IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 23, 2020, accepted November 6, 2020, date of publication November 16, 2020,
date of current version November 30, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3037451

Multispectral Facial Recognition: A Review

LUÍS LOPES CHAMBINO[1,2], JOSÉ SILVESTRE SILVA[1,3,4], AND ALEXANDRE BERNARDINO[5,6], (Member, IEEE)

[1]Portuguese Military Academy, 1169-203 Lisbon, Portugal
[2]Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisbon, Portugal
[3]Military Academy Research Center (CINAMIL), 1169-203 Lisbon, Portugal
[4]Laboratory for Instrumentation, Biomedical Engineering and Radiation Physics (LIBPhys-UC), 3004-516 Coimbra, Portugal
[5]Department of Electrical and Computer Engineering, Instituto Superior Técnico, 1049-001 Lisbon, Portugal
[6]Institute for Systems and Robotics (ISR), 1049-001 Lisbon, Portugal

Corresponding author: José Silvestre Silva (jose.silva@academiamilitar.pt)

This work was supported in part by the Military Academy Research Center (CINAMIL) through the Project Multi-Spectral Facial Recognition, and in part by the FCT with the Laboratory of Robotics and Engineering Systems of Fundação para a Ciência e a Tecnologia (LARSyS)–FCT Project under Grant UIDB/50009/2020.

ABSTRACT Multispectral images are images with more than one channel acquired in different bands or spectral ranges of the electromagnetic spectrum. Each one has specific details that can be exploited in facial recognition applications. In particular, to detect facial expression variations, pose variations and presentation attacks, a facial analysis system can benefit not only of images from the visible spectral band but also of infrared images. In this paper we perform a review of the state of the art methods used in multispectral facial recognition using images from the visible spectral band and also from the Near Infrared, Short Wavelength Infrared and Long Wavelength Infrared sub-bands. The public multispectral databases for facial analysis are identified, and a comparison is made, taking into consideration their specifications. The multispectral facial recognition methods are classified according to their basic working principle, from the traditional Fusion and Subspace methods to the more recent Deep Neural Networks.

INDEX TERMS Face recognition, multispectral image, infrared image.

I. INTRODUCTION

Now-a-days it is possible to see a growth of applications that use facial recognition systems, whether for collective use, as in companies, or for personal use, as in smartphones. There is also an increasing use of more than one spectral range, to improve results in facial recognition.

There are two main modes of image acquisition in facial recognition systems: in a controlled environment, where a person cooperates in acquiring images, and in an uncontrolled environment, also known as ''in the wild'', where a person does not cooperate or has no knowledge during the phase of image acquisition. Systems that use only the visible spectrum (VIS) have several obstacles, such as occlusions, pose variation, cooperation of the person and, the most problematic, changes in the luminosity. As a result, it is necessary to complement these facial recognition systems, either with the use of other biometric sensors (e.g. fingerprint or iris) or other spectral bands, in order to minimize these problems.

The use of the infrared spectrum, namely the Near Infrared (NIR), Short Wavelength Infrared (SWIR), Medium Wavelength Infrared (MWIR) and Long Wavelength Infrared (LWIR) spectral bands, has been used successfully in facial recognition systems, as a complement of the visible spectrum [1], [2]. These systems, which use more than one spectral band, are called multispectral. Table 1 shows the most used spectral groups applied in facial recognition.

TABLE 1. Spectral ranges [3] used in facial recognition.

Spectral Band Name	Wavelength (μm)
Visible	0.38 – 0.75
Near Infrared (NIR)	0.75 – 1.40
Short Wavelength Infrared (SWIR)	1.40 – 3.00
Mid Wavelength Infrared (MWIR)	3.00 – 8.00
Long Wavelength Infrared (LWIR)	8.00 – 15.00

The infrared spectral band has several advantages when compared to the visible spectrum; it is imperceptible to the human eye and, at the same time, less sensitive to differences in luminosity. For instance, the night cameras used in video surveillance use LEDs with emission in the infrared spectrum

The associate editor coordinating the review of this manuscript and approving it for publication was Michele Nappi.

to illuminate the scene and perform night surveillance without people realizing it.

The spectral bands NIR and SWIR are very close to the visible spectrum, thus afford an easy adaptation of automatic learning methods trained with images of the visible spectrum. The spectral bands MWIR and LWIR (or thermal, as is also know) allow the use of facial recognition systems at night, when the luminosity is very low or even zero.

In general, multispectral facial recognition systems are composed of the following phases, illustrated in Figure 1: image acquisition, face detection, face alignment, feature extraction and, for last, classification.

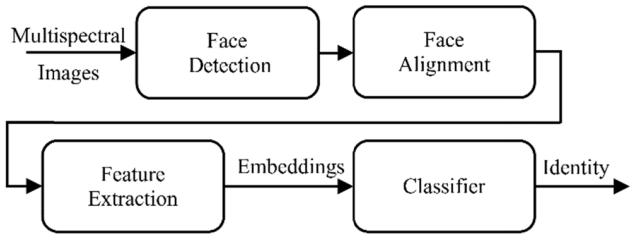

FIGURE 1. Generic multispectral facial recognition system.

The system starts with the acquisition of multispectral images. After, a face detection is performed on the image to obtain a face bounding box, so that it is possible to discard what does not belong to the person to be identified (i.e., background or other persons).

A facial alignment is performed afterwards. To do so it is first necessary to extract facial landmarks. Facial landmarks are well defined facial structures on the face (i.e., nose, eyes, jaw). After the extraction of the facial landmarks it is possible to perform a facial alignment, this alignment can be done through the eyes or through the eyes together with the mouth.

Facial detection and facial landmarks extraction can be performed on all spectral band images or only on the visible spectral band image. Visible images obtain better results than other spectral bands, because facial detection and facial landmarks extractors models are trained only with visible images. Therefore, if the images are acquired in the same location at the same time (i.e., aligned), it is preferred to use the visible image, and then share the bounding box and facial landmarks for the other spectral bands images.

The feature extraction phase has the main objective to extract the intrinsic characteristics of each identity. This phase depends on the method employed. Obtained the features, or embeddings, representing the identity it is possible to classify those features in order to obtain the identity of the person in the image.

Multispectral facial recognition systems, in comparison with only visible facial recognition systems, can be used as a method to add an extra security layer, to recognize a person more accurately, in accessing a high security place, in order to guarantee access only to authorized people. These places can be hospitals, schools, laboratories and military buildings [1].

Through the development of an improved facial recognition system, it is possible to guarantee a more reliable and more robust access control, protecting property and increasing people's safety.

The aim of this paper is to present a thorough literature review of the ongoing growth of the multispectral facial recognition. In comparison with other already published reviews in the field [4]–[6], the present paper makes several important contributions, highlighted below.

Firstly, we make a review of the state-of-the-art methods in multispectral facial recognition using only international journals with impact factor. Secondly, we provide compact and summarized information from the studied databases in great detail, having all the information in one place to assist researchers in finding the most suitable databases for their studies. Third, a real study was carried out in relation to the frequency of use of the databases, other review papers use Google Scholar cite to perform such task [5]. Fourth, we are clear on how our research was carried out, allowing a follow-up to the paper by other authors, something that is not possible to do with other articles. Finally, a summary of the methods and databases used, jointly with the results obtained for each database used and the conclusions by the author.

The remainder of this paper is organized as follows: in Section II an explanation of the systematic review procedure performed by us, with an analysis of the distribution by years and areas of research; Section III illustrates the most used databases and provides an summarized information from the databases studied; Section IV describes the performance evaluation methods used in multispectral facial recognition used by the papers studied; Section V comprises the main part of the paper, here we describe different methods used in the studied papers, grouped by methodology and year of publication; Section V provides a conclusion and future trends in the field.

II. SYSTEMATIC REVIEW

This section presents a systematic review of articles in multispectral facial recognition, as well as an analysis of its distribution by years and areas of research.

This systematic review was carried out in June of 2020 with the aid of the Web of Science database. Were selected all articles published during the period of 2000 to 2020, in journals with impact factor (works published in conferences were not considered). In the advanced search interface of Web of Science database were inserted the following search parameters:

((TS=multispectral OR TS=spectral OR TS=thermal OR TS=SWIR) AND (TS="facial recognition" OR TS="facial detection" OR TS="face recognition" OR TS="face detection") NOT (TS=chemical) NOT (TS=vein) NOT (TS=emotion) NOT (TS=expression) NOT (TS=pedestrian) NOT (TS=palmprint) NOT (TS=eye) NOT (TS=alcohol) NOT (TS=blood))

This search located 283 articles published in 132 scientific journals with impact factor. For the present analysis, only articles that perform facial recognition or facial detection with

two or more spectral bands (e.g. VIS-NIR, VIS-LWIR, VIS-NIR-LWIR, NIR-LWIR, among other possible combinations) were considered, reducing the number of articles to 47; these papers were considered the most relevant to this survey.

