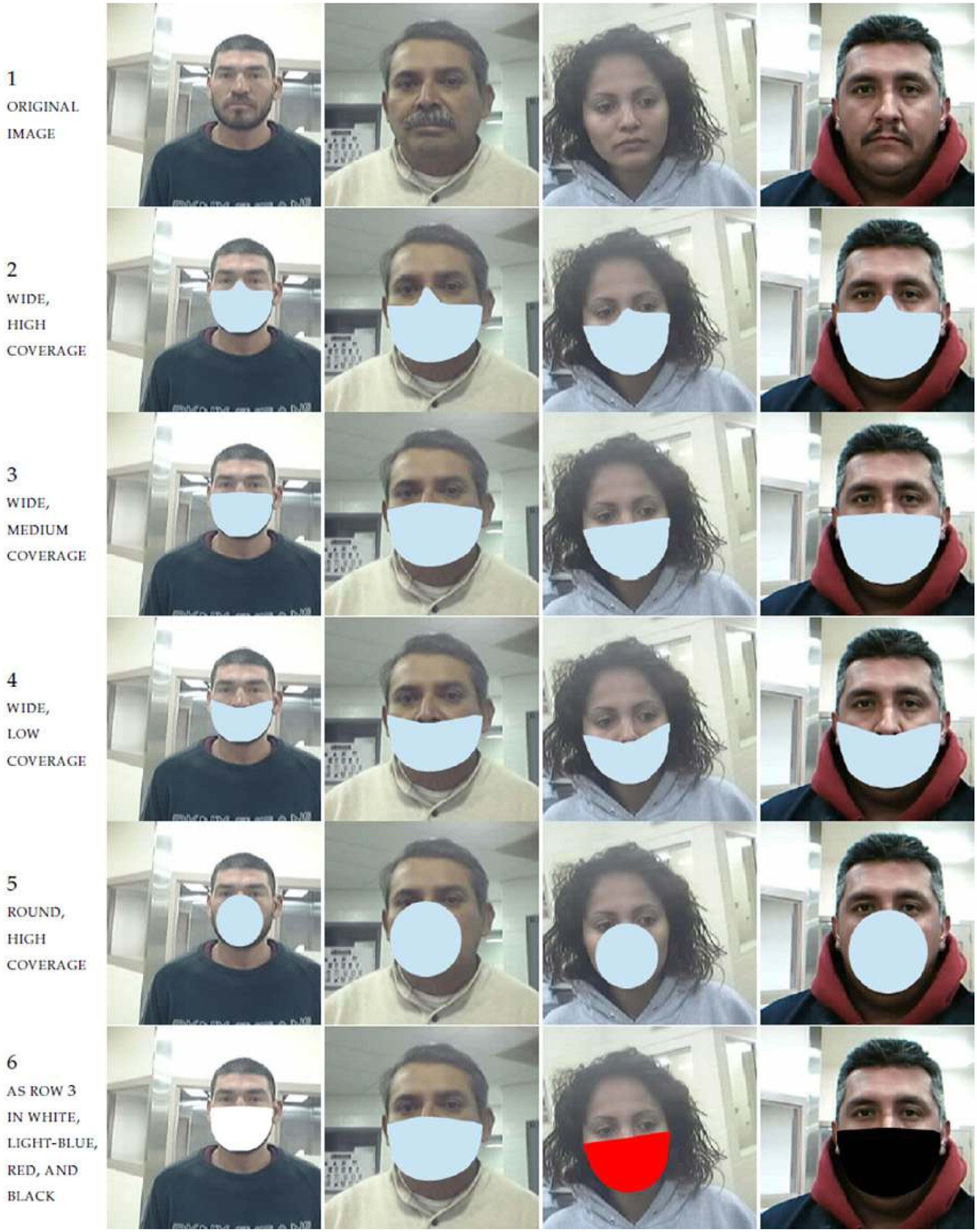

Figure 4-2. Synthetically Generated Face Masks Used in NIST Face Recognition Vendor Test Studies [43].

feature descriptors like local binary pattern, histogram of oriented gradients, and scale-invariant feature transform via techniques like wavelets or ordinal measures [45].

