

Spectraface

A Fast and Simple Thermal / Visual Face Recognition System

Ziyuan Zhu

Computer Science Department

Lehigh University

zachzhu2016@gmail.com

1. Introduction

Visual face recognition methods utilize images captured in human-visible spectrum to recognize individuals which in itself is limited by external factors such as illumination variations, occlusions, and blurred outlines. [1] [5] It can only achieve results as good as a pair of human eyes do. Thermal face alleviates most of the problems faced by visual faces by capturing radiation information fall in a higher wavelength. Developing a solution that recognizes thermal faces expands the spectrum of facial imageries to be recognized. However, thermal face recognition has its own challenges such as a lower resolution, sensitive to varying skin temperatures, and inability to see through glasses.

2. Related work

One of the most common algorithms used for face recognition is Eigenface, which achieves near-real-time face recognition by using Principal Component Analysis (PCA) and maps high-dimension face images to a lower-dimension "face space". However, it has high-expected requirements for the face images such that the seen image and probe image must reside in a similar environment, otherwise it begins to fail considerably. This is due to the nature of the algorithm that it encodes the most relevant information in a face image. [3][4] Eigenface has inspired the development of Spectraface by the nature that Eigenface essentially captures and encodes the most variant features in an image, it can be used for recognizing any objects as well as thermal faces beyond the scope of faces. Another method widely used to recognize thermal faces is using fused image, which comprises of multiple images, but it is not adopted in Spectraface [5] (it could be a nice extension given both visual and thermal faces are provided in the IRIS dataset, however, it is not very likely that a production system could provide both types of faces, hence it is not the priority here). [5] There's also a solution based on Generative Adversarial Network to generate visual faces from thermal faces. However, this approach is complicated and computationally expensive. [2] The method described in this paper focuses on not only accuracy but also simplicity and speed, thus the techniques used in TV-GAN are not considered but compared against. [2]

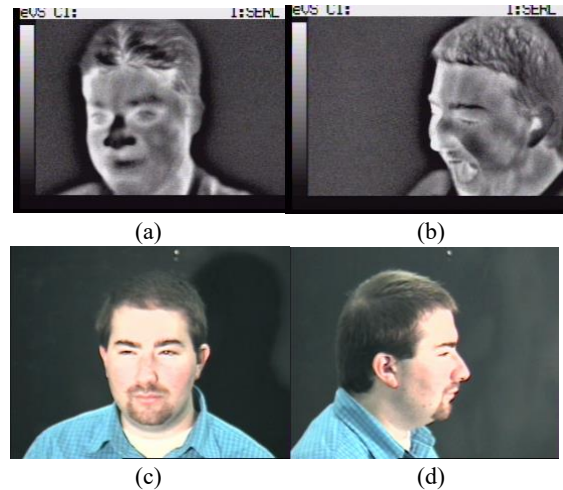


Figure 1: (a) and (b) are thermal faces with different facial expressions, (c) and (d) are visual faces with different rotations.

3. Dataset

IRIS Thermal/Visual Database is used as the dataset to test Spectraface. It consists of 30 subjects (28 men and 2 women of fair diversity in race) and their faces in three facial expressions - surprised, laughing, and anger - and five illumination conditions - left light on, right light on, both lights on, dark room, and both lights off. There are equal number of thermal and visual faces, all with resolution 320x240 pixels. (embed the chair image and a few sample images) This dataset requires Spectraface work on faces across various spectrums, expressions, illuminations, rotations, as well as faces with occlusions. The only minor limitation presented by this dataset is low resolution associated with its images. It would have been more ideal to include more women as the subjects and some NIR faces. [5][9] Although IRIS does not include specification in which spectrum these thermal images belongs to, observations on the dataset tell that they are imaged in between mid-wave and long-wave infrared. There are very few thermal face image databases available in public (the other being Terravic Facial Infrared Database). [5]

4. Spectraface Algorithm

Spectraface revolves around two phases, training and production. The training phase involves creating a set of representative faces used to compare against a new input face, while the production phase is responsible for assigning an existing identity to a face or creating a new identity to represent it.