Multispectral facial recognition is a topic whose relevance has grown exponentially, as shown in figure 2. In the last five years, from 2016 to 2020, there was a significant increase in papers, when compared to previous years. This phenomenon can be justified by three reasons: (i) the price reduction of infrared cameras, namely the cameras that acquire images in the spectral bands of NIR and LWIR; (ii) the need to reduce human intervention in the control of accesses, thus allowing the allocation of human resources to other tasks; (iii) the implementation of deep learning in facial recognition systems, and consequently, the achievement of very promising results.

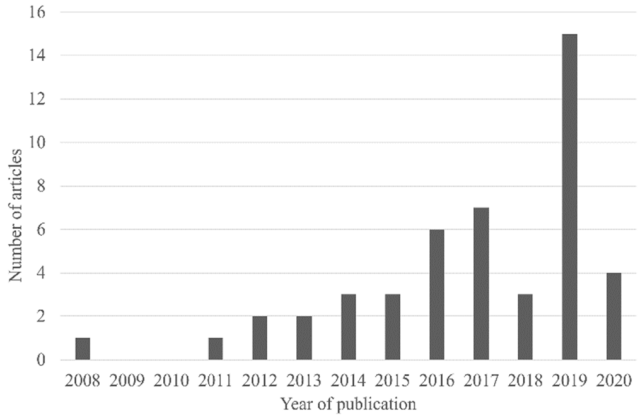

FIGURE 2. Distribution of articles by year of publication, until June 2020.

Figure 3 shows the paper distribution of the selected articles according to their research area. From this figure, it is seen that the main areas of research related to multispectral facial recognition are Computer Science (32%) and Engineering (28%). The increasing use of deep neural networks, being this a recent and innovative method, is also spreading to multispectral facial recognition systems. Therefore, it is reasonable that papers related to Computer Science and Engineering take a high percentage of the total papers shown on figure 3.

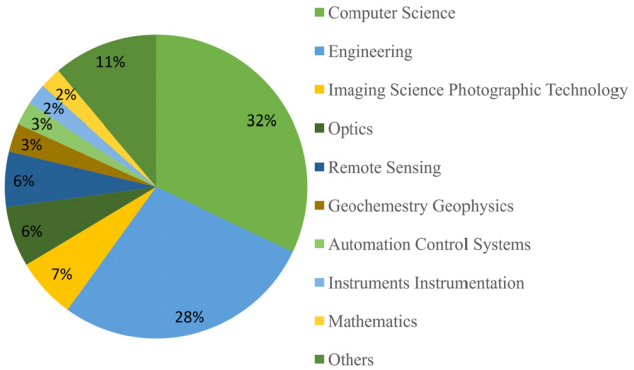

FIGURE 3. Paper distribution according to research areas.

III. MULTISPECTRAL DATABASES

This section addresses the public databases used in the most relevant papers. In a first phase, an analysis is carried out on the database frequency use. Then, an analysis of their properties enhances their differences and similarities.

Public databases (dark green bars in figure 4) are more frequently used, as they allow the comparison of different methods, making easy the researcher's task in choosing the best database for its purpose. Private databases (light blue bars in figure 4) are only used by their authors, and consequently the methods used are developed by the authors, making it difficult to compare different methods.

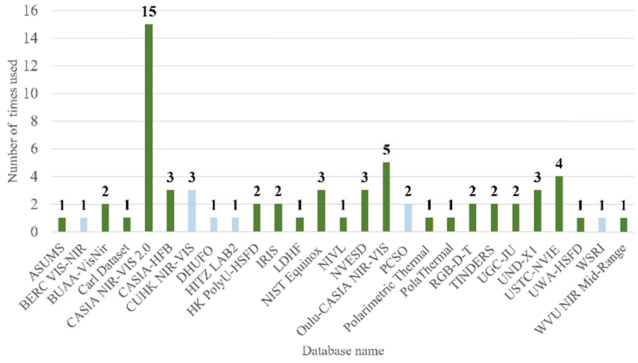

FIGURE 4. Database distribution used in the most relevant papers.

Figure 4 shows that the most used databases are the Chinese Academy of Sciences (CASIA) NIR-VIS 2.0 [7] (used 15 times in the 47 surveyed papers), the Oulu-CASIA NIR-VIS [8] (used 5 times) and the University of Science and Technology of China - Natural Visible and Infrared (USTC-NVIE) [9] (used 4 times).

The CASIA NIR-VIS 2.0 database stands out from other databases because of: (i) the protocols from the dataset are already defined (i.e., which images to use in the training and testing phase), this way it is easier to compare different methods, some databases do not have previously defined protocols, and (ii) the database is composed of two databases, the original images and the same images in the resolution 128×128 pixels with the facial detection and facial alignment already performed, with this processed subset, the dataset is more straightforward, as it is easier to implement the proposed methodology.

The databases shown in table 2 are public and are available for academic purposes. They are classified by their name, the year of creation, the spectral bands used, and the number of people in it, the number of images per person, and the number of images present in the database. Some comments were added with relevant information. It should be noted that, in the construction of this table, were not considered other types of images, such as sketches [10]–[12] or images with information about depth [13], [14].

From table 2, it can be seen that most of the databases have several years old, with an average existence of around 8 years. Another relevant detail extracted from table 2 is that the

TABLE 2. Properties of the Public Databases Used in Most Relevant Papers.

Name	Cited in	Year	Spectral band	Number of people	Images / people	Number of images	Best Rank-1	Comments
ASUMS [15]	[15]	2011	VIS, LWIR	96	6	576	97.9 [15]	With 4 luminosity variations.
BUAA-VisNir [16]	[17] [18]	2012	VIS, NIR	150	162	24 300	97.4 [17]	With 9 different facial expressions.
Carl Dataset [19]	[20]	2013	VIS, NIR, LWIR	41	180	7 380	75.6 [20]	With 3 luminosity variations and 5 different facial expressions.
CASIA NIR-VIS 2.0 [7]	[10] [12] [17] [18] [21] [22] [23] [24] [25] [26] [27] [28] [29] [30] [31]	2013	VIS, NIR	725	24	17 580	99.4 [31]	-
CASIA-HFB [32]	[23] [25] [33]	2009	VIS, NIR	100	16	1 616	95.2 [23]	With changes in facial expressions.
HK PolyU-HSFD [34]	[35] [36]	2010	VIS, NIR	25	900	22 500	99.8 [35]	With 2 luminosity variations, 3 different poses and 2 different facial expressions.
IRIS [37]	[38] [39]	-	VIS, LWIR	30	141	4 228	96.0 [39]	With 5 luminosity variations and 3 different facial expressions.
LDHF [40]	[41]	2014	VIS, NIR	100	8	800	78.0 [41]	With 2 luminosity variations and 4 different distances human-camera.
NIST Equinox [42]	[36] [43] [44]	2007	VIS, SWIR, MWIR, LWIR	95	-	-	99.6 [43]	With 3 luminosity variations and 3 different facial expressions.
NIVL [45]	[10]	2012	VIS, NIR	574	43	24 605	94.5 [10]	-
NVESD [46]	[47] [48] [49]	2013	VIS, MWIR, LWIR	50	-	-	82.3 [47]	-
Oulu-CASIA NIR-VIS [8]	[17] [18] [26] [29] [30]	2009	VIS, NIR	80	36	2 880	99.9 [18]	With 3 luminosity variations and 6 different facial expressions.
Polarimetric Thermal [50]	[50]	2019	VIS, LWIR	111	-	-	98.0 [50]	With 2 luminosity variations.
PolaThermal [51]	[10]	2016	VIS, LWIR	60	video	video	76.3 [10]	With several different facial expressions 3 different distances human-camera.
RGB-D-T [13]	[13] [14]	2016	VIS, LWIR	51	900	45 900	86.9 [14]	-
TINDERS [52]	[53] [54]	2009	VIS, NIR, SWIR	48	26	1 255	97.8 [54]	With 2 different facial expressions.
UGC-JU [55]	[56] [57]	2015	VIS, LWIR	84	39	6 552	99.2 [56]	With 22 different poses (one with glasses and 4 with occlusions) and 7 different facial expressions.
UND-X1 [58]	[43] [47] [48]	2004	VIS, LWIR	82	56	4 584	99.1 [43]	-
USTC-NVIE [9]	[22] [27] [38] [59]	2010	VIS, LWIR	215	162	34 830	97.4 [38]	With 3 luminosity variations, 9 different poses and 3 different facial expressions.
UWA-HSFD [60]	[35]	2013	VIS, NIR	70	57	3 960	99.8 [35]	-
WVU NIR Mid-Range [61]	[61]	2015	VIS, NIR, SWIR, LWIR	103	5 videos	515 videos	56.0 [61]	With 2 luminosity variations.

average number of people present in a multispectral database is 138 people, which is small number when compared to the databases containing images from visible spectrum.