Spatial and channel attention (types of self-attention mechanisms) were integrated into feature extraction to enable more learning of representative features. Discriminative feature discovery substantially contributed to the highly accurate models developed by Tan and Kumar [45] where the utility of CNNs for nonfacial recognition use cases was successfully applied for improved recognition power.

4.5 GAIT RECOGNITION

Human gait recognition (HGR) is a subdiscipline within computer vision. Sharif et al. (2020) characterized HGR biometric detection by using four key steps:

> (a) enhancement of motion region in frame by the implementation of linear transformation with HSI (hue, saturation, and intensity) color space; (b) Region of Interest (ROI) detection based on parallel implementation of optical flow and background subtraction; (c) shape and geometric features extraction and parallel fusion; (d) multi-class support vector machine (MSVM) utilization for recognition features [46].

There are two types of HGR ML methods: model-based and model-free. Model-based types focus on structural human features without any motion, while model-free types are focused on examining the images of the body's silhouette. The research presented in Sharif et al. (2020) is novel in its ability to perform parallel features fusion and Euclidean distance-based best selection [46]. This approach produces highly accurate results. Similar work in HGR using biometric score fusion is an alternative approach [47], but Sharif et al.'s (2020) work shows a unique difference in persona recognition due primarily to the four-step data processing approach they detail (Figure 4-3).

4.6 DATA MANAGEMENT OPTIMIZATION APPROACHES

Biometric identification and recognition systems have recently benefited from overlapping research contributions in the areas of big data, indexing and retrieval, and graph feature fusion. Zhu and Jiang (2020) used Two-Dimensional Principal Component Analysis (2DPCA) for global feature extraction (a big data problem) with Local Binary Pattern-supported local feature extraction [48]. While the processing of local features was challenging, this work focused on fusing these two feature sources to yield optimal recognition results using big data methods.

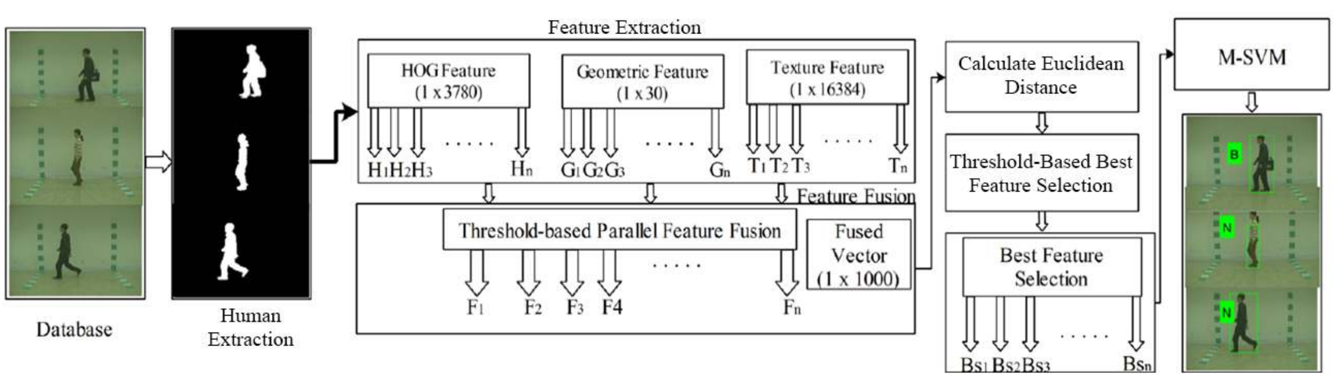

Figure 4-3. HGR Fusion and Feature Selection [46].

Drozdowski et al. (2021) describe an approach using an indexing method that focuses on facial parent template pairing in an intelligent configuration, using similarity characteristics like nonmated comparison scores [49]. Their work focused on identification retrieval where accessed candidates are reduced in successive query steps. Both studies leveraged standard (big) data management methods as applied to biometric identification to yield fusion improvements.

