# Multispectral Images Applied to Face Recognition

Luis Lopes Chambino
luis.chambino@tecnico.ulisboa.pt
Military Academy, Portugal,
Instituto Superior Técnico, Universidade de Lisboa, Portugal

José Silvestre Silva
jose.silva@academiamilitar.pt
Military Academy & CINAMIL, Lisbon, Portugal,
LIBPhys-UC, Coimbra, Portugal

Alexandre Bernardino
alex@isr.tecnico.ulisboa.pt
Institute for Systems and Robotics (ISR), Portugal
Instituto Superior Técnico, Universidade de Lisboa, Portugal

## Abstract

Facial recognition is a method of identifying or authenticating the identity of individuals through their faces. Systems that use multispectral images in face recognition obtain better results that those who use only visible images. In this work, we propose a multi-channel deep convolutional neural network approach for facial recognition using multispectral images. A study is carried out to assess the performance of Support Vector Machines and k-Nearest Neighbor classifiers to classify the 256-d embeddings obtained by adapting the Domain Specifc Units in the LightCNN. Experimental results in the Tufts face dataset show competitive performance in facial recognition obtaining a rank-1 score of 99.5%.

## 1 Introduction

Nowadays 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 usage 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, non-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 other biometric sensors (e.g. fingerprint or iris) or other spectral bands, in order to minimize these problems.

The infrared spectrum, namely the Near Infrared (NIR), Short Wave 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]. These systems, which use more than one spectral band, are called multispectral.

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 have LEDs with emission in the infrared spectrum 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 MWIR and LWIR spectral bands (also known as thermal bands) allow the use of facial recognition systems at night, when the luminosity is very low or even zero.

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.

## 2 Methods

Each multispectral image may have several channels, one for each spectral band. When an monospectral band is in RGB, this image is converted to greyscale, as it is a requirement in the next phases.

Face detection is performed in all channels using the OpenCV [2] deep neural network (DNN) to obtain a face bounding box. If face detection was inconclusive (normal in LWIR channels, since the DNN used was trained in VIS images) the face bounding box from the VIS channel is used.

After the face bounding box is obtained, face landmark detection is performed using the DNN provided by Dlib [3]. A face alignment is accomplished by transforming the image such that the eyes centres are horizontal and in a predefined coordinate.

Finally, the aligned images are resized to a size of 144 x 144 pixels, to perform data augmentation on the images in order to better generalize our model, later explained in section 2.3. Figure 1 shows images in each spectral band.

Figure 1: Illustrative images of the VIS, NIR and LWIR spectral bands.

### 2.1 Model Architecture

The method used to extract the face embeddings to perform facial recognition is based on the concept of Pereira et al. [4] (the Domain Specific Units (DSU)), and we use the LightCNN [5] as the deep convolutional neural network, in contrary with Pereira work where he uses the Inception-ResNet-V2 [6] neural network.

Pereira et al. [4] showed that low level features in deep convolutional neural networks can be adapted to satisfy a specific spectral band, doing so it is not necessary to re-train the entire convolutional neural network. For this reason, we use a neural network that was previously trained for the task of facial recognition.

The LigthCNN model used in this work was trained with several face images, in the visible band, for face recognition. This architecture was chosen because of the small number of parameters, when comparing with other networks in face recognition. The reduced set of parameters was achievable because of the use of the Max-Feature Map (MFM) activation function as an alternative for the Rectified Linear Units (ReLu), which supresses low activation neurons in each layer [5].

The proposed architecture is shown in the Figure 2. The LightCNN takes as input a 128 x 128 pixels image and produces a 256-dimensional embedding, which can be used as a face representation. Each 256-d embedding represent the identity of the person through the spectral band of the channel used. Then all the embeddings produced by the channels are concatenated. In the last fully connected layer, the linear activation function was used.

The last fully connected layer it is added to produce the final 256-dimensional embedding, which can be used as a face representation by all the channels.

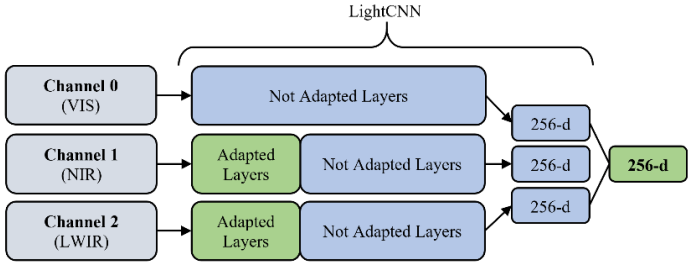


Figure 2: Overview of the proposed architecture: adapted layers (green) and not adapted layers (blue).

Using the pre-trained weights from the LightCNN, trained for face recognition in the visible band, it is possible to avoid a possible overfit, since the available multispectral datasets have a very limited number of images.

Obtained the 256-d embeddings it is now necessary to train a classifier to identify the person in the image. For comparison purposes, two classifiers were also used: the Support Vector Machines (SVM) and the k-Nearest Neighbour (kNN).

## 2.2 Dataset

The Tufts face dataset [7] was used to test the architecture. First it was necessary to clean and pre-process the dataset before using it. Initially the missing and corrupted images were excluded, then a facial detection was done. If facial detection was inconclusive a manual face detection was performed. Then all images where cropped and resized to a predefined size of 144x144 pixels. After this pre-processing task, the final dataset had 7 715 images from 109 persons.

This dataset was split into three subsets: 60% for training, 20% for validation and the last 20% for testing. It was performed a stratified split in the dataset so that each person is equally represented in each split. This step is necessary since the number of images per person in the dataset is not equal.