According to Masi *et al.* [62] a database that has a high number of images of different people is advantageous for training deep neural networks, to cover the high variety of human appearance. Masi concluded that a database with several images of the same person with different luminosity variations and pose conditions allows a better learning through the neural network. Thus, a database composed of a larger number of images of each individual has the advantage of allowing the retraining of neural networks already trained with databases with images of several people.

Figure 5 shows the distribution of spectral bands (NIR, SWIR, MWIR, LWIR) in the public multispectral databases. This figure shows that the number of databases containing

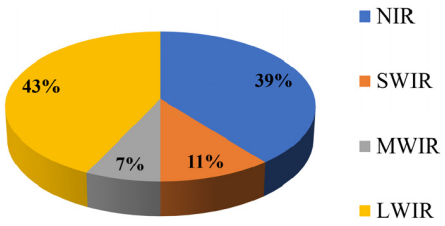

FIGURE 5. Distribution of spectral bands used in databases.

SWIR and MWIR images are very small, 11% and 7% respectively. This fact is due to the expensive price of SWIR and MWIR cameras, when compared to the NIR or LWIR cameras.

IV. PERFORMANCE EVALUATION

Facial recognition is mainly used for the identification or authentication of people [2].

Identification is the process of determining the identity of an individual, through a comparison with other identities in a database, accomplish a one-to-many (1:N) comparison. Identification is considered as a close-set problem, if we know that the person to be identified exists in the given database (i.e., there are no outsiders), and an open-set problem if it is not known if the person appears in the database.

Authentication is the process of confirming a person's identity as correct, or not, when comparing with the offered claim of identity. To prove the identity, a person must type his ID number or pass the ID card whose information validates the identity. As such, we are performing a one to one authentication (1:1).

The identification method, being automatic, does not require any user intervention. However, it has the disadvantage that if the database is large, this process can be time consuming, since it will have to go through N images.

It is necessary to use a benchmark to evaluate the performance of a given method in comparison to others. The most used methods are the Rank-N, the False Acceptance Rate (FAR), and the computational time used by the algorithm [35].

Matching performance (assigning a face image to an identity) is measured as the percentage of identification attempts for which the face image prediction is returned in the top N ranked results. Rank-1 refers to the percentage of predicted identities that return their matching as correct (predicted correctly the person identity), as the highest scoring result (the 1st result). Rank-10 refers to the percentage of face image predictions that correspond correctly to their equivalent identities in the top 10 highest-scoring results.

Rank-1 is computed by dividing the total number of images correctly identified (Correctly Identified) with the total number of identifications made (Identification Attempts):

$$\text{Rank} - 1\,(\%) = \frac{\text{CI}}{\text{IA}} \times 100 \tag{1}$$

Rank-N is an extension of rank-1, but in this case, instead of checking if the most probable image is the correct one, it is checked whether the correct image is among the most probable N images.

In closed-set identification, the Cumulative Match Characteristic (CMC) curve is frequently used [63]. This is a plot of the identification rate at rank-N, where the N values most used are 5 and 10. In open-set identification the Receiver Operating Characteristic (ROC) curve is commonly used [64], where the verification rate is plotted versus the FAR. From the ROC curve it is possible to obtain the Area Under the ROC Curve (AUC), this corresponds to the area bellow the ROC curve.

Verification rate, or true positive rate, measures the proportion of actual positives that are correctly classified.

$$\text{Verification Rate}\,(\%) = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \tag{2}$$

where TP is the number of true positives, and FN is the number of false negatives in the data.

The FAR, or false positive rate, is an empirical estimate of the probability (the percentage of times) at which the system incorrectly classifies a biometric sample belonging to the claimed identity (impostor) when the sample actually belongs to a different subject (correct person). This personification is one of the most serious biometric security error, as it provides undue authorization to users who should not have it [2]. In access control systems, FAR quantifies the probability of the biometric system, e.g. the facial recognition system, to give access to an unauthorized user:

$$\text{FAR}\,(\%) = \frac{\text{FP}}{\text{FP} + \text{TN}} \times 100\% \tag{3}$$

where FP is the number of undue authorizations (False Positive), and TN is the number of true negatives. For example, a system that contains a FAR rate of 1% imply that out of 100 classifications considered correct, 99 were truly correct, and one was incorrect.

A system with reduced FAR values is more secure, preventing impostors from entering. However, reduced FAR values are accompanied by lower verification rates. There is a trade-off between the FAR and the verification rate, being necessary to fine-tune the algorithm to comply with the application requirements of the face recognition system. More secure systems have associated a lower FAR. The most common FAR values are 1% and 0.1%.

The performance evaluation is also measured by computing the processing time of the algorithm. When several algorithms achieve similar rank values, with fixed FAR values, authors compute the time needed to identify a given number of people to show the superiority of their algorithms.

V. METHODS

This section highlights the most relevant papers in multispectral facial recognition area. These papers are grouped according to the methods used. The analysis of each method

includes a description of the different approaches adopted by each author, the databases used, the results produced, and the conclusions achieved.

During the analysis of each paper it was possible to see that there are three distinct approaches during the implementation of face recognition methods: (i) multi-channel to multi-channel, (ii) multi-channel to single-channel, (iii) and single-channel to single-channel, where a channel can be a spectral band or a spectral range inside a spectral band.

The first approach uses the same channels during the training and testing phase. Using this approach, it is possible to use all information available from all channels, having as disadvantage the higher costs of the setup.

The multi-channel to single-channel approach uses all the channels in the training phase and, in the testing phase uses only one channel. This approach is helpful to reduce costs during the implementation of the facial recognition system since we only use a single camera. The approach is also known as heterogeneous face recognition.

The last approach is the most limited one, with the advantages and limitations associated with the channel used. In this case, a single channel is used in the training and testing phase. The approach is also known as homogeneous face recognition.

Through the systematic analysis, the most relevant papers were grouped into five main methods: feature representation, coupled subspace learning, image synthesis, fusion, and deep neural networks. Table 3 shows the five methods and the papers that used those methods. The most used method is deep neural networks since it is a recent method and it has produced promising results in multispectral facial recognition systems.

TABLE 3. Most used methods and the corresponding papers.

Method	Cited in	Percentage
Feature Representation	[22] [27] [41] [53] [54] [65]	13 %
Coupled Subspace Learning	[15] [21] [33] [38] [47] [11] [66] [67] [68] [69]	21 %
Image Synthesis	[49] [50] [14] [59] [70] [18]	13 %
Fusion	[20] [36] [43] [13] [56] [57] [61] [71] [72] [39]	21 %
Deep Neural Networks	[12] [17] [23] [24] [25] [26] [28] [10] [29] [30] [35] [48] [73] [74] [31]	32 %

Figure 6 shows that, until 2017, there was a predominance of papers that use the fusion method for multispectral facial recognition. Since then, most of the papers used deep neural networks, once it provides better results.

In figure 7 it is plotted a boxplot diagram in which it shows the performance obtained in each method. This way is possible to provide a performance comparison of each method. As it is possible to see from the figure, Deep Neural Network obtains the best results, thus justifying the appearance of new

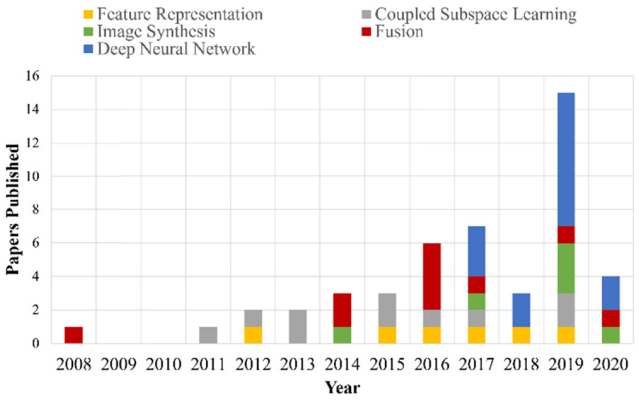

FIGURE 6. Distribution of used methods by year of publication.

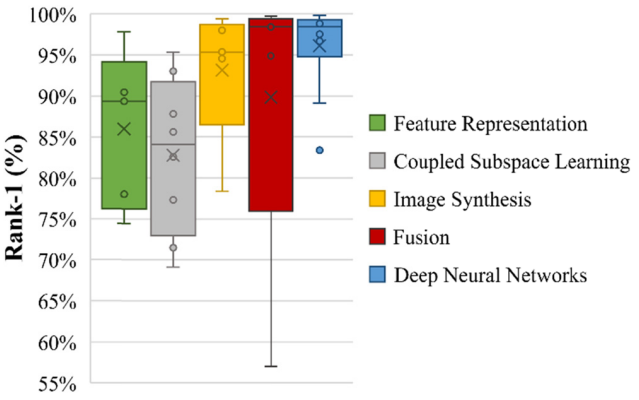

FIGURE 7. Comparison of performance by each method.

neural networks and methods within this area (also proven by figure 6).

