Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2023.1120000

Multimodal Face Data Sets—A Survey of Technologies, Applications, and Contents

KAMELA AL-MANNAI[1], KHALED AL-THELAYA[1], JENS SCHNEIDER[1], AND SPIRIDON BAKIRAS[2]

[1]Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, 00000 Doha, Qatar (e-mail: {kaalmannai | khalthelaya | jeschneider}@ hbku.edu.qa)

[2]Infocomm Technology Cluster, Singapore Institute of Technology, 567739 Singapore. (e-mail: spiridon.bakiras@ singaporetech.edu.sg)

Corresponding author: Kamela A. Al-Mannai (e-mail: kaalmannai@ hbku.edu.qa).

ABSTRACT Due to their ease-of-use, biometric verification methods to control access to digital devices have become ubiquitous. Many rely on supervised machine learning, a process that is notoriously data-hungry. At the same time, biometric data is sensitive from a privacy perspective, and a comprehensive review from a data set perspective is lacking. In this survey, we present a comprehensive review of multimodal face data sets (e.g., data sets containing RGB color plus other channels such as infrared or depth). This follows a trend in both industry and academia to use such additional modalities to improve the robustness and reliability of the resulting biometric verification systems. Furthermore, such data sets open the path to a plethora of additional applications, such as 3D face reconstruction (e.g., to create avatars for VR and AR environments), face detection, registration, alignment, and recognition systems, emotion detection, anti-spoofing, etc. We also provide information regarding the data acquisition setup and data attributes (ethnicities, poses, facial expressions, age, population size, etc.) as well as a thorough discussion of related applications. Readers may thus use this survey as a tool to navigate the existing data sets both from the application and data set perspective. To existing surveys we contribute, to the best of our knowledge, the first exhaustive review of multimodalities in these data sets.

INDEX TERMS Anti-spoofing, Face detection, Face verification, Face verification systems, Multimodal face data sets.

I. INTRODUCTION

OWED to their ease of use and difficulty to replicate, biometrics are nowadays ubiquitously used for access management of digital devices. As of this writing, all major device manufacturers (e.g., Microsoft, Apple, Google, etc.) have adopted fingerprint readers, face scanners, or a combination of both to provide users with an easy and quick way to verify their access rights to devices. This mass adoption of biometric verification was made possible, in large parts, due to advances in imaging technology and artificial intelligence, with methodologies such as manifold learning paving the way for scalable and discriminative biometric verification.

From a technical perspective, a biometric verification system is presented with a sample that is then compared to a data base of biometric samples obtained during a sign up phase. If a positive match between the presented sample and the data base can be established, access is granted, otherwise, it is denied. A popular variant of this verification task is to use images of faces. The reason is that the mere presence of an authorized user in front of a device is sufficient to unlock and access the device, making this form of biometric verification arguably the easiest to use.

However, technical challenges include the robust identification of the user's face under a variety of lighting conditions (e.g., near darkness or sharp highlights), under facial poses, and from different viewing angles. Typical modern face recognition systems overcome low light conditions using infrared or near-infrared sensors to illuminate the scene, albeit at the disadvantage that successful systems must be able to process both color (RGB) and grayscale images.

Many systems tackle this task by using machine learning [1], and deep learning, in particular. The process usually

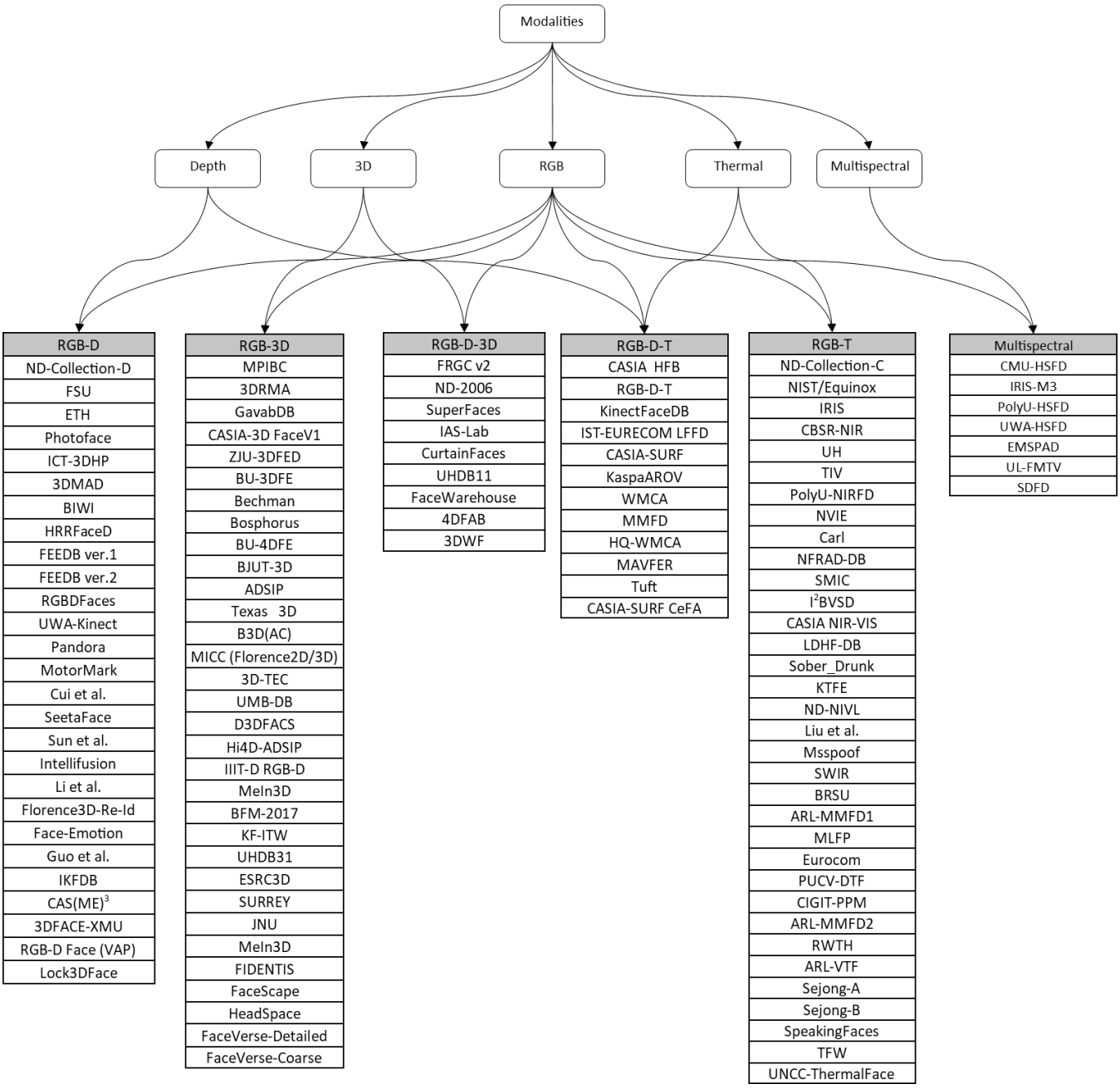

FIGURE 1: Taxonomy of the multimodal data sets based on modality type.

consists of image acquisition, often adding modalities other than visible-spectrum (VIS) RGB color channels (e.g., depth or infrared) followed by face detection, feature alignment or funneling, computing an embedding in a high-dimensional latent space, before a typically simple thresholding is applied (e.g., cosine metric or squared distances between the data obtained during sign up and the authentication attempt).

Such verification systems are often trained in a supervised fashion, requiring vast amounts of training data. However, images of faces are sensitive from a privacy perspective, and, as a result, such data often requires licensing, if available in the first place.

In this survey, we present a review of facial data sets currently available. The purpose is twofold—not only to present a comprehensive overview of existing data sets including a thorough discussion of their features/attributes, but also to review the eligibility of the data for different tasks such as face detection, alignment, registration, verification, and recognition, as well as 3D face recognition, liveness detection, and emotion recognition. Following recent trends in both industry and academia to develop biometric verification systems that are robust against various forms of anti-spoofing, lighting conditions, and facial expressions, we place particular emphasis on *multimodal* data sets that com-

bine RGB color channels with additional data such as depth, (near-)infrared or even geometry. We explicitly exclude pure RGB data from this survey and refer the reader to recent reviews instead [2]–[5].

II. SURVEY METHODOLOGY

Our focus is on multimodal data sets designed for facial analysis applications. Our primary aim is to compile the most widely used data sets accessible to the research community, thus simplifying the process of selecting the most suitable data set for various types of applications. To achieve this goal, we emphasize four key criteria: scanning technologies, modalities, demographics, and applications. Throughout this survey, we analyze and categorize data sets considering these four criteria.

In this study, we include peer-reviewed articles spanning from 1990 to 2023. We exclude extended abstracts and papers written in languages other than English. Our paper selection process involved gathering articles from both Scopus and Google Scholar using specific search queries, such as "multimodal data sets for facial analysis," "RGB Depth Thermal 3D 2.5D data sets for facial analysis," "face data set modalities," and "facial scanning and analysis databases/data sets." We considered only the initial search results until the titles and keywords indicated a lack of relevance. This initial phase yielded approximately 1,000 papers.

Subsequently, we conducted a thorough review, eliminating duplicate papers and screening the remaining ones based on their abstracts, adhering to our exclusion criteria. Following this review, we engaged in extensive forward and backward referencing to identify the most pertinent publications. This meticulous process led us to retain more than 200 papers. While we cite all these remaining papers in this survey, our focus then shifted to identifying about 150 data sets that encompass two or more modalities. It is essential to highlight that we excluded plain RGB data sets, since there are numerous surveys available that provide comprehensive coverage of this specific type of data sets.

III. TAXONOMY

We focus on multimodal data sets that comprise one or more types of data sources. Following our search criteria, we have gathered data from four distinct modalities for facial scanning and analysis: RGB, Depth, 3D, and thermal data (e.g., infrared and near-infrared). These data sets are categorized based on these four modalities, as illustrated in Figure 1. Furthermore, we have compiled and compared various demographic attributes for each of these data sets, as detailed in Table 1. These attributes include age, gender, ethnicity, the number of data collection sessions, and the time intervals between these sessions. We would like to emphasize that if the demographic information is not readily available in the primary data set documentation, we exclude the publication from our demographic analysis. Furthermore, in Section IV, we explore and compare various facial scanning technologies. These technologies are categorized into three groups: Visible, 3D/Depth, and thermal. To facilitate organization, we classify the collected data sets based on the respective scanning technologies, and this classification can be found in Tables 3 and 4. Subsequently, in Section V, we delve into a detailed analysis of these data sets based on attributes associated with the scanning technology and data source. These attributes include the number of samples, camera specifications, data resolution, and wavelength characteristics. Finally, we approach the collected data sets from the perspective of their applicability in facial analysis, as discussed in Section VI. In this context, we identify the specific types of applications each data set is most suitable for.

IV. FACIAL SCANNING TECHNOLOGIES

The previous few decades have seen a surge in the use of facial scanning technologies. These technologies play a pivotal role in applications ranging from biometric authentication to entertainment and medical diagnostics. The array of facial scanning techniques encompasses various approaches, each leveraging distinct principles to achieve accurate and detailed 3D representations of facial features. The huge amount of data generated daily paired with a marked increase in computational capacity has presented researchers with an exciting chance to use these technologies to record and analyze human bahavior and visual characteristics.

Structured light scanning projects controlled patterns of light to capture facial contours, allowing for precise measurements. Time-of-Flight (ToF) cameras emit light and measure its return time, enabling rapid depth sensing. Stereo vision systems use multiple cameras to triangulate facial geometry, while photometric stereo analyzes lighting variations to infer depth information. Additionally, passive techniques rely on natural light and image analysis to create 3D models without emitting light actively. This section delves into the spectrum of facial scanning technologies, shedding light on their unique attributes and applications.

A. VISIBLE LIGHT SCANNING

A visible camera sensor is a specialized scanner designed to capture visible light within the 400 to 700 nanometers spectrum. It seamlessly converts this light into an electrical signal to form images and video streams. These cameras aim to render images similar to human perception by capturing light in the red, green, and blue (RGB) wavelengths. This approach facilitates a true representation of colors, resulting in realistic imagery. In contemporary contexts, advanced security and surveillance cameras offer the ability to identify targets and objects within the scene using high-definition (HD) resolution or greater. These cameras come equipped with a range of lens options, further enhancing their versatility.

Similar to the human eye, visible light cameras require lighting to operate effectively. Environmental factors like fog, haze, smoke, heat waves, and smog exert notable influence on their performance. As a result, their practical use is con-

TABLE 1: Demographic information of the multimodal data sets included in this study.

Data Set	Year	Subjects	Male	Female	Samples	Modalities	Age Range	Sessions	Time-lapse	Ethnicities	Application
MPIBC [6]	1996	200	100	100	1,400	RGB/Mesh	20–40			Caucasian	Reconstruction, Recognition
3DRMA [7]	1998	120	106	14	360	PointCloud	20–60			European	Verification
CMU-HSFD [8]	2002	45			147	Hyperspectral		1–5	Several Weeks		Recognition
ND-Collection-C [9]	2002	240				Gray/LWIR		10	10 Weeks		Recognition
ND-Collection-D [10]	2003	275			953	RGB/Depth		2	6–13 Weeks		Matching, Recognition
CASIA-3D FaceV1 [11]	2004	123			4,624	RGB/Mesh				East-Asian	Expressions
FRGCv2 [12]	2004	466	266	200	50,000	RGB/Depth/Mesh	18–28			Asian, White, Others	Detection, Recognition
GavabDB [13]	2004	61	45	16	427	RGB/PointCloud	18–40			Caucasian	Detection, Recognition
FRAV3D [14]	2005	106	79	27	1,696	RGB/Depth/Mesh					Verification
BU-3DFE [15]	2006	100			2,500	RGB/Mesh				5 Ethnicities	Expressions
IRIS-M3 [16]	2006	82	62	20	2,624	Multispectral				Caucasian, Asian, African	Recognition
BU-4DFE [17]	2008	101			60,600	RGB/Mesh				4 Ethnicities	Expressions
Bosphorus [18]	2008	105	60	45	4,652	RGB/PointCloud	25–35			Caucasian	Expressions
BJUT-3D [19]	2009	500	250	250	46,500	RGB/Mesh	16–49			East-Asian	Pose Estimation, Recognition
Polikovsky [20]	2009	10			13	Gray/HS				Asian, Caucasian	Expressions
CASIA HFB [21]	2009	100	57	43	992	RGB/NIR/Depth					Matching, Recognition
ADSIP [22]	2009	10	2	8	210	RGB/Mesh					Expressions
Texas 3D [23]	2010	118			2,298	RGB/PointCloud	22–75			Caucasian, African, Asian, Hispanic	Detection, Recognition
PolyU-HSFD [24]	2010	25	17	8	300	Hyperspectral	21–33	4	3–10 Months	East-Asian	Recognition
B3D(AC) [25]	2010	14	6	8	1,109	RGB/Mesh	21–53				Expressions
NVIE [26]	2010	215	157	58	436	Mono/LWIR	17–31			White, Black, Asian, Latino	Expressions
Carl [27]	2010	41	32	9	7,380	RGB/Thermal			6 Weeks		Recognition
Photoface [28]	2011	261	227	34	7,356	RGB/Depth/Albedo		1–20	1–5 Weeks	Caucasian	Recognition, Verification
NFRAD-DB [29]	2011	50	37	13	600	RGB/NIR				East-Asian	Recognition
D3DFACS [30]	2011	10	4	6	534	2D/Mesh	23–41			Caucasian	Expressions
Hi4D-ADSIP [31]	2012	80	32	48	3,360	2D/Mesh	18–60			Variable	Expressions
CASIA NIR-VIS [32]	2013	725			17,580	RGB/NIR		4	4 Years	East-Asian	Matching, Recognition
SMIC [33]	2013	20	14	6	306	RGB/NIR	22–34			Asian, Caucasian, African	Expressions
3DMAD [34], [35]	2013	17	10	7	255	RGB/Depth		3	2 Weeks		Anti-Spoofing
LDHF-DB [36]	2013	100	70	30	1,600	RGB/NIR					Matching, Recognition
I^2BVSD [37]	2013	75	60	15	1,362	RGB/LWIR				Asian	Verification
Sober Drunk [38], [39]	2013	41	31	10	4,100	LWIR					Pose Estimation, Expressions
KinectFaceDB [40]	2014	52	38	14		RGB/IR/PointCloud	25–32		5–14 Days	Caucasian, Middle East, Maghreb, East-Asian, Indian, Hispanic, African-American	Detection, Recognition
KTFE [41]	2014	26	16	10		RGB/NIR/LWIR	11–32			7 Ethnicities	Expressions
Liu et al. [42]	2015	77			181	RGB/LWIR				White, Black, Asian, Latino	Expressions
Lock3DFace [43]	2015	499	377	122	5,711	RGB/Depth					Detection, Recognition
ND-NIVL [44]	2015	574			22,264	RGB/NIR			6 Months		Recognition
MeIn3D [45]	2016	9,663	4,638	5,025	12,000	RGB/Mesh	1–83				Morphable Models
IST-EURECOM LFFD [46]	2017	100			4,000	RGB/Depth				19 Ethnicities	Expressions
Pandora [47]	2017	22	10	12	110	RGB/Depth					Pose Estimation
UHDB31 [48]	2017	77	53	24	25,872	2D/Mesh					Recognition
ESRC3D [49]	2018	99	45	54		RGB/Mesh					Expressions
FIDENTIS [50]	2018	2,476	1,154	1322		RGB/Mesh	6–60			53 Ethnicities	Recognition, Reconstruction,
4DFAB [51]	2018	180	120	60	1,800,000	RGB/3D/Depth	5–50		5 Years	30 Ethnicities	Expressions
UL-FMTV [52]	2018	134	86	48		Multispectral			4 Years		Pose Estimation, Expression
Eurocom [53]	2018	50			4,200	RGB/Thermal			3–4 Months		Expressions
PUCV-DTF [54]	2018	46	40	6	11,500	Thermal	18–29				Pose Estimation, Expressions
SDFD [55]	2018	54	54	0	6,480	RGB/NIR	20–50				Recognition
MAVFER [56]	2020	17	7	10	17	RGB/Depth/LWIR	21–38			Korean	Expressions
HeadSpace [57]	2020	1,519			1,519	RGB/Mesh	1–89			White, Asian, Black	Morphable Models
Tuft [58]	2020	113	39	74	10,000	RGB/NIR/LWIR/PointCloud	4–70			15 Ethnicities	Recognition
Sejong-A [59]	2021	30	16	14	1,500	RGB/SWIR/NIR			2 Weeks	Russian, African, Caucasian Southeast- & Southcenter-Asian	Verification
Sejong-B [59]	2021	70	44	26	23,000	RGB/SWIR/NIR			2 Weeks	Russian, African, Caucasian Southeast- & Southcenter-Asian	Verification
UNCC-ThermalFace [60]	2022	10	5	5	10,000	LWIR					Recognition

strained to daylight hours and clear weather conditions. To avoid these limitations, visible light cameras are often coupled with illumination or thermal infrared counterparts. This combination can function during nighttime or in low-light scenarios, as well as in situations involving haze, fog, smoke, or sandstorms—conditions that can otherwise compromise the functionality of cameras capturing the visible spectrum. Due to these distinct advantages, Infiniti Electro-Optics, a leader in the surveillance industry, strongly advocates the adoption of multi-sensor EO/IR systems for tasks demanding long-range surveillance and mission-critical applications.

B. 3D AND DEPTH FACIAL SCANNING

The rapid advancement of 3D sensors presents a good opportunity for facial analysis, potentially bypassing the inherent constraints of 2D technologies. The sophisticated geometric details within 3D facial data hold the potential to significantly enhance facial analysis output, especially in scenarios where 2D technologies are found to be inefficient. While advances have been made in 2D face recognition research in recent years, its precision remains heavily reliant on lighting conditions and the alignment of human poses [61]. The accuracy of 2D facial analysis output tends to decrease when facing low light or improperly aligned facial poses within the camera's field of view [62]. This has prompted numerous researchers to shift their attention toward 3D facial analysis. Various types of 3D scanning technologies are utilized to capture detailed and accurate representations of the human face. These technologies enable the extraction of geometric and textural information, which are essential for tasks such as face recognition, expression analysis, and virtual avatar creation [63].

Passive and active 3D facial acquisition systems are two categories of technologies used to capture three-dimensional representations of human faces. They differ in their approach to acquiring depth information and the type of interaction

with the subject being scanned [64].

1) Active Acquisition Systems

Active systems involve the emission of external stimuli, typically light or infrared radiation, onto the subject's face. These emitted signals are then measured after being reflected or scattered off the face's surface to calculate the distances and to create a 3D representation [65]. The system actively engages with the subject through the emission and detection of these signals, where an active light source rotates around an item or face to scan the full object surface. These sensors provide comprehensive 3D readings, although the majority of them are restricted to static situations [66]. Examples of these technologies include structured light scanners, ToF scanners, and laser triangulation scanners. This type of scanning technology operates in varying lighting conditions and is suitable for various facial analysis applications, including biometrics and medical diagnostics [67]. In the following, we discuss some of the prominent types of 3D active scanning technologies used in facial analysis applications.

Structured Light Scanning. Structured light scanners project a known pattern of light onto the subject's face and then capture how the pattern is distorted by the surface. The pattern's deformation is then recorded using a CCD (charge-coupled device). The depth data is derived from the way the camera's sensor interprets the pattern within the environment. As an illustration, when a sequence of stripes is projected onto a spherical object, these stripes exhibit a specific distortion and curvature as they conform to the contours of the object's surface. The reconstruction of the 3D image is accomplished through sophisticated software [66].

Recent advances in structured light scanning make it a practical approach for 3D facial capturing [68]. There are several structured light scanning devices that are commonly used for facial scanning. For instance, the Kinect v1, a depth sensor developed by Microsoft and introduced in 2010, is based on structured light technology. This device is equipped with a visible light camera and a depth sensor that consists of an IR projector and detector. The projector emits a pattern of infrared light in the form of a grid or speckle pattern onto the scene. This pattern is carefully designed and controlled, with distinct features that can be easily tracked. The infrared pattern gets distorted by objects and surfaces in the environment. The distortion of the pattern is due to the varying distances of different points on the objects from the sensor. The infrared camera observes and captures the distorted pattern as it is reflected back by the objects. By analyzing the observed distortion of the pattern, the Kinect software calculates a relatively sparse depth information for object points covered by the speckle pattern, and interpolates this information into a full depth frame. This scanning technology was used for facial scanning in several data sets [69]–[72]. The Intel RealSense depth camera can also be categorized as a structured light scanning technology since it employs coded light to capture images [73]. This technology relies on the precise interpretation of a projected light pattern. Coded and structured light cameras excel when used indoors and within relatively limited ranges, which can vary depending on the camera's light intensity. The Intel RealSense depth camera is used by many researchers to collect facial data sets for many applications, such as face spoofing attack detection [74]–[79], facial expression recognition [80], and face recognition [81].

However, a challenge with such systems is their susceptibility to interference from environmental factors like direct sun light, other cameras, or devices emitting infrared signals. Moreover, a sophisticated algorithm is required to compute the distance at every point within the pattern. Many earlier structured light scanning systems lack modularity. This implies that a subject's entire facial data, spanning from ear to ear, cannot be captured from multiple viewpoints simultaneously. Consequently, a secondary capture from an alternative angle is required. Although KinectFusion cite-Newcombe:2011:Fusion addresses this problem, such movements introduce potential discrepancies in the resulting 3D data, as both the system and the subject may need to adjust positions [67].