## 2.3 Training Procedure

Data augmentation was used to obtain a more generalized model. In the training set it was used random horizontal mirroring and random cropping to the size of 128x128 pixels. With the validation set it was only applied a center cropping to a size of 128x128 pixels, to comply with the required size of the LightCNN.

During the training procedure the proposed architecture was trained with the cross-entropy loss function. As the architecture was implemented in Pytorch, the cross-entropy loss function combines the Logarithmic SoftMax (LogSoftMax) and the negative log likelihood (NLLLoss) into a single function. Was used the Adam optimizer with a batch size of 16 and a learning rate of $10^{-3}$.

# 3 Experiments and Results

To evaluate the performance of the proposed architecture several analyses were performed.

The first analysis allowed us to choose the appropriate classifier. Three classifiers were implemented: SVM with radial base function (rbf) kernel and the linear kernel and the kNN with the Euclidian distance. To obtain the best hyperparameters a stratified 5-fold cross-validation (CV) was performed during the training of each classifier with the training and validation set.

The hyperparameters fine-tuned were the regularization parameter (C), the kernel coefficient (γ) and the number of neighbours (k). The best hyperparameters, the range used in each hyperparameter, and the rank-1 value obtained with it are displayed in Table 1.

Table 1: Best hyperparameters obtained for each classifier.

| Classifier | Hyperparameters | | | Rank-1 |
|---|---|---|---|---|
| | $10^{-10} \leq C \leq 10^{5}$ | $10^{-10} \leq \gamma \leq 10^{2}$ | $1 \leq k \leq 25$ | |
| SVM - Linear | > 0.01 | - | - | 99.8 % |
| SVM - rbf | 10 | $10^{-4}$ | - | 99.8 % |
| kNN | - | - | 1 | 99.5 % |

Determined the more suitable hyperparameters for each classifier they were trained only with the training set. Afterwards the classifiers were used to classify the 256-d embeddings from the test set. It was achieved a rank-1 score 99.70 %, 99.24 % and 99.24% for the SVM-Linear, SVM-rbf and kNN, respectively.

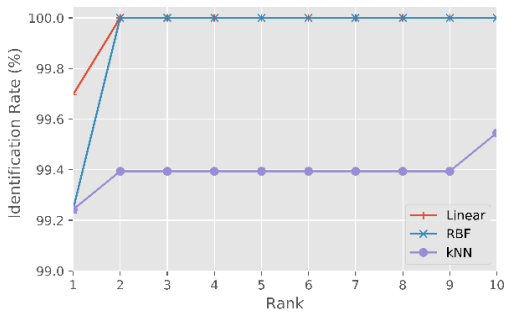


Figure 3: Cumulative matching curves for each classifier.

Figure 3 presents the cumulative matching curve (CMC) for each classifier up to rank-10. It is possible to observe that both SVM classifiers obtain a 100% score at rank-2 and have similar results. Further studies concluded that kNN obtains a 100% score at rank-102. These experimental results indicate that the SVM classifiers are more useful in identifying a person identity.

A comparison is made with other state of the art methods for face recognition that used the same dataset. This comparison can be seen in Table 2. When compared with other methods the proposed architecture proves to be a viable choice for multispectral face recognition, obtaining higher results.

Table 2: Comparison with state-of-the-art face recognition methods for the Tufts face dataset.

| Method | Rank-1 |
|---|---|
| Circular HOG [8] | 94.5 % |
| TR-GAN [9] | 88.7 % |
| Proposed methodology | 99.7 % |

# 4 Conclusions

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.

In this work, it is proposed a new architecture for facial recognition using multispectral images. The architecture produces 256-d embeddings that represent the identity of a person through multispectral images. To test and compare this architecture it is used the Tufts face dataset. To classify the 256-d embeddings an SVM-linear classifier proved to be the best classifier, obtaining the higher rank-1 score. Experimental results verify the effectiveness of the proposed architecture in multispectral face recognition when comparing with other state-of-the-art methods.

# Acknowledgements

This work was supported in part by the Military Academy Research Center (CINAMIL) under project Multi-Spectral Facial Recognition, and by FCT with the LARSyS – FCT Project UIDB/50009/2020.

# References

[1] W. Zhang, X. Zhao, J. Morvan, and L. Chen, "Improving Shadow Suppression for Illumination Robust Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, pp. 611–624, 2019.

[2] G. Bradski, "*The OpenCV Library*", Dr. Dobb's Journal of Software Tools, 2000.

[3] D. King, "*Dlib-ml: A Machine Learning Research*", Journal of Machine Learning Research, vol. 10, pp. 1755-1758, 2009.

[4] T. D. Pereira, A. Anjos, and S. Marcel, "Heterogeneous Face Recognition Using Domain Specific Units," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 7, pp. 1803-1816, Jul 2019.

[5] X. Wu, R. He, Z. Sun, and T. Tan, "A Light CNN for Deep Face Representation with Noisy Labels," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, pp. 2884-2896, 2018.

[6] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-V4, Inception-ResNet and the Impact of Residual Connections on Learning," in Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. San Francisco, United States of America: AAAI Press, 2017, p. 4278–4284.

[7] K. Panetta, et al. "A Comprehensive Database for Benchmarking Imaging Systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 3, pp. 509-520, 2018.

[8] S. Rajeev, K. Shreyas, Q. Wan, K. Panetta and S. Agaian, "Illumination Invariant NIR Face Recognition Using Directional Visibility ", *Electronic Imaging, Image Processing: Algorithms and Systems XVII*, 2019, pp. 273-1-273-7.

[9] L. Kezebou, V. Oludare, K. Panetta, and S. Agaian, "TR-GAN: thermal to RGB face synthesis with generative adversarial network for cross-modal face recognition", Proceedings SPIE, *Mobile Multimedia/Image Processing, Security, and Applications*, vol. 11399, 2020.