A. FEATURE REPRESENTATION

The methods based on Feature Representation seek to extract, during the feature extraction phase, the best features that are invariant to the spectral range used in each image. By extracting the features (e.g., edges, corners, eyes, mouth, etc) we reduce the initial image information, by eliminating the irrelevant information, doing so we simplify the computation done by the classifier. In this way, it is possible to reduce the modality gap between the different spectral bands [21].

This method can be used in standalone (along with a classifier), meaning that this is the only method used to do facial recognition, or (as will be presented in the next subsections) as a basis for other methods of facial recognition [21]. One disadvantage of this method is that some features extractors, such as Local Binary Patterns (LBP), ignores the face spatial structure, crucial to obtain a good performance in heterogeneous facial recognition systems [22].

Nicolo and Schmid [54] presents a heterogeneous facial recognition system that associates an image of the SWIR spectral band to an image of the VIS spectral band. The magnitudes and phases of the Gabor filtered image are then processed by three local operators separately: Simplified Weber Local Descriptor, LBP and a generalized LBP.

Each local operator produces a histogram containing 135 bins. Afterwards, the three histograms are concatenated into a single feature vector (of histograms). To compare feature vectors extracted from the two images the authors used the symmetric Kullback–Leibler divergence distance. The proposed method was tested with TINDERS database [52] and achieved a rank-1 classification rate of 97.8%.

Cao *et al.* [53] proposed the use of Composite Multilobe Descriptor (CMLD) to extract features, using the TINDERS database [52]. To compare the features extracted from the two images the authors used the symmetric Kullback–Leibler divergence distance. A heterogeneous NIR-VIS facial recognition was performed, achieving a 91.54% verification rate at 1% FAR and a 70.14% rank-1 verification rate. Heterogeneous facial recognition was performed with the same database, but with spectral bands SWIR-VIS, obtaining the following scores: a 99.46% verification rate at 1% FAR and a 78.65% rank-1 verification rate. The authors compared their results with the ones produced with other methods, such as LBP, Gabor and HOG, and concluded that the proposed method achieves better results.

Shamia and Chandy [41] uses a combination of Histogram of Oriented Gradients (HOG) and LBP to extract facial features from NIR images to perform facial recognition at distances of 1, 60, 100 and 150 meters. To compare the features extracted from the two images the authors used the Euclidian distance. The LDHF database [40] was used, which includes VIS and NIR images captured at several distances (1, 60, 100 and 150 meters). In this database, the rank-1 score of 72%, 78% and 32% were obtained for distances of 60, 100 and 150 meters, respectively.

Peng *et al.* [22] developed a graphical representation based HFR (G-HFR) that uses a Markov network (an undirected graph whose links represent symmetrical probabilistic dependencies [75]) to represent heterogeneous image patches separately, which takes the spatial compatibility between neighboring image patches into consideration. The CASIA NIR-VIS 2.0 [7] and USTC-NVIE [9] databases were used to test the method, achieving a rank-50 score of 83.32% and 95.38% in the first and second databases, respectively.

In a second work of Peng *et al.* [27], he proposed the use of Sparse Graphical Representation based Discriminant Analysis (SGR-DA) to represent heterogeneous facial images of different modalities (different spectral bands). The adaptive sparse vectors were generated through Markov networks, and are considered very effective for heterogeneous facial recognition. With the use of spatial partition strategy, the discrimination of heterogeneous facial images was improved. The proposed method processed the CASIA NIR-VIS 2.0 [7] database and the USTC-NVIE [9] database achieving a rank-50 score of 87.84% and 93.08%, for the first and second database, respectively. The author also compared his method with G-HFR [22] and concluded that the SGR-DA obtained an increase of 4.52% and a decrease of 2.30%, in rank-50 score, in the first and second databases, respectively.

B. COUPLED SUBSPACE LEARNING

Coupled Subspace Learning methods project the features of different spectral bands in a common subspace. This subspace allows to identify the most relevant information, using the redundant information that is common to different spectral bands. With this approach, it is possible to reduce the difference between the images from distinct spectral bands. However, the discriminative power of the learned common space is heavily reduced if the modality gap is large [21]. This method has the disadvantage that during the projection of the image on the subspace there is always information that is discarded, and may decrease the performance of the facial recognition system [22].

Huang *et al.* [69] proposed a new method, the Discriminative Spectral Regression, that maps the facial images of VIS and NIR in a common discriminative subspace. With the proposed method, it was achieved a rank-1 score of 95.33% on the CASIA-HFB database [32].

Jin *et al.* [21] presented a method of feature extraction, the Coupled Discriminative Feature Learning, and applied it to heterogeneous facial recognition. This method maximizes the interclass variations and minimize the intraclass variations. The CASIA NIR-VIS 2.0 [7] database was used to perform tests and obtained the following scores for rank-1, verification rate with FAR at 1%, and verification rate with FAR at 0.1%: 71.5%, 55.1%, 67.7%, respectively.

Li *et al.* [11] proposed the Mutual Component Analysis (MCA) to study the features that are mutual (common) to the two types of images, in this case VIS and LWIR. MCA was tested on CASIA NIR-VIS 2.0 [7], obtaining a rank-1 score of 69.10%, a verification rate of 76.21% and 92.71%, for a FAR of 0.1% and 1%, respectively.

Hu *et al.* [47] used the Difference of Gaussian (DoG) filter (a band-pass filter that enhances the edges of the images and removes noise) in the preprocessing phase to reduce luminosity variations between VIS images and to reduce local variations in LWIR images. To extract image features, the Histogram of Oriented Gradients with a size of 16×16 pixels was used. The pre-processing and feature extraction phases were designed to reduce the modality gap between VIS and LWIR images. This approach facilitates the one-versus-all facial recognition model based on the Partial Least Squares (PLS) model. The NVESD [46] database was used to perform heterogeneous facial recognition LWIR-VIS at distances of 1 m, 2 m and 4 m, producing a rank-1 score of 82.3%, 70.8% and 33.3%, respectively. In this database, heterogeneous facial recognition MWIR-VIS was also performed at distances of 1 m, 2 m and 4 m, achieving a rank-1 score of 92.7%, 81.3% and 64.6%, respectively. The UND-X1 [58] database was used for heterogeneous facial recognition LWIR-VIS and produced a rank-1 score of 72.7%.

Gong *et al.* [67] applied a new feature descriptor called Common Encoding Feature Discriminant Approach to perform heterogeneous facial recognition, reducing the large

modality gap between the NIR and VIS images. The CASIA NIR-VIS 2.0 [7] database was used to perform tests, in which a rank-1 score of 85.6% was obtained.

Lei *et al.* [33] proposed the Coupled Discriminant Analysis Method for heterogeneous facial recognition between VIS and NIR images. They propose two implementations of Locality Constraint in Kernel Space (LCKS): the LCKS coupled discriminant analysis (LCKS-CDA) and the LCKS coupled spectral regression (LCKS-CSR). The CASIA-HFB database [32] was used to test both methods. The LCKS-CDA method obtained a rank-1 score of 73.18%, a verification rate with a FAR at 1% and 0.1% of 31.21% and 16.61%, respectively. The LCKS-CSR method obtained a rank-1 score of 81.43%, a verification rate with a FAR at 1% and 0.1% of 54.81% and 35.69%, respectively.

Klare and Jain [66] proposed the use of the Prototype Random Subspace (P-RS) to perform heterogeneous facial recognition between VIS-NIR images and VIS-LWIR images. Using P-RS it is possible to use different feature descriptors to represent the probe and gallery images. When compared with the FaceVACS[1] (a commercial off-the-shelf facial recognition system), the proposed method produced better results for LWIR images than for NIR. Using the CBSR database [76] the proposed method (and the commercial FaceVACS method) produced 87.8% (87.8%) rank-1 scores, 95.8% (92.0%) verification rate at 0.1% FAR and 98.2% (93.7%) verification rate at 1% FAR.

Bhowmik *et al.* [38] employed a variant of the Independent Component Analysis (ICA), applying the Logarithmic transformation on the basic ICA, named as Log-ICA. Two architectures were developed, Log-ICA I and Log-ICA II; the last one presented better results. The author concluded that the proposed method achieved good results in heterogeneous facial recognition, it can also be applied in facial expression recognition and recognize facial images with noise. Two VIS-LWIR databases were used, IRIS [37] and USTC-NVIE [9]. In the first database, the rank-1 score of 88.18% and 90.51% was obtained for Log-ICA I and Log-ICA II, respectively. In the second database, the rank-1 score of 95.92% and 97.4% was achieved for Log-ICA I and Log-ICA II, respectively.

C. IMAGE SYNTHESIS

The image synthesis methods transform an image from one spectral band to another, allowing to compare two images more easily. These methods enable to synthesize an image in the visible spectral range, using as starting point an image from another spectral band (e.g. LWIR spectral band). The main advantage of image synthesis is that, as soon as a LWIR image is synthesized as a VIS image, it is possible to apply facial recognition method designed for VIS images [66].

