

Multispectral Facial Recognition in the Wild

Subjects: **Computer Science**, **Artificial Intelligence**

Contributor: Pedro Martins , José Silvestre Silva , Alexandre Bernardino

Face recognition systems in uncontrolled environments have shown impressive performance improvements. However, most are limited to the use of a single spectral band in the visible spectrum. The use of multi-spectral images makes it possible to collect information that is not obtainable in the visible spectrum when certain occlusions exist (e.g., fog or plastic materials) and in low- or no-light environments. The state of the art regarding face recognition systems in an uncontrolled environment has led to the conclusion that image synthesis methods, mainly with GANs, have been used to combat intrapersonal variations, such as the difference in pose and facial expression. On the other hand, in the area of multispectral face recognition, with a plurality of solutions presented by the use of multispectral images, fusion methods are those that make the most use of images captured in different spectral bands in order to make a decision. The main problem encountered is the limited number of images (and people) in multispectral databases in an uncontrolled environment, which makes it challenging to train convolutional neural networks, which are the most used method for feature extraction.

deep neural networks

multispectral imaging

face recognition

in the wild

1. Introduction

The sense of sight allows people to observe dangers, identify objects, and recognize people. This last task is fundamental for human beings as social beings. It enables people to differentiate the level of trust someone can give to a specific person, with this being at the base of the construction of communities. Such is the importance of this task that it has become one of the main topics of research with the emergence of machine learning, thus allowing machines to incorporate this biological capacity.

Multi-spectral images have several military applications, from detection of camouflaged people ^[1], classification of vegetation types in military regions ^[2], landmine detection ^[3] and face recognition ^[4]. The current face recognition systems operating in the visible (VIS) domain have reached a significant level of maturity. It is possible to observe their wide use nowadays, from security mechanisms to unlocking electronic devices such as smartphones and personal computers to population control systems ^[5].

However, most current face recognition systems ^[6] require the cooperation of the user to ensure that pictures are taken in favorable conditions (frontal postures, good illumination, no occlusion) and have trouble dealing with uncontrolled scenarios. Uncontrolled environment scenarios, such as riots and violent demonstrations, can often be used by criminals and terrorist cell members to move around and cause damage to Homeland Security, as this type of environment adds difficulty to their detection. The uncontrolled environment is mainly characterized by a

variety of lighting, pose, facial expressions and the existence of occlusions [5]. These features are challenges to face recognition systems due to the multiple intrapersonal variations they provide, making it difficult to correctly identify an individual's identity based on a collaborative image of the individual.

2. Deep Neural Networks

2.1. Face Recognition

In general, a face recognition system is described in several phases. The first phase consists of acquiring the facial images and pre-processing them, such as locating the faces and cropping them. In a second phase, the features are extracted from the facial image, for instance, the position of facial landmarks, eye distance or even the face tones. Finally, these features are used in a classifier for identification or verification purposes.

Face recognition can be performed in a controlled or uncontrolled environment. The controlled environment, also known as consent recognition, is one in which the user cooperates in the recognition by facilitating it through correct and static posture in a place with good lighting. In the uncontrolled environment, recognition is dynamic, without the user cooperating in acquiring an image, making the face recognition process very difficult due to the diversity of the surrounding environment (e.g., low visibility), facial poses and expressions.

2.2. Multispectral Imaging in an Uncontrolled Environment

The databases of the VIS domain and the use of image synthesizers, which generate multiple poses and facial expressions from the obtained images, have allowed the difficulties associated with the variety of poses and facial expressions to be circumvented. However, two points have proved more difficult to overcome: the change of illumination and occlusions. This has led to the use of multiple spectral bands, with particular emphasis on the infrared (IR) spectral band, which can acquire images in environments with little or no brightness and overcome occlusions such as smoke and fog. In short, multispectral analysis allows a face recognition system to extract facial features that would be impossible to obtain with images from the VIS spectral band.

The IR bands can be categorized according to several spectral bands [7]. The active bands are the near-infrared (NIR) and short-wavelength infrared (SWIR). To acquire images in these bands, the object must receive illumination, even if scarce, because it is through reflection that the image is acquired. Such a fact means these images are commonly used in night vision devices. The NIR band allows the difficulties posed by the variation of illumination to be overcome, while the SWIR has the advantage of obtaining images through smoke and fog. The passive bands are the mid-wavelength infrared (MWIR) and long-wavelength infrared (LWIR). Unlike the active bands, the passive bands allow people to acquire images using only the thermal radiation emitted by a body, commonly known as thermal images.

The use of IR images for automatic face recognition is not without challenges, as these images are sensitive to the emotional, physical and health conditions of the individual, as well as the surroundings, and do not serve as an

absolute alternative to the use of the VIS spectrum, but rather as a complement [8]. Another difficulty arises from the low number of public databases with images from both spectral ranges and in an uncontrolled environment [9], which limit the creation of rich classification models and the ability to characterize the performance of those systems in realistic conditions.