Time-of-Flight (ToF). ToF sensors operate similarly to other laser scanners, yet their distinctive advantage lies in their ability to capture entire scenes instantaneously, making them well suited for dynamic environments. ToF offers full-range, full-frame distance data at impressive frame rates which positions these sensors as potential alternatives to conventional 3D acquisition systems. The distance calculation depends on measuring the phase difference between emitted near-infrared light from a LED, and the subsequently received near-infrared signal [67].

Microsoft Kinect v2, introduced by Microsoft in 2013, is one of the most common ToF sensors used for depth estimation in many applications, including facial scanning. It is equipped with a laser IR transmitter and a depth sensor. The emitter projects modulated IR light into the observed area. The depth sensor then captures the light that is reflected back. A timing generator is utilized to ensure precise synchronization between the IR emitter and the depth sensor. By analyzing the phase shift between the emitted and reflected light, Kinect v2 can accurately calculate the depth for each individual pixel. Several facial analysis data sets employ this device to capture the depth and 3D shape of human faces [56], [82]–[84]. The quality of the generated models shows higher accuracy, quality, and resolution compared to Microsoft Kinect v1.

Because of their compact design, ToF sensors can seamlessly integrate into real-time facial analysis systems similar to standard 2D cameras. Nevertheless, these sensors are not devoid of limitations. Challenges encompass restricted resolution, susceptibility to noise in the data, exclusive grayscale outputs, high cost, and inherent limitations in resolution [68].

Laser Triangulation Scanners. A triangulation-based 3D

TABLE 2: Comparison of the most common 3D and depth scanning technologies.

	ToF	**Structured Light**	**Stereo Vision**	**Laser Triangulation**
Type	Active	Active	Passive	Active
Distance	0.4–5m	0.5–1.2m	$\leq$ 2m	$\leq$ 2.5m
Environment	Indoor/Outdoor	Indoor	Controlled lighting	Indoor/Outdoor
Software Overhead	Low	Medium	High	Medium
Accuracy	Medium	High	Low	High
Resolution	High	High	Low	High
Response Time	High	Low	Medium	High
Cost	Low	Medium	Low	High
Depth Range	Scalable	Scalable	Limited	Scalable
Compactness	High	High	Low	High
Advantages	• Depth scalability. • Suitable for scanning large objects.	• Captures a large area of the object at once. • Provides higher detail levels compared to ToF and Stereo Vision and higher safety compared to Laser Triangulation technology.	• Copes well with long distances and moving objects. • The hardware implementation cost is very low. • Well-suited for capturing images for intuitive presentation to humans.	• Less sensitive to environmental lighting conditions and mechanical alignment. • Suitable for scanning large objects. • Compact-sized and portable.
Disadvantages	• Trade accuracy for speed and depth scalability.	• Struggles with dark, transparent, or shiny objects. • Not suitable for very large objects. • Sensitive to optical interference. • Requires to remain still when taking the scan of an object.	• Scan quality can be affected by the lighting environment, the quality of cameras, and the software. • Less effective in measuring distance. • Requires careful calibration. • Requires sufficient intensity and color variation.	• Not safe. • Sensitive to ambient light.

scanning technology, such as the Minolta Vivid scanner [85], detects the laser beam's emitting and receiving angles before using triangulation methods to establish the exact point of reflection. A precise map is generated by calculating and grouping multiple reflection locations as the laser beam scans through the face. The scanning speed of triangulation-based devices is sacrificed for precision. The target individual would have to remain still for several minutes before a 3D facial map can be obtained. As a result, this method is impractical for 3D video recording [63]. Laser triangulation scanners are commonly used for industrial and precision applications, but they may not be the primary choice for capturing facial data due to their nature and the need for precise positioning. As a result, there are fewer face data sets captured using laser triangulation scanners compared to other 3D scanning technologies [86].

2) Passive Acquisition Systems

The second type of 3D system used for capturing human faces are passive vision systems, which contain solely cameras using only the ambient light [64]. Because passive vision systems for 3D data acquisition relies on 2D images, it suffers from a correspondence problem—it is difficult to discover a set of correct corresponding points in different images captured using multiple cameras for the same object at the same moment. The problem is normally addressed using sparse matching of feature points, e.g., by using SIFT citeLowe:1999:SIFT and RANSAC citeFischler:1981:RANSAC, followed by bundle adjustment citeTriggs:1999:Bundle to refine the reconstruc-

TABLE 3: 3D and depth scanning technologies with corresponding data sets.

Data Set	Structred Light			Stereo Vision				ToF	Laser Triangulation	
	Kinect v1	Intel RealSense	other	3dMD	DI3D	DI4D	other	Kinect v2	Minolta Vivid	CyberWare 3030
CASIA-3D FaceV1 [11]									✓	
RGB-D Face (VAP) [69]	✓									
FRGCv2 [12]									✓	
BU-3DFE [15]				✓						
BU-4DFE [17]					✓					
3D-TEC [87]									✓	
UMB-DB [88]									✓	
SuperFaces [89]	✓			✓						
CurtainFaces [70], [90]	✓									
FaceWarehouse [71]	✓									
Lock3DFace [43]	✓									
IIIT-D RGB-D [72]	✓									
UHDB11 [91]				✓						
CAS(ME)^3 [92]		✓								
3DMAD [34], [35]	✓									
CASIA-SURF [79]		✓								
WMCA [74]		✓								
HQ-WMCA [93]		✓								
CASIA-SURF CeFA [94]		✓								
ND-2006 [95]									✓	
GavabDB [13]									✓	
UoY [96]				✓						
BJUT-3D [19]										✓
FRAV3D [14]									✓	
Pandora [47]								✓		
MPIBC [6]										✓
BIWI [97]	✓									
ICT-3DHP [98]	✓									
KaspaAROV [82]	✓							✓		
HRRFaceD [83]								✓		
IST-EURECOM LFFD [46]						✓				
FaceVerse-Detailed [99]						✓				
FaceVerse-Coarse [99]			✓							
Cui et al. [100]		✓								
ESRC3D [49]					✓					
SURREY [101]				✓						
JNU [101]				✓						
3DWF [102]							✓			
Intellifusion [103], [104]										
Li et al. [105]								✓		
SeetaFace [106]		✓								
MotorMark [107]								✓		
Sun et al. [108]							✓			
IAS-Lab [109]	✓									
RGBDFaces [110]	✓									
MICC (Florence2D/3D) [111]				✓						
Face-Emotion [112]								✓		
Florence3D-Re-Id [84]								✓		
IKFDB [113]								✓		
MMFD [114]		✓								
RGB-D-T [115]		✓								
FIDENTIS [50]							✓			
FaceScape [116]						✓				
CASIA HFB [21]									✓	
4DFAB [51]						✓		✓		
ND-Collection-D [10]									✓	
3DFACE-XMU [117]							✓			
ZJU-3DFED [118]			✓							
FSU [119]									✓	
B3D(AC) [25]			✓				✓			
D3DFACS [30]				✓						
Hi4D-ADSIP [31]					✓					
ADSIP [22]				✓						
MAVFER [56]								✓		
HeadSpace [57]				✓						
MeIn3D [45]				✓						
Tuft [58]							✓			
UHDB31 [48]				✓						
Bechman [120]									✓	
Eurocom [53]	✓									

tion. Still, the reconstructed 3D data resulting from these systems may be exceedingly noisy, incomplete, and inconsistent [65]. In the following, we discuss various types of 3D passive scanning technologies used in the context of facial analysis.

Stereo Vision Systems. Stereo vision utilizes two or more cameras placed slightly apart to capture images of the same scene from different angles. By analyzing the disparities between these images, the system can calculate depth information. This method is suitable for capturing dynamic facial expressions and subtle depth changes due to movement [121]. The 3DMD corporation sells various types of stereo-vision scanners. The 3dMDface system is one of the stereo scanners that are specifically developed by 3DMD for facial scanning [122]. In this system, multiple cameras are used to generate a high-quality color texture map that is registered with the 3D data. Typically, the collected shape and texture data capture the whole face, resulting in a texture-mapped mesh with high coverage and precision. This system is used in many facial analysis data sets [15], [45], [89], [91], [101], [111]. Similarly, 3dMDhead is another capturing system developed by 3DMD, based on Stereo Vision technology to capture the whole space of the head. It is used by Dai et al. [57] to generate a shape-and-texture 3D morphable model of the full head. The DI3D imaging is another stereo vision system that uses two or more readily available digital SLR cameras, making it easily accessible [123]. What sets it apart is its freedom from the need for intense white lighting, intricate pattern projections, or lasers. Instead, it employs the technique of triangulating, whereby the high-resolution images captured by these paired cameras are used to generate real-time 3D surfaces. This system is used by Zhang et al. [17] to capture 3D spatio-temporal features in subtle facial expressions to understand the relation between pose and motion dynamics in facial action units. The Stirling ESRC data set was also captured using the DI3D scanning system to collect 3D face scans of 100 subjects under seven different expression variations. This data set was used by Feng et al. [49] to develop an approach for dense 3D reconstruction from 2D face images in the wild. The DI3D system was also used by Bogdan et al. [31] to collect the Hi4D-ADSIP data set, which consists of 3,360 facial scans captured from 80 subjects. As technology advances, this method extends into capturing dynamic motion over time. By employing three or more industrial-grade video cameras, another scanning technology, named DI4D, can create complete 3D color video sequences of moving surfaces. Each frame in the sequence is treated as an individual stereo pair of images which are then automatically processed to produce detailed 3D color surfaces. The resulting data streams are seamlessly merged to craft a series of high-resolution 3D polygonal images that can be played back as a dynamic movie sequence. The frame rate can vary depending on the hardware, but the system readily achieves a smooth capture of at least 25 frames per second. In practical terms, the DI4D Capture System is well-suited for applications where capturing not only the static 3D shape but also the motion and changes in the object's surface over time is crucial. This could be particularly valuable in fields such as facial animation, biomechanics, and medical imaging, where capturing and analyzing dynamic facial expressions, body movements, or deformations are essential for research and creative endeavors. This dynamic scanning system was used by Cheng et al. [51] to collect the 4DFAB data set, which includes 4D facial scanning of 180 subjects for both posed and spontaneous facial behaviors collected over a period of five years under multiple sessions.

Shape from Shading (Photoclinometry). Shape-from-Shading (SFS) estimates surface orientation by using shading from a single image. It is a popular topic of research because of its obvious uses and ease of capture—the goal is to rebuild an accurate 3D model from a 2D snapshot, which eliminates the need for expensive and/or complex capture hardware. Because it is extremely difficult to separate gradient information from color or texture information in a single image, there will always be ambiguity as to whether an intensity gradient is due to a slope or some color, pattern shift, or shadowing [124].

Photometric Stereo (PS). PS is an improved SFS method that aims to eliminate the ambiguities associated with the standard SFS methodology of estimating 3D shape from a single image by separating the 3D morphology from the 2D texture. It creates a 3D form from three or more photos of the same item, each light from a different and known direction, and estimates surface normals at each pixel [28], [125].

C. THERMAL AND MULTISPECTRAL SCANNING

Most conventional image-based algorithms have an excellent performance in terms of accuracy when the face image is recorded under controlled conditions. However, these methods fail when presented with images captured under an uncontrolled environment with high distortions resulting from changes in illumination. A nighttime situation is an example of a condition where human recognition, based exclusively on visible spectrum pictures, may be impractical. Infrared imaging can be used to overcome these challenges by capturing the temperature of the skin [143]. On the other hand, multispectral and hyperspectral camera sensors excel in capturing both spatial and spectral data from human faces. Progress in imaging technologies has led to the emergence of multispectral imaging devices boasting broader technology options, enhanced quality, and reduced cost. Among the spectrum-recording apparatuses are UV-visible, near-infrared (NIR), short-wave IR (SWIR), mid-wave IR (MWIR), and long-wave IR (LWIR) devices. Despite substantial variation in their price points, a trend of decreasing costs persists as technological advancements drive higher pixel densities, improved pixel yields, and increasing demand by applications [144].

Infrared thermal sensors facilitate the imaging of scenes and objects through two methods: IR light reflectance and

TABLE 4: Thermal infrared and multispectral/hyperspectral scanning with corresponding data sets.

Data Set	VIS	NIR	SWIR	MWIR	LWIR	other	Multi-/Hyperspectral
SMIC [33]	✓	✓					
I^{2}BVSD [37]	✓	✓					
Msspoof [126]	✓	✓					
SWIR [127]	✓		✓	✓			
BRSU [128]	✓		✓				
EMSPAD [129]							✓
MLFP [130]	✓	✓					
CASIA-SURF [79]		✓					
CIGIT-PPM [131]	✓	✓					
PolyU-HSFD [24]							✓
CMU-HSFD [8]							✓
ND-Collection-C [9]	✓				✓		
ND-NIVL [44]	✓	✓					
CASIA HFB [21]	✓	✓					
CASIA NIR-VIS [32]	✓	✓					
LDHF-DB [36]	✓	✓					
NFRAD [29]	✓	✓					
PolyU-NIRFD [132]	✓	✓					
NVIE [26]	✓				✓		
Liu et al. [42]	✓				✓		
IRIS [133]	✓				✓		
UH [134]				✓			
Carl [27]	✓	✓					
ARL-MMFD1 [135]	✓				✓		
ARL-MMFD2 [136]	✓				✓		
UL-FMTV [52]							✓
Eurocom [53]	✓					✓	
Tuft [58]	✓	✓			✓		
Sejong-A [59]	✓	✓			✓		
Sejong-B [59]	✓	✓			✓		
Sober Drunk [38], [39]					✓		
PUCV-DTF [54]					✓		
TFW [137]	✓					✓	
SpeakingFaces [138]	✓				✓		
KTFE [41]	✓				✓		
NIST/Equinox [139]	✓					✓	
SDFD [55]							✓
CBSR-NIR [140]	✓	✓					
RWTH [141]					✓		
UNCC-ThermalFace [60]					✓		
IRIS-M3 [16]							✓
UWA-HSFD [142]							✓

IR radiation emittance. This utilization of IR radiation stems from its correlation with the heat generated or reflected by an object, a concept known as thermal imaging. Since IR radiation wavelengths are longer than those of visible light, it falls outside the spectrum of human vision. The infrared (IR) spectrum can be categorized based on wavelength into the following bands [145] (for reference, the visible spectrum ranges approximately from 0.4 to 0.7 μm):

- NIR: Spanning from 0.7 to 1 μm.
- SWIR: Including the range of 1 to 3 μm.
- MWIR: Covering wavelengths from 3 to 5 μm.
- LWIR: Extending between 8 to 14 μm.
- FWIR: Comprising far wavelengths greater than 14 μm.

The NIR and SWIR bands are commonly termed "reflected infrared radiation," while the MWIR and LWIR bands are referred to as "thermal infrared radiation." Notably, the latter bands do not require an additional light or heat source; thermal radiation sensors can create images of the environment or objects solely by detecting the thermal energy emitted by observed elements in the scene [143].

NIR cameras exhibit heightened sensitivity to temperature changes but offer less detailed information than visible light

cameras. This is because the amount of colors captured in the visible spectrum delivers more comprehensive data and is easier to interpret [144]. Variations in facial images between the visible and infrared bands increase as wavelength increases. Therefore, the LWIR band is frequently employed to achieve complete lighting condition invariance since lighter areas in infrared images indicate higher temperatures.

V. MODALITIES

Facial analysis is a critical domain in biometrics and computer vision, offering applications in security, human-computer interaction, and emotional analysis. Understanding the diversity of modalities used for this purpose is fundamental. These modalities provide various ways to capture and interpret facial data. In this section, we explore five distinct modalities: RGB, Depth, 3D, Thermal, and Multispectral imaging, and their roles in facial recognition, expressions analysis, and verification.

A. RGB DATA

RGB imaging captures the color information of the face using red, green, and blue channels. It is the most commonly used modality and provides valuable visual appearance details. Modern RGB-based facial analysis systems have indeed achieved remarkable results in various applications, such as face recognition and authentication. However, they primarily rely on the visual appearance of faces captured through standard RGB cameras. This dependence on visual appearance poses a significant challenge in scenarios where lighting conditions are less than ideal, like poorly lit rooms, outdoor environments during nighttime, or overcast days. In such situations, the quality of the RGB images can deteriorate, leading to decreased accuracy and reliability in facial analysis tasks. Therefore, the fusion of RGB images with other modalities such as depth, thermal, and 3D meshes, represents a significant advancement in facial analysis technology. It not only mitigates the challenges posed by variable lighting conditions but also opens up new possibilities for applications in security, entertainment, healthcare, etc., where accurate facial analysis is paramount. For example, the output of applications such as face anti-spoofing, can be highly improved by augmenting the RGB images with other modalities, such as depth and thermal channels, to acquire more geometric and biometric features that can help detect several attacks.

There are several RGB data sets that are widely used for facial analysis applications, such as MS-Celeb-1M [146], one of the most common data sets with one million celebrity face images collected from the web. It covers a wide range of poses, ages, and ethnicities. VGGFace2 [147] is a data set with over 3 million face images from 9,000 individuals. It provides diverse poses, lighting conditions, and ages. CASIA-WebFace [148], [149] contains over 490,000 images from 10,575 subjects, including images captured in uncontrolled conditions from the web. The MEVIEW (Micro ExpressionsVIdEos in the Wild) data set [150] contains 40 micro-expression video clips at 25 fps with an image resolution of 1280×720. The average length of the video clips in the data set is 3 seconds, and the camera shot is often switched. The emotion types in MEVIEW are divided into seven classes: happiness, contempt, disgust, surprise, fear, anger, and ambiguous/unclear emotions. IJB-A [151] is an RGB in the wild data set containing 500 subjects with manually localized face images. In our survey, we target multimodal data sets that incorporate multiple modalities for facial analysis applications. Data sets that are composed of plain RGB data sets are not covered by this survey. For more details about such data sets, we refer the reader to Castaneda et al. [4] and Chihaoui et al. [5], which provide a more comprehensive discussion of this type of data sets.

B. 3D DATA

3D data sets specifically focus on capturing the 3D shape of the face. These data sets typically consist of 3D facial scans or point clouds that represent the facial geometry. Instead of using RGB images, they directly capture the facial structure in a three-dimensional space. This allows for precise analysis of facial features and more robust face recognition algorithms that can handle pose variations and other geometric deformations. Therefore, the field of facial analysis has seen a significant shift towards 3D face recognition techniques [10]. This shift is primarily driven by the need to address the limitations and challenges posed by conventional 2D face analysis systems. One of the key advantages of 3D face recognition is the wealth of geometric information it offers. By capturing the three-dimensional structure of the face, including the contours, shape, and spatial relationships of facial features, 3D systems can create a more accurate and unique facial signature for each individual. When comparing 2D and 3D facial analysis accuracy under identical pose and lighting conditions, 3D face recognition often outperforms its 2D counterpart [159]. Table 5 summarizes 3D data sets whereas Table refRGB-D-3D presents data sets that include both depth and 3D facial scanning.

The point cloud representation is the most fundamental way to depict the facial surface. It is also the most common output generated by 3D scanners. It encompasses an unorganized collection of 3D coordinates corresponding to points on the facial surface [160]. In the past, it was viewed as a sparse approximation of the actual surface, but with the advent of point-based rendering and increases in storage and processing capabilities this perception is diminishing. Nowadays, facial analysis can delve into increasingly finer levels of detail without concerns about memory limitations. This ease of use has also recently accelerated the development of point cloud algorithms. Additionally, there are suggestions to employ sparser representations of the complete point cloud, such as contour and profile curves, to approximate the facial shape [161].

On the other hand, mesh representations are achieved by tesselating 3D point clouds, typically using triangular facets. This connectivity or topology data eases the retrieval of neighboring points and, thus, enables the measurement of

TABLE 5: RGB-3D data sets.

Data set	Year	Subjects (M/F)	Samples	Camera	Modalities	RGB Resolution	Demographics	3D Resolution	Application
MPIBC [6]	1996	200 (100/100)	1,400	CyberWare TM	RGB/Mesh		Age, Gender	70,000 vertices	Reconstruction, Recognition
3DRMA [7]	1998	120 (106/14)	360	Structured Light	PointCloud		Age, Gender	4,000 vertices	Recognition, Verification
CASIA-3D FaceV1 [11]	2004	123	4,624	Minolta Vivid 910	RGB/Mesh				Expressions
GavabDB [13]	2004	61 (45/16)	427	Minolta Vivid 700	RGB/PointCloud		Age, Gender	10k-20k vertices	Detection, Recognition
BU-3DFE [15]	2006	100	2,500	3DMD	RGB/Mesh	512×512	Age, Gender, Ethnicity		Micro Expressions
ZJU-3DFED [118]	2006	40	360	Structured Light System	RGB/Mesh			2,000 vertices, 4,000 triangles	Recognition, Posed Expressions
Bechman [120]	2007			Cyberware 3030	RGB/Mesh				Dynamic Expressions
BU-4DFE [17]	2008	101	60,600	Di3D	RGB/Mesh	$1,040 \times 1,329$	Age, Gender, Ethnicity		Alignment, Dynamic Expressions
Bosphorus [18]	2008	105 (60/45)	4,652	Inspeck Mega Capturor II	RGB/PointCloud	$1,600 \times 1,200$	Age, Gender	35,000 vertices	Detection, Recognition, Verification
ADSIP [22]	2009	10 (2/8)	210	3dMD	RGB/Mesh	640×480			Anti-Spoofing
BJUT-3D [19]	2009	500 (250/250)	46,500	CyberWare 3030 RGB/PS	RGB/Mesh		Age, Gender		Recognition
Texas 3D [23]	2010	118	2,298	MU-2 stereo	RGB/PointCloud	751×501	Age, Gender		Detection, Recognition
B3D(AC) [25]	2010	14 (6/8)	1,109	Structured light, Stereo vision	RGB/Mesh		Age, Gender		Micro Expressions
MICC (Florence2D/3D) [111]	2011	53		3dMD	RGB/Mesh	$3,341 \times 2,027$		40,000 vertices, 80,000 triangles	Recognition
3D-TEC [87]	2011	214	428	Vivid 910	RGB/PointCloud	480×640	Age, Gender		Matching, Recognition
UMB-DB [88]	2011	143	1,473	Vivid 900	RGB/PointCloud	640×480	Age, Gender		Detection, Recognition
D3DFACS [30]	2011	10 (4/6)	534	3DMD	RGB/Mesh	$1,024 \times 1,280$	Age, Gender	30,000 vertices	Dynamic Expressions
Hi4D-ADSIP [31]	2012	80 (32/48)	3,360	DI3D	RGB/Mesh	$2,352 \times 1,728$	Age, Gender, Ethnicity		Dynamic Expressions
IIIT-D RGB-D [72]	2013	106	4,605	Kinect v1	RGB/PointCloud	640×480	Age, Gender		Detection, Recognition
MeIn3D [45]	2016	9,663 (4,638/5,025)	12,000	3dMD	RGB/Mesh		Age, Gender, Ethnicity	60,000 vertices, 120,000 triangles	Recognition, 3D Morphable Models
BFM-2017 [152]	2017	200	360	ABW-3D	RGB/Mesh				Registration, 3D Morphable Models
KF-ITW [153]	2017	17		Kinect v1	RGB/Mesh				3D Morphable Models, Recognition
UHDB31 [48]	2017	77 (53/24)	25,872	3dMD	RGB/Mesh	Multiple		25,000 vertices, 49,500 triangles	Recognition
ESRC3D [49]	2018	99 (45/54)		DI3D	RGB/Mesh				Recognition, Expressions
SURREY [101]	2018	168	168	3dMD	RGB/Mesh		Age, Gender, Ethnicity		3D Morphable Models
JNU [101]	2018	774	774	3dMD	RGB/Mesh		Age, Gender		3D Morphable Models
FIDENTIS [50]	2018	2,476 (1,154/1,322)		Vectra M1/XT/H1	RGB/Mesh		Age, Gender, Ethnicity	20k-60k vertices	Recognition, Verification
FaceScape [116]	2020	938	18,760	DSLR System	RGB/Mesh		Age, Gender	2M vertices, 4M triangles	3D Morphable Model
HeadSpace [57]	2020	1,519	1,519	3dMD	RGB/Mesh		Age, Gender, Ethnicity	180,000 vertices	3D Morphable Models, Reconstruction
FaceVerse-Detailed [99]	2022	128	2,688	DSLR System	RGB/Mesh		Age, Gender		Posed Expressions
FaceVerse-Coarse [99]	2022		60,000	Structured Light	RGB/Mesh		Age, Gender		Posed Expressions

geodesic distances between facial locations and simplifies rendering for viewing. Numerous techniques can be used to create a mesh consisting of triangles, quadrilaterals, or other simple convex polygons from a point cloud, with the power crust algorithm standing out as the most effective. Furthermore, Dharavath et al. [162] describe a technique for constructing a regular facial mesh model based on the scattered point cloud.