The main problem with this method is that image synthesis is a difficult process and, in most cases, the performance of the facial recognition system is highly dependent on the accuracy of the synthesized image [22].

Osia and Bourlai [49] used LWIR images to produce, through synthesis, equivalent images in the VIS spectral band. In order to demonstrate the advantages of the proposed method, LBP was used to perform facial recognition on the synthesized images. The method was tested on the NVESD database [46] that includes VIS, MWIR and LWIR images. Heterogeneous facial recognition MWIR-VIS was performed resulting in 75.32% rank-1 score. Heterogeneous facial recognition LWIR-VIS was also done and achieved 81.41% rank-1 score.

Zhang *et al.* [50] applied a new multi-level dense-residual fusion Generative Adversarial Networks (GAN) to synthesize VIS images from LWIR images, producing better qualitative results, when comparing the synthesized image with the original image from VIS. The author creates a new database, Polarimetric Thermal with VIS and LWIR images. His method applied to this database achieved an Area Under ROC Curve value of 98%.

Cao *et al.* [59] used a data augmentation-based joint learning to introduce synthesized images into the learning process. The aggregated data (the original images plus the synthesized images) augments the size of the intraclass set, which may increase discriminative information. Using the USTC-NVIE database [9] composed of VIS and LWIR images, it was achieved a 95.35% rank-50 score.

Litvin *et al.* [14] proposed the use of a convolutional neural network to perform the synthesis of LWIR images to VIS images. He modified the FusionNet architecture [77] and its training algorithm to decrease overfitting, adding dropout after bridge connections, randomized leaky Rectified Linear Units (ReLUs) and orthogonal regularization. The method was tested for each of the three image variations present in the RGB-D-T [13] database: pose, expressions and luminosity variations; producing a rank-1 score of 86.94%, 97.52% and 99.19%, respectively, for each variation.

He *et al.* [18] proposed the Adversarial Cross-spectral Face Completion (CFC) that uses a generative adversarial network that synthesizes VIS images from NIR images. This approach is different from other methods, since it uses an inpainting component that synthesizes and inpaints VIS image textures from NIR image textures. The method converts any pose in NIR images to a frontal pose in VIS images, resulting in paired NIR and VIS textures. Then, a warping procedure is applied to integrate the two components into an end-to-end deep network. The last step is to perform facial recognition on the synthesized images using the LightCNN [78]. The CFC was tested on three databases: (i) CASIA NIR-VIS 2.0 [7], (ii) Oulu-CASIA NIR-VIS [8] and (iii) BUAA-VisNir [16], achieving the following rank-1 scores: 98.6%, 99.9% and 99.7%, respectively for each database. Using the same databases, the CFC produced the following verification rate for a FAR of 1% and 0.1%: (i) 99.2% and 97.3%, (ii) 98.1% and 90.7%, (iii) 98.7% and 97.8%, respectively for the three databases.

[1] https://www.cognitec.com/facevacs-videoscan.html

D. FUSION

The most relevant methods for image fusion applied to facial recognition are feature fusion and score fusion. Feature fusion combines the features of several image sources, acquired through a feature extractor, into a feature vector. These features include information about, e.g. edges, corners, lines and textures, are computed and are concatenated into a single feature vector, to be used to perform segmentation or facial detection [79]. Feature fusion is employed to reduce the dimensionality of the final feature vector [2].

The score fusion improves the overall performance of the classification, combining the output of several classifiers into a global classifier. The most used method for score fusion is majority voting, in which the classification obtained for each classifier is taken into account and then a vote is used, which consists of finding out which classification occurs more frequently, assigning it to the global classifier. Another used method in score fusion is the adapted weighted, where each classifier is assigned a dynamic weight. In this manner classifiers that exhibit low performance will have assigned a low weight and consequently, less importance in the global classification.

The application of image fusion in facial recognition systems has several advantages, such as, the reduction of error rate and the cost of implementation through the application of several low-cost cameras, instead of a single more expensive camera [80].

Singh *et al.* [43] performs an image fusion of the VIS and LWIR images using a Granular Support Vector Machine[2] to compute both dynamically and locally the weights to generate the fused image. The fused image is used to obtain the scores, a 2D Log-Polar Gabor Transform is used to extract the global facial features, and the Local Binary Pattern (LBP) is used to extract the local facial features. Then, the score fusion is applied. Tests are performed on the UND-X1 [58] and NIST Equinox [42] databases achieving a verification rate of 99.91% and 99.54%, respectively, with a FAR at 0.01%.

Seal *et al.* [56] proposed a VIS and LWIR image fusion algorithm, which uses translation the invariant wavelet transform and Random Forests. This algorithm combines the useful information present in visible and thermal images using image entropy and achieved rank-1 score of 99.07% in the UGC-JU database [55].

Bourlai *et al.* [61] used a Multi-Feature Scenario Dependent Fusion (MFSDF). First, the features were extracted using LBP, GABOR and HOG. Then, a fusion of scores was made with eleven scenarios: the three individual features (LBP, GABOR and HOG), six combinations of two features (e.g. LBP + GABOR), the sum of the three individual features, and for the last, a weighted fusion scheme where weights were assigned to each descriptor based on the performance scores (distance scores). Afterwards an empirical evaluation was done to determine which scenario obtains the best rank-1 score. The MFSDF was compared with other face recognition methods, such as, PCA and linear discriminant analysis (LDA). The WVU NIR Mid-Range database [61] was used.

Seal *et al.* [57] applied a fusion process that calculates the weighted sum of LWIR and VIS facial information with two weighting factors. In order to evaluate the method, initially, two independent facial recognition were performed, the first one on the VIS image and the second one on the LWIR image; each one produced a score that is equal to the probability of correct classification for each image. In a second phase, facial recognition is done using the fused image created by using the proposed fusion process where the weights are the scores previously computed. The UGC-JU [55] was used and produced an accuracy value of 98.42%.

Simón *et al.* [13] used LBP, HOG, HAAR and HOGOM to extract the features of the VIS, LWIR and depth images, then concatenated into a single feature vector for training the Weighted-Nearest Neighbor classifier (W-kNN). The intuition behind W-kNN is to give more weight to the points which are nearby and less weight to the points that are farther away. After that, a Convolutional Neural Network (CNN) processed each initial image. The next step was merging the three images. Finally, the final classifier was obtained by fusing W-kNN and CNN classifiers with different weights. The author introduces the RGB-D-T database [13], composed of VIS, LWIR and depth images.

Kanmani and Narasimhan [39] proposed three optimization based fusion methods that aid the heterogeneous face recognition problem. In the first and second methods, the input image was decomposed into high and low frequency coefficients through dual tree discrete wavelet transform. Then a population-based optimization technique [81] was applied to find the optimal weights to perform the fusion of VIS and LWIR images. The third method applies a Self Tuning Particle Swarm Optimization to prevent premature convergence of the particle swarm. It uses a curvelet transform to perform image decomposition preserving the edges along the curves, and to improve the searching of optimal weight coefficients, a Brain storm optimization algorithm is used for optimization. Using the IRIS [37] database, was achieved a rank-1 score of 94.17%, 94.50% and 96.00% for the first, second and third methods, respectively.

E. DEEP NEURAL NETWORKS

Artificial neural networks, inspired on human's neural networks, had produced promising results, surpassing the methods previously described.

The use of neural networks in facial recognition is relatively simple. Initially, an image is sent to a neural network that extracts a set of features. When this network receives another image from the same person, it must produce a set of very similar features, whereas the opposite should happen when the input is an image of a different person. The neural networks most used today are the deep neural networks,

[2]In granular computing, the information is divided into sub-problems, called granules, and these sub-problems are solved individually at different granularity levels.

which comprise a higher number of decision layers than traditional artificial neural networks.

The current Deep Neural Networks have as disadvantage the training time, which is very dependent on the Graphic Processing Unit (GPU) performance. Sometimes, neural networks are compared not only by the score obtained, since they may be very similar, but also by the computation time of the training and the classification stages [35] [12].

Sarfraz and Stiefelhagen [48] reduces the modality gap between LWIR and VIS images using a deep neural network that captures the non-linear relationship between the two modalities. When compared with Partial Least Squares (PLS) based models, the proposed method achieves an increase in rank-1 of 10% in UND-X1 [58] database and 15% to 30% in NVESD [46] database.

Jin *et al.* [23] used a Multi-Task Clustering Extreme Learning Machine (MTC-ELM) in order to improve the feature learning between the two spectral bands: VIS and NIR images. The MTC-ELM was designed to classify large amounts of data. The CASIA HFB [32] and CASIA NIR-VIS 2.0 [7] databases were used and produced a rank-1 score of 95.2% and 89.1%, respectively.

Oh *et al.* [24] proposed the use of a single hidden-layer Gabor-based network to perform heterogeneous facial recognition. When applied to CASIA NIR-VIS 2.0 [7] database achieved a rank-1 score of 97.52%.