SECTION 05 FRONTIERS OF BIOMETRIC RECOGNITION

New technological breakthroughs will undoubtedly increase the speed of biometric identification while simultaneously increasing the number of personas that can be enrolled and recognized. Some of the novel technologies that promise to revolutionize biometric processing are discussed in this section.

5.1 DATA STORAGE AND COMPUTATIONAL SPEED

DL models that employ multiple layers require large volumes of data to significantly outperform traditional analytical methods based on convex optimizations (e.g., linear and kernel methods). SVM performance for both unimodal and multimodal biometric systems scale, but only to a certain point. The use of CNNs is a natural advancement in model evolution driven by fusion of data and their increased availability. As a result, storage and computational resources are often integrated into performance and scaling models as multicore architectures can offer high performance at smaller form factors.

Due to the physical limitations of transistors, the central processing unit (CPU) becomes a "core," which is multiplied on the motherboard in a "multicore" architecture. Computer processing speed is based on hardware elements such as the field-programmable gate array, CPU, GPU, and multicore architecture. GPUs are optimized for the 4×4 matrix-vector products used in a CNN [50]. Storage can be volatile (random-access memory) or nonvolatile (e.g., a "hard disk"). Solid-state drive (SSD) technology is nonvolatile, meaning that the data written persist when the SSD is turned off and back on, and yet functions nearly as fast as random-access memory modules.

5.2 LATENCY REDUCTION

Latency is the delay in a process based on processing time or information travel. For data access to be efficient, this delay must be minimized. As noted in the beginning of this section, processing and memory hardware is becoming more effective. Cloud and distributed storage is an area of active research. Improved indexing to support faster search drives Internet search, voice-activated search (e.g., Alexa, Siri), and social media. Fast access to databases with optimally configured "shards" (predesigned and well-organized data partitions) provides speed of access to, and redundancy of, data. For DoD operations, long distances between network nodes can add latency or delays in communication beyond the optimal.

5.3 NEURAL COMPUTATION AND THE ATTENTION MECHANISM

Research in biological neural computation is focused on attention (or memory) for noise reduction and is based on relative focus/unfocus in human attention patterns. The Attention Mechanism has arguably become one of the most important concepts in the DL field. It is inspired by human biological systems, which tend to focus on sharp contrasts when processing large amounts of information. With the development of DNNs, attention mechanisms have been used in diverse application domains [51]. Four variables of attention mechanisms are the softness (soft focus, or "smoothness") of attention, forms of input features, input representation, and output representation. Neuromorphic computing software imitates spike signals in the brain and is used to create energy-efficient hardware for information processing, which is capable of highly sophisticated tasks [52]. Combining attention mechanisms with neural computation would increase the theoretical power of a DNN. Systems built with standard electronics achieve gains in speed and energy by mimicking the distributed topology of the brain. Scaling-up such systems and improving their energy usage, speed, and performance by several orders of magnitude require a revolution in hardware.

5.4 UPSTREAM FUSION

Upstream fusion of data from multiple sensor types and modalities improves target detection and classification by forming data vectors prior to degradation due to statistical algorithms applied to the data. Level-based classifications generally categorize fusions as low-level, mid-level, and high-level sensor fusion, corresponding to the common terms data-level, feature-level, and decision-level fusion, respectively [53]. Heterogeneous data sources for fusion involve signals such as infrared full-motion, video, passive radio-frequency (P-RF), pressure, radar, acoustic, chemical, electromagnetic, thermal, proximity, optical, and electro-optical (EO) vision. These signals are mostly independent of each other and can be combined into data vectors usable by ML [53].