The exploration of 3D facial surface acquisition began approximately two decades ago, marking a significant milestone in computer vision research. One of the earliest data sets in 3D facial scanning is MPIBC [6]. This data set was collected using a CyberWare scanning system. It includes seven views of 200 laser-scanned faces taken with different poses. Among the pioneering data sets in this field, the 3DRMA data set [7] stands out as one of the earliest endeavors to capture and represent human facial shapes as point clouds. Employing structured light technology, this data set was meticulously compiled from 120 individuals, each posing twice in front of the scanning system. Another noteworthy contribution, the GavabDB data set [13], emerged during the same era and was captured using the Minolta VI-700 device. The FRAV3D [14] data set is also captured using a Minolta VI-700 3D laser light-stripe triangulation range-finder, which provides a polygonal 3D mesh model. It contains around 1696 images from 106 subjects. Every face was scanned several times. Frontal views were preferred, although little turns were allowed in the acquisition process. Due to these changes in the face pose, normalization has to be done [14].

It is important to note that, given the nascent stage of 3D face scanning technologies at that time, the data collected often exhibited various imperfections, including artifacts, noise, and missing regions. Consequently, extensive preprocessing became a necessity to rectify these imperfections and attain satisfactory results [163]. The landscape of 3D face scanning evolved with the introduction of more advanced scanners such as the Vivid 910 3D, which is renowned for its superior scanning capabilities, subsequently leading to the creation of several high-quality data sets. Notable data sets generated using this advanced device include CASIA-3D [11], ND-Collection-D [10], ND-2006 [95], 3D-TEC [87], and UMB-DB [88].

The 3dMD system stands out as one of the most widely used devices for acquiring 3D human facial data. Early data sets like BU-3DFE [15] were among the first to be captured using this system, and it continues to be the preferred choice for generating 3D data sets. This system is notably

TABLE 6: RGB-Depth data sets.

Data Set	Year	Subjects (M/F)	Samples	Camera	Resolution		Demographics	Application
					RGB	Depth		
ND-Collection-D [10]	2003	275	953	Minolta Vivid 900	640×480	640×480		Matching, Recognition
FSU [119]	2003	37	222	Minolta Vivid 700	242×347	242×347		Recognition
ETH [154]	2008	26	10,545	Structured Light Camera	640×480	150×200	Gender	Pose Estimation
Photoface [28]	2011	261 (227/34)	7,356	Photometric Stereo	$1,280 \times 1,024$		Gender	Recognition, Verification
3DFACE-XMU [117]	2011	15	118	Stereo Vision				Recognition
RGB-D Face (VAP) [69]	2012	31	1,581	Kinect	$1,280 \times 960$	640×480		Detection, Recognition
ICT-3DHP [98]	2012		10	Kinect v1	640×480			Pose Estimation
3DMAD [34], [35]	2013	17 (10/7)	255	VIS/Kinect v1	640×480	640×480	Gender	Anti-Spoofing
BIWI [97]	2013	20	15,000	Kinect	640×480	640×480		Pose Estimation
HRRFaceD [83]	2014	18	22	Kinect v2	512×424			Detection, Recognition
FEEDB ver.1 [155]	2014	50	1,650	Kinect	640×480		Age, Gender	Expressions
FEEDB ver.2 [156]	2014	50	1,550	Kinect	640×480		Age, Gender	Expressions
RGBDFaces [110]	2014	28		Kinect				Recognition, Reconstruction
Lock3DFace [43]	2015	509 (377/122)	5,711	Kinect	$1,920 \times 1,080$	512×424	Age, Gender	Detection, Recognition
UWA-Kinect [157]	2016	48	15,000	Kinect				Recognition
Pandora [47]	2017	22 (10/12)	110	Kinect v1	$1,920 \times 1,080$	512×424	Gender	Pose Estimation
MotorMark [107]	2017	35	30,000	Kinect	$1,280 \times 720$	515×424		Pose Estimation
IST-EURECOM LFFD [46]	2017	100	4,000	Lytro ILLUM			Age, Gender, Ethnicity	Detection, Recognition, Expressions
Cui et al. [100]	2018	747	845,000	RealSense II				Detection, Recognition
SeetaFace [106]	2018	747	845,000	RealSense II				Identification
Sun et al. [108]	2018	35	142,10	HD Dual Camera	$1,280 \times 720$			Anti-Spoofing
Intellifusion [103], [104]	2019	1,205	403,068					Reconstruction
Li et al. [105]	2019	15	9,800	Kinect			Age, Gender	Reconstruction
Florence3D-Re-Id [84]	2019	16	2,471	Kinect v2				Identification
Face-Emotion [112]	2020	69	1,000	Kinect	150×110			Expressions
Guo et al. [158]	2021		800	Iphone X	480×640			Reconstruction
IKFDB [113]	2021			Kinect				Reconstruction
CAS(ME)^3 [92]	2022	216	4,950	RealSense	$1,280 \times 720$		Age, Gender	Macro/Micro Expressions

employed in the creation of data sets such as MeIn3D [164], HeadSpace [57], UHDB31 [91], and D3DFACS [30]. The primary reason for its widespread use is the innovative "hybrid" stereo vision approach integrated into its systems. This cutting-edge technique seamlessly combines both active and passive stereo vision triangulation strategies, resulting in exceptionally advanced 3D imaging outputs [165].

In contrast, the Di3D and Di4D systems rely solely on passive stereo vision, an advanced imaging technique that captures 3D data without the need for structured light or lasers. Passive stereo vision involves capturing multiple images of an object from various angles and utilizing the disparities between these images to calculate the 3D coordinates of points on the object's surface. The BU-4DFE [17] and ESRC3D [49] data sets were acquired using the Di3D system, while the Di4D system was employed for the 4DFAB [51] data set.

Initially, most 3D data sets were primarily designed to address issues related to face recognition [23], [111] and static facial expression recognition [15], [120]. However, in recent years, the applications of 3D data sets have expanded to encompass more advanced applications, including face verification [51] and dynamic spontaneous facial expression analysis [17], [31]. Furthermore, several 3D data sets have been collected specifically to facilitate 3D statistical morphable model analysis and 3D face reconstruction from 2D data, such as HeadSpace [57], MeIn3D [45], KF-ITW [153], and FaceScape [116]. These data sets include a greater number of 3D scans collected from a more extensive range of subjects, encompassing various ages, genders, and ethnic backgrounds.

C. DEPTH (RGB-D) DATA

In recent years, there have been significant advancements in technology, particularly in the field of computer vision. One notable development is the increased availability and affordability of Red, Green, Blue, and Depth (RGB-D) sensors. These sensors are capable of capturing both color and depth information from the environment. Unlike traditional RGB sensors, which only capture color information, RGB-D sensors like those found in devices such as the Microsoft Kinect [67] and Intel RealSense [73] offer an additional dimension of data representing the depth. This information represents the distance from the sensor to various points on the subject's face, creating a three-dimensional representation of the facial structure. This added dimensionality is what sets RGB-D face recognition apart and contributes to its superior accuracy.

The key advantage of RGB-D face analysis lies in its ability to leverage spatial features. By incorporating depth data, algorithms can discern not only the colors and textures of facial features but also their positions in three-dimensional space. This spatial awareness enables more accurate and robust analysis, as it accounts for variations in pose, lighting conditions, and even the presence of occlusions, such as eyeglasses or facial hair. This additional depth information helps

TABLE 7: RGB-Depth-3D Data Sets.

Data Set	Year	Subjects (M/F)	Samples	Camera	Modalities	Resolution RGB	Resolution Depth	Resolution 3D	Demographics	Application
FRGCv2 [12]	2004	466 (266/200)	50,000	Minolta Vivid 910	RGB/Depth/Mesh	$1,704 \times 2,272$ $1,200 \times 1,600$	640×480		Age, Gender, Ethnicity	Detection, Recognition
FRAV3D [14]	2005	106 (79/27)	1,696	Minolta Vivid 700	RGB/Depth/Mesh	242×347		10,000 vertices, 15,000 triangles	Age, Gender	Verification
ND-2006 [95]	2006	888	13,450	Minolta Vivid 910	RGB/Depth/PointCloud	640×480	640×480	112,000 vertices	Age, Gender	Detection, Recognition, Expression
SuperFaces [89]	2012	50	50	3dMD/Kinect	RGB/Depth/Mesh	$3,341 \times 2,027$		40,000 vertices, 80,000 triangles	Age, Gender	Detection, Superresolution
IAS-Lab [109]	2013	45	315	Kinect	RGB/Depth/PointCloud	$1,920 \times 1,080$	960×540			Recognition
CurtainFaces [70], [90]	2013	52	4,784	Kinect	RGB/Depth/PointCloud	128×128			Age, Gender	Detection, Recognition, Verification
FaceWarehouse [71]	2014	150	3,000	Kinect	RGB/Depth/Mesh	640×480			Age, Gender	Detection, Recognition
UHDB11 [91]	2014	23	1,625	3dMD	RGB/Depth/Mesh	$3,888 \times 2,592$			Age, Gender	Detection, Recognition
4DFAB [51]	2018	180 (120/60)	1,800,000	DI4D/Kinect	RGB/Depth/Mesh	640×480	640×480	60k-75k vertices	Age, Gender, Ethnicity	Expressions
3DWF [102]	2019	92		Asus Xtion	RGB/Depth/PointCloud			Age, Gender		Recognition

TABLE 8: RGB-Thermal Data Sets.

Data Set	Year	Subjects (M/F)	Samples	Camera	Modalities	Resolution RGB/Gray	Resolution Thermal	Demographics	Wavelength	Application
ND-Collection-C [9]	2002	240		VIS/Merlin	Gray/LWIR	$1,200 \times 1,600$	320×240		7,000–14,000nm	Recognition
NIST/Equinox [139]	2004	90	3,244	VIS/NIR	RGB/LWIR	320×240	320×240		8,000–12,000nm	Posed Expressions
IRIS [133]	2006	32	4,228	Raytheon PalmIR Pro	RGB/LWIR	320×240	320×240			Posed Expressions
UH [134]	2007	138	7,590	VIS/Flir	MWIR		640×512	Age, Gender, Ethnicity	3,000–5,000nm	Recognition
CBSR-NIR [140]	2007	197	3,940	VIS/NIR	RGB/NIR	640×480	640×480		780–1,100nm	Recognition
TIV [166]	2009	20	21,676	RaytheonL-3 T hermal-Eye2000AS	LWIR		320×240			Recognition
PolyU-NIRFD [132]	2010	350	35,000	NIR LED, JAI camera	RGB/NIR			Gender	780–850nm	Recognition, Verification
NVIE [26]	2010	215 (157/58)	436	VIS/SAT-HY6850	Mono/LWIR	704×480, 320×240		Age, Gender		Spontaneous and Posed Expressions
Carl [27]	2010	41 (32/9)	7,380	VIS/TESTO880-3	RGB/NIR	640×480	160×120		820–1,000nm	Recognition
NFRAD-DB [29]	2011	50 (37/13)	600	VIS/DSLR	RGB/NIR	$3,872 \times 2,592$	$3,872 \times 2,592$	Gender	810–960nm	Recognition
SMIC [33]	2013	20 (14/6)	306	VIS/NIR/HS	RGB/NIR	640×480	640×480	Age, Gender		Micro Expressions
I^2BVSD [37]	2013	75	681	VIS/LWIR	RGB/LWIR	$4,288 \times 2,848$	720×576	Age, Gender		Verification
CASIA NIR-VIS [32]	2013	725	17,580	VIS/ENIR	RGB/NIR	640×480	640×480	Gender		Matching, Recognition
LDHF-DB [36]	2013	100 (70/30)	1,600	VIS/RayMax300	RGB/NIR	$5,184 \times 3,456$		Gender	850nm	Matching, Recognition
Sober Drunk [38], [39]	2013	41 (31/10)	4,100	FLIR A10	LWIR	-	128×160		750–1300nm	Pose Estimation
KTFE [41]	2014	26 (16/10)		VIS/NECR300	RGB/NIR/LWIR	320×240	336×256		8,000–14,000nm	Spontaneous Expressions
ND-NIVL [44]	2015	574	341	Nikon D90/ CanonEOS50D	RGB/NIR	$4,770 \times 3,177$ $4,288 \times 2,848$				Recognition
Liu et al. [42]	2015	77	181	VIS/FLIR	RGB/LWIR	640×480		Ethnicity	8,000–14,000nm	Spontaneous Expressions
ARL-MMFD1 [135]	2016	60	960	VIS/Polaris Sensor	RGB/LWIR	640×480	640×480		7,500–11,100nm	Recognition
Msspoof [126]	2016	21	4,704	VIS/NIR	RGB/NIR			Age, Gender		Anti-Spoofing
SWIR [127]	2016	5	141	VIS/M-SWIR	RGB/SWIR			Age, Gender		Anti-Spoofing
BRSU [128]	2016	50+	660	VIS/AM-SWIR	RGB/SWIR			Age, Gender		Anti-Spoofing
MLFP [130]	2017	10	1,350	VIS/NIR/LWIR	RGB/NIR			Age, Gender		Anti-Spoofing
CIGIT-PPM [131]	2019	72	93,358	VIS/NIR	RGB/NIR			Age, Gender		Anti-Spoofing
Eurocom [53]	2018	50	4,200	VIS/FLIR Duo R	RGB/LWIR	160×120	160×120	Age, Gender, Ethnicity	7,500–13,500nm	Recognition, Verification, Expressions
PUCV-DTF [54]	2018	46 (40/6)	11,500	FLIR TAU 2	LWIR	-	640×480	Age, Gender	7,500–13,500nm	Pose Estimation, Expressions
ARL-MMFD2 [136]	2019	111	111	VIS/Polaris Sensor	RGB/LWIR	640×480	640×480		7,500–11,200nm	Reconstruction, Synthesis
RWTH [141]	2019	90	10,000	Infratec HD820	LWIR		$1,024 \times 768$			Expressions
ARL-VTF [167]	2021	395	500,000	VIS/FLIR Boson/ FLIR Grasshopper3	Mono/RGB/LWIR	658×492	640×512		7,500–13,500nm	Verification
Sejong-A [59]	2021	30 (16/14)	1,500	VIS/Pi NoIR/ Therm-App	RGB/SWIR/NIR	$4,032 \times 3,024$	$1,680 \times 1,050$, 768×756	Gender, Ethnicity	700-1,000nm, 750-1,400nm	Recognition, 3D Morphable Models
Sejong-B [59]	2021	70 (44/26)	23,000	VIS/Pi NoIR/ Therm-App	RGB/SWIR/NIR	$4,032 \times 3,024$	$1,680 \times 1,050$, 768×756	Gender, Ethnicity	700-1,000nm, 750-1,400nm	Recognition, 3D Morphable Models
SpeakingFaces [138]	2021	142	4,581,595	VIS/FLIR T540	RGB/LWIR	$1,920 \times 1,080$	464×348			Expressions
TFW [137]	2022	147	9,982	FLIR T540	LWIR	-	464×348		7,500–14,000nm	Detection
UNCC-ThermalFace [60]	2022	10 (5/5)	10,000	Flir A700	LWIR		Multiple			Recognition

address some of the challenges faced by traditional RGB-based face analysis systems, such as variations in lighting conditions, pose, and occlusions [168]. In Table 6, we list some of the RGB-D data sets that are publicly available. Some of the RGB-D data sets provide 3D mesh or point cloud data, as summarized in Table 7.

Depth sensors capture the 3D structure of the face. Genuine faces exhibit depth variations caused by the facial features, including the nose, eyes, and mouth. Depth information allows the system to verify the presence of these natural 3D features. Therefore, depth information is crucial in face verification and anti-spoofing applications, because depth sensors can differentiate between the texture of a printed image or a screen display and the actual 3D contours of a face. While a high-quality image might fool a purely texture-based system, depth information reveals the absence of true facial structure. The 3DMAD [34], [35] data set was one of the earliest to capture depth information for face anti-spoofing applications. The depth information was collected from 17 individuals in three different sessions within two weeks. The data set contains mask attacks in order to assess the spoofing performance of 3D masks against RGB and depth information. Similarly, the Intel RealSense device was used to capture depth information in the CASIA-SURF [79] data set for face anti-spoofing. The data set includes several attacks, such as printing and face features cutting. Depth information is also used in several other data sets such as WMCA [74], MMFD [114], HQ-WMCA [93], and CASIA-SURF CeFA [94].

The integration of depth information proves highly bene-

ficial across various applications, with particular significance in the domain of pose estimation analysis. A case in point is the Biwi data set [97], specially designed for head pose estimation. Comprising RGB-D data captured with a Kinect sensor, this data set offers comprehensive head pose annotations for every frame. Its utility extends to tasks like face detection and head pose estimation. Similarly, the ICT-3DHP data set [98], collected for pose estimation applications, leverages Microsoft Kinect's depth-sensing capabilities. Distinguished by its inclusion of uncontrolled pose variations, this data set broadens the scope of pose analysis. Notably, the Pandora data set [47] stands as a recent addition to the pose estimation data sets. It includes contribution from 22 individuals who perform diverse poses and occlusions, mirroring real-life scenarios. Jiang et al. [169] construct a large-scale RGB-D face data set including more than 100k identities, mainly in frontal pose, and a relatively small RGB-D data set with 952 identities in various poses. Collectively, these data sets emphasize the role of depth information for pose estimation applications.

Facial expression analysis is improved by depth information integration. The RGB-D Face (VAP) [69] face data set is one of the common data sets collected using a Kinect sensor. It includes facial images of various individuals under different lighting conditions, poses, and expressions. Curtin-Faces [70], [90] is captured using a Microsoft Kinect Sensor. A total of over 5,000 samples were captured from 52 subjects, including a mix of male/female and with and without glasses. These images have varying facial expressions, viewpoints, illumination, and occlusion, simulating a real-world, uncontrolled face recognition problem. KinectFaceDB [40] offers a range of expressions, poses, and occlusions that can be utilized to develop robust face verification and recognition algorithms. FaceWarehouse [71] is a data set of RGB-D facial expressions for visual computing applications captured with a Kinect RGB-D camera. It is composed of 3000 facial images from 150 individuals, aged from 7 to 80, of various ethnic backgrounds with neutral expressions and 19 other actions, such as mouth opening, smile, etc. The Lock3DFace data set [43] contains 5,711 RGB-D face videos from 509 subjects with variations in facial expression, pose, occlusion, and time-lapses. It provides a standard evaluation protocol with the aforementioned four variations. The CAS(ME)^3 [92] data set provides around 80 hours of videos, including 1,109 manually labeled micro-expression and 3,490 macro-expressions. It also provides depth information as an additional modality and elicits micro-expression with high ecological validity using stimuli following the mock crime paradigm as well as physiological and voice signals.

Similarly, the depth information can be utilized for face reconstruction and recognition applications. The IIITD RGB-D data set [72] consists of 106 male and female subjects with multiple RGB-D images of each subject. All the images are captured using a Microsoft Kinect sensor. Since the images are unsegmented, the data set can be used for both face detection and recognition in RGB-D space. The HRRFaceD [83] data set consists of high-resolution depth images captured using the Microsoft Kinect v2 device. This data set includes facial images from 18 individuals captured in various poses, including frontal and lateral views. Furthermore, the data set comprises facial images of certain individuals both with and without glasses. Guo et al. [158] use the depth information provided by an Iphone X sensor to capture 800 samples for 3D face reconstruction. zhang et al. [106] collected an RGB-D data set for face recognition using RealSense II as opposed to the Kinect sensor. The data set comprises approximately 845,000 RGB-D images featuring 747 subjects. The images exhibit consistent variations in pose and only minor alterations in lighting conditions.

D. THERMAL INFRARED IMAGING

The field of facial analysis has gained significant attention, particularly with the use of various imaging technologies, such as IR imaging sensors [41]. The human body is responsive to electromagnetic wavelengths that are not visible to the naked eye. In this imaging technology, special cameras equipped with infrared sensors capture thermal radiation within the range of 0.7-14.0 μm, which falls within the infrared spectrum. This differs from traditional visual cameras, which capture electromagnetic energy in the visible spectrum range of 0.4-0.7 μm. Two key factors influence the amount of radiation released: the temperature of the material and its emissivity, which is a measure of how efficiently it emits radiation. However, creating images in certain portions of the thermal IR spectrum can be quite challenging. Specifically, there are significant limitations in imaging within the strong atmospheric absorption bands in the wavelength range of 2.4-3.0 μm between the SWIR and MWIR regions, and in the range of 5.0-8.0 μm between the MWIR and LWIR spectrum. The human face and torso emit both the MWIR and LWIR bands within the thermal IR spectrum. Thermal infrared cameras can detect changes in facial temperature from a distance and produce 2D images known as thermograms. Notably, the LWIR band is preferred for facial recognition within the thermal IR spectrum, due to the considerably higher emissions in this band compared to the other bands [170].

The human face is a valuable biometric feature that can be used in security systems for the purpose of person identification and verification. However, in the context of thermal face verification, there are specific challenges and considerations that need to be addressed. One of the primary challenges in face verification is to accurately match the input face with a stored face image of the same person already present in the system's database. This process involves complex algorithms that analyze facial features and patterns to make a positive identification. In the case of thermal face verification, the methods focus on analyzing facial thermograms, which are representations of the heat patterns emitted by the face [21]. In the context of thermal face verification, there is a need to represent a thermal face image using biometric features that not only capture the unique thermal characteristics of the face but are also compact and suitable for use in classification

TABLE 9: RGB-Depth-Thermal Data Sets.