Guei and Akhloufi [25] applied the Deep Convolutional Generative Adversarial Network (DCGAN), which increased the size of the images while preserving important facial details. The initial images were 16×16, and the final images were 64×64. The CASIA NIR-VIS 2.0 [7] database was used to validate the method.

Hu *et al.* [26] developed the Multiple Deep Network with Scatter Loss and Diversity Combination (MDNDC) to reduce intra-class variations and increase inter-class variations. With scatter loss, it was possible to reduce the modality gap, thus preserving the information of the person to be identified. The Multiple Deep Network extracted the features and the Diversity Combination (DC) was used to adaptively adjust the weights of each deep net. MDNDC was tested on the CASIA NIR-VIS 2.0 [7] database obtaining a rank-1 score of 98.9%, a verification rate of 99.6% and 97.6%, for a FAR of 1% and 0.1%, respectively. MDNDC was also tested on Oulu-CASIA NIR-VIS [8] database, producing a rank-1 score of 99.8%, a verification rate of 88.1% and 65.3%, for a FAR of 1% and 0.1%, respectively.

Peng *et al.* [28] proposed the use of a deep local descriptor learning framework applied in heterogeneous facial recognition systems, which was able to learn discriminant and compact local information directly from facial images. A novel cross-modality enumeration loss is proposed to eliminate the modality gap on local patch level, which is then integrated into a convolutional neural network for deep local descriptor extraction. The method was tested on CASIA NIR-VIS 2.0 [7] database achieving a 96.68% rank-1 score.

Pereira *et al.* [10] adapted low-level features from deep convolutional neural networks (DCNN) to Domain Specific Units (DSU). These units behave as low-level feature detectors that are domain specific. While the low-level layers are adapted, the networks share the same set of high-level features from the source domain without re-training them. The author used two different methods to train DCNN, the Siamese and the Triplet Neural Network. DCNN was tested in three databases: (i) CASIA NIR-VIS 2.0 [7], (ii) NIVL [45] and (iii) PolaThermal [51]. The following rank-1 scores for Siamese and Triplet Neural Network were obtained from these databases: (i) 96.3% and 90.1%, (ii) 94.5% and 92.2%, (iii) 76.3% and 50.9%, respectively.

He *et al.* [17] implemented the Wasserstein distance in a convolutional neural network, called Wasserstein CNN (WCNN). The Wasserstein distance is the distance between two probability distributions in a given space, and was used to reduce the modality gap between the VIS and NIR images. WCNN was tested on three databases: (i) CASIA NIR-VIS 2.0 [7], (ii) Oulu-CASIA NIR-VIS [8] and (iii) BUAA-VisNir [16], achieving the following rank-1 scores: 98.7%, 98.0% and 97.4%, respectively for each database. Using the same databases, the WCNN produced the following verification rate for a FAR of 1% and 0.1%: (i) 99.5% and 98.4%, (ii) 81.5% and 54.6%, (iii) 96.0% and 91.9%, respectively for the three databases.

Hu and Hu [29] proposed the use of a new heterogeneous facial recognition method, the Discriminant Deep Feature Learning Based on Joint Supervision Loss and Multi-layer Feature Fusion (DDFLJM). The author also made a comparative study with WCNN [17] using (i) CASIA NIR VIS 2.0 [7] database and (ii) Oulu-CASIA NIR-VIS [8] database. The results were the following rank-1 scores (i) 98.8% and (ii) 99.3%, the verification rate at 1% FAR of (i) 99.4% and (ii) 86.1%, and also the verification rate at 0.1% FAR of (i) 97.3% and (ii) 63.5%, respectively, for the first and second databases. The author also studied the cost functions using CASIA NIR-VIS 2.0 [7] database and concluded that the loss function Scatter Loss (SL) achieved better results (98.5%) when compared with Softmax results (84.5%), both values for the rank-1 score.

In another work of Peng *et al.* [30], he proposed a high-dimensional deep local representation re-ranking method to perform heterogeneous VIS-NIR facial recognition. The ranking results are refined through the use of a Locally Linear Re-Ranking (LLRe-Rank) technique. Tests were performed on (i) CASIA NIR VIS 2.0 [7] database and (ii) Oulu-CASIA NIR-VIS [8] database and achieved the rank-1 scores of 98.7% and 98.9%, respectively, and the following verification rate for a FAR of 1% and 0.1%: (i) 99.4% and 96.6% for the first database, (ii) 86.1% and 61.7% for the second database, respectively.

Wu *et al.* [35] developed a deep convolutional neural network for multispectral facial recognition that explores the intraspectrum discriminant information and interspectrum correlation information. The author calls this network

intraspectrum discrimination and inter-spectrum correlation analysis deep network (IDICN) and performed tests on two databases, HK PolyU-HSFD [34] and UWA-HSFD [60], achieving a 99.76% and 99.85% rank-1 scores, respectively.

Bae *et al.* [31] introduced two modules to improve heterogeneous facial recognition, with the final image being the VIS spectral band. The first module consists of (i) the pre-processing chain, to guarantee that the range of intensity is similar between the translated image and the target image; (ii) the CycleGAN to learn the mapping between an input NIR image and an output VIS image using a training set of aligned image pairs; (iii) the Siamese network to simultaneously learn a latent space by adding constrains in the learning procedure of mapping functions. In the second module, images of the training database and its translated images are used to fine-tune the pre-trained backbone model (a ResNet-101 [82] trained with the Celeb-1M database [83]) to obtain a discriminative 512-dimensional embedding vector. In the testing phase, the CASIA NIR-VIS 2.0 [7] database was used. Without the pre-processing module, it was achieved a rank-1 score of 99.07%, and a verification rate of 98.67% for a FAR at 0.1%. With the pre-processing module, the authors obtained better results: a rank-1 score of 99.40% and a verification rate of 98.74% for a FAR at 0.1%.

VI. CONCLUSION

After a systematic study, it was possible to conclude that the most used methods for facial recognition and the ones that achieved best results are based on neural networks. In fact, 36% of the most relevant papers use neural networks as a multispectral facial recognition method. It should be noted that since 2019 there was a reappearance of the image synthesis methods due to the use of neural networks, mainly the Generative Adversarial Networks (GAN), to carry out the image synthesis.

It was also possible to conclude that the most used metric to compare methods in different databases is the rank-1.

The main problem of current multispectral facial recognition systems is the availability of multispectral databases. Through this work, it was observed that the most widely used public database is CASIA NIR-VIS 2.0 [7]. When compared to databases with images from the visible spectral band, the current public multispectral databases are very small (in terms of total number of images), which may lead to an overfitting of the neural network during the training phase. The multispectral databases have several limitations, such as: the reduced number of images; the fact that there is no public database with facial images of the same individual at different spectral bands (e.g. VIS, NIR, SWIR and LWIR); the non-existence of pose, luminosity and distance variations among images in the same database.

In general, multispectral facial recognition methods achieve better performance when compared to facial recognition systems that use only images from visible spectral band. Through the use of multispectral images in facial recognition, it is possible to overcome some characteristic gaps in the spectral bands. As is the case with LWIR spectral band that, because they are not influenced by differences in luminosity, are able to complement the VIS images that are, as several authors had stated in their works [39], [43], [49].

However, the use of deep neural networks as a method to perform multispectral facial recognition is still limited due to the reduced number of images (and people) in the current multispectral databases. Nevertheless, deep neural networks are most used methods to perform multispectral facial recognition, able to produce very promising results. When applied to CASIA NIR-VIS 2.0 [7] database the best results are up to 99.4% rank-1 score.

Multispectral facial recognition still has plenty of space to evolve and improve. The main targets of multispectral facial recognition systems continue to be security and surveillance, especially in critical locations, such as airports or military classified areas.

REFERENCES

[1] W. Zhang, X. Zhao, J.-M. Morvan, and L. Chen, "Improving shadow suppression for illumination robust face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 3, pp. 611–624, Mar. 2019.

[2] A. K. Jain, A. A. Ross, and K. Nandakumar, *Introduction to Biometrics*. New York, NY, USA: Springer-Verlag, 2011, doi: 10.1007/978-0-387-77326-1.

[3] A. D'Amico, C. Natale, F. Castro, S. Iarossi, A. Catini, and E. Martinelli, "Volatile compounds detection by IR acousto-optic detectors," in *Unexploded Ordnance Detection and Mitigation* (NATO Science for Peace and Security Series B: Physics and Biophysics), J. Byrnes, Ed. Dordrecht, The Netherlands: Springer, 2009.

[4] K. R. Kakkirala, S. R. Chalamala, and S. K. Jami, "Thermal infrared face recognition: A review," in *Proc. UKSim-AMSS 19th Int. Conf. Comput. Modeling Simulation (UKSim)*, Apr. 2017, pp. 55–60.

[5] R. Munir and R. A. Khan, "An extensive review on spectral imaging in biometric systems: Challenges & advancements," *J. Vis. Commun. Image Represent.*, vol. 65, Dec. 2019, Art. no. 102660.