3. Current Work

Multi-spectral face recognition in an uncontrolled environment can be subdivided into two areas. The first is face recognition in an uncontrolled environment, which is already challenging. The second is multi-spectral face recognition, i.e., using different spectral bands in face recognition.

3.1. Face Recognition in an Uncontrolled Environment

The uncontrolled environment, strongly characterized by pose-light-expression factors, emerges as a problem for current recognition systems. A significant step was taken towards solving this type of problem by introducing very large databases to train Deep Convolutional Neural Networks (DCNN) in combination with the emergence of image synthesis methods [5]. The two main image synthesis methods are: (i) one-to-many augmentation, which consists of generating different poses of a face from a canonical face image; (ii) many-to-one normalization, which consists of normalizing any pose of the face to a canonical face pose [5]. The use of Generative Adversarial Networks (GAN), introduced by Goodfellow et al. [10], is characterized by the use of a generator and a discriminator (see **Figure 1**). The generator is responsible for producing samples given an input image so that the discriminator cannot discern which of the samples is real and which is false.

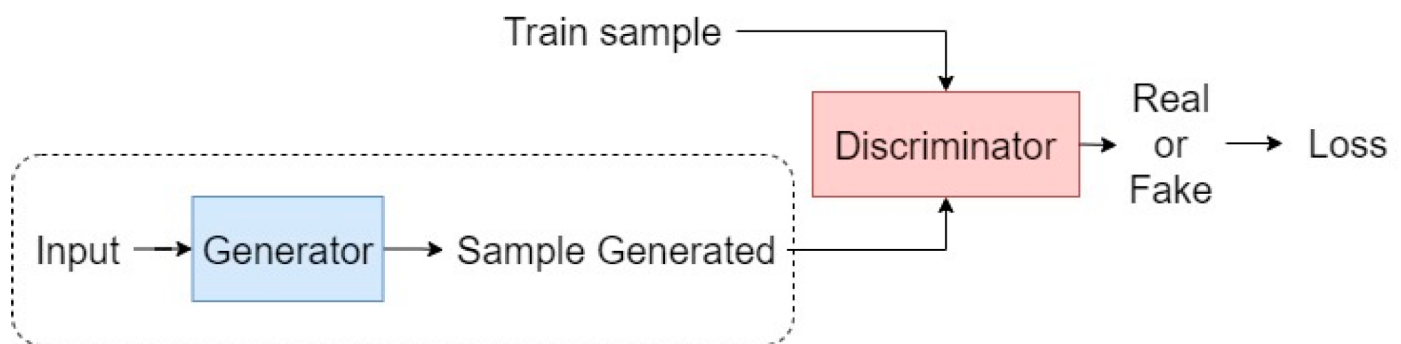


Figure 1. Schematic of the training of a GAN. The dashed line shows the process of sample generation.

Since their appearance in face normalization, with DR-GAN [11], GANs have taken the lead in solving the problem of pose and facial expression variation. As for one-to-many augmentation using GANs, as is the case with the DA-GAN network [12], their image production power also gives them an advantage compared to other algorithms.

Normalization of many-to-one images is an extreme image synthesis problem due to the pose differences of a face. Cao et al. [13] proposed HF-PIM, normalizing the face to a frontal pose through a texture fusion deformation

procedure leveraging a dense matching field to interconnect the 2D and 3D surface spaces. Qian et al. [14] presented Face Normalization Module (FNM), which encodes images using a pre-trained network for feature extraction and generates realistic images.

One-to-many augmentation is another approach to achieve face recognition regardless of the pose. Tran et al. [15] synthesized different poses through 3D modeling and then trained a DCNN to perform face recognition with varied poses. The DA-GAN proposed by Zhao et al. [12] created 2D images through 3D modeling and then refined the obtained 2D images to be as realistic as possible, using a GAN to try to preserve the identity of the face. Thus, the DA-GAN network was also used to augment the training data.

3.2. Multispectral Face Recognition

The main multi-spectral face recognition methods can be characterized by three important features: Image Synthesis Methods, Fusion Methods and Loss Functions.

Fusion methods are subdivided into feature fusion and score fusion. In the first, a fusion of features from the different spectral bands of the facial image is performed, allowing the most relevant features to be extracted from the different bands and joining them in a vector. The second method combines the scores obtained from each classifier uni-band (e.g., a classifier operating only in the LWIR band and another operating only in the NIR band) [16].

The image synthesis methods allow an image of a spectral band to be transformed into another, helping to compare two images. The main advantage of image synthesis is that it enables an image to be passed from any spectral band to the VIS band, making it possible to use classifiers implemented to process images of the VIS spectrum [17]. One of the most recent works in this area synthesizes VIS images from NIR images using GANs [18].

Finally, all neural networks have cost functions for the training moment to update the network weights. However, certain cost functions have been proposed to proceed specifically to the classification of multi-spectral images. Examples of these cost functions are the Scatter Loss [19] and the Wasserstein Distance [20].