Data Set	Year	Subjects (M/F)	Samples	Camera	Modalities	Resolution		Demographics	Wavelength	Applications
						RGB	Thermal			
CASIA HFB [21]	2009	100 (57/43)	992	VIS/ENIR, Minolta vivid 910	RGB/IR/Depth	640 × 480	640 × 480	Gender	700-880nm	Matching, Recognition
KinectFaceDB [40]	2014	52 (38/14)		Kinect	RGB/IR/PointCloud	256 × 256		Age, Gender, Ethnicity		Detection, Recognition
RGB-D-T [115]	2014	51	45,900	Kinect, AXIS Q1922	RGB/IR/Depth	640 × 480	384 × 288			Recognition
CASIA-SURF [79]	2018	1,000	21,000	RealSense	RGB/IR/Depth			Age, Gender		Anti-Spoofing
KaspaAROV [82]	2018	108	831	Kinect v1/v2	RGB/IR/Depth	640 × 480, 1,920 × 1,080, 320 × 240,	512 × 424			Detection, Recognition
WMCA [74]	2019	72	6,716	RealSense/STC-PRO	RGB/IR/Depth			Age, Gender		Anti-Spoofing
MMFD [114]	2019	15	43,853	RealSense II	RGB/IR/Depth	1,280 × 720	640 × 480			Anti-Spoofing
HQ-WMCA [93]	2020	51	2,904	Basler acA1921-150uc, Intel RealSense D415.	RGB/IR/Depth			Age, Gender		Anti-Spoofing
MAVFER [56]	2020	17 (7/10)	17	Kinect v2 and FLIR A65	RGB/LWIR/Depth	1408 × 792	640 × 512	Age, Gender	7,500-13,000nm	Expressions
Tuft [58]	2020	113 (39/74)	10,000	VIS, LYTRO ILLUM 40, FLIR Vue Pro	RGB/LWIR/PointCloud	336 × 256	336 × 256	Age, Gender, Ethnicitiy		Recognition
CASIA-SURF CeFA [94]	2021	1,607	23,538	RealSense	RGB/IR/Depth			Age, Gender		Anti-Spoofing

TABLE 10: Multispectral/Hyperspectral Datasets

Data Set	Year	Subjects (M/F)	Samples	Camera	Resolution	Demographics	Wavelength	Bands	Step Size	Applications
CMU-HSFD [8]	2002	45	147	Spectro-polarimetric	640 × 480		450–1,100nm	65	10nm	Recognition
IRIS-M3 [16]	2006	82 (62/20)	2,624	VIS/CRI's VariSpec/ Raytheon Palm-IR-Pro	2,272 × 1,704 640 × 480	Age, Gender, Ethnicity	480–720nm	25	10nm	Matching, Recognition
PolyU-HSFD [24]	2010	25 (17/8)	300	CRI's VariSpec LCTF			400–720nm	33	10nm	Recognition
UWA-HSFD [142]	2015	70	120	CRI's VariSpec LCTF			400–720nm		10nm	Recognition
EMSPAD [129]	2017	50	14,000	SpectraCam		Age, Gender		7		Anti-Spoofing
UL-FMTV [52]	2018	238 (86/48)		VIS/Jenoptik/ FLIR Phoenix Indigo IR/Goodrish	640 × 512	Age, Gender, Ethnicity	8,000–14,000nm 3,000–5,000nm 900–1,700nm 750–1,100nm	4		Pose Estimation, Expressions
SDFD [55]	2018	54 (54)	6480	RGB/NIR			530–1,000nm	8		Recognition

algorithms [145]. Unlike visible-spectrum images, thermal face images reveal different details about the face, primarily related to the heat patterns on the skin's surface. Therefore, it is more difficult for attackers to spoof the system with printed photos or screens displaying facial images.

Numerous publicly available thermal imaging data sets have been developed for applications in face verification and identification. Some of these data sets are shown in Table 8. The ND-Collection-C data set [9], as one of the early contributions in this domain, was created with a specific focus on such applications. It was captured employing a Merlin-Uncooled camera in 2002, yielding 2,492 frontal long-wave infrared (LWIR) thermal images sourced from 241 individuals. In 2007, the CSBR-NIR [140] data set was primarily designed to achieve illumination-invariant face verification. This data set encompasses a total of 3,940 near-infrared (NIR) facial images, featuring 197 individuals. These images are organized into two distinct sets: a gallery set and a probe set. Within the gallery set, each individual is represented by eight images, while the probe set comprises a comprehensive set of twelve images for each subject. Similarly, in the same year, the University of Houston [134] data set was created to assess the impact of physiological information on face recognition, specifically focusing on the permanency of innate characteristics beneath the skin. The data set was captured during multiple sessions, with a six-month time gap between sessions, introducing variations in poses and facial expressions. Fast-forwarding to 2010, the PolyU-NIRFD [132] data set was collected using a custom-designed camera, significantly augmenting the volume of available face images in comparison to previous data sets. This extensive data set includes 35,000 thermal facial images gathered from 350 subjects. It spans a diverse range of poses, expressions, and scales, enhancing its utility for research purposes.

Subsequently, several other data sets, including NFRAD-DB [29], LDHF-DB [36], ND-NIVL [44], ARL-MMFD1 [135], SDFD [55], and UNCC-ThermalFace [60] have been made accessible to the research community. Each of these data sets captures thermal imaging across different wave bands, encompasses multiple illumination variations, and introduces variability in facial expressions and poses. These data sets collectively facilitate advancements in thermal face verification by offering diverse and comprehensive resources for researchers. They cater to the exploration of thermal imaging in different contexts, thereby contributing to the ongoing progress in this field.

The need to develop robust face verification and recognition systems increases as face spoofing attacks evolve. The CIGIT-PPM [131] data set includes VIS and NIR face image pairs of real access and attack attempts from 72 subjects, comprised of 61 live persons and 11 masks. The BVSD [37] data set contains images capturing both visible and thermal spectra from 75 individuals with varying disguises. Each participant has multiple images, ranging from 6 to 10, including at least one neutral face image and several images with different disguises. The data set consists of 681 images for each spectrum, with visible images taken using a Nikon D-90 camera and thermal images captured with a thermal camera. SWIR [127] is the first data set of corresponding SWIR and RGB color images incorporating various types of masks and facial disguises.

Combining thermal infrared and depth sensors for facial scanning and analysis offers several advantages. These include improved robustness in varying lighting conditions, the ability to analyze facial expressions and emotions, enhanced security through liveness detection, improved face detection, privacy benefits, adaptability to different environments, and a range of applications from health monitoring to gaming. In essence, the synergy between thermal and depth data enhances the accuracy, versatility, and reliability of facial analysis systems. Table 9 summarize data sets that combine both thermal infrared and depth information.

E. MULTISPECTRAL/HYPERSPECTRAL IMAGING

Multispectral and hyperspectral imaging are advanced imaging techniques used in various fields, including facial analysis, to capture and analyze the spectral information of an object or scene beyond what is visible to the human eye. They involve capturing images at different wavelengths across the electromagnetic spectrum. These techniques provide a wealth of data that can be valuable for a range of applications, including facial analysis and recognition. It can provide additional information about the face beyond what is captured by RGB images alone. Multispectral imaging involves capturing data from multiple discrete spectral bands, typically spanning a range of wavelengths beyond the visible spectrum (e.g., ultraviolet and infrared) [171]. In contrast, hyperspectral imaging is more advanced than multispectral imaging, involving the capture of data in many narrow and contiguous spectral bands, providing highly detailed spectral information [172].

There are several multispectral and hyperspectral data sets designed for several facial analysis applications, as presented in Table 10. The Multispectral Latex Mask-based Video Face Presentation attack (MLFP) data set combines thermal and visible videos with and without wearing face masks for ten individuals. Videos are captured in indoor and outdoor environments. The data set contains 1,350 videos, of which 1,200 videos are of faces wearing masks and 150 videos are of faces without masks [130]. The Multispectral-Spoof data set (MSSPOOF) [126] contains paired images of VIS and NIR modality, which are captured under various environments. It covers genuine face images, printed VIS, and NIR images. BRSU [128] consists of 130 participants and combines spectral measurements at several points on faces and limbs with pictures taken using both an RGB camera and the presented multispectral camera system. The Extended Multispectral Presentation Attack Face Dataset, EMSPAD [129], comprises face scans of 50 subject collected by a multispectral camera for both the evaluation of presentation attack detection and the analysis of face presentation attack vulnerability.

VI. APPLICATIONS

Facial multimodal data sets have a wide range of applications across different domains. These data sets offer a holistic view of individuals' facial features and characteristics, allowing for more comprehensive insights and enhancing the accuracy of facial analysis. There are several data sets that are mainly designed for face detection, recognition, and verification applications. A selection of these data sets are summarized in Table 11. Similarly, there are several data sets that are mainly designed to address problems related to facial expression and pose estimation applications. We present some of these data sets in Table 13. Data sets that are designed for face anti-spoofing applications are shown in Table 12. In the following, we discuss several key applications of facial analysis multimodal data sets.

A. FACE DETECTION

Face detection is a crucial research topic in computer vision, focusing on developing algorithms and techniques to identify human faces within images or video frames. It serves as a critical pre-processing step for various face-related applications, including face recognition, facial expression analysis, age estimation, and more. Early face detection methods relied on traditional computer vision techniques. These methods often involved analyzing image features such as edges, color, or textures to identify potential face regions [173]–[175]. Other solutions relied on feature-based face detection, by identifying specific facial features, such as mouth, nose, and eyes, leveraging their geometry to identify faces. These methods can be effective but are sensitive to variations in pose, lighting conditions, and occlusions [176]. The last decade has seen an increasing interest in machine learning approaches that significantly advanced face detection research. In particular, deep learning has revolutionized the field by automatically learning discriminative features from raw image data [177], [178].

Variations in illumination, pose, and occlusion are examples of challenges that face detection algorithms have to handle to improve robustness and generalization. The collection of a balanced data set that leverages these challenges is one of the main research objectives in face detection. Data sets play a crucial role in the development of research in this area. Driven by new multimodel data sets and deep learning advances, research in face detection continues to evolve rapidly [179]. There are several data sets that researchers use for face detection. For instance, Mian et al. [180] employed the FRGC v2 [12] data set to develop a 3D face detection approach. Similarly, Pamplona Segundo et al. [181] used 3D data sets like BU-3DFE [15], Bosphorus [18], Texas 3D [23], and RGB-D Face [69] to evaluate a real-time 3D facial detection system. Furthermore, there is the TFW [137] data set, which was created specifically for detecting faces in thermal images. This data set includes manually marked boxes around faces and precise locations of five key facial points: the centers of the eyes, the tip of the nose, and the corners of the mouth. Adding these facial points to the face detection process serves as an extra guide, significantly improving detection accuracy, especially in situations where facial images are complex. The NVIE [26] data set is utilized by Basbrain et al. [182] for face detection in thermal imaging.

TABLE 11: Face Detection, Recognition, Verification, and Reconstruction Data Sets.

Data Set	Year	Subjects (M/F)	Modalities	Landmarks	Expressions	Poses	Occlusions	Distance	Illumination	Applications
MPIBC [6]	1996	200 (100/100)	RGB/Mesh			3				Reconstruction, Recognition
3DRMA [7]	1998	120 (106/14)	PointCloud			3				Recognition, Verification
ND-Collection-C [9]	2002	240	Gray/LWIR		2				3 Settings	Recognition
ND-Collection-D [10]	2003	275	RGB/Depth					150cm		Matching, Recognition
FSU [119]	2003	37	RGB/Depth		6					Recognition
FRGCv2 [12]	2004	466 (266/200)	RGB/Depth/Mesh		2			15m	2 Settings	Detection, Recognition
GavabDB [13]	2004	61 (45/16)	RGB/PointCloud		3	6				Detection, Recognition
FRAV3D [14]	2005	106 (79/27)	RGB/Depth/Mesh		Variable	Variable				Verification, Recognition
UoY [96]	2006	350	RGB/Depth/Mesh		5	3				Detection, Recognition
IRIS-M3 [16]	2006	82 (62/20)	Multispectral		1	25		120cm	3 Settings	Recognition
UH [134]	2007	138	MWIR		5	5	Variable			Recognition
CBSR-NIR [140]	2007	197	RGB/NIR	4				50–100cm	Variable	Recognition
Bechman [120]	2007		RGB/Mesh	65						Recognition
CASIA HFB [21]	2009	100 (57/43)	RGB/NIR/Depth	Eye Coordinates	2		1	80–120cm	2 Settings	Matching, Recognition
TIV [166]	2009	20	LWIR			3	2		2	Recognition
Texas 3D [23]	2010	118	RGB/PointCloud	25	5					Detection, Recognition
PolyU-HSFD [24]	2010	25 (17/8)	Multispectral							Recognition
PolyU-NIRFD [132]	2010	350	RGB/NIR		Variable	Variable		80–120cm		Recognition, Verification
Carl [27]	2010	41 (32/9)	RGB/Thermal		1	1	1	135cm	3	Recognition
3D-TEC [87]	2011	214	RGB/PointCloud		1					Detection, Recognition
UMB-DB [88]	2011	143	RGB/PointCloud		3	1	3			Detection, Recognition
Photoface [28]	2011	261 (227/34)	2D/Depth/Albedo	11	5				4 Settings	Recognition, Verification
MICC (Florence2D/3D) [111]	2011	53	RGB/Mesh					Variable	2 Settings	Recognition
NFRAD-DB [29]	2011	50 (37/13)	RGB/NIR		Neutral	2		1–50m	4 Settings	Recognition
3DFACE-XMU [117]	2011	15	RGB/Depth							Recognition
RGB-D Face (VAP) [69]	2012	31	RGB/Depth		4	17		85cm	Variable	Detection, Recognition
CurtainFaces [70], [90]	2013	52	RGB/Depth/ PointCloud		7	7	1		5	Detection, Recognition, Verification
IIIT-D RGB-D [72]	2013	106	RGB/PointCloud		Variable	Variable	Variable			Detection, Recognition
IAS-Lab [109]	2013	45	RGB/Depth/ PointCloud		2	2		2	2	Recognition
CASIA NIR-VIS [32]	2013	725	RGB/NIR	Eye Coordinates	Variable	Variable	1	Variable		Matching, Recognition
LDHF-DB [36]	2013	100 (70/30)	RGB/NIR					1–150m	2 Settings	Matching, Recognition
I2BVSD [37]	2013	75 (60/15)	RGB/LWIR		-		9			Verification
KinectFaceDB [40]	2014	52 (38/14)	RGB/IR/ PointCloud	Annotations	9	Variable	5		Variable	Detection, Recognition
FaceWarehouse [71]	2014	150	RGB/Depth/Mesh		20					Detection, Recognition
UHDB11 [91]	2014	23	RGB/Depth/Mesh		Variable	12			6	Detection, Recognition
HRRFaceD [83]	2014	18	RGB/Depth			Variable	1			Detection, Recognition
RGBDFaces [110]	2014	28	RGB/Depth			11		3		Recognition, Reconstruction
RGB-D-T [115]	2014	51	RGB/Depth/ Thermal		5	7		1.3	6	Recognition
Lock3DFace [43]	2015	509 (377/122)	RGB/Depth	Annotations	6	2	1		Variable	Detection, Recognition
ND-NIVL [44]	2015	574	RGB/NIR				6	152–213cm	Indoor	Recognition
UWA-Kinect [157]	2016	48	RGB/Depth		Variable	Variable				Recognition
ARL-MMFD1 [135]	2016	60	RGB/LWIR	6	1	1		2.5–7.5m	1	Recognition
IST-EURECOM LFFD [46]	2017	100	RGB/Depth	5	3	6	6		2	Recognition
KF-ITW [153]	2017	17	RGB/Mesh	Annotations	2				Variable	3D Morphable Models
UHDB31 [48]	2017	77 (53/24)	RGB/Mesh	12					3	Recognition
KaspaAROV [82]	2018	108	RGB/Depth/NIR		Variable	Variable			Variable	Detection, Recognition
Cui et al. [100]	2018	747	RGB/Depth			Variable			Variable	Detection, Recognition
SeetaFace [106]	2018	747	RGB/Depth			Variable			Yes	Identification
FIDENTIS [50]	2018	2,476 (1,154/1,322)	RGB/Mesh	42						Recognition, Reconstruction
SDFD [55]	2018	54	RGB/NIR			1	3			Recognition
3DWF [102]	2019	92	RGB/Depth/ PointCloud	Annotations		10				Recognition
Intellifusion [103], [104]	2019	1,205	RGB/Depth							Detection, Recognition
Li et al. [105]	2019	15	RGB/Depth		6					Recognition
Florence3D-Re-Id [84]	2019	16	RGB/Depth	Annotations		Variable	Variable	Variable		Identification
ARL-MMFD2 [136]	2019	111	RGB/LWIR	6	1	1		2.5m		Synthesis
FaceScape [116]	2020	938	RGB/Mesh		20					3D Morphable Models
HeadSpace [57]	2020	1,519	RGB/Mesh	23						3D Morphable Models
Tuft [58]	2020	113 (39/74)	RGB/NIR/LWIR/ PointCloud		5	9		150cm	2	Recognition
Guo et al. [158]	2021	-	RGB/Depth		Variable				Variable	Reconstruction
ARL-VTF [167]	2021	395	Mono/RGB/LWIR	6	2	3	1	210cm	1	Verification
Sejong-A [59]	2021	30 (16/14)	RGB/SWIR/NIR			1	13	200cm	2	Verification
Sejong-B [59]	2021	70 (44/26)	RGB/SWIR/NIR			15	13	200cm	2	Verification
TFW [137]	2022	147	Thermal	9		Variable	Variable	Variable	3	Detection
UNCC-ThermalFace [60]	2022	10 (5/5)	LWIR	72		25		Variable		Recognition

B. FACE ALIGNMENT

Facial alignment applications have witnessed significant advancements in recent years, largely owing to innovations in imaging technologies. The primary goal of facial alignment is to accurately locate and align key facial landmarks or features, such as eyes, nose, and mouth, within an image or

TABLE 12: Anti-Spoofing Data Sets.

Data Set	Year	Subjects	Samples	Camera	Attack Type					
					Print	Elec. Screen	Cut	2D Mask	3D Mask	Replay
3DMAD [34], [35]	2013	17	255	VIS/Kinect			✗		✗	
Msspoof [126]	2016	21	4,704	VIS/NIR	✗					
SWIR [127]	2016	5	141	VIS/M-SWIR	✗				✗	
BRSU [128]	2016	≥50		VIS/AM-SWIR	✗				✗	
EMSPAD [129]	2017	50	14,000	SpectraCam		✗				
MLFP [130]	2017	10	1,350	VIS/NIR/Thermal				✗		
CASIA-SURF [79]	2018	1,000	21,000	RealSense	✗		✗			
Sun et al. [108]	2018	35	14,210	HD dual camera	✗					✗
CIGIT-PPM [131]	2019	72	93,358	VIS/NIR	✗					
MMFD [114]	2019	15	43,853	RealSense II	✗		✗			✗
WMCA [74]	2019	72	6,716	RealSense/STC-PRO		✗		✗		✗
HQ-WMCA [93]	2020	51	2,904	RealSense	✗	✗	✗			✗
CASIA-SURF CeFA [94]	2021	1,607	23,538	RealSense	✗			✗		✗

TABLE 13: Facial Expressions and Pose Estimation Data Sets.

Data Set	Year	Subjects (M/F)	Modalities	Landmarks	Emotions	Poses	Occlusions	Distance	Illuminations	Applications
CASIA-3D FaceV1 [11]	2004	123	RGB/Mesh		3	5	1		5	Posed Expressions
NIST/Equinox [139]	2004	90	RGB/Thermal		3				3 Settings	Posed Expressions
BU-3DFE [15]	2006	100	RGB/Mesh		6	2				Action Unit, Static Posed Expressions
ND-2006 [95]	2006	888	RGB/Depth/PointCloud		5					Posed Expressions
ZJU-3DFED [118]	2006	40	RGB/Mesh		4					Posed Expressions
IRIS [133]	2006	32	RGB/LWIR		3	11		183cm	5 Settings	Posed Expressions
Bosphorus [18]	2008	105 (60/45)	RGB/PointCloud	24	34	13	4	150cm		Posed Expressions
BU-4DFE [17]	2008	101	RGB/Mesh		6	Variable				Dynamic Expressions
ETH [154]	2008	26	RGB/Depth			Variable				Pose Estimation
ADSIP [22]	2009	10 (2/8)	RGB/Mesh		7			100cm		Posed Expressions
B3D(AC) [25]	2010	14 (6/8)	RGB/Mesh		11	Variable				Dynamic Expressions
NVIE [26]	2010	215 (157/58)	Mono/LWIR		8			75cm	3 Settings	Spontaneous and Posed Expressions
D3DFACS [30]	2011	10 (4/6)	RGB/Mesh	47	6					Dynamic Expression, Action Unit
ICT-3DHP [98]	2012		RGB/Depth	Annotations		Variable				Pose Estimation
SMIC [33]	2013	20 (14/6)	RGB/NIR		3					Spontaneous Expressions
BIWI [97]	2013	20	RGB/Depth			Variable				Pose Estimation
Sober Drunk [38], [39]	2013	41 (31/10)	LWIR		2					Pose Estimation, Posed Expressions
FEEDB ver.1 [155]	2014	50	RGB/Depth		33					Posed Expressions
FEEDB ver.2 [156]	2014	50	RGB/Depth	22	33					Posed Expressions
KTFE [41]	2014	26 (16/10)	RGB/NIR/LWIR		6		1	85cm	7	Spontaneous Expressions
Liu et al. [42]	2015	77	RGB/LWIR		8			75cm		Spontaneous Expressions
Pandora [47]	2017	22 (10/12)	RGB/Depth	Annotations		Variable	5		Indoor	Pose Estimation
MotorMark [107]	2017	35	RGB/Depth	68					Controlled	Pose Estimation
ESRC3D [49]	2018	99 (45/54)	RGB/Mesh		7	4			7	Posed Expressions
4DFAB [51]	2018	180 (120/60)	RGB/Depth/3D	79	6		2	140cm	Variable	Posed Expressions
UL-FMTV [52]	2018	238 (86/48)	Multispectral		7	4		100cm	7 Settings	Pose Estimation, Posed Expressions
PUCV-DTF [54]	2018	46 (40/6)	Thermal	22	4					Pose Estimation, Posed Expressions
Eurocom [53]	2018	50	RGB/Thermal		7	4	6	150cm		Posed Expressions
RWTH [141]	2019	90	LWIR	68	8	9		90cm	2 Settings	Posed Expressions
Face-Emotion [112]	2020	69	RGB/Depth		6					Posed Expressions
MAVFER [56]	2020	17 (7/10)	RGB/Depth/LWIR	Annotations	2					Posed Expressions
SpeakingFaces [138]	2021	142	RGB/LWIR					100cm		Posed Expressions
CAS(ME)^3 [92]	2022	216	RGB/Depth		7					Macro/Micro Expressions
FaceVerse-Detailed [99]	2022	128	RGB/Mesh		21					Posed Expressions

video frame. These applications find relevance in a variety of domains, including computer vision, healthcare, augmented reality, and biometrics. The precision of facial alignment is essential for tasks such as face recognition, expression analysis, 3D face reconstruction, and facial feature tracking. 3D geometry, depth, and thermal imaging have significantly elevated the precision and adaptability of facial alignment applications. These modalities empower algorithms to accurately locate facial landmarks, even in challenging conditions, making facial alignment more robust and versatile across various domains.