4.1. Training phase

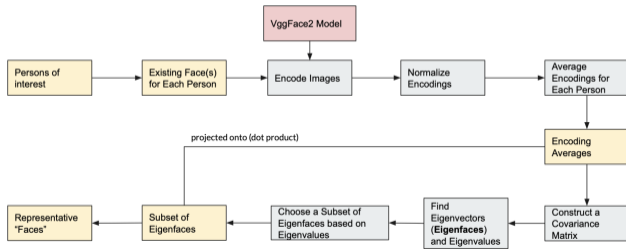


Figure 2: training phase pipeline figure.

The training phase starts with a set of pre-selected identities of interest along with their thermal/visual face images (one identity tends to have multiple face images taken in with different conditions). The first step is to crop out the face of from the images with face detection. The face detection in Spectraface is achieved by using a neural network pre-trained with thermal face datasets in [8], achieving almost 100% accuracy. This model had to be adopted due to the low accuracy (about 30%) of thermal face detection resulted from using the model [11], which was designed to detect visual faces. Note that the detection step is spared in Figure 2 and 4. These cropped out faces become the training face images. Each of these training face images' latent features are encoded into a 512-dimension vector by Inception Resnet (V1) models in PyTorch, pretrained on VGGFace2. [10] This step serves two purposes: noisy reduction and dimension reduction, which allows these encodings to be averaged later on. Next, each encoding representing each face image are normalized by mean subtraction and given that each identity owns a set of face images (feature vectors / encodings), they are averaged and reduced down to a single feature vector / encoding. Now, there is only one encoding for each identity (counterpart of a frontal face in classic Eigenface algorithm) and a covariance matrix is created from those. The eigenvectors, also called eigenfaces, and eigenvalues are extracted from this covariance matrix, sorted in descending order by eigenvalues, and the top eigenvectors that cover 95% (95% is the threshold value used in Spectraface) of the eigenvalues are selected to be used to represent the variance in the training set. [3][4] Each eigenface can now be used to represent both existing and new faces. The eigenvalues associated with each eigenface represent how much the images in the training set vary from the mean image in that direction. [3][4] To establish a set of "faces", namely encodings, to represent each identity in the training set, each identity's encoding is projected onto the set of eigenfaces to create so-called "representative faces" in the context of this paper.

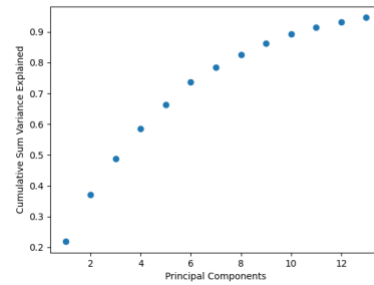


Figure 3: principal component and variance plot. The top 95% of variance is already captured by 14 out of 30 principal components.

4.2 Production phase

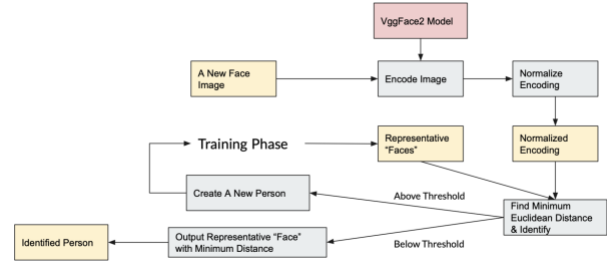


Figure 4: production phase pipeline figure.

The second phase is production which accepts faces not only from seen identities but also unseen ones, a new identity gets created whenever the difference between the new encoded face and a "representative face" exceeds certain threshold. The pipeline works as follows: encode the new face image as before into a 512-dimension vector and normalize it by mean subtraction. To find out if this new face belongs to a seen identity, the Euclidean distances between the new face encoding and all the "representative faces" are calculated, from which the "representative face" with the minimum distance is returned. If the minimum Euclidean distance is below a pre-set threshold, the identity corresponds to that "representative face" is returned. However, if it is above that value, a new identity object is created in the system. Spectraface uses 0.3 (30%) as its threshold value for tolerating seen faces. This value is selected based on the statistics of the minimum, average, and maximum Euclidean distances between samples in the dataset. Once a new identity has entered into the system, the training phase need to rerun to reflect the new identity for future jobs.

5. Experimental results

Several experiments were conducted to measure the accuracy of Spectraface. The faces of unseen identities (strangers) are subtracted from the original dataset so that they can be used to test accuracy for recognizing unseen identities later on, and the

remaining faces are randomized and split into 80% training and 20% testing.