3.3. Gaps

Although several scientific works address multi-spectral face recognition, few of these demonstrate its power in an uncontrolled environment due to the limitations in current databases of multi-spectral face images. In existing datasets, the variations of conditions are not extreme, as they are usually semi-controlled environments and not *in the wild* (uncontrolled environment). For example, the most studied database in multi-spectral face recognition, CASIA NIR-VIS 2.0 [21], uses images in which the pose has few deviations from the frontal position, which does not reliably characterize the uncontrolled environment. Thus, the fact that these databases are incomplete (compared to those of the VIS band) is still a barrier to improving the capability of multi-spectral face recognition systems in an uncontrolled environment.

References

1. Bento, N.A.; Silva, J.S.; Dias, J.B. Detection of Camouflaged People. *Int. J. Sens. Netw. Data Commun.* 2016, 5, 143–148.
2. Gonçalves, M.; Silva, J.S.; Bioucas-Dias, J. Classification of Vegetation Types in Military Region. In *Proceedings of the SPIE Security and Defence 2015 Europe: Electro-Optical Remote Sensing, Photonic Technologies, and Applications*, Toulouse, France, 21–22 September 2015; p. 9649.
3. Silva, J.S.; Guerra, I.F.L.; Bioucas-Dias, J.; Gasche, T. Landmine Detection Using Multispectral Images. *IEEE Sens. J.* 2019, 19, 9341–9351.
4. Chambino, L.L.; Silva, J.S.; Bernardino, A. Multispectral Face Recognition Using Transfer Learning with Adaptation of Domain Specific Units. *Sensors* 2021, 21, 4520.
5. Masi, I.; Wu, Y.; Hassner, T.; Natarajan, P. Deep face recognition: A survey. In *Proceedings of the 31st SIBGRAPI Conference on Graphics, Patterns and Images*, Foz do Iguaçu, Paraná, Brazil, 29 October–1 November 2018; pp. 471–478.
6. Chambino, L.L.; Silva, J.S.; Bernardino, A. Multispectral Facial Recognition: A Review. *IEEE Access* 2020, 8, 207871–207883.
7. Munir, R.; Khan, R.A. An extensive review on spectral imaging in biometric systems: Challenges & advancements. *J. Vis. Commun. Image Represent.* 2019, 65, 102660.
8. Zhang, W.; Zhao, X.; Morvan, J.; Chen, L. Improving Shadow Suppression for Illumination Robust Face Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 2019, 41, 611–624.
9. Panetta, K.; Wan, Q.; Agaian, S.; Rajeev, S.; Kamath, S.; Rajendran, R.; Rao, S.P.; Kaszowska, A.; Taylor, H.A.; Samani, A.; et al. A Comprehensive Database for Benchmarking Imaging Systems. *IEEE Trans. Pattern Anal. Mach. Intell.* 2020, 42, 509–520.
10. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In *Proceedings of the Advances in Neural Information Processing Systems*, Montreal, QC, Canada, 8–13 December 2014; Volume 27.
11. Tran, L.; Yin, X.; Liu, X. Disentangled representation learning gan for pose-invariant face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Juan, PR, USA, 21–26 July 2017; pp. 1415–1424.
12. Zhao, J.; Xiong, L.; Li, J.; Xing, J.; Yan, S.; Feng, J. 3d-aided dual-agent gans for unconstrained face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 2018, 41, 2380–2394.
13. Cao, J.; Hu, Y.; Zhang, H.; He, R.; Sun, Z. Towards high fidelity face frontalization in the wild. *Int. J. Comput. Vis.* 2019, 128, 1485–1504.

14. Qian, Y.; Deng, W.; Hu, J. Unsupervised face normalization with extreme pose and expression in the wild. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 9851–9858.
15. Tuan Tran, A.; Hassner, T.; Masi, I.; Medioni, G. Regressing robust and discriminative 3D morphable models with a very deep neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Juan, PR, USA, 21–26 July 2017; pp. 5163–5172.
16. Kanmani, M.; Narasimhan, V. Optimal fusion aided face recognition from visible and thermal face images. *Multimed. Tools Appl.* 2020, 79, 17859–17883.
17. Peng, C.; Gao, X.; Wang, N.; Li, J. Graphical Representation for Heterogeneous Face Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 2016, 39, 301–312.
18. He, R.; Cao, J.; Song, L.; Sun, Z.; Tan, T. Adversarial cross-spectral face completion for NIR-VIS face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 2019, 42, 1025–1037.
19. Hu, W.P.; Hu, H.F.; Lu, X.L. Heterogeneous Face Recognition Based on Multiple Deep Networks with Scatter Loss and Diversity Combination. *IEEE Access* 2019, 7, 75305–75317.
20. He, R.; Wu, X.; Sun, Z.N.; Tan, T.N. Wasserstein CNN: Learning Invariant Features for NIR-VIS Face Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 2019, 41, 1761–1773.
21. Li, S.; Yi, D.; Lei, Z.; Liao, S. The CASIA NIR-VIS 2.0 Face Database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Portland, OR, USA, 23–28 June 2013; pp. 348–353.

Retrieved from <https://encyclopedia.pub/entry/history/show/64158>