The BIWI [97] data set contains RGB-D data of human faces captured using the Microsoft Kinect sensor. This data set is proposed mainly for face alignment across large poses. It includes RGB images, depth maps, and skeleton information [183]. The 3D Face Alignment in the Wild (3DFAW) [184] data set is widely used for face alignment. It contains an annotated corpus of over 23,000 multi-view images from a wide range of conditions, captured in both controlled and in-the-wild settings. The data set includes images from MultiPIE and BP4D [185] as well as images collected from the Internet. All images were annotated in a consistent way with 66 3D facial points. The Florence data set [111] is used by Guo et al. [186] for 3D dense face alignment. The BU-4DFE [17] and BP4D [185] data sets are used by Jeni et al. [187] for dense 3D face alignment from

2D videos in real time.

C. FACE REGISTRATION

Face registration is the process of aligning or registering multiple facial images or 3D face models into a common coordinate system or reference frame. The goal is to ensure that all faces are in a consistent pose, scale, and orientation, making it easier to compare, analyze, or combine them for various applications [188].

Face registration is particularly important when working with a database of facial images or 3D face models, where faces may vary in pose, expression, or illumination. It helps bring these faces into a canonical form such that subsequent processing or comparisons can be performed accurately [189].

The relationship between face alignment and face registration lies in the fact that face alignment is often an integral step in the face registration process. Before registering faces, it is common to perform face alignment on each individual face to ensure that the facial landmarks are correctly positioned. These landmarks can then be used as reference points during the registration process to align and normalize the faces. Gerig et al. introduced the BFM-2017 [152] data set for evaluating face registration algorithms. They used the data set to develop a pipeline for face registration based on Gaussian processes. The FRGC v1 [12] data set was used by Tena et al. [190] for 3D Dense Registration. Ayyagari et al. [188] utilized the IRIS 3D data set for face registration. Ma et al. [191] used the IRIS [133] data set to evaluate non-rigid registration of visible and infrared face images.

D. FACE RECOGNITION

Face recognition goes beyond face verification and aims to identify or recognize individuals from a set of known identities. It involves comparing a given face image against a database or a gallery of known face images and determining the most likely identity for the input face.

The process of face recognition typically involves the following steps: localization of faces in the input image, alignment of the detected face to a standard pose or configuration, extracting discriminative features from the face image, comparing the feature representation of the input face against a database of known face features, and identifying the best match with the highest similarity score. Face recognition has numerous security, surveillance, biometrics, and human-computer interaction applications.

Solutions for face recognition include feature extraction-based approaches and deep learning-based approaches. Algorithms developed based on feature extraction methods aim to extract robust and discriminative features that can capture the unique characteristics of each person. Traditional methods utilize handcrafted features such as Local Binary Patterns (LBP) [192], Histogram of Oriented Gradients (HOG) [193], or Eigenfaces citeSirovich:1987:Eigenface. On the other hand, deep learning techniques, particularly Convolutional Neural Networks (CNNs), have shown remarkable performance in automatically learning highly effective feature representations from raw face images. Various deep learning architectures have been employed to enhance face recognition performance [1]. Models like FaceNet [194], VGGFace [195], and DeepFace [196] have achieved state-of-the-art results on benchmark data sets.

Face recognition algorithms often face challenges in handling variations in pose, illumination, expression, and occlusion. Researchers have explored domain adaptation techniques to improve the generalization capability of models across different environments or data sets. Several benchmark data sets are widely used for evaluating and comparing the performance of face recognition algorithms in RGB images. Examples include the Labeled Faces in the Wild (LFW), MegaFace, CelebA, MS-Celeb-1M, and IARPA Janus data sets. These data sets provide standardized protocols, labeled identities, and a diverse range of face images to facilitate the fair evaluation of algorithms. However, RGB images are highly sensitive to variations in lighting conditions. Changes in ambient lighting, shadows, or harsh illumination can significantly affect the appearance of a face, making it challenging for a face recognition system to perform consistently in diverse lighting environments. Therefore, interest in using thermal, depth, and 3D images for face recognition is rapidly increasing during the last few years. Numerous data sets have been created to provide input data for face recognition algorithms. Some of the early 3D data sets that were mainly collected for face recognition include FRGCv2 [12], Bosphorus [18], UoY [96], BJUT-3D [19], Texas 3D [23], and Photoface [28]. These data sets include between 100 and 500 subjects. The FRGCv2 and BJUT-3D data sets include more than 40,000 samples whereas UoY [96], Texas 3D [23], Bosphorus [18], and Photoface [28] provide less than 10,000 samples. The FIDENTIS [50] is a 3D data set recently released with more than 2,400 subjects accompanied by fundamental demographic and descriptive information. This data set is organized based on individual subjects and contains both single-scan entries and a smaller subset of multi-scan entries. The multi-scan entries vary in terms of the time elapsed between recording sessions and the types of 3D data capture devices used. This data set is considered one of the largest 3D data set available for face recognition.

Depth data plays a crucial role in face recognition applications and several RGB-D data sets are commonly used in this context. Examples include FEEDB [156], Lock3DFac [43], and HRRFaceD [83]. In addition, the UHDB31 [48] data set was introduced specifically to allow researchers to evaluate the impact of pose, illumination, and resolution on their face recognition algorithms. Despite its relatively small number of subjects, UHDB31 provides challenging data samples for face recognition due to its wide range of poses and diverse lighting conditions. The data set meticulously distributes its data samples across 21 different poses and three distinct illuminations. In the study by Jiang et al. [103], the Intellifusion RGB-D data set is employed for face recognition. This

data set includes RGB-D images of 1,205 individuals, each represented by multiple images, resulting in a comprehensive data set comprising a total of 403,068 images that include both RGB and depth data.

Over the past decade, there has been a growing focus on face recognition using thermal imaging data sets, leading to the collection of numerous data sets to serve as input for training recognition models. The UH data set, as described in Buddharaju et al. [134], offers thermal scans for 138 subjects and grants access to over 7,000 MWIR images. The ND-NIVL [44] data set comprises NIR scans of more than 500 subjects captured indoors. Similarly, the CIGIT-PPM [131] data set encompasses over 93,000 NIR scans of 72 subjects, all acquired under controlled illumination conditions. The Eurocom [53] data set provides scans of 50 subjects in various poses, with occlusions and different lighting settings. Additionally, there are data sets that combine thermal and depth images for face recognition, such as RGB-D-T [115], KinectFaceDB [40], and Tuft [58].

E. FACE VERIFICATION

Face verification, also known as face authentication, aims to verify whether two face images belong to the same person or not. In face verification, the system compares the facial features extracted from two images and determines whether they represent the same person or different individuals. The goal is to determine if there is a match or a mismatch between the faces [197]. The process of face verification typically involves the following steps: locating and extracting faces from the input images, aligning the detected faces to a standard pose or configuration, extracting distinctive features from the aligned faces, comparing the feature representations of the two faces, and making a binary decision (match or non-match) based on a predefined threshold or similarity metric.

There are several types of multimodal data sets that are specifically collected for training and evaluating face verification algorithms. FRAV3D [14] is a multimodal data set used by Conde et al. [198] for face verification. The data set is collected over a ten-month period involving 105 volunteers. All of the participants fall within the young adult age range (18–35 years), are of Caucasian ethnicity. The data set also exhibits a gender bias, with 81 males and 24 females included. McCool et al. [199] use the FRGC v2 data set for face verification by dividing the 3D face into separate parts. The same data set is later used by McCool et al. [200] for 3D face verification using feature distribution modeling techniques. Križajet al. also use the FRGC v2 data set to evaluate a 3D face verification approach developed using Gaussian mixture models. Ouamane et al [201] introduce an innovative method for face verification, wherein they represent 2D and 3D face images as a high-dimensional tensor. They evaluate the proposed approach using FRGC v2 [12], Bosphorus [18], and CASIA 3D [11]. Yu et al. [202] use both the FRGC v2 and Bosphorus data sets to develop a 3D face verification approach using sparse ICP With resampling and denoising.

Depth data represents a crucial input for improving face verification approaches. Lin et al. [203] use RGB-D input data to develop a deep learning approach for face verification where the III-D RGB-D face data set [72] is used for training the proposed model. This data set was generated using a Kinect sensor and comprises RGB-D images of 106 individuals. The EUROCOM [53] and CurtinFaces [70] RGB-D data sets are also used for face verification evaluation experiments conducted by Xu et al. [204].

Face verification in thermal images has received great interest during the last few years. The ARL [135] and Tuft [58] data sets are utilized by Di et al. [205] to develop a multi-scale visible to thermal face verification approach using attribute-guided synthesis. Peri et al. [206] collected a data set, named MILAB-VTF(B), for face verification evaluation. The data set consists of matched thermal and visible video recordings. This data set includes data from 400 subjects of indoor and long-range outdoor thermal-visible face imagery.

F. FACE ANTI-SPOOFING

Face verification technology has wide applications in serving as a reliable means of authenticating individuals based on their facial features. However, conventional face verification systems, while highly effective, often lack the ability to distinguish between genuine faces and spoofed face attacks. These spoof attacks can take various forms, including the use of printed photos, digital images, masks meticulously crafted to resemble live faces (presentation attacks), and videos of faces (replay attacks). The vulnerability of traditional face verification systems to such spoof attacks poses a significant security risk. If these systems cannot differentiate between real individuals and fraudulent attempts, unauthorized access can occur, potentially leading to security breaches or compromised authentication processes. This is where face anti-spoofing solutions become a demand.

Face anti-spoofing is a critical and evolving field within biometrics and computer vision. Its primary objective is to discern between genuine, live faces and spoofed face attacks, effectively identifying and thwarting fraudulent attempts. In essence, anti-spoofing measures act as a protective layer, ensuring that only legitimate access is granted and that spoofed faces fail the verification process. Implementing effective face anti-spoofing techniques is crucial for improving face verification systems' overall safety and reliability. By bolstering the system's ability to detect and reject spoof attacks, it not only safeguards against security breaches but also enhances user trust and confidence in the technology. This is particularly essential in applications where security is paramount, such as access control, secure banking transactions, and identity verification in critical infrastructure settings [207]. Therefore, many data sets are specifically designed for face anti-spoofing applications.

Face-antispoofing data sets are collected using different visible light scanning technologies such as mobile and conventional webcams. CASIA-FASD [208] (CASIA Face Anti-Spoofing Dataset) contains 1,000 genuine face images and 4,000 spoofing face images from 50 subjects. The spoofing

attacks include printed photos, mobile phone displays, iPad displays, and computer displays. The Spoof in the Wild (SiW) [209] data set includes real and spoof face images captured with mobile devices. It contains various spoofing attacks, including printed photos, replay attacks, and masks. The mobile face spoof (FSD) [210] data set is a face spoof data set, created using the cameras of a laptop (MacBook Air3) and a mobile phone (Google Nexus 54) and three types of attack medium (iPad, iPhone, and printed photo). The Replay-Mobile [211] data set consists of short video recordings of both real-access and attack attempts of 40 different identities. Each video is approximately 10 seconds long (300 frames at 30 fps), and is captured at HD resolution ($720 \times 1,280$). The Oulu-NPU [212] data set is designed explicitly for face anti-spoofing. It consists of both real access attempts and spoofing attacks using various materials. The Replay-Attack [213] data set contains videos of real access attempts and spoofing attacks performed using various materials and techniques.

The evolution of new advanced scanning technologies such as depth and thermal imaging has enabled new robust methods for face anti-spoofing. Many data sets have been created to provide input data for these methods. The NUAA Imposter [149] data set is a face anti-spoofing data set created by researchers at the Nanjing University of Aeronautics and Astronautics (NUAA). This data set is designed for the purpose of developing and evaluating face liveness detection methods. The 3DMAD data set contains 76,500 frames of 17 different users, recorded using a Microsoft Kinect sensor for both real-access and spoofing attacks using 3D facial masks [34], [35]. The Wide Multi-Channel presentation Attack (WMCA) data set includes genuine faces and seven categories of attack samples. Each data sample contains images of four modalities: VIS, NIR, thermal, and depth [74]. The CASIA-SURF data set consists of 1,000 subjects and 21,000 video clips with 3 modalities (RGB, Depth, IR). It has six types of photo attacks involving multiple operations, e.g., cropping, bending the print paper, and stand-off distance [79]. In the CASIA-SURF CeFA data set, the Intel RealSense is used to capture the RGB, Depth, and IR videos simultaneously at 30 fps and a resolution of 1280×720 pixels for each frame in the video. Subjects are asked to move their head smoothly so as to have a maximum of around 30° deviation of head pose in relation to frontal view [94], [214]. The GUC-LiFFAD data set is a new face artifact data set collected using LFC. It comprises 80 subjects. Face artifacts are generated by simulating two widely used attacks, such as photo print and electronic screen attacks [215]. The 3D Mask [216] is a recent 3D mask anti-spoofing data set with more variations to simulate the real-world scenario. This data set contains 12 masks from two companies with different appearance qualities. Seven cameras from stationary and mobile devices and six lighting settings that cover typical illumination conditions are included. Therefore, each subject contains 42 (seven cameras $\times$ six lighting conditions) genuine and 42 mask sequences, totaling 1,008 videos. The silicone mask attack data set SMAD [217] comprises 130 videos, including 65 real samples and 65 silicone-masked samples. The attack samples in SMAD wear vivid silicone masks that fit well with holes in the eyes and mouth regions. Some silicone masks also have hair, mustaches, and beards for life-like impressions.

G. FACIAL EXPRESSIONS AND EMOTIONS DETECTION

Human communication is profoundly shaped by facial expressions—a rich range of emotional cues that convey human thoughts and feelings. Understanding these expressions has been a pursuit of great significance, both in psychology and technology. In the domain of facial expression recognition, researchers and engineers have developed various modalities to dissect and comprehend this intricate non-verbal language [163].

Among these modalities the dynamic, static, posed, action units, and spontaneous forms of expression recognition each offer a distinct perspective and set of challenges [33]. Static expression recognition relies on a single image frame to capture facial expressions. It aims to identify emotions by scrutinizing the facial configuration at a specific moment in time. Static recognition is vital in scenarios where a snapshot of emotional state suffices, such as security systems and still image analysis. Similarly, posed expression recognition is Often used in controlled environments. Posed expression recognition involves individuals intentionally mimicking specific emotions. This modality serves as a foundation for understanding basic emotional archetypes, laying the groundwork for more comprehensive analyses. On the other hand, spontaneous expression recognition, the most challenging and truest reflection of human emotion, involves detecting emotions as they naturally occur without prompting or preparation. This modality seeks to unveil unscripted, authentic emotional responses, providing valuable insights in fields like psychology, market research, and human-computer interaction. The early research in facial expression recognition mainly focused on posed static expressions. Several data sets were created to meet the requirements of this kind of facial analysis. The NIST/Equinox [139] data set was released in 2004. It contains three distinct facial expressions of 90 subjects captured using thermal imaging. The FRGCv2 [12] data set, which was released in the same year, contains 3D scanning of 466 subjects. The latter data set includes a variety of posed static facial expressions. The ND-2006 [95] data set contains 3D scanning of 888 subjects performing five facial expressions. The BU-3DFE [15] data set is one of the most common multi-modal data sets that provide 3D face models with seven expressions: happiness, disgust, fear, anger, surprise, sadness, and neutral, with different levels of intensity. There are 100 subjects, of which 56 are male and 44 are female. The majority of subjects were undergraduates of various ethnicities. For each subject, there are 25 face models. The facial expressions and emotions data set (FEEDB) [155] consists of 1,650 recordings of 50 persons posing for 33 different facial expressions and emotions. The

second version of FEEDB consists of 1,550 recordings of 50 persons recorded in two separate video streams, separately for RGB and depth channels.

In contrast, dynamic expression recognition captures the temporal evolution of facial expressions, providing insight into the progression of emotions. Dynamic recognition is particularly valuable in applications that demand real-time emotion tracking, such as human-computer interaction and emotion-aware technology [42]. Several facial expression data sets provide a rich source of input data that can be used to excel research in this field. The MMI [218] data set contains recordings of facial expressions performed by multiple individuals. It includes video sequences and high-quality 3D facial scans. The data set provides annotations for different expressions, intensity levels, and onset and apex frames of expressions. The FRGCv2 [12] data set combines audio and dense 3D facial deformations of effective communication. It contains images from 466 subjects collected in 4,007 scans with two facial expression variances. The BU-4DFE data set, as described in Yin et al. [219], comprises 606 sequences of facial expressions obtained from 101 individuals. In each sequence, one of the six fundamental facial expressions is demonstrated, beginning with a neutral expression, reaching the apex of the expression, and then returning to neutrality. The data set includes seven frames captured around the moment of the most intense expression, and these frames are associated with the corresponding sequence labels. This arrangement results in a total of 4,272 images (101 individuals $\times$ 6 expressions $\times$ 7 frames). The B3D(AC) or ETH-3DAV data set [25] is a collection of high-quality, realistic 3D facial scans. The scans were obtained using a 3D scanner while individuals pronounced a set of 40 predetermined sentences under both neutral and deliberately induced emotional conditions.

Spontaneous expression refers to genuine and uncontrolled emotional reactions or facial expressions that occur naturally in response to one's emotions, thoughts, or immediate surroundings. These expressions are not consciously planned, rehearsed, or posed but instead emerge instinctively in reaction to a particular stimulus, situation, or internal emotional state. The Binghamton-Pittsburgh 4D spontaneous expression data set (BP4D) [17], [185] provides RGB-D data for facial expression analysis. It includes spontaneous 3D facial expressions captured using a Di3D dynamic face-capturing system. The Multimodal Spontaneous Emotion data set (MMSE/BP4D+) [220] is an extension of the BP4D data set, which provides RGB-D data for spontaneous facial expression analysis, captured using RGB and depth sensors.

Similarly, action unit recognition processes deeper intricacies of facial expressions. Action unit recognition breaks down the face into its constituent movements. It dissects the nuanced muscular changes underlying expressions, allowing for a granular understanding of emotional subtleties. This modality finds applications in psychology, clinical assessment, and animation. The Dynamic 3D Facial Action Coding System (FACS) data set (D3DFACS) [30] presents the first dynamic 3D FACS data set for facial expression research, portraying ten subjects performing between 19 and 97 different AUs both individually and in combination. Abbasnejad et al. [221] creates different, synthetic action units and expressions to generate a large-scale synthetic facial expression data set geared towards training neural networks. The 3D Relightable Facial Expressions (ICT-3DRFE) [222] data set comprises RGB-D data captured by a structured light scanner. It includes high-resolution 3D face scans, RGB images, and depth maps of human faces expressing various emotions. It contains 3D models for 23 subjects and 15 expressions, as well as photometric information that allows for photo-realistic rendering.

H. POSE ESTIMATION

Face pose estimation, also known as facial pose estimation or head pose estimation, is a computer vision task that involves determining the orientation or pose of a person's face in a three-dimensional space relative to a reference coordinate system. This estimation typically includes the angles representing the pitch (up-down tilt), yaw (left-right rotation), and roll (sideways tilt) of the face. The result is often expressed as a set of angles or a transformation matrix that describes the face's position and orientation. The accurate estimation across a full range of head poses is challenging since faces can exhibit a wide range of poses. Faces may also appear at different scales and resolutions in images or videos, making it challenging to detect and track facial landmarks accurately. Moreover, variations in lighting conditions, in addition to partial occlusions, increase the complexity of estimating face poses. To address these challenges, several multimodal data sets are designed with high variability in illumination, poses, and scales.

Depth data offers significant advantages, particularly in the field of pose estimation analysis. A notable example is the Biwi data set [97], which has been specifically created for head pose estimation. This data set incorporates RGB-D information captured through a Kinect sensor, providing extensive annotations for head poses in each frame. Likewise, the ICT-3DHP data set [98] is designed for pose estimation purposes using depth information. What sets it apart is its inclusion of uncontrolled variations in poses, significantly expanding the scope of pose analysis. The Pandora data set [47] is a pose estimation data set with a diverse range of poses and occlusions, mimicking real-life scenarios.

I. 3D MORPHABLE MODELS(3DMM)

A 3D morphable face model represents facial shape and appearance as a generative model. It establishes dense point-to-point correspondence across all faces through a registration process on a set of example faces. This correspondence enables the meaningful combination of faces in a linear manner, resulting in the creation of morphologically realistic faces. It also involves the separation of facial shape and color by eliminating factors from external variables like illumination and camera parameters. 3D morphable models are statistical

models that capture the inherent variability and structure of 3D facial shapes within a population. They are often built using a large data set of 3D facial scans or models. Most of the 3D data sets that are publicly available are commonly used as input to train 3D morphable models [223]. However, few data sets are particularly proposed for training 3DMMs. FaceScape [116] is a large-scale data set that includes high-quality 3D scanning of 938 subjects with a variety of expressions, which makes it a perfect fit for training 3DMMs. It employs a multi-view 3D reconstruction system to acquire the initial mesh models using 68 DSLR cameras, with 30 of them dedicated to capturing high-resolution images of the front side, while the remaining cameras capture images of the side part. Geriag et al. [152] introduced the Basel face model data set, known as BFM-2017, which incorporates facial expressions and age distribution. This data set was mainly created for processing non-rigid registration of faces, which is a crucial step for designing 3DMMs. Headspace [57] is also mainly designed for generating 3DMMs. This data set stands as the first publicly available data set that provides both the shape and texture components for the entire human head. Booth et al. [45] collected a large-scale 3D data set composed of 10,000 high-quality 3D facial scans to develop 3DMM. As of the time of writing and to the best of our knowledge, this data set stands as the most expansive 3D data set in terms of the number of individuals it includes.

VII. ANALYSIS

All findings in this section are based on the data sets presented in this work. Comparisons between different data sets could yield different results.

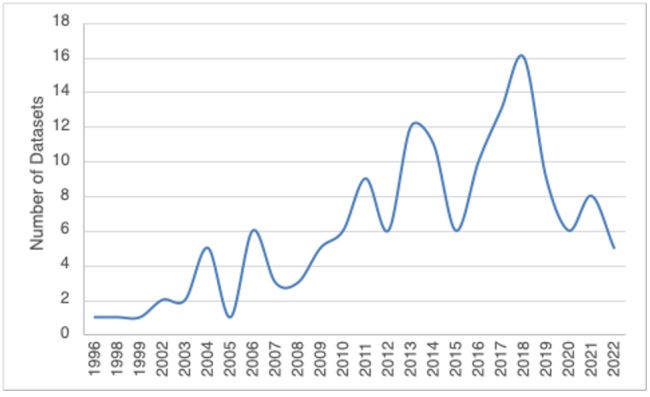

FIGURE 2: Overview of researchers' interest in collecting facial data sets over time.

The data sets presented in this review span the time between 1996 and 2022. The graph in Figure 2 illustrates the trend in the number of collected data sets over time. Before 2010, the maximum number of collected data sets was approximately six. However, after 2010, there was a significant surge in the number of generated data sets, establishing six papers as the new annual minimum. We attribute this in part to the Kinect v1, which was released by Microsoft in 2010 and marked the advent of consumer-grade (thus, affordable) depth sensors.