[6] M. Kristo and M. Ivasic-Kos, "An overview of thermal face recognition methods," in *Proc. 41st Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, May 2018, pp. 1098–1103.

[7] S. Z. Li, D. Yi, Z. Lei, and S. Liao, "The CASIA NIR-VIS 2.0 face database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2013, pp. 348–353.

[8] G. Zhao, X. Huang, M. Taini, S. Z. Li, and M. Pietikäinen, "Facial expression recognition from near-infrared videos," *Image Vis. Comput.*, vol. 29, no. 9, pp. 607–619, Aug. 2011.

[9] 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 Trans. Multimedia*, vol. 12, no. 7, pp. 682–691, Nov. 2010.

[10] T. de Freitas Pereira, A. Anjos, and S. Marcel, "Heterogeneous face recognition using domain specific units," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 7, pp. 1803–1816, Jul. 2019.

[11] Z. Li, D. Gong, Q. Li, D. Tao, and X. Li, "Mutual component analysis for heterogeneous face recognition," *ACM Trans. Intell. Syst. Technol.*, vol. 7, no. 3, pp. 1–23, Apr. 2016.

[12] W. Hu and H. Hu, "Fine tuning dual streams deep network with multi-scale pyramid decision for heterogeneous face recognition," *Neural Process. Lett.*, vol. 50, no. 2, pp. 1465–1483, Oct. 2019.

[13] M. O. Simón, "Improved RGB-DT based face recognition," *IET Biometrics*, vol. 5, no. 4, pp. 297–303, 2016.

[14] A. Litvin, K. Nasrollahi, S. Escalera, C. Ozcinar, T. B. Moeslund, and G. Anbarjafari, "A novel deep network architecture for reconstructing RGB facial images from thermal for face recognition," *Multimedia Tools Appl.*, vol. 78, no. 18, pp. 25259–25271, Sep. 2019.

[15] Y. Zheng, "Orientation-based face recognition using multispectral imagery and score fusion," *Opt. Eng.*, vol. 50, no. 11, Nov. 2011, Art. no. 117202.

[16] D. Huang, J. Sun, and Y. Wang, "The BUAA-VisNir face database instructions," Beihang Univ., Beijing, China, Tech. Rep. IRIP-TR-12-FR-001, 2012.
[17] R. He, X. Wu, Z. Sun, and T. Tan, "Wasserstein CNN: Learning invariant features for NIR-VIS face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1761–1773, Jul. 2019.
[18] R. He, J. Cao, L. Song, Z. Sun, and T. Tan, "Adversarial cross-spectral face completion for NIR-VIS face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 5, pp. 1025–1037, May 2020.
[19] V. Espinosa-Duró, M. Faundez-Zanuy, and J. Mekyska, "A new face database simultaneously acquired in visible, near-infrared and thermal spectrums," *Cognit. Comput.*, vol. 5, no. 1, pp. 119–135, Mar. 2013.
[20] C. Chen and A. Ross, "Matching thermal to visible face images using hidden factor analysis in a cascaded subspace learning framework," *Pattern Recognit. Lett.*, vol. 72, pp. 25–32, Mar. 2016.
[21] Y. Jin, J. Lu, and Q. Ruan, "Coupled discriminative feature learning for heterogeneous face recognition," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 3, pp. 640–652, Mar. 2015.
[22] C. Peng, X. Gao, N. Wang, and J. Li, "Graphical representation for heterogeneous face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 2, pp. 301–312, Feb. 2017.
[23] Y. Jin, J. Li, C. Lang, and Q. Ruan, "Multi-task clustering ELM for VIS-NIR cross-modal feature learning," *Multidimensional Syst. Signal Process.*, vol. 28, no. 3, pp. 905–920, Jul. 2017.
[24] B.-S. Oh, K. Oh, A. B. J. Teoh, Z. Lin, and K.-A. Toh, "A Gabor-based network for heterogeneous face recognition," *Neurocomputing*, vol. 261, pp. 253–265, Oct. 2017.
[25] A.-C. Guei and M. Akhloufi, "Deep learning enhancement of infrared face images using generative adversarial networks," *Appl. Opt.*, vol. 57, no. 18, p. D98, 2018.
[26] W. Hu, H. Hu, and X. Lu, "Heterogeneous face recognition based on multiple deep networks with scatter loss and diversity combination," *IEEE Access*, vol. 7, pp. 75305–75317, 2019.
[27] C. Peng, X. Gao, N. Wang, and J. Li, "Sparse graphical representation based discriminant analysis for heterogeneous face recognition," *Signal Process.*, vol. 156, pp. 46–61, Mar. 2019.
[28] C. Peng, N. Wang, J. Li, and X. Gao, "DLFace: Deep local descriptor for cross-modality face recognition," *Pattern Recognit.*, vol. 90, pp. 161–171, Jun. 2019.
[29] W. Hu and H. Hu, "Discriminant deep feature learning based on joint supervision loss and multi-layer feature fusion for heterogeneous face recognition," *Comput. Vis. Image Understand.*, vol. 184, pp. 9–21, Jul. 2019.
[30] C. Peng, N. Wang, J. Li, and X. Gao, "Re-ranking high-dimensional deep local representation for NIR-VIS face recognition," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4553–4565, Sep. 2019.
[31] H. B. Bae, T. Jeon, Y. Lee, S. Jang, and S. Lee, "Non-visual to visual translation for cross-domain face recognition," *IEEE Access*, vol. 8, pp. 50452–50464, 2020.
[32] S. Z. Li, Z. Lei, and M. Ao, "The HFB face database for heterogeneous face biometrics research," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2009, pp. 1–8.
[33] Z. Lei, S. Liao, A. K. Jain, and S. Z. Li, "Coupled discriminant analysis for heterogeneous face recognition," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 6, pp. 1707–1716, Dec. 2012.
[34] W. Di, L. Zhang, D. Zhang, and Q. Pan, "Studies on hyperspectral face recognition in visible spectrum with feature band selection," *IEEE Trans. Syst., Man, Cybern., A, Syst. Humans*, vol. 40, no. 6, pp. 1354–1361, Nov. 2010.
[35] F. Wu, X.-Y. Jing, X. Dong, R. Hu, D. Yue, and L. Wang, "Intraspectrum discrimination and interspectrum correlation analysis deep network for multispectral face recognition," *IEEE Trans. Cybern.*, vol. 50, no. 3, pp. 1009–1022, Mar. 2020.
[36] H. Zhao and S. Sun, "Sparse tensor embedding based multispectral face recognition," *Neurocomputing*, vol. 133, pp. 427–436, Jun. 2014.
[37] B. Abidi, S. Huq, and M. Abidi, "Fusion of visual, thermal, and range as a solution to illumination and pose restrictions in face recognition," in *Proc. 38th Annu. Int. Carnahan Conf. Secur. Technol.*, Oct. 2004, pp. 325–330.
[38] M. K. Bhowmik, P. Saha, A. Singha, D. Bhattacharjee, and P. Dutta, "Enhancement of robustness of face recognition system through reduced Gaussianity in log-ICA," *Expert Syst. Appl.*, vol. 116, pp. 96–107, Feb. 2019.
[39] M. Kanmani and V. Narasimhan, "Optimal fusion aided face recognition from visible and thermal face images," *Multimedia Tools Appl.*, vol. 79, pp. 17859–17883, Feb. 2020.
[40] D. Kang, H. Han, A. K. Jain, and S.-W. Lee, "Nighttime face recognition at large standoff: Cross-distance and cross-spectral matching," *Pattern Recognit.*, vol. 47, no. 12, pp. 3750–3766, Dec. 2014.
[41] D. Shamia and D. A. Chandy, "Intelligent system for cross-spectral and cross-distance face matching," *Comput. Electr. Eng.*, vol. 71, pp. 915–924, Oct. 2018.
[42] G. Bebis, A. Gyaourova, S. Singh, and I. Pavlidis, "Face recognition by fusing thermal infrared and visible imagery," *Image Vis. Comput.*, vol. 24, no. 7, pp. 727–742, Jul. 2006.
[43] R. Singh, M. Vatsa, and A. Noore, "Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition," *Pattern Recognit.*, vol. 41, no. 3, pp. 880–893, Mar. 2008.
[44] S. Sun, H. Zhao, and B. Jin, "Robust tensor preserving projection for multispectral face recognition," *Math. Problems Eng.*, vol. 2014, Aug. 2014, Art. no. 597245.
[45] 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 *Proc. IEEE 7th Int. Conf. Biometrics Theory, Appl. Syst. (BTAS)*, Sep. 2015, pp. 1–8.
[46] K. A. Byrd, "Preview of the newly acquired NVESD-ARL multimodal face database," *Proc. SPIE*, vol. 8734, Mar. 2013, Art. no. 018734.
[47] S. Hu, J. Choi, A. L. Chan, and W. R. Schwartz, "Thermal-to-visible face recognition using partial least squares," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 32, no. 3, pp. 431–442, 2015.
[48] M. S. Sarfraz and R. Stiefelhagen, "Deep perceptual mapping for cross-modal face recognition," *Int. J. Comput. Vis.*, vol. 122, no. 3, pp. 426–438, May 2017.
[49] N. Osia and T. Bourlai, "Bridging the spectral gap using image synthesis: A study on matching visible to passive infrared face images," *Mach. Vis. Appl.*, vol. 28, nos. 5–6, pp. 649–663, Aug. 2017.
[50] 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," *Int. J. Comput. Vis.*, vol. 127, nos. 6–7, pp. 845–862, Jun. 2019.
[51] 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 *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2016, pp. 187–194.
[52] R. B. Martin, M. Sluch, K. M. Kafka, R. Ice, and B. E. Lemoff, "Active-SWIR signatures for long-range night/day human detection and identification," *Proc. SPIE*, vol. 8734, May 2013, Art. no. 87340J.
[53] Z. Cao, N. A. Schmid, and T. Bourlai, "Composite multilobe descriptors for cross-spectral recognition of full and partial face," *Opt. Eng.*, vol. 55, no. 8, 2016, Art. no. 083107.
[54] F. Nicolo and N. A. Schmid, "Long range cross-spectral face recognition: Matching SWIR against visible light images," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 6, pp. 1717–1726, Dec. 2012.
[55] A. Seal, D. Bhattacharjee, M. Nasipuri, and D. K. Basu, "UGC-JU face database and its benchmarking using linear regression classifier," *Multimedia Tools Appl.*, vol. 74, no. 9, pp. 2913–2937, May 2015.
[56] A. Seal, D. Bhattacharjee, and M. Nasipuri, "Human face recognition using random forest based fusion of à-trous wavelet transform coefficients from thermal and visible images," *AEU-Int. J. Electron. Commun.*, vol. 70, no. 8, pp. 1041–1049, 2016.
[57] A. Seal, D. Bhattacharjee, M. Nasipuri, C. Gonzalo-Martin, and E. Menasalvas, "Fusion of visible and thermal images using a directed search method for face recognition," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 31, no. 4, Apr. 2017, Art. no. 1756005.
[58] X. Chen, P. J. Flynn, and K. W. Bowye, "Visible-light and Infrared Face Recognition," in *Proc. ACM Workshop Multimodal User Authentication*, 2003, pp. 48–55.
[59] B. Cao, N. Wang, J. Li, and X. Gao, "Data augmentation-based joint learning for heterogeneous face recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 6, pp. 1731–1743, Jun. 2019.
[60] M. Uzair, A. Mahmood, and A. Mian, "Hyperspectral face recognition using 3D-DCT and partial least squares," in *Proc. Brit. Mach. Vis. Conf.*, 2013, pp. 1–10.
[61] T. Bourlai, N. Mavridis, and N. Narang, "On designing practical long range near infrared-based face recognition systems," *Image Vis. Comput.*, vol. 52, pp. 25–41, Aug. 2016.