5.5 COMPETITIVE, COMPLEMENTARY, AND COOPERATIVE FUSION

Competitive fusion is independent classification from separate sources such as video or radar. Complementary fusion exploits the overlap between modalities such as EO/RF through methods such as decision-level fusion or voting. Cooperative fusion of EO, RF, and other modalities provides a complete picture of the environment that the individual features and input data alone cannot, for instance, with triangulation [53]. An example of cooperative fusion is a digital imaging and remote-sensing image generation (DIRSIG) simulation performed in 2017 by Michigan Tech Research Institute. The inputs consisted of medium-wave infrared full-motion video (IR FMV) input and three emulated corresponding P-RF sensors. The targets included automotive vehicles; and obstacles were used to obscure the targets at times. The DIRSIG dataset contained 13 simulations that cover a variety of visual obscuration scenarios, while receiving RF signals at three different locations [53].

The researchers found that traditional ML methods faltered due to large amounts of data and that DL would work better for data fusion. This research clearly indicates that for next-generation, multimodal, biometric data-fusion efforts, multistream, multisensor, and multimodal sources may be best fused upstream. In addition, the application of DL methods is necessary to produce a complete biometric recognition system that integrates all available identity information.

5.6 QUANTUM COMPUTING

Quantum computing promises to revolutionize next-generation computing applications due to its radically different approach as compared to current computing assets. Quantum computers do not rely on bits with the value of 1 or 0; rather, they are designed to use quantum computing bits, or qubits, which can exist at intermediate values. In addition, quantum computers may be superior to classical computers for solving certain problems, such as the factoring of very large numbers and tackling some of the long-standing challenges in science, such as modeling the human brain or the millions of objects in space. Moret-Bonillo (2015) shows how "The Quantum Circuit Model" can point the way to increased biometric identification performance and lower energy consumption [54]. However, at this time, quantum computing is still in the theoretical stage.

SECTION 06 CONCLUSION

Data fusion requires attention to causal inference: how the data supports the result, which might be a biometric identification or another biometric classification. The problem of inference depends heavily on which stage of the process data are fused or considered together. It is clear that different parts of the body and even different parts of the face are independent in biometric applications. Hence, it follows that biometric recognition will benefit from AI/ML algorithms conducting postclassification independently, with thresholds set at the match, rank, or decision level. However, some algorithms may run faster if data are fused into feature vectors at an early stage (preclassification) and run with the same AI/ML algorithm, such as a CNN [21, 22].

Biometric systems have been significantly improved by 1) the increased predictive power of their algorithms through using fusion methods in conjunction with DL techniques; 2) the ability to employ novel approaches to data management; and 3) through employing computational power and storage to support improvements in search, sort, optimization, and processing. Note that while the latter two approaches contribute to improvements in biometric systems, these improvements are not novel to other industries like cloud computing or enterprise relational databases. Further improvements may come from increasing a system's predictive power through novel applications of DL along with applying optimized hardware to support applying increased computational power and storage. The unique blend of DL's predictive power combined with computational and storage improvements may drive substantial opportunity to deploy extremely robust algorithms that can scale and perform.