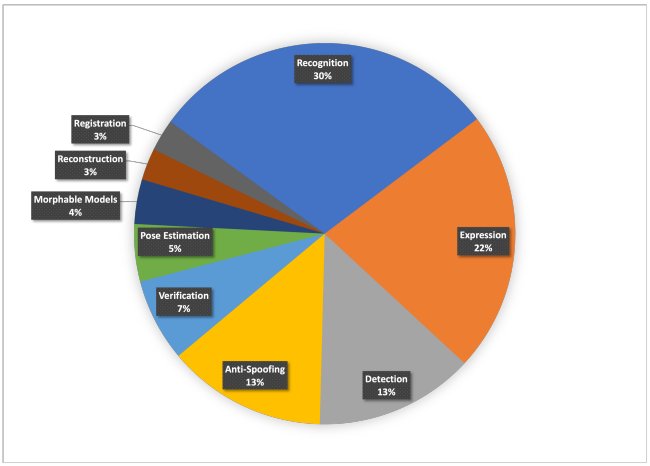

FIGURE 3: Distribution of data sets acoss tasks.

We also examined the types of tasks for which these data sets were collected. However, it is crucial to recognize that the number of data sets alone does not necessarily indicate the level of advancement in the field. Ideally, we would compare data set sizes based on the number of samples. Yet, due to variations in the sample types collected by different researchers, we opted to compare data set sizes in Figure 3 based on the number of subjects.

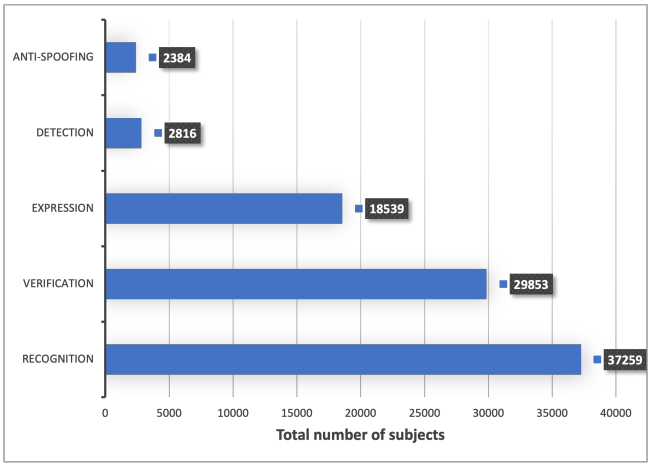

FIGURE 4: Total number of subjects aggregated across multiple data sets per task.

Surprisingly, data sets for verification tasks were the least numerous compared to the top four tasks. However, as shown in Figure 4, these data sets contributed substantially to the verification research field by encompassing a significant number of subjects and, consequently, samples.

We further analyzed data sets collected for the top five tasks over time. Figure 5 reveals that the initial interest in collecting facial data sets was primarily for recognition tasks, with verification tasks emerging around 2003. Detection and expression tasks followed suit in 2004, while anti-spoofing

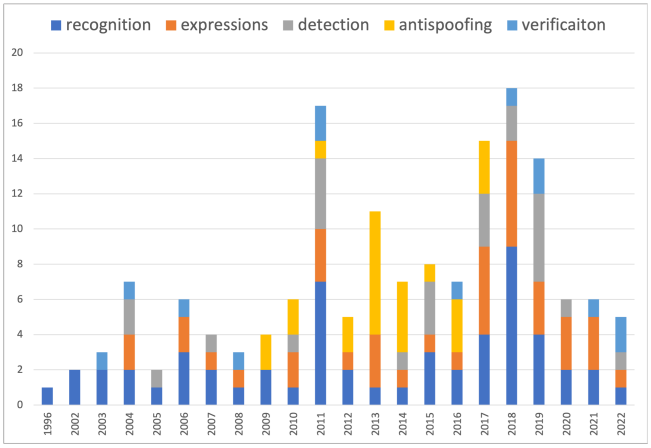

FIGURE 5: Trends in data collection for different tasks over time.

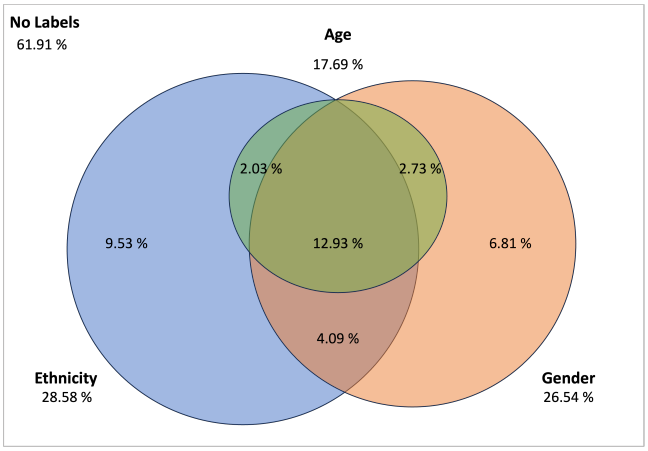

FIGURE 6: Statistics on labeling participant information across data sets

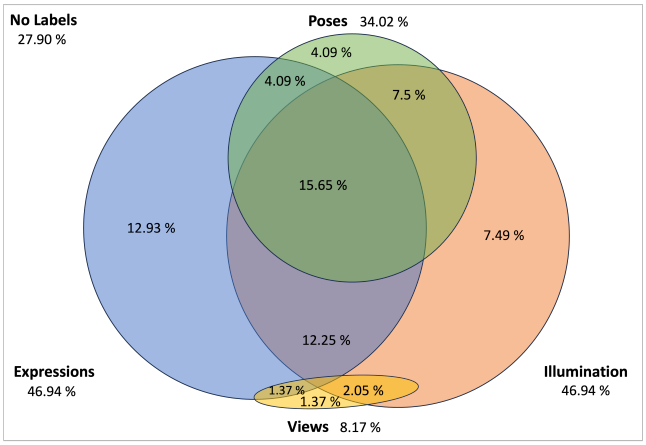

FIGURE 7: Statistics on labeling experimental settings across data sets.

data sets made their appearance in 2009. Notably, data sets for recognition and expressions have continued to be generated consistently, while anti-spoofing data sets have not been as prevalent in the last five years. This timeline roughly coincides with, but pre-dates, face biometrics becoming a consumer-level feature in 2011 (FaceUnlock, introduced with the Android 4 OS), sparking a wider interest in protecting digital devices from spoofing attacks.

It is evident that the focus on different tasks reached peaks at different times. For instance, anti-spoofing peaked in 2013, while data sets for recognition and expressions were most prolific in 2018, and for detection in 2019. However, data sets specifically designed for verification tasks did not gain as much traction over time, possibly due to the ease of using recognition-focused data sets for verification. In this context we would also like to note that face verification may not need as much data as other tasks. The reason is that the verification process is, essentially, an outlier detection or one-class classification problem, in which a face is compared with another face obtained during the registration process. As a result, existing manifold learning models or embeddings that have been trained on data sets designed for different purposes may be used.

Regardless of the intended purpose behind data set generation, 38% of data sets were annotated with participant information. Figure 6 indicates that the participants' age is rarely considered in isolation; it mostly appears in papers that also disclose ethnicity and gender. Among these attributes, ethnicity is the most frequently considered label, even when analyzed separately. Gender is also highly considered, often appearing alongside ethnicity and age attributes. We believe these statistics to be crucial for future research as cross-domain generalization (e.g., training on a limited variety of ethnicities before general deployment) and class imbalances remain a fundamental challenge of modern AI.

In examining experimental settings for participants (Figure 7), we found that 72.1% of the data sets include variations in facial expressions, poses/views, or illumination settings. 49.95% of the data sets take expressions and illumination into consideration. The graph clearly shows no relation between views and poses. The reason behind the previous relation is that views refer to placing sensors at different angles from the subject, while poses refer to changing the participant's viewing angle from the sensor's perspective. Therefore, it is logical to have either one of the two attributes. However, existing data sets prefer poses over views, with a presence of 30% in data sets compared to 8.17%, respectively. The graph also indicates a good size of data sets, 15.65%, that consider all top three attributes at the same time.

In our analysis, we investigated the types of sensors adopted in collecting the 3D data sets. Figure 8 shows that one-third of the 3d data sets are captured using by Microsoft Kinect sensor (36%), followed by the 3dMD system (16%), Intel RealSense (11%), and Minolta Vivid 910 (7%). However, the combined percentage of the three Minolta Vivid versions represents approximately 14% , placing it in the third position after the 3dMD system.

Figure 9 shows that VIS is the oldest type of sensor among the top four. 3dMD and FLIR started to appear in 2006 and

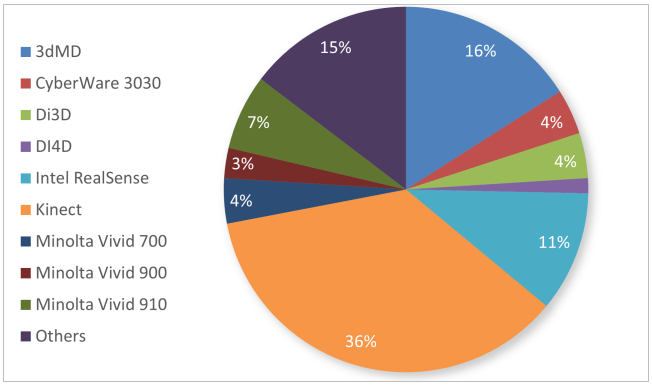

FIGURE 8: Distribution of sensor types across data sets

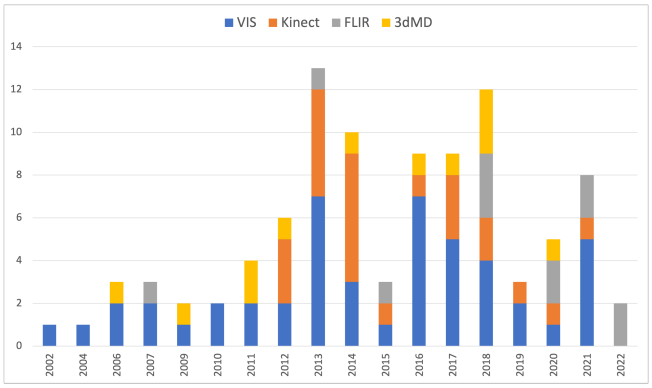

FIGURE 9: Temporal analysis of data set usage of common sensor types.

2007, respectively. Both of them reached their peak in 2018. Despite the fact that the Kinect v1 appeared first in 2010, it attracted researchers from its early stages considering that Microsoft released the non-commercial version of the Kinect SDK only in 2011.

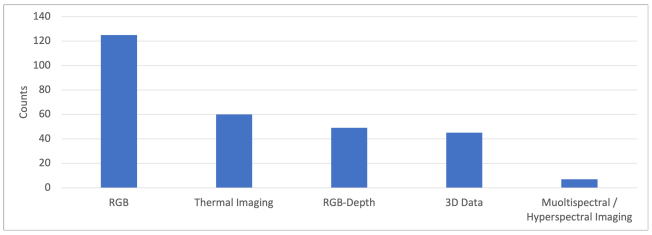

FIGURE 10: Counting and ranking different modality types based on their occurrence

The second part of the experimental settings we are looking into is the type of modalities used in collecting the data sets. Figure 9 shows that RGB ranks first, Which is a justifiable conclusion since the RGB modality is incorporated in most of the multimodal datasets, followed by thermal imaging , followed by RGB-Depth and 3D Data. In Figure 11, thermal imaging is decomposed into thermal and infrared (NIR, SWIR, MWIR, and LWIR). The graph shows that LWIR is favored in facial tasks due to its high emission compared to MWIR, for instance.

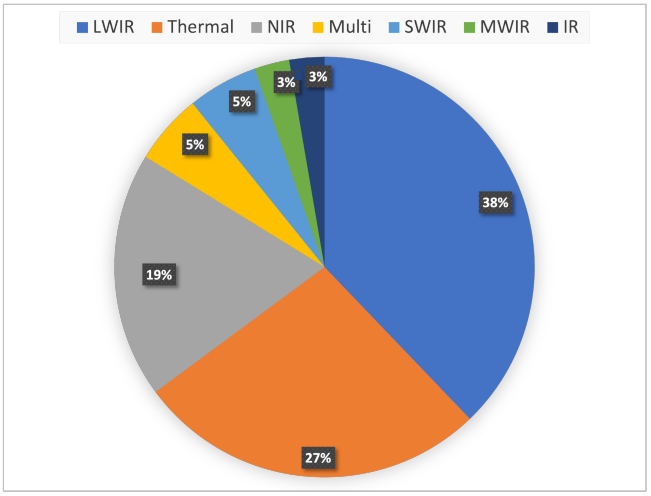

FIGURE 11: The distribution of thermal, infrared, and multi-modal data sets.

VIII. DISCUSSION

Large-scale data sets. Large-scale multimodal data sets are predominantly accessible to well-funded companies and resource-rich institutions. These data sets have the potential to contain a wealth of information. In the context of facial analysis, this means that these data sets provide valuable contextual information for understanding and interpreting facial expressions, identities, or emotions. While there are some large-scale multimodal data sets available, they are relatively small in terms of both the population represented and the modalities they cover. In other words, the available data sets may not capture the full diversity of individuals and the wide variety of data sources that are encountered in the real world. In addition, different cultures may have different reservations regarding taking pictures of faces out of privacy concerns, resulting in yet another source for biases. Despite the presence of a few existing large-scale data sets, there is a continued need to collect more extensive and diverse data sets for facial analysis applications. This is because the effectiveness and fairness of facial analysis algorithms often depend on the diversity of the training data. Collecting data sets with a broader range of input samples, including different ethnicities, ages, genders, and environmental conditions, is crucial to improving the performance and fairness of facial analysis models. The availability of large-scale multimodal data sets allows for more comprehensive analysis. Researchers, auditors, regulators, and policymakers can study the data sets to better understand their capabilities and limitations, identify potential risks, and address any harms associated with the use of facial analysis in various applications, such as surveillance, identity verification, or emotion recognition.

General purpose data sets. In the past, data sets were typically collected with specific goals and tasks in mind. This means that data collection efforts focused on gathering

information directly relevant to solving a particular problem or addressing a specific use case. With the advent of deep learning, there has been a significant shift in how models are trained. Instead of using purpose-specific data sets, the current state-of-the-art involves training large-scale, "general-purpose" AI models. These models are initially trained on vast data sets collected from multiple data sources, which may contain diverse and unfiltered information. These large-scale AI models, such as deep neural networks, can be thought of as compressed representations of the data they are trained on. In essence, they encapsulate the patterns, features, and knowledge present in the massive training data sets. This makes them versatile, as they can be fine-tuned or specialized for various tasks.

Privacy and bias issues. The weights of deep neural networks carry a representation of the training data sets, which raises privacy and bias issues. There is a concern raised about the failure to account for the rights, welfare, and interests of vulnerable individuals and communities. In multimodal facial analysis data sets, this concern could relate to issues like consent for using facial images, potential harm from AI technology misuse, and the impact of biased algorithms on marginalized groups. Variations in age, gender, and ethnicity within facial analysis data sets can introduce bias in the representation of human faces. Bias occurs when certain demographic groups are overrepresented or underrepresented in the data set, leading to inaccurate or unfair results in facial analysis applications. If a data set lacks diversity in terms of ethnicity and primarily includes faces from a single ethnic group, facial analysis algorithms may struggle to perform accurately on faces from underrepresented groups, a problem known as cross-domain generalization. This can lead to misidentification, poorer facial recognition, and inaccurate ethnicity-based analyses.

Facial and environmental variations. The challenges in face analysis data sets persist due to factors such as different facial expressions, poses, lighting conditions, age-related changes, and the use of makeup. To potentially address these issues, both three-dimensional (3D) and infrared (IR) face recognition technologies have been explored. Likewise, image-based color/albedo decompositions have been proposed, both for the purpose of extracting makeup [224] and highlight removal [225] to help multiview image alignment. However, their effectiveness in improving performance regarding these factors has yet to be explored. 3D face scans offer advantages by capturing facial shape information and representing facial geometry, making them less susceptible to variations in lighting and viewpoint changes when compared to (multi-view) 2D images. Nevertheless, they may be sensitive to changes in facial expressions, and the challenge of handling age-related variations is also pertinent in 3D face recognition. However, it is important to note that 3D-based approaches are not without challenges of their own. These include computational complexity, potential issues with precise alignment of 3D scans, and the generation of undesirable artifacts when creating virtual views based on 3D models.

Recent advances in the field of facial analysis have been directed towards making the most of thermal images. These innovations seek to use the unique properties of thermal imaging, especially its capability to capture distinctive patterns of superficial blood vessels on the face. These patterns contain information about a person's physiological information that does not change with time and can be accurately extracted from thermal images. This is quite advantageous because it remains reliable even when the environment or conditions change, making thermal imaging a standout choice compared to other ways of capturing facial information.

When it comes to handling different lighting conditions, NIR imaging has shown great promise in delivering accurate results. NIR images have specific advantages over visible light, especially when dealing with various lighting angles and situations. Because of these advantages, infrared (IR) imaging is now being considered as a promising option for biometric applications (e.g., for face verification on smartphones). This is because it has the capability to illuminate low light scenatios and the potential to provide consistent results for an individual over time and in diverse lighting situations, making it extremely useful in applications where robust and dependable facial recognition is needed.

Multimodal data sets. Assessing which of the three modalities (RGB, 3D, or thermal) is superior for facial analysis depends on the development of algorithms and rigorous evaluations. Visible light imagery is relatively easy to obtain at high quality, making it a reliable choice. Three-dimensional facial data closely mimics the way human vision works, especially when combined with visual texture information. On the other hand, the IR modality, particularly thermal images, can reveal a unique facial vascular network for each individual, which is challenging to alter. Each of these modalities has its own strengths and weaknesses. However, as prior research shows, combining multiple modalities generally leads to better performance than relying on a single modality alone.

Dynamic and static data sets. Dynamic data sets, which involve capturing changes in facial expressions, movements, and temporal variations, offer distinct advantages over static data sets in various contexts within facial analysis and scanning. For example, dynamic data sets are essential for accurately recognizing and analyzing emotions because they capture the temporal evolution of facial expressions. Emotions often involve changes in expression over time, and static images may not convey the full emotional context. Dynamic data sets are also crucial for training and validating facial action recognition systems based on facial action coding system, which involves categorizing facial muscle movements and their temporal patterns. In biometric applications, such as liveness detection or anti-spoofing, dynamic data sets are essential for verifying the authenticity of a person's face. Static images can be easily spoofed, but dynamic analysis of

facial movements can help distinguish real faces from fake ones.

IX. CONCLUSION

In this work, we provided a comprehensive review of existing multimodal face data sets. Our work assumes a data-centric approach, categorizing existing data based on the technology used, the data contents, and the applications. This allows readers to browse through data sets relevant to their work from multiple perspectives. Our findings show that multimodal data sets can boost performance and robustness in many applications. A concern that remains (as with most data sets) are cross-domain generalization problems due to biases in ethnicity, age, and gender, as well as class imbalances that may lead to mispredictions or underrepresentation of minorities. We believe that the latter point, in particular, deserves future research, not only into work of representing minorities accurately but also into societal and infrastructural biases (e.g., the need for funding and recruiting volunteers) to ensure that work based on such data sets remains fair and equitable.

X. DECLARATIONS

Data. The bibfile as well as the tables (in Excel format) can be downloaded from this dropbox link.[1]

Conflices of Interest. The authors do not have any competing interests regarding this work.

Funding. The authors did not receive any external funding for this work.

Guarantor. Jens Schneider.

Contributorship. Kamela Al-Mannai and Khaled Al-Thelaya contributed to the data collection and analysis in equal parts. All authors contributed equally to the design of the methodology, and contributed equally to the writing. Jens Schneider and Spiridon Bakiras guided the research and helped with the analysis.

Acknowledgements. The author's would like to thank their respective academic institution for funding their time to work on this manuscript.

REFERENCES

[1] I. Masi, Y. Wu, T. Hassner, and P. Natarajan, "Deep face recognition: A survey," in 2018 31st SIBGRAPI conference on graphics, patterns and images (SIBGRAPI). IEEE, 2018, pp. 471–478.

[2] Z. Ming, M. Visani, M. M. Luqman, and J.-C. Burie, "A survey on anti-spoofing methods for facial recognition with RGB cameras of generic consumer devices," MDPI Journal of Imaging, vol. 6, p. #129, 2020. [Online]. Available: https://doi.org/10.3390/jimaging6120139

[3] X. Ben, Y. Ren, J. Zhang, S.-J. Wang, K. Kpalma, W. Meng, and Y.-J. Liu, "Video-based facial micro-expression analysis: A survey of datasets, features and algorithms," IEEE transactions on pattern analysis and machine intelligence, vol. 44, no. 9, pp. 5826–5846, 2021.

[4] G. Castaneda and T. M. Khoshgoftaar, "A survey of 2d face databases," in 2015 IEEE International Conference on Information Reuse and Integration, 2015, pp. 219–224.

[5] M. Chihaoui, A. Elkefi, W. Bellil, and C. Ben Amar, "A survey of 2d face recognition techniques," Computers, vol. 5, no. 4, 2016. [Online]. Available: https://www.mdpi.com/2073-431X/5/4/21

[1] https://www.dropbox.com/scl/fo/zkhgubep2gi3e4mhf2dfx/h?rlkey=0139pouq5l41s9kta759c6knz&dl=0

[6] N. F. Troje and H. H. Bülthoff, "Face recognition under varying poses: The role of texture and shape," Vision research, vol. 36, no. 12, pp. 1761–1771, 1996.

[7] C. Beumier and M. Acheroy, "Sic db: multi-modal database for person authentication," in Proceedings 10th International Conference on Image Analysis and Processing. IEEE, 1999, pp. 704–708.

[8] L. J. Denes, P. Metes, and Y. Liu, "Hyperspectral face database," Carnegie Mellon University, Pittsburgh, PA, Tech. Rep. CMU-RI-TR-02-25, October 2002. [Online]. Available: https://www.ri.cmu.edu/publications/hyperspectral-face-database/

[9] X. Kevin and W. Bowyer, "Visible-light and infrared face recognition," in Workshop on Multimodal User Authentication, vol. 48. Citeseer, 2003.

[10] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "Face recognition using 2d and 3d facial data," in Workshop in Multidimonal User Authentication pp25-32. Citeseer, 2003.

[11] [Online]. Available: http://biometrics.idealtest.org/

[12] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol. 1. IEEE, 2005, pp. 947–954.

[13] A. Moreno, "Gavabdb: a 3d face database," in Proc. 2nd COST275 Workshop on Biometrics on the Internet, 2004, 2004, pp. 75–80.

[14] C. Conde and A. Serrano, "3d facial normalization with spin images and influence of range data calculation over face verification," in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops, 2005, pp. 115–115.

[15] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3d facial expression database for facial behavior research," in 7th international conference on automatic face and gesture recognition (FGR06). IEEE, 2006, pp. 211–216.

[16] H. Chang, H. Harishwaran, M. Yi, A. Koschan, B. Abidi, and M. Abidi, "An indoor and outdoor, multimodal, multispectral and multi-illuminant database for face recognition," in 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06), 2006, pp. 54–54.

[17] X. Zhang, L. Yin, J. F. Cohn, S. Canavan, M. Reale, A. Horowitz, and P. Liu, "A high-resolution spontaneous 3d dynamic facial expression database," in 2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG). IEEE, 2013, pp. 1–6.