[62] I. Masi, Y. Wu, T. Hassner, and P. Natarajan, "Deep face recognition: A survey," in *Proc. 31st SIBGRAPI Conf. Graph., Patterns Images (SIBGRAPI)*, Oct. 2018, pp. 471–478.
[63] Y.-B. Zhao, J.-W. Lin, Q. Xuan, and X. Xi, "HPILN: A feature learning framework for cross-modality person re-identification," *IET Image Process.*, vol. 13, no. 14, pp. 2897–2904, Dec. 2019.
[64] M. Song, X. Shang, and C.-I. Chang, "3-D receiver operating characteristic analysis for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 8093–8115, Nov. 2020.
[65] N. Short, S. W. Hu, P. Gurram, K. Gurton, and A. Chan, "Improving cross-modal face recognition using polarimetric imaging," *Opt. Lett.*, vol. 40, no. 6, pp. 882–885, 2015.
[66] B. F. Klare and A. K. Jain, "Heterogeneous face recognition using kernel prototype similarities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1410–1422, Jun. 2013.
[67] D. Gong, Z. Li, W. Huang, X. Li, and D. Tao, "Heterogeneous face recognition: A common encoding feature discriminant approach," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2079–2089, May 2017.
[68] S. A. Angadi and S. M. Hatture, "Face recognition through symbolic modeling of face graphs and texture," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 33, no. 12, Nov. 2019, Art. no. 1956008.
[69] X. Huang, Z. Lei, M. Fan, X. Wang, and S. Z. Li, "Regularized discriminative spectral regression method for heterogeneous face matching," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 353–362, Jan. 2013.
[70] K. P. Gurton, A. J. Yuffa, and G. W. Videen, "Enhanced facial recognition for thermal imagery using polarimetric imaging," *Opt. Lett.*, vol. 39, no. 13, pp. 3857–3859, 2014.
[71] Z. Xie, S. Zhang, X. Yu, and G. Liu, "Infrared and visible face fusion recognition based on extended sparse representation classification and local binary patterns for the single sample problem," *J. Opt. Technol.*, vol. 86, no. 7, pp. 408–413, 2019.
[72] J. Gui, Z. Sun, J. Cheng, S. Ji, and X. Wu, "How to estimate the regularization parameter for spectral regression discriminant analysis and its kernel version?" *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 2, pp. 211–223, Feb. 2014.
[73] K. Siwek and S. Osowski, "Deep neural networks and classical approach to face recognition–comparative analysis," *Przegląd Elektrotechniczny*, vol. 94, no. 4, pp. 1–4, 2018.
[74] J. Chmielinska and J. Jakubowski, "Face recognition based on deep learning techniques and image fusion," *Przegląd Elektrotechniczny*, vol. 95, no. 11, pp. 150–154, 2019.
[75] J. Pearl, "Markov and Bayesian networks: Two graphical representations of probabilistic knowledge," in *Probabilistic Reasoning in Intelligent Systems*. San Mateo, CA, USA: Morgan Kaufmann, 1998, ch. 3.
[76] S. Z. Li, R. Chu, S. Liao, and L. Zhang, "Illumination invariant face recognition using near-infrared images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 627–639, Apr. 2007.
[77] T. Minh Quan, D. G. C. Hildebrand, and W.-K. Jeong, "FusionNet: A deep fully residual convolutional neural network for image segmentation in connectomics," 2016, *arXiv:1612.05360*. [Online]. Available: http://arxiv.org/abs/1612.05360
[78] X. Wu, R. He, Z. Sun, and T. Tan, "A light CNN for deep face representation with noisy labels," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 11, pp. 2884–2896, Nov. 2018.
[79] B. Na and S. Js, "Detection of camouflaged people," *Int. J. Sensor Netw. Data Commun.*, vol. 5, no. 3, pp. 143–148, 2016.
[80] M. Liggins II, D. Hall, and J. Llinas, *Handbook of Multisensor Data Fusion: Theory and Practice*. Boca Raton, FL, USA: CRC Press, 2017.
[81] M. Omran, A. P. Engelbrecht, and A. Salman, "Particle swarm optimization method for image clustering," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 19, no. 03, pp. 297–321, May 2005.
[82] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[83] Y. Xu, Y. Cheng, J. Zhao, Z. Wang, L. Xiong, K. Jayashree, H. Tamura, T. Kagaya, S. Pranata, S. Shen, J. Feng, and J. Xing, "High performance large scale face recognition with multi-cognition softmax and feature retrieval," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 1898–1906.

LUÍS LOPES CHAMBINO was born in Leiria, Portugal, in 1995. He is currently pursuing the master's degree in electrical and computer engineering with the Instituto Superior Técnico, Lisbon. He joined the Portuguese Military Academy in September 2014. He has completed the first four of seven years of his officer training with the Military Academy.

JOSÉ SILVESTRE SILVA received the Ph.D. degree in electrical engineering from the University of Aveiro, in 2005. He is currently an Assistant Professor in algorithms and data structures, computer programming, information security and cyber-defence, and operating systems with the Exact Sciences and Engineering Department, Portuguese Military Academy. His research interests include biomedical engineering, medical imaging, image processing, pattern recognition, image fusion, and multispectral image analysis.

ALEXANDRE BERNARDINO (Member, IEEE) received the Ph.D. degree, in 2004. He is currently an Associate Professor with the Department of Electrical and Computer Engineering and a Senior Researcher with the Computer and Robot Vision Laboratory, Institute for Systems and Robotics, Instituto Superior Técnico (IST), and the Faculty of Engineering with Lisbon University. He has participated in several national and international research projects as a principal investigator and a technical manager. He has graduated 12 Ph.D. students and more than 80 M.Sc. students. He has published more than 40 research articles in peer-reviewed journals and more than 100 papers on peer-reviewed conferences in robotics, vision, and cognitive systems. His primary research interests include application of computer vision, machine learning, cognitive science, and control theory to advanced robotics and automation systems.