This Page Intentionally Left Blank

REFERENCES

1. United States Joint Chiefs of Staff. Joint Doctrine Note 1–20: Joint Identity Activities. Washington, D.C. https://www.jcs.mil/Portals/36/Documents/Doctrine/jdn_jg/jdn1_20.pdf?ver=cvudeXvFG01_3F9XoWF9K-w%3D%3D. 24 November 2020.
2. Director, Operational Test and Evaluation (DOT&E). "Department of Defense (DOD) Automated Biometric Identification System (ABIS) Version 1.2: Initial Operational Test and Evaluation Report." https://apps.dtic.mil/sti/pdfs/ADA626558.pdf. 2015.
3. Drozdowski, P., F. Struck, C. Rathgeb, and C. Busch. "Detection of Glasses in Near-Infrared Ocular Images." 2018 International Conference on Biometrics (ICB), pp. 202–208, doi: 10.1109/ICB2018.2018.00039, 2018.
4. Department of the Army. Request for Information: Biometrics Enabling Capability (BEC) – Increment 1 (Amendment 0003). Notice ID: W909MY-21-R-C001. https://sam.gov/opp/495d5ee2d157443a92ff390a4f-633b5a/view. 20 November 2020.
5. Burt, C. "US Military Integrates Biometrics-Enabled Watchlist with DoD ABIS." Biometric Update. https://www.biometricupdate.com/202105/us-military-integrates-biometrics-enabled-watchlist-with-dod-abis. 14 May 2021.
6. ISO/IEC. International Organization for Standardization (ISO) Information Technology – Vocabulary – Part 37: Biometrics ISO/IEC 2382-37. https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:-37:ed-2:v1:en:term:3.3.21. 2017.
7. Jain, A., K. Nandakumar, and A. Ross. "50 Years of Biometric Research: Accomplishments, Challenges, and Opportunities." *Pattern Recogn. Lett.* 79, 80–105. 2016.
8. Ross, A. and A. Jain. "Multimodal Biometrics: An Overview." 2004 12th European Signal Processing Conference, pp. 1221–1224. 2004.
9. Jain, A., P. Flynn, and A. Ross. *Handbook of Biometrics* (1st ed.). Springer Publishing Company, Incorporated. 2010.
10. Thakkar, D. "What is a Biometric Template? Is it Secure?" Bayometric. https://www.bayometric.com/biometric-template-security/ n.d.
11. Grother, P., M. Ngan, K. Hanaoka. MISTIR 8238 - Ongoing Face Recognition Vendor Test (FRVT) Part 2: Identification. U.S. Department of Commerce, National Institutes of Standards and Technology. 2020.
12. Finizola, J., J. Targino, F. Teodoro, and C. Lima. "Comparative Study Between Deep Face, Autoencoder and Traditional Machine Learning Techniques Aiming at Biometric Facial Recognition." *2019 International Joint Conference on Neural Networks (IJCNN)*, 1–8. 2019.
13. Yang, M. H., D. J. Kriegman, and N. Ahuja. "Detecting Faces in Images: A Survey." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1), 34-58. 2002.
14. Minaee, S., P. Luo, Z. Lin, and K. Bowyer. "Going Deeper Into Face Detection: A Survey." https://arxiv.org/abs/2103.14983. 2021.
15. Taigman, Y., M. Yang, M. Ranzato, and L. Wolf. "DeepFace: Closing the Gap to Human-Level Performance in Face Verification." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701–1708. 2014.
16. Pranav, K. B., and J. Manikandan. "Design and Evaluation of a Real-Time Face Recognition System Using Convolutional Neural Networks." Third International Conference on Computing and Network Communications, *Procedia Computer Science* 171. 1651–1659. 2020.
17. Pai, A. "CNN vs. RNN vs. ANN – Analyzing 3 Types of Neural Networks in Deep Learning." PSG College of Technology. 17 February 2020.
18. Hong, L., A. Jain, and S. Pankanti. "Can Multibiometrics Improve?" *Proc. AutoID*, 99, 59–64, 1999.
19. Mahesh, K., and K. Sharma. "Biometric System: Unimodal Versus Multibiometric Fusion and Its Current Applications: Review." In: Smys, S., R. Palanisamy, Á. Rocha, and G. N. Beligiannis (eds.), *Computer Networks and Inventive Communication Technologies. Lecture Notes on Data Engineering and Communications Technologies*, 58, Springer, Singapore. 2021.
20. Bharadwaj, S., H. Bhatt, M. Vatsa, and R. Singh. "Periocular Biometrics: When Iris Recognition Fails." *Proc. 4th IEEE Int. Conf. Biometrics:Theory, Appl. Syst.* 1–6, 2010.
21. Tiong, L., S. Kim, and Y. Ro. "Implementation of Multimodal Biometric Recognition via Multi-Feature Deep Learning Networks and Feature Fusion." *Multimed Tools Appl.*, 78, 22743–22772. 2019.
22. Tiong, L., S. Kim, and Y. Ro. "Multimodal Facial Biometrics Recognition: Dual-Stream Convolutional Neural Networks with Multi-Feature Fusion Layers." *Image and Vision Computing*, 102(103977). 2020.