[18] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3d face analysis," in Biometrics and Identity Management: First European Workshop, BIOID 2008, Roskilde, Denmark, May 7-9, 2008. Revised Selected Papers 1. Springer, 2008, pp. 47–56.

[19] Y. Baocai, S. Yanfeng, W. Chengzhang, and G. Yun, "Bjut-3d large scale 3d face database and information processing," Journal of Computer Research and Development, vol. 6, no. 020, p. 4, 2009.

[20] S. Polikovsky, Y. Kameda, and Y. Ohta, "Facial micro-expressions recognition using high speed camera and 3d-gradient descriptor," in 3rd International Conference on Imaging for Crime Detection and Prevention (ICDP 2009). IET, 2009.

[21] S. Z. Li, Z. Lei, and M. Ao, "The hfb face database for heterogeneous face biometrics research," in 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2009, pp. 1–8.

[22] C. D. Frowd, B. J. Matuszewski, L.-K. Shark, and W. Quan, "Towards a comprehensive 3d dynamic facial expression database," in Proceedings of the 9th WSEAS international conference on signal, speech and image processing, and 9th WSEAS international conference on Multimedia, internet & video technologies, 2009, pp. 113–119.

[23] S. Gupta, K. R. Castleman, M. K. Markey, and A. C. Bovik, "Texas 3d face recognition database," in 2010 IEEE Southwest Symposium on Image Analysis & Interpretation (SSIAI). IEEE, 2010, pp. 97–100.

[24] W. Di, L. Zhang, D. Zhang, and Q. Pan, "Studies on hyperspectral face recognition in visible spectrum with feature band selection," IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 40, no. 6, pp. 1354–1361, 2010.

[25] G. Fanelli, J. Gall, H. Romsdorfer, T. Weise, and L. Van Gool, "A 3-d audio-visual corpus of affective communication," IEEE Transactions on Multimedia, vol. 12, no. 6, pp. 591–598, 2010.

[26] S. Wang, Z. Liu, S. Lv, Y. Lv, G. Wu, P. Peng, F. Chen, and X. Wang, "A natural visible and infrared facial expression database for expression recognition and emotion inference," IEEE Transactions on Multimedia, vol. 12, no. 7, pp. 682–691, 2010.

[27] V. Espinosa-Duró, M. Faundez-Zanuy, and J. Mekyska, "A new face database simultaneously acquired in visible, near-infrared and thermal spectrums," Cognitive Computation, vol. 5, pp. 119–135, 2013.
[28] S. Zafeiriou, M. Hansen, G. Atkinson, V. Argyriou, M. Petrou, M. Smith, and L. Smith, "The photoface database," in CVPR 2011 WORKSHOPS, 2011, pp. 132–139.
[29] H. Maeng, H.-C. Choi, U. Park, S.-W. Lee, and A. K. Jain, "Nfrad: Near-infrared face recognition at a distance," in 2011 International Joint Conference on Biometrics (IJCB). IEEE, 2011, pp. 1–7.
[30] D. Cosker, E. Krumhuber, and A. Hilton, "A facs valid 3d dynamic action unit database with applications to 3d dynamic morphable facial modeling," in 2011 International Conference on Computer Vision, 2011, pp. 2296–2303.
[31] B. J. Matuszewski, W. Quan, L.-K. Shark, A. S. McLoughlin, C. E. Lightbody, H. C. Emsley, and C. L. Watkins, "Hi4d-adsip 3-d dynamic facial articulation database," Image and Vision Computing, vol. 30, no. 10, pp. 713–727, 2012, 3D Facial Behaviour Analysis and Understanding. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0262885612000170
[32] S. Li, D. Yi, Z. Lei, and S. Liao, "The casia nir-vis 2.0 face database," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2013, pp. 348–353.
[33] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikäinen, "A spontaneous micro-expression database: Inducement, collection and baseline," in 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2013, pp. 1–6.
[34] N. Erdogmus and S. Marcel, "Spoofing in 2d face recognition with 3d masks and anti-spoofing with kinect," in 2013 IEEE sixth international conference on biometrics: theory, applications and systems (BTAS). IEEE, 2013, pp. 1–6.
[35] ——, "Spoofing face recognition with 3d masks," IEEE Transactions on Information Forensics and Security, vol. 9, no. 7, pp. 1084–1097, 2014.
[36] H. Maeng, S. Liao, D. Kang, S.-W. Lee, and A. K. Jain, "Nighttime face recognition at long distance: Cross-distance and cross-spectral matching," in Computer Vision–ACCV 2012: 11th Asian Conference on Computer Vision, Daejeon, Korea, November 5-9, 2012, Revised Selected Papers, Part II 11. Springer, 2013, pp. 708–721.
[37] T. I. Dhamecha, A. Nigam, R. Singh, and M. Vatsa, "Disguise detection and face recognition in visible and thermal spectrums," in 2013 International Conference on Biometrics (ICB), 2013, pp. 1–8.
[38] S. Menon, S. J., A. S.K., A. P. Nair, and S. S., "Driver face recognition and sober drunk classification using thermal images," in 2019 International Conference on Communication and Signal Processing (ICCSP), 2019, pp. 0400–0404.
[39] G. Koukiou and V. Anastassopoulos, "Drunk person identification using thermal infrared images," International journal of electronic security and digital forensics, vol. 4, no. 4, pp. 229–243, 2012.
[40] R. Min, N. Kose, and J.-L. Dugelay, "Kinectfacedb: A kinect database for face recognition," Systems, Man, and Cybernetics: Systems, IEEE Transactions on, vol. 44, no. 11, pp. 1534–1548, Nov 2014.
[41] H. Nguyen, K. Kotani, F. Chen, and B. Le, "A thermal facial emotion database and its analysis," in Image and Video Technology: 6th Pacific-Rim Symposium, PSIVT 2013, Guanajuato, Mexico, October 28-November 1, 2013, Proceedings, vol. 8333. Springer, 2014, p. 397.
[42] P. Liu and L. Yin, "Spontaneous facial expression analysis based on temperature changes and head motions," in 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), vol. 1, 2015, pp. 1–6.
[43] J. Zhang, D. Huang, Y. Wang, and J. Sun, "Lock3dface: A large-scale database of low-cost kinect 3d faces," in 2016 International Conference on Biometrics (ICB). IEEE, 2016, pp. 1–8.
[44] J. Bernhard, J. Barr, K. W. Bowyer, and P. Flynn, "Near-ir to visible light face matching: Effectiveness of pre-processing options for commercial matchers," in 2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS), 2015, pp. 1–8.
[45] J. Booth, A. Roussos, S. Zafeiriou, A. Ponniah, and D. Dunaway, "A 3d morphable model learnt from 10,000 faces," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.
[46] A. Sepas-Moghaddam, V. Chiesa, P. L. Correia, F. Pereira, and J.-L. Dugelay, "The ist-eurecom light field face database," in 2017 5th International Workshop on Biometrics and Forensics (IWBF), 2017, pp. 1–6.
[47] G. Borghi, M. Venturelli, R. Vezzani, and R. Cucchiara, "Poseidon: Face-from-depth for driver pose estimation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4661–4670.
[48] H. A. Le and I. A. Kakadiaris, "Uhdb31: A dataset for better understanding face recognition across pose and illumination variation," in Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 2555–2563.
[49] Z.-H. Feng, P. Huber, J. Kittler, P. Hancock, X.-J. Wu, Q. Zhao, P. Koppen, and M. Raetsch, "Evaluation of dense 3d reconstruction from 2d face images in the wild," in 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), 2018, pp. 780–786.
[50] P. Urbanová, Z. Ferková, M. Jandová, M. Jurda, D. Černỳ, and J. Sochor, "Introducing the fidentis 3d face database," AnthropologicAl review, vol. 81, no. 2, pp. 202–223, 2018.
[51] S. Cheng, I. Kotsia, M. Pantic, and S. Zafeiriou, "4dfab: A large scale 4d database for facial expression analysis and biometric applications," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018.
[52] R. S. Ghiass, H. Bendada, and X. Maldague, "Université laval face motion and time-lapse video database (ul-fmtv)," in Proceedings of the 14th International Conference on Quantitative Infrared Thermography, 2018.
[53] K. Mallat and J.-L. Dugelay, "A benchmark database of visible and thermal paired face images across multiple variations," in 2018 International Conference of the Biometrics Special Interest Group (BIOSIG), 2018, pp. 1–5.
[54] G. Hermosilla, J. L. Verdugo, G. Farias, E. Vera, F. Pizarro, and M. Machuca, "Face recognition and drunk classification using infrared face images," Journal of Sensors, vol. 2018, pp. 1–8, 2018.
[55] R. Raghavendra, N. Vetrekar, K. B. Raja, R. S. Gad, and C. Busch, "Detecting disguise attacks on multi-spectral face recognition through spectral signatures," in 2018 24th International Conference on Pattern Recognition (ICPR), 2018, pp. 3371–3377.
[56] J. Lee, S. Kim, S. Kim, and K. Sohn, "Multi-modal recurrent attention networks for facial expression recognition," IEEE Transactions on Image Processing, vol. 29, pp. 6977–6991, 2020.
[57] H. Dai, N. Pears, W. Smith, and C. Duncan, "Statistical modeling of craniofacial shape and texture," International Journal of Computer Vision, vol. 128, pp. 547–571, 2020.
[58] K. Panetta, Q. Wan, S. Agaian, S. Rajeev, S. Kamath, R. Rajendran, S. P. Rao, A. Kaszowska, H. A. Taylor, A. Samani, and X. Yuan, "A comprehensive database for benchmarking imaging systems," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 3, pp. 509–520, 2020.
[59] U. Cheema and S. Moon, "Sejong face database: A multi-modal disguise face database," Computer Vision and Image Understanding, vol. 208-209, p. 103218, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S107731422100062X
[60] R. Ashrafi, M. Azarbayjani, and H. Tabkhi, "Charlotte-thermalface: A fully annotated thermal infrared face dataset with various environmental conditions and distances," Infrared Physics & Technology, vol. 124, p. 104209, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1350449522001906
[61] C. Xu, Y. Wang, T. Tan, and L. Quan, "Depth vs. intensity: which is more important for face recognition?" in Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., vol. 1, 2004, pp. 342–345 Vol.1.
[62] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "Face recognition using 2d and 3d facial data," in Workshop on Multimodal User Authentication. Citeseer, 2003.
[63] S. Zhou and S. Xiao, "3d face recognition: a survey," Human-centric Computing and Information Sciences, vol. 8, no. 1, pp. 1–27, 2018.
[64] S. K. Mada, M. L. Smith, L. N. Smith, and P. S. Midha, "Overview of passive and active vision techniques for hand-held 3d data acquisition," in Opto-Ireland 2002: Optical Metrology, Imaging, and Machine Vision, vol. 4877. SPIE, 2003, pp. 16–27.
[65] D. Beltran and L. Basañez, A Comparison between Active and Passive 3D Vision Sensors: BumblebeeXB3 and Microsoft Kinect. Springer International Publishing, 2014, pp. 725–734. [Online]. Available: https://doi.org/10.1007/978-3-319-03413-3_54
[66] H. Chen and W. Cui, "A comparative analysis between active structured light and multi-view stereo vision technique for 3d reconstruction of face

model surface," Optik, vol. 206, p. 164190, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0030402620300243

[67] H. Sarbolandi, D. Lefloch, and A. Kolb, "Kinect range sensing: Structured-light versus time-of-flight kinect," Computer Vision and Image Understanding, vol. 139, pp. 1–20, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1077314215001071

[68] S. Giancola, M. Valenti, and R. Sala, A Survey on 3D Cameras: Metrological Comparison of Time-of-Flight, Structured-Light and Active Stereoscopy Technologies. Springer, 2018.

[69] R. Hg, P. Jasek, C. Rofidal, K. Nasrollahi, T. B. Moeslund, and G. Tranchet, "An rgb-d database using microsoft's kinect for windows for face detection," in 2012 Eighth International Conference on Signal Image Technology and Internet Based Systems. IEEE, 2012, pp. 42–46.

[70] P. Peursum and W. Liu, "Curtinfaces database," Apr 2013. [Online]. Available: https://researchdata.edu.au/curtinfaces-database/3640

[71] C. Cao, Y. Weng, S. Zhou, Y. Tong, and K. Zhou, "Facewarehouse: A 3d facial expression database for visual computing," IEEE Transactions on Visualization and Computer Graphics, vol. 20, no. 3, pp. 413–425, 2014.

[72] G. Goswami, S. Bharadwaj, M. Vatsa, and R. Singh, "On rgb-d face recognition using kinect," in 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS), 2013, pp. 1–6.

[73] A. Zabatani, V. Surazhsky, E. Sperling, S. B. Moshe, O. Menashe, D. H. Silver, Z. Karni, A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Intel realsense sr300 coded light depth camera," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 10, pp. 2333–2345, 2020.

[74] A. George, Z. Mostaani, D. Geissenbuhler, O. Nikisins, A. Anjos, and S. Marcel, "Biometric face presentation attack detection with multi-channel convolutional neural network," IEEE Transactions on Information Forensics and Security, vol. 15, pp. 42–55, 2019. [Online]. Available: https://doi.org/10.1109/TIFS.2019.2916652

[75] S. Bhattacharjee, A. Mohammadi, and S. Marcel, "Spoofing deep face recognition with custom silicone masks," in 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS), 2018, pp. 1–7.

[76] X. Chen, S. Xu, Q. Ji, and S. Cao, "A dataset and benchmark towards multi-modal face anti-spoofing under surveillance scenarios," IEEE Access, vol. 9, pp. 28 140–28 155, 2021.

[77] S. Zhang, X. Wang, A. Liu, C. Zhao, J. Wan, S. Escalera, H. Shi, Z. Wang, and S. Z. Li, "A dataset and benchmark for large-scale multi-modal face anti-spoofing," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2019.

[78] G. Te, W. Hu, and Z. Guo, "Exploring hypergraph representation on face anti-spoofing beyond 2d attacks," in 2020 IEEE International Conference on Multimedia and Expo (ICME), 2020, pp. 1–6.

[79] S. Zhang, X. Wang, A. Liu, C. Zhao, J. Wan, S. Escalera, H. Shi, Z. Wang, and S. Z. Li, "A dataset and benchmark for large-scale multi-modal face anti-spoofing," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2019.

[80] J. V. Patil and P. Bailke, "Real time facial expression recognition using realsense camera and ann," in 2016 International Conference on Inventive Computation Technologies (ICICT), vol. 2, 2016, pp. 1–6.

[81] Y. Delaere, "Implementation and optimization of facial recognition with the intelrealsense sr300," B.S. thesis, University of Twente, 2017.

[82] P. Chhokra, A. Chowdhury, G. Goswami, M. Vatsa, and R. Singh, "Unconstrained kinect video face database," Information Fusion, vol. 44, pp. 113–125, 2018.

[83] T. Mantecon, C. R. del Bianco, F. Jaureguizar, and N. García, "Depth-based face recognition using local quantized patterns adapted for range data," in 2014 IEEE International Conference on Image Processing (ICIP), 2014, pp. 293–297. [Online]. Available: https://sites.google.com/site/hrrfaced/

[84] P. Pala, L. Seidenari, S. Berretti, and A. Del Bimbo, "Enhanced skeleton and face 3d data for person re-identification from depth cameras," Computers & Graphics, vol. 79, pp. 69–80, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0097849319300135

[85] Z. Majid, H. Setan, and A. Chong, "3d modeling of human face with noncontact three dimensional digitizer," in International Symposium and Exhibition on Geoinformation, 2004.

[86] L. J. Spreeuwers, "Multi-view passive 3d face acquisition device," BIOSIG 2008: Biometrics and Electronic Signatures, 2008.

[87] V. Vijayan, K. W. Bowyer, P. J. Flynn, D. Huang, L. Chen, M. Hansen, O. Ocegueda, S. K. Shah, and I. A. Kakadiaris, "Twins 3d face recognition challenge," in 2011 International Joint Conference on Biometrics (IJCB), 2011, pp. 1–7.

[88] A. Colombo, C. Cusano, and R. Schettini, "Umb-db: A database of partially occluded 3d faces," in 2011 IEEE international conference on computer vision workshops (ICCV workshops). IEEE, 2011, pp. 2113–2119.

[89] S. Berretti, A. Del Bimbo, and P. Pala, "Superfaces: A super-resolution model for 3d faces," in Computer Vision–ECCV 2012. Workshops and Demonstrations: Florence, Italy, October 7-13, 2012, Proceedings, Part I 12. Springer, 2012, pp. 73–82.

[90] B. Y. Li, A. S. Mian, W. Liu, and A. Krishna, "Using kinect for face recognition under varying poses, expressions, illumination and disguise," in 2013 IEEE Workshop on Applications of Computer Vision (WACV), 2013, pp. 186–192.

[91] G. Toderici, G. Evangelopoulos, T. Fang, T. Theoharis, and I. A. Kakadiaris, "Uhdb11 database for 3d-2d face recognition," in Proc. 6th Pacific-Rim Symposium on Image and Video Technology, 2014, pp. 73–86.

[92] J. Li, Z. Dong, S. Lu, S.-J. Wang, W.-J. Yan, Y. Ma, Y. Liu, C. Huang, and X. Fu, "Cas(me)^3: A third generation facial spontaneous micro-expression database with depth information and high ecological validity," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.

[93] G. Heusch, A. George, D. Geissbühler, Z. Mostaani, and S. Marcel, "Deep models and shortwave infrared information to detect face presentation attacks," IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 2, no. 4, pp. 399–409, 2020.

[94] A. Liu, Z. Tan, J. Wan, S. Escalera, G. Guo, and S. Z. Li, "Casia-surf cefa: A benchmark for multi-modal cross-ethnicity face anti-spoofing," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021, pp. 1179–1187.

[95] T. C. Faltemier, K. W. Bowyer, and P. J. Flynn, "Using a multi-instance enrollment representation to improve 3d face recognition," in 2007 First IEEE International Conference on Biometrics: Theory, Applications, and Systems, 2007, pp. 1–6.

[96] T. Heseltine, N. Pears, and J. Austin, "Three-dimensional face recognition using combinations of surface feature map subspace components," Image and Vision Computing, vol. 26, no. 3, pp. 382–396, 2008.

[97] G. Fanelli, T. Weise, J. Gall, and L. Van Gool, "Real time head pose estimation from consumer depth cameras," in Pattern Recognition: 33rd DAGM Symposium, Frankfurt/Main, Germany, August 31–September 2, 2011. Proceedings 33. Springer, 2011, pp. 101–110.

[98] T. Baltrušaitis, P. Robinson, and L.-P. Morency, "3d constrained local model for rigid and non-rigid facial tracking," in 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2610–2617.

[99] L. Wang, Z. Chen, T. Yu, C. Ma, L. Li, and Y. Liu, "Faceverse: A fine-grained and detail-controllable 3d face morphable model from a hybrid dataset," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2022, pp. 20 333–20 342.

[100] J. Cui, H. Zhang, H. Han, S. Shan, and X. Chen, "Improving 2D face recognition via discriminative face depth estimation," in 2018 International Conference on Biometrics (ICB), 2018, pp. 140–147.

[101] P. Koppen, Z.-H. Feng, J. Kittler, M. Awais, W. Christmas, X.-J. Wu, and H.-F. Yin, "Gaussian mixture 3d morphable face model," Pattern Recognition, vol. 74, pp. 617–628, 2018.

[102] M. Quintana, S. Karaoglu, F. Alvarez, J. M. Menendez, and T. Gevers, "Three-d wide faces (3dwf): Facial landmark detection and 3d reconstruction over a new rgb–d multi-camera dataset," Sensors, vol. 19, no. 5, 2019. [Online]. Available: https://www.mdpi.com/1424-8220/19/5/1103

[103] L. Jiang, J. Zhang, C. Li, and J. Zhou, "Rgb-d face recognition via spatial and channel attentions," in 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), vol. 5, 2021, pp. 2037–2041.

[104] W. Gao, C. Luo, L. Wang, X. Xiong, J. Chen, T. Hao, Z. Jiang, F. Fan, M. Du, Y. Huang et al., "Aibench: towards scalable and comprehensive datacenter ai benchmarking," in Benchmarking, Measuring, and Optimizing: First BenchCouncil International Symposium, Bench 2018, Seattle, WA, USA, December 10-13, 2018, Revised Selected Papers 1. Springer, 2019, pp. 3–9. [Online]. Available: https://www.benchcouncil.org/aibench/training/download.html

[105] J. Li, Y. Mi, G. Li, and Z. Ju, "Cnn-based facial expression recognition from annotated rgb-d images for human–robot interaction," International Journal of Humanoid Robotics, vol. 16, no. 04, p. 1941002, 2019.