The first experiment involves recognizing thermal faces using the Inception Resnet (V1) pre-trained with VggFace2, the encoding model, without fine-tuning. This setup was able to achieve 40 – 50% accuracy for recognizing seen identities across different ratio of unseen identities. The accuracy for recognizing unseen faces fluctuates between 30 – 60%. In another experiment with an attempt to improve this set of accuracies, the encoding model was further trained (fine-tuned) with IRIS dataset with 100 epochs. However, results have shown that fine-tuning does not make a difference in accuracy. This may be due to the small size of the dataset, having only 30 subjects, and there may be a significant domain gap between thermal and visual faces (VggFace2 contains faces in visible spectrum only).

Given that IRIS dataset also contains visual faces in correspondence with its thermal faces, Spectraface was tested against those visual faces with the same encoding model without fine-tuning. A set of much higher accuracies were achieved for both seen (90%) and unseen (40 – 70%) identities' faces. This may be due to the fact that the encoding model had been pre-trained with visual faces.

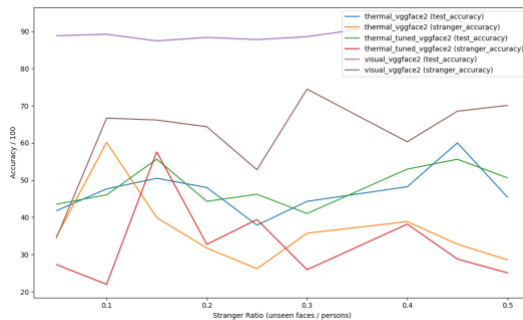


Figure 5: recorded accuracies. Unseen identity ratio (stranger ratio) refers to the ratio of unseen identities in the test dataset.

6. Future work

There is a lot of room for improvements and further validations. One possible extension could be replacing the encoding model with another model trained from scratch using the IRIS dataset. This may allow a set of more representative features to be extracted from thermal faces. Given that IRIS dataset is not significant big, more rigorous validation methods like ten-fold cross-validation may be used.

7. References

- [1] Daniel Sáez Trigueros, Li Meng, and Margaret Hartnett. 2018. Face Recognition: From Traditional to Deep Learning Methods. (October 2018). Retrieved December 4, 2020 from <https://arxiv.org/abs/1811.00116>
- [2] Teng Zhang, Arnold Wiliem, Siqi Yang, and Brian C. Lovell. 2017. TV-GAN: Generative Adversarial Network Based Thermal to Visible Face Recognition. (December 2017). Retrieved December 4, 2020 from <https://arxiv.org/abs/1712.02514>

- [3] G.Md. Zafaruddin and H.S. Fadewar. 1970. Face Recognition Using Eigenfaces. (January 1970). Retrieved December 4, 2020 from https://link.springer.com/chapter/10.1007/978-981-13-1513-8_87
- [4] Anon. Eigenfaces for Recognition - MIT Press Journals. Retrieved December 4, 2020 from <https://www.mitpressjournals.org/doi/pdf/10.1162/jocn.1991.3.1.71>
- [5] Mrinal Kanti Bhowmik et al. 2011. Thermal Infrared Face Recognition – A Biometric Identification Technique for Robust Security system. (July 2011). Retrieved December 4, 2020 from <https://www.intechopen.com/books/reviews-refinements-and-new-ideas-in-face-recognition/thermal-infrared-face-recognition-a-biometric-identification-technique-for-robust-security-system>
- [6] Anon. FaceNet: A Unified Embedding for Face Recognition and ... Retrieved December 4, 2020 from <https://arxiv.org/pdf/1503.03832.pdf>
- [7] Timesler. timesler/facenet-pytorch. Retrieved December 4, 2020 from <https://github.com/timesler/facenet-pytorch>
- [8] Maxbbbraun. Maxbbbraun / thermal face. Retrieved December 4, 2020 from <https://github.com/maxbbbraun/thermal-face>
- [9] James W. Davis. Retrieved December 4, 2020 from <http://vcipl-okstate.org/pbvs/bench/Data/02/download.html>
- [10] Timesler. timesler/facenet-pytorch. Retrieved December 4, 2020 from <https://github.com/timesler/facenet-pytorch>
- [11] Ageitgey. ageitgey/face_recognition. Retrieved December 5, 2020 from https://github.com/ageitgey/face_recognition
- [12] Vutsalsinghal. vutsalsinghal/EigenFace. Retrieved December 5, 2020 from <https://github.com/vutsalsinghal/EigenFace>