REFERENCES, *continued*

23. Li, H., Z. Lin, X. Shen, J. Brandt, and G. Hua. "A Convolutional Neural Network Approach for Face Detection." *Int. Conf. Comput. Vis. Pattern Recognit*. (CVPR). IEEE, Boston, MA, USA, 5325–5334. 2015.

24. Ross, A., K. Nandakumar, and A. Jain. *Handbook of Multibiometrics*, Springer Publishers. 2006.

25. Singh, M., R. Singh, and A. Ross. "A Comprehensive Overview of Biometric Fusion." *Information Fusion*, 187–205. 2019.

26. Ross, A., A. Jain, and J.-Z. Qian. "Information Fusion in Biometrics." *Proc. of 3rd International Conference on Audio- and Video-Based Person*, pp. 354–359. Sweden. 2001.

27. Ordonez, F., and D. Roggen. "Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition." *MDPI Sensors*. https://www.mdpi.com/1424-8220/16/1/115. 2016.

28. Shepherd, T. "An Exploration into Biometrics, Security." GCFW Practical Version 4.1. SANS Institute. 9 April 2005.

29. Tiong, L., Y. Lee, and A. Teoh. "Periocular Recognition in the Wild: Implementation of RGB-OCLBCP Dual-Stream CNN." *Applied Sciences* 9(13): 2709. 2019.

30. Gupta, K., G. S. Walia, and K. Sharma. "Quality Based Adaptive Score Fusion Approach for Multimodal Biometric System." *Appl Intell.*, 50, 1086–1099. 2020.

31. Kuśmierczyk, A., M. Sławińska, K. Żaba, and K. Saeed. "Biometric Fusion System Using Face and Voice Recognition." In: Chaki, R., A. Cortesi, K. Saeed, and N. Chaki (eds). *Advanced Computing and Systems for Security. Advances in Intelligent Systems and Computing*, 996, Springer, Singapore, 2020.

32. Alsaade, F. "Score Level Fusion for Multimodal Biometrics." Ph.D. Thesis, University of Hertfordshire, Hatfield, UK. 2008.

33. Ross, A., and A. Jain. "Multimodal Biometrics: An Overview. *Proceedings of the12th European Signal Processing Conference (EUSIPCO)*, Vienna, Austria, 1221–1224. 2004.

34. Ross, A. "Fusion, Feature-Level." https://link.springer.com/referenceworkentry/10.10072F978-0-387-73003-5_157. 2009.

35. Jain, A., A. Ross, and S. Prabhakar. "An Introduction to Biometric Recognition." *IEEE Transactions on Circuits and Systems for Video Technology*, 14, 4–19. 2004.

36. Jyothi, C., and K. Ramanjaneyulu. "IOT-Based Occlusion Invariant Face Recognition System." In: Bhateja, V., S. C. Satapathy, C. M. Travieso-Gonzalez, and W. Flores-Fuentes (eds). *Computer Communication, Networking and IoT. Lecture Notes in Networks and Systems*, 423–430. Springer, Singapore, 2021.

37. Luo, Z., J. Li, and Y. Zhu. "A Deep Feature Fusion Network Based on Multiple Attention Mechanisms for Joint Iris-Periocular Biometric Recognition." *IEEE Signal Processing Letters*, 28, pp. 1060–1064. 2021.

38. Cuzzocrea, A., and E. Mumolo. "Dempster-Shafer-Based Fusion of Multi-Modal Biometrics for Supporting Identity Verification Effectively and Efficiently." *2021 IEEE 2nd International Conference on Human-Machine Systems (ICHMS)*, 1–8. 2021.

39. Wang, Y., Y. Wang, and T. Tan. "Combining Fingerprint and Voiceprint Biometrics for Identity Verification: An Experimental Comparison." In: Zhang, D., and A. K. Jain, eds., *Biometric Authentication, First International Conference, ICBA 2004, Proceedings, ser. Lecture Notes in Computer Science*, 3072, Springer,. 663–670. 2004.