[106] H. Zhang, H. Han, J. Cui, S. Shan, and X. Chen, "Rgb-d face recognition via deep complementary and common feature learning," in 2018 13th

IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), 2018, pp. 8–15.
[107] E. Frigieri, G. Borghi, R. Vezzani, and R. Cucchiara, "Fast and accurate facial landmark localization in depth images for in-car applications," in Image Analysis and Processing-ICIAP 2017: 19th International Conference, Catania, Italy, September 11-15, 2017, Proceedings, Part I 19. Springer, 2017, pp. 539–549. [Online]. Available: http://imagelab.ing.unimore.it/landmarkdepth
[108] X. Sun, L. Huang, and C. Liu, "Multimodal face spoofing detection via rgb-d images," in 2018 24th International Conference on Pattern Recognition (ICPR), 2018, pp. 2221–2226.
[109] M. Munaro, S. Ghidoni, D. T. Dizmen, and E. Menegatti, "A feature-based approach to people re-identification using skeleton keypoints," in 2014 IEEE International Conference on Robotics and Automation (ICRA), 2014, pp. 5644–5651.
[110] G.-S. J. Hsu, Y.-L. Liu, H.-C. Peng, and P.-X. Wu, "Rgb-d-based face reconstruction and recognition," IEEE Transactions on Information Forensics and Security, vol. 9, no. 12, pp. 2110–2118, 2014. [Online]. Available: https://github.com/AvLab-CV/RGB-D-Face-Database
[111] A. D. Bagdanov, A. Del Bimbo, and I. Masi, "The florence 2d/3d hybrid face dataset," in Proceedings of the 2011 Joint ACM Workshop on Human Gesture and Behavior Understanding, ser. J-HGBU '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 79–80. [Online]. Available: https://doi.org/10.1145/2072572.2072597
[112] S. Liu, S. Guo, W. Wang, H. Qiao, Y. Wang, and W. Luo, "Multi-view laplacian eigenmaps based on bag-of-neighbors for rgb-d human emotion recognition," Information Sciences, vol. 509, pp. 243–256, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025519307753
[113] S. M. H. Mousavi and S. Y. Mirinezhad, "Iranian kinect face database (ikfdb): a color-depth based face database collected by kinect v. 2 sensor," SN Applied Sciences, vol. 3, no. 1, p. 19, 2021.
[114] P. Zhang, F. Zou, Z. Wu, N. Dai, S. Mark, M. Fu, J. Zhao, and K. Li, "Feathernets: Convolutional neural networks as light as feather for face anti-spoofing," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019, pp. 0–0.
[115] O. Nikisins, K. Nasrollahi, M. Greitans, and T. B. Moeslund, "Rgb-d-t based face recognition," in 2014 22nd International Conference on Pattern Recognition, 2014, pp. 1716–1721.
[116] H. Yang, H. Zhu, Y. Wang, M. Huang, Q. Shen, R. Yang, and X. Cao, "Facescape: A large-scale high quality 3d face dataset and detailed riggable 3d face prediction," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2020.
[117] L. Yunqi, L. Haibin, and L. Qingmin, "Geometric features of 3d face and recognition of it by pca." Journal of multimedia, vol. 6, no. 2, 2011.
[118] Y. Wang, G. Pan, Z. Wu, and Y. Wang, "Exploring facial expression effects in 3d face recognition using partial icp," in Asian Conference on Computer Vision. Springer, 2006, pp. 581–590.
[119] C. Hesher, A. Srivastava, and G. Erlebacher, "A novel technique for face recognition using range imaging," in Seventh International Symposium on Signal Processing and Its Applications, 2003. Proceedings., vol. 2, 2003, pp. 201–204 vol.2.
[120] Y. Hu, Z. Zhang, X. Xu, Y. Fu, and T. S. Huang, "Building large scale 3d face database for face analysis," in Multimedia Content Analysis and Mining: International Workshop, MCAM 2007, Weihai, China, June 30-July 1, 2007. Proceedings. Springer, 2007, pp. 343–350.
[121] N. Uchida, T. Shibahara, T. Aoki, H. Nakajima, and K. Kobayashi, "3d face recognition using passive stereo vision," in IEEE International Conference on Image Processing 2005, vol. 2, 2005, pp. II–950.
[122] J. White, A. Ortega-Castrillon, C. Virgo, K. Indencleef, H. Hoskens, M. Shriver, and P. Claes, "Sources of variation in the 3dmdface and vectra h1 3d facial imaging systems." Scientific Reports, vol. 10, no. 1, pp. 4443–4443, 2020.
[123] R. Winder, T. Darvann, W. McKnight, J. Magee, and P. Ramsay-Baggs, "Technical validation of the di3d stereophotogrammetry surface imaging system," British Journal of Oral and Maxillofacial Surgery, vol. 46, no. 1, pp. 33–37, 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0266435607004445
[124] D. Samaras, D. Metaxas, P. Fua, and Y. Leclerc, "Variable albedo surface reconstruction from stereo and shape from shading," in Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662), vol. 1, 2000, pp. 480–487 vol.1.
[125] X. Wang, Y. Guo, B. Deng, and J. Zhang, "Lightweight photometric stereo for facial details recovery," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2020.
[126] I. Chingovska, N. Erdogmus, A. Anjos, and S. Marcel, "Face recognition systems under spoofing attacks," Face Recognition Across the Imaging Spectrum, pp. 165–194, 2016.
[127] H. Steiner, A. Kolb, and N. Jung, "Reliable face anti-spoofing using multispectral SWIR imaging," in International Conference on Biometrics (ICB), 2016, pp. 1–8. [Online]. Available: https://doi.org/10.1109/ICB.2016.7550052
[128] H. Steiner, S. Sporrer, A. Kolb, and N. Jung, "Design of an active multispectral swir camera system for skin detection and face verification," Journal of Sensors, vol. 2016, 2016.
[129] R. Raghavendra, K. B. Raja, S. Venkatesh, F. A. Cheikh, and C. Busch, "On the vulnerability of extended multispectral face recognition systems towards presentation attacks," in 2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA), 2017, pp. 1–8.
[130] A. Agarwal, D. Yadav, N. Kohli, R. Singh, M. Vatsa, and A. Noore, "Face presentation attack with latex masks in multispectral videos," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 81–89.
[131] F. Jiang, P. Liu, and X. Zhou, "Multilevel fusing paired visible light and near-infrared spectral images for face anti-spoofing," Pattern Recognition Letters, vol. 128, pp. 30–37, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016786551830583X
[132] B. Zhang, L. Zhang, D. Zhang, and L. Shen, "Directional binary code with application to polyu near-infrared face database," Pattern Recognition Letters, vol. 31, no. 14, pp. 2337–2344, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865510002291
[133] I. O. W. S. Bench, IRIS Thermal/Visible Face Database, DOE University Research Program in Robotics under grant DOE-DE-FG02-86NE37968; DOD/TACOM/NAC/ARC Program under grant R01-1344-18; FAA/NSSA grant R01-1344-48/49; Office of Naval Research under grant #N000143010022., 2006. [Online]. Available: http://vcipl-okstate.org/pbvs/bench/
[134] P. Buddharaju, I. T. Pavlidis, P. Tsiamyrtzis, and M. Bazakos, "Physiology-based face recognition in the thermal infrared spectrum," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 4, pp. 613–626, 2007.
[135] S. Hu, N. J. Short, B. S. Riggan, C. Gordon, K. P. Gurton, M. Thielke, P. Gurram, and A. L. Chan, "A polarimetric thermal database for face recognition research," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2016.
[136] H. Zhang, B. S. Riggan, S. Hu, N. J. Short, and V. M. Patel, "Synthesis of high-quality visible faces from polarimetric thermal faces using generative adversarial networks," International Journal of Computer Vision, vol. 127, pp. 845–862, 2019.
[137] A. Kuzdeuov, D. Aubakirova, D. Koishigarina, and H. A. Varol, "Tfw: Annotated thermal faces in the wild dataset," IEEE Transactions on Information Forensics and Security, vol. 17, pp. 2084–2094, 2022.
[138] M. Abdrakhmanova, A. Kuzdeuov, S. Jarju, Y. Khassanov, M. Lewis, and H. A. Varol, "Speakingfaces: A large-scale multimodal dataset of voice commands with visual and thermal video streams," Sensors, vol. 21, no. 10, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/10/3465
[139] J. Heo, S. Kong, B. Abidi, and M. Abidi, "Fusion of visual and thermal signatures with eyeglass removal for robust face recognition," in 2004 Conference on Computer Vision and Pattern Recognition Workshop, 2004, pp. 122–122.
[140] S. Z. Li, R. Chu, S. Liao, and L. Zhang, "Illumination invariant face recognition using near-infrared images," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 4, pp. 627–639, 2007.
[141] M. Kopaczka, R. Kolk, J. Schock, F. Burkhard, and D. Merhof, "A thermal infrared face database with facial landmarks and emotion labels," IEEE Transactions on Instrumentation and Measurement, vol. 68, no. 5, pp. 1389–1401, 2019.
[142] M. Uzair, A. Mahmood, and A. Mian, "Hyperspectral face recognition with spatiospectral information fusion and pls regression," IEEE Transactions on Image Processing, vol. 24, no. 3, pp. 1127–1137, 2015.
[143] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2d and 3d face recognition: A survey," Pattern Recognition Letters, vol. 28, no. 14, pp. 1885–1906, 2007, image: Information and Control. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865507000189
[144] T. Bourlai and L. A. Hornak, "Face recognition outside the visible spectrum," Image and Vision Computing, vol. 55, pp. 14–17, 2016,

recognizing future hot topics and hard problems in biometrics research. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0262885616300531

[145] M. Krišto and M. Ivasic-Kos, "An overview of thermal face recognition methods," in 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2018, pp. 1098–1103.

[146] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition," in Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14. Springer, 2016, pp. 87–102.

[147] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), 2018, pp. 67–74.

[148] X. Liu, Y. Wu, and X. Zhou, "A deep neural network model for face recognition," in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2015, pp. 3722–3726.

[149] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," arXiv preprint arXiv:1411.7923, 2014.

[150] P. Husák, J. Cech, and J. Matas, "Spotting facial micro-expressions "in the wild"," in 22nd Computer Vision Winter Workshop (Retz), 2017, pp. 1–9.

[151] B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, and A. K. Jain, "Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1931–1939.

[152] T. Gerig, A. Morel-Forster, C. Blumer, B. Egger, M. Luthi, S. Schoenborn, and T. Vetter, "Morphable face models - an open framework," in 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), 2018, pp. 75–82.

[153] J. Booth, E. Antonakos, S. Ploumpis, G. Trigeorgis, Y. Panagakis, and S. Zafeiriou, "3d face morphable models "in-the-wild"," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

[154] M. D. Breitenstein, D. Kuettel, T. Weise, L. van Gool, and H. Pfister, "Real-time face pose estimation from single range images," in 2008 IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.

[155] M. Szwoch, "Feedb: a multimodal database of facial expressions and emotions," in 2013 6th International Conference on Human System Interactions (HSI). IEEE, 2013, pp. 524–531.

[156] ——, "On facial expressions and emotions rgb-d database," in Beyond Databases, Architectures, and Structures: 10th International Conference, BDAS 2014, Ustron, Poland, May 27-30, 2014. Proceedings 10. Springer, 2014, pp. 384–394.

[157] M. Hayat, M. Bennamoun, and A. A. El-Sallam, "An rgb–d based image set classification for robust face recognition from kinect data," Neurocomputing, vol. 171, pp. 889–900, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231215010048

[158] Y. Guo, L. Cai, and J. Zhang, "3d face from x: Learning face shape from diverse sources," IEEE Transactions on Image Processing, vol. 30, pp. 3815–3827, 2021.

[159] Y. Guo, H. Wang, L. Wang, Y. Lei, L. Liu, and M. Bennamoun, "3d face recognition: Two decades of progress and prospects," ACM Comput. Surv., aug 2023. [Online]. Available: https://doi.org/10.1145/3615863

[160] Y. Jing, X. Lu, and S. Gao, "3d face recognition: A comprehensive survey in 2022," Computational Visual Media, pp. 1–29, 2023.

[161] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2d and 3d face recognition: A survey," Pattern Recognition Letters, vol. 28, no. 14, pp. 1885–1906, 2007, image: Information and Control. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865507000189

[162] K. Dharavath, R. H. Laskar, and F. A. Talukdar, "Qualitative study on 3d face databases: A review," in 2013 Annual IEEE India Conference (INDICON), 2013, pp. 1–6.

[163] X. Li, T. Jia, and H. Zhang, "Expression-insensitive 3d face recognition using sparse representation," in 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 2575–2582.

[164] J. Booth, A. Roussos, A. Ponniah, D. Dunaway, and S. Zafeiriou, "Large scale 3d morphable models," International Journal of Computer Vision, vol. 126, no. 2, pp. 233–254, 2018.

[165] C.-H. J. Tzou, N. M. Artner, I. Pona, A. Hold, E. Placheta, W. G. Kropatsch, and M. Frey, "Comparison of three-dimensional surface-imaging systems," Journal of Plastic, Reconstructive & Aesthetic Surgery, vol. 67, no. 4, pp. 489–497, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1748681514000047

[166] N. K. Benamara, E. Zigh, T. B. Stambouli, and M. Keche, "Towards a robust thermal-visible heterogeneous face recognition approach based on a cycle generative adversarial network," IJIMAI, vol. 7, no. 4, pp. 132–145, 2022.

[167] D. Poster, M. Thielke, R. Nguyen, S. Rajaraman, X. Di, C. N. Fondje, V. M. Patel, N. J. Short, B. S. Riggan, N. M. Nasrabadi et al., "A large-scale, time-synchronized visible and thermal face dataset," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021, pp. 1559–1568.

[168] G. Pitteri, M. Munaro, and E. Menegatti, "Depth-based frontal view generation for pose invariant face recognition with consumer rgb-d sensors," in Intelligent Autonomous Systems 14: Proceedings of the 14th International Conference IAS-14 14. Springer, 2017, pp. 925–937.

[169] L. Jiang, J. Zhang, and B. Deng, "Robust rgb-d face recognition using attribute-aware loss," IEEE transactions on pattern analysis and machine intelligence, vol. 42, no. 10, pp. 2552–2566, 2019.

[170] R. Shoja Ghiass, O. Arandjelović, A. Bendada, and X. Maldague, "Infrared face recognition: A comprehensive review of methodologies and databases," Pattern Recognition, vol. 47, no. 9, pp. 2807–2824, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320314001137

[171] L. L. Chambino, J. S. Silva, and A. Bernardino, "Multispectral facial recognition: A review," IEEE Access, vol. 8, pp. 207 871–207 883, 2020.

[172] X. Zhang and H. Zhao, "Hyperspectral-cube-based mobile face recognition: A comprehensive review," Information Fusion, vol. 74, pp. 132–150, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253521000695

[173] Y.-Q. Wang, "An analysis of the viola-jones face detection algorithm," Image Processing On Line, vol. 4, pp. 128–148, 2014.

[174] R. Ranjan, V. M. Patel, and R. Chellappa, "A deep pyramid deformable part model for face detection," in 2015 IEEE 7th international conference on biometrics theory, applications and systems (BTAS). IEEE, 2015, pp. 1–8.

[175] T. Mita, T. Kaneko, and O. Hori, "Joint haar-like features for face detection," in Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, vol. 2. IEEE, 2005, pp. 1619–1626.

[176] Z. Jin, Z. Lou, J. Yang, and Q. Sun, "Face detection using template matching and skin-color information," Neurocomputing, vol. 70, no. 4, pp. 794–800, 2007, advanced Neurocomputing Theory and Methodology. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231206002840

[177] D. Garg, P. Goel, S. Pandya, A. Ganatra, and K. Kotecha, "A deep learning approach for face detection using yolo," in 2018 IEEE Punecon, 2018, pp. 1–4.

[178] X. Sun, P. Wu, and S. C. Hoi, "Face detection using deep learning: An improved faster rcnn approach," Neurocomputing, vol. 299, pp. 42–50, 2018.

[179] A. Kumar, A. Kaur, and M. Kumar, "Face detection techniques: a review," Artificial Intelligence Review, vol. 52, pp. 927–948, 2019.

[180] A. Mian, M. Bennamoun, and R. Owens, "Automatic 3d face detection, normalization and recognition," in Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06), 2006, pp. 735–742.

[181] M. Pamplona Segundo, L. Silva, O. Bellon, and S. Sarkar, "Orthogonal projection images for 3d face detection," Pattern Recognition Letters, vol. 50, pp. 72–81, 2014, depth Image Analysis. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865513003693

[182] A. M. Basbrain, J. Q. Gan, and A. Clark, "Accuracy enhancement of the viola-jones algorithm for thermal face detection," in Intelligent Computing Methodologies: 13th International Conference, ICIC 2017, Liverpool, UK, August 7-10, 2017, Proceedings, Part III 13. Springer, 2017, pp. 71–82.

[183] X. Zhu, Z. Lei, X. Liu, H. Shi, and S. Z. Li, "Face alignment across large poses: A 3d solution," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 146–155.

[184] L. A. Jeni, S. Tulyakov, L. Yin, N. Sebe, and J. F. Cohn, "The first 3d face alignment in the wild (3dfaw) challenge," in Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part II 14. Springer, 2016, pp. 511–520.

[185] X. Zhang, L. Yin, J. F. Cohn, S. Canavan, M. Reale, A. Horowitz, P. Liu, and J. M. Girard, "Bp4d-spontaneous: a high-resolution spontaneous

3d dynamic facial expression database," Image and Vision Computing, vol. 32, no. 10, pp. 692–706, 2014.
[186] J. Guo, X. Zhu, Y. Yang, F. Yang, Z. Lei, and S. Z. Li, "Towards fast, accurate and stable 3d dense face alignment," in European Conference on Computer Vision. Springer, 2020, pp. 152–168.
[187] L. A. Jeni, J. F. Cohn, and T. Kanade, "Dense 3d face alignment from 2d videos in real-time," in 2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG), vol. 1. IEEE, 2015, pp. 1–8.
[188] V. Ayyagari, F. Boughorbel, A. Koschan, and M. Abidi, "A new method for automatic 3d face registration," in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops, 2005, pp. 119–119.
[189] S. Bodhi and S. Naveen, "Face detection, registration and feature localization experiments with rgb-d face database," Procedia Computer Science, vol. 46, pp. 1778–1785, 2015, proceedings of the International Conference on Information and Communication Technologies, ICICT 2014, 3-5 December 2014 at Bolgatty Palace & Island Resort, Kochi, India. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050915001969
[190] J. R. Tena, M. Hamouz, A. Hilton, and J. Illingworth, "A validated method for dense non-rigid 3d face registration," in 2006 IEEE International Conference on Video and Signal Based Surveillance, 2006, pp. 81–81.
[191] J. Ma, J. Zhao, Y. Ma, and J. Tian, "Non-rigid visible and infrared face registration via regularized gaussian fields criterion," Pattern Recognition, vol. 48, no. 3, pp. 772–784, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320314003471
[192] T. Ojala, M. Pietikäinen, and D. Harwood, "Performance evaluation of texture measures with classification based on kullback discrimination of distributions," in IAPR International Conference on Pattern Recognition, 1994, pp. 582–585.
[193] R. K. McConnell, "Method of and apparatus for pattern recognition," USA Patent US4 567 610A, 1982.
[194] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015.
[195] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in BMVC 2015 - Proceedings of the British Machine Vision Conference 2015. British Machine Vision Association, 2015.
[196] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1701–1708. [Online]. Available: https://doi.org/10.1109/CVPR.2014.220
[197] C. Wang, Y. Wang, Y. Chen, H. Liu, and J. Liu, "User authentication on mobile devices: Approaches, threats and trends," Computer Networks, vol. 170, p. 107118, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389128618312799
[198] C. Conde, A. Serrano, and E. Cabello, "Multimodal 2d, 2.5d & 3d face verification," in 2006 International Conference on Image Processing, 2006, pp. 2061–2064.
[199] C. McCool, V. Chandran, S. Sridharan, and C. Fookes, "3d face verification using a free-parts approach," Pattern Recognition Letters, vol. 29, no. 9, pp. 1190–1196, 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865508000329
[200] C. McCool, J. Sanchez-Riera, and S. Marcel, "Feature distribution modelling techniques for 3d face verification," Pattern Recognition Letters, vol. 31, no. 11, pp. 1324–1330, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865510000449
[201] A. Ouamane, A. Chouchane, E. Boutellaa, M. Belahcene, S. Bourennane, and A. Hadid, "Efficient tensor-based 2d+3d face verification," IEEE Transactions on Information Forensics and Security, vol. 12, no. 11, pp. 2751–2762, 2017.
[202] Y. Yu, F. Da, and Y. Guo, "Sparse icp with resampling and denoising for 3d face verification," IEEE Transactions on Information Forensics and Security, vol. 14, no. 7, pp. 1917–1927, 2019.
[203] T.-Y. Lin, C.-T. Chiu, and C.-T. Tang, "Rgb-d based multi-modal deep learning for face identification," in ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020, pp. 1668–1672.
[204] X. Xu, W. Li, and D. Xu, "Distance metric learning using privileged information for face verification and person re-identification," IEEE Transactions on Neural Networks and Learning Systems, vol. 26, no. 12, pp. 3150–3162, 2015.
[205] X. Di, B. S. Riggan, S. Hu, N. J. Short, and V. M. Patel, "Multi-scale thermal to visible face verification via attribute guided synthesis," IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 3, no. 2, pp. 266–280, 2021.
[206] N. Peri, J. Gleason, C. D. Castillo, T. Bourlai, V. M. Patel, and R. Chellappa, "A synthesis-based approach for thermal-to-visible face verification," in 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021), 2021, pp. 01–08.
[207] F. Jiang, P. Liu, X. Shao, and X. Zhou, "Face anti-spoofing with generated near-infrared images," Multimedia Tools and Applications, vol. 79, pp. 21 299–21 323, 2020.
[208] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in 2012 5th IAPR international conference on Biometrics (ICB). IEEE, 2012, pp. 26–31.
[209] Y. Liu, A. Jourabloo, and X. Liu, "Learning deep models for face anti-spoofing: Binary or auxiliary supervision," in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 389–398. [Online]. Available: https://doi.org/10.1109/CVPR.2018.00048
[210] D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," IEEE Transactions on Information Forensics and Security, vol. 10, no. 4, pp. 746–761, 2015. [Online]. Available: https://doi.org/10.1109/TIFS.2015.2400395
[211] A. Costa-Pazo, S. Bhattacharjee, E. Vazquez-Fernandez, and S. Marcel, "The replay-mobile face presentation-attack database," in 2016 International Conference of the Biometrics Special Interest Group (BIOSIG), 2016, pp. 1–7.
[212] A. Anjos and S. Marcel, "Counter-measures to photo attacks in face recognition: a public database and a baseline," in 2011 international joint conference on Biometrics (IJCB). IEEE, 2011, pp. 1–7.
[213] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in 2012 BIOSIG-proceedings of the international conference of biometrics special interest group (BIOSIG). IEEE, 2012, pp. 1–7.
[214] S. Zhang, A. Liu, J. Wan, Y. Liang, G. Guo, S. Escalara, H. J. Escalante, and S. Z. Li, "CASIA-SURF: A large-scale multi-model benchmark for face anti-spoofing," IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 2, no. 2, pp. 182–193, 2020. [Online]. Available: https://doi.org/10.1109/TBIOM.2020.2973001
[215] R. Raghavendra, K. B. Raja, and C. Busch, "Presentation attack detection for face recognition using light field camera," IEEE Transactions on Image Processing, vol. 24, no. 3, pp. 1060–1075, 2015. [Online]. Available: https://doi.org/10.1109/TIP.2015.2395951
[216] S. Liu, B. Yang, P. C. Yuen, and G. Zhao, "A 3d mask face anti-spoofing database with real world variations," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2016, pp. 100–106.
[217] I. Manjani, S. Tariyal, M. Vatsa, R. Singh, and A. Majumdar, "Detecting silicone mask-based presentation attack via deep dictionary learning," IEEE Transactions on Information Forensics and Security, vol. 12, no. 7, pp. 1713–1723, 2017.
[218] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in 2005 IEEE International Conference on Multimedia and Expo, 2005, pp. 5 pp.–.
[219] L. Yin, Y. Sun, T. Worm, and M. Reale, "A high-resolution 3d dynamic facial expression database (2008)," in 8th IEEE International Conference on Automatic Face & Gesture Recognition, 2008.
[220] Z. Zhang, J. M. Girard, Y. Wu, X. Zhang, P. Liu, U. Ciftci, S. Canavan, M. Reale, A. Horowitz, H. Yang et al., "Multimodal spontaneous emotion corpus for human behavior analysis," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 3438–3446.
[221] I. Abbasnejad, S. Sridharan, D. Nguyen, S. Denman, C. Fookes, and S. Lucey, "Using synthetic data to improve facial expression analysis with 3d convolutional networks," in Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 1609–1618.
[222] G. Stratou, A. Ghosh, P. Debevec, and L.-P. Morency, "Exploring the effect of illumination on automatic expression recognition using the ict-3drfe database," Image and Vision Computing, vol. 30, no. 10, pp. 728–737, 2012.
[223] B. Egger, W. A. P. Smith, A. Tewari, S. Wuhrer, M. Zollhoefer, T. Beeler, F. Bernard, T. Bolkart, A. Kortylewski, S. Romdhani, C. Theobalt, V. Blanz, and T. Vetter, "3d morphable face models—past, present,

and future," ACM Trans. Graph., vol. 39, no. 5, jun 2020. [Online]. Available: https://doi.org/10.1145/3395208

[224] X. Yang, T. Taketomi, and Y. Kanamori, "Makeup extraction of 3D representation via illumination-aware image decomposition," Computer Graphics Forum, vol. 42, no. 2, pp. 293–307, 2023.

[225] T. Su, Y. Zhou, Y. Yu, and S. Du, "Highlight removal of multi-view facial images," Sensors, vol. 22, no. 17, p. #6656, 2022.