40. Soviany, S., S. Pușcoci, and V. Săndulescu. "A Biometric Identification System with Kernel SVM and Feature-level Fusion." *2020 12th International Conference on Electronics, Computers and Artificial Intelligence (ECAI),* 1–6. 2020.

41. Park, U., A. Ross, and A. Jain. "Periocular Biometrics in the Visible Spectrum: A Feasibility Study." *3rd IEEE International Conference on Biometrics: Theory Applications and Systems*, Washington, DC, USA, 153–158. 2009.

42. Kumari, P., and K. R. Seeja. "A Novel Periocular Biometrics Solution for Authentication During Covid-19 Pandemic Situation." *J. Ambient Intell. Human Comput.*, 12, 10321–10337. 2021.

43. Ngan, M., P. Grother, and K. Hanaoka. NISTIR 8331 – Ongoing Face Recognition Vendor Test (FRVT) Part 6B: Face Recognition Accuracy with Face Masks Using Post-COVID-19 Algorithms, U.S. Department of Commerce, National Institutes of Standards and Technology. 2020.

44. Zhao, Z., and A. Kumar. "Improving Periocular Recognition by Explicit Attention to Critical Regions in Deep Neural Network." *IEEE Trans. Inf. Forensics Secur. 13*(12): 2937–2952. 2018.

REFERENCES, *continued*

45. Tan, C., and A. Kumar. "Towards Online Iris and Periocular Recognition Under Relaxed Imaging Constraints." *IEEE Trans. Image Process*., 22(10), 3751–3765. 2013.

46. Sharif, M., M. Attique, M. Z. Tahir, M. Yasmim, T. Saba, and U. J. Tanik. "A Machine Learning Method with Threshold Based Parallel Feature Fusion and Feature Selection for Automated Gait Recognition." *Journal of Organizational and End User Computing (JOEUC)*, 32(2), 67–92. 2020.

47. Wasnik, P., K. Schafer, R. Ramachandra, C. Busch, and K. Raja. "Fusing Biometric Scores Using Subjective Logic for Gait Recognition on Smartphone." *2017 International Conference of the Biometrics Special Interest Group (BIOSIG)*, 1–5. 2017.

48. Zhu, Y., and Y. Jiang. "Optimization of Face Recognition Algorithm Based on Deep Learning Multi Feature Fusion Driven by Big Data." *Image and Vision Computing*, 104, 104023, 1–8. 2020.

49. Drozdowski, P., F. Stockhardt, C. Rathgeb, D. Osorio-Roig, and C. Busch. "Feature Fusion Methods for Indexing and Retrieval of Biometric Data: Application to Face Recognition With Privacy Protection." *IEEE Access*, 9, 139361–139378. 2021.

50. Alom, M. "The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches." arxiv.org/abs/1803.01164v2. 2018.

51. Niu, Z., G. Zhong, and H. Yu. "A Review on the Attention Mechanism of Deep Learning." *Neurocomputing*, 452, 48–62. 2021.

52. Marković, D., A. Mizrahi, D. Querlioz, and J. Grollier. "Physics for Neuromorphic Computing." *Nat. Rev. Phys.*, 2, 499–510. 2020.

53. Vakil, A., J. Liu, P. Zulch, E. Blasch, R. Ewing, and J. Li. "A Survey of Multimodal Sensor Fusion for Passive RF and EO Information Integration,. *IEEE Aerospace and Electronic Systems Magazine*, 36(7), pp. 44–61. 2021.

54. Moret-Bonillo, V. "Can Artificial Intelligence Benefit from Quantum Computing?" *Prog. Artif. Intell.* 3, 89–105. 2015.

ARTIFICIAL INTELLIGENCE (AI) AND MACHINE LEARNING (ML) IN BIOMETRIC DATA FUSION

By Abdul Rahman, Ph.D., Steven R. Knudsen, Ph.D., Deanna C. Milonas, Daniel Fleming, and John Clements

HDIAC-BCO-2021-192