얼굴인식을 위한 다중입력 CNN 의 기본 구현

Basic Implementation of Multi Input CNN for Face Recognition

usman@sju.ac.kr

Usman Cheema* and Seungbin Moon*
*Dept. of Computer Engineering, Sejong University

Abstract

Face recognition is an extensively researched area of computer vision. Visible, infrared, thermal, and 3D modalities have been used against various challenges of face recognition such as illumination, pose, expression, partial information, and disguise. In this paper we present a multi-modal approach to face recognition using convolutional neural networks. We use visible and thermal face images as two separate inputs to a multi-input deep learning network for face recognition. The experiments are performed on IRIS visible and thermal face database and high face verification rates are achieved.

1. Introduction

Face recognition holds potential as an easy and non-intrusive form of biometric identification. Its applications in digital entertainment, media album management, human computer interaction, security, and forensics has gained it much attention over the previous decades. Illumination, pose, expression, occlusion, and recently, disguise are challenges faced by commercial applications of face recognition in open world scenarios. Infrared, thermal and 3D modalities have shown increased performance against some of these challenges independently. Fusion of multiple modalities has achieved equal or improved performance over single modal recognition rates. This is because additional information about the scene is captured in multiple modalities. Visible modality excels at capturing image and texture information, infrared and thermal modalities can capture material and temperature information, respectively, and 3D imaging captures shape as a third a coordinate. Decision fusion has shown improved results over feature or image fusion with traditional machine learning approaches. Convolutional Neural networks (CNN) eliminate the need for feature extraction and image co-registering as required in traditional machine learning approaches.

In this paper we present a feature level fusion technique for visible and thermal images using a multi-input CNN. Section 2 reviews the current methods in modality fusion. In Section 3 we present our experimental setup followed by concluding remarks and future work in Section 4.

2. Multi-Modal Fusion

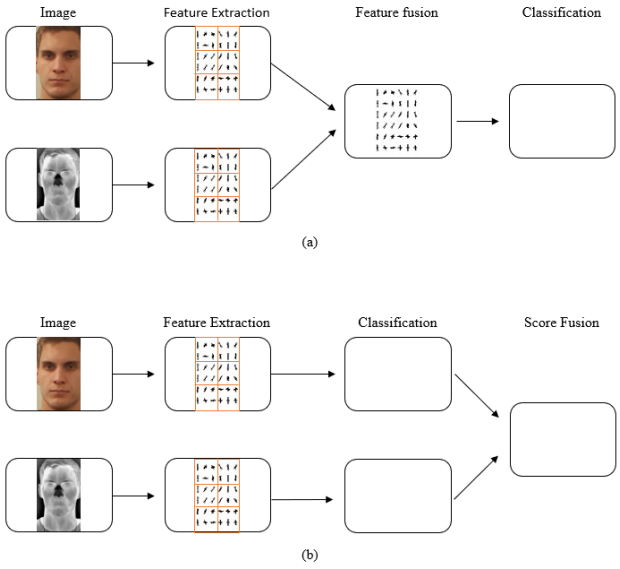
Near human level face recognition performance has been achieved using visible, thermal, and infrared images. Using multi-modal approaches over using single modality for identity recognition has shown various degrees of improvement depending on the challenges presented in the database. Visible + Infrared, Visible + thermal, Visible + 3D are some of the modality pairs used for face recognition in the literature. As each modality helps overcome the disadvantage of other modality, fusing multiple modalities can achieve higher recognition rates. Fusion of two or modalities can be performed at feature or score level. Fig. 1 shows an example pipeline for both scenarios.

Feature level fusion

In feature level fusion, separate features are extracted from all images and combined to create hybrid features. These hybrid features are further used by a classification algorithm for facial recognition. Feature level fusion is performed by [1] using traditional classifiers on visible, thermal, and depth face images.

Score level fusion

In score fusion independent matching scores for the two modalities are calculated and the results are combined to obtain and final mean score. Local Binary Patterns descriptor is used with SVM classifier and scores are fused for thermal and visible images for face recognition by [2].



(Figure 1) Fusion approaches for multi-modal data, a) feature fusion, b) score fusion.

The advent of Neural Networks (NN) has eliminated the need for independent feature extraction and classification algorithms. Multi-input NN's can be designed to ingest multiple inputs of same category e.g. visible and thermal images, or different classes e.g. image and text data. Various algorithms have been proposed for multi-biometric classification such as face image and iris, face image and fingerprint etc.

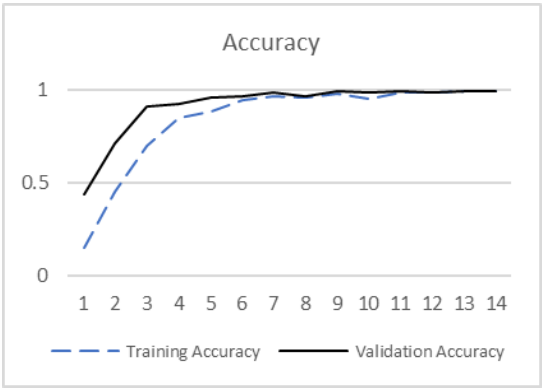
3. Experimental Work

We propose a multi-input CNN based on VGG [3] architecture. The network takes a visible and thermal image pair as input for training and classification. The inherent nature of CNN eliminates the need for co-registering the image pair, a difficult problem for visible and thermal image pairs. Our proposed architecture consists of three segments, namely S_V, S_T, and S_C. S_V take a visible image as input, S_T takes a thermal image as input, and are identical otherwise. The output from S_V and S_T is added and fed to S_C which contains two blocks of three convolution layers and a max pooling layer, each. A fully connected layer followed by a SoftMax layer performs classification in S_C.

The network was trained and tested on IRIS Thermal/Visible Face Database. The database contains images of 29 subjects with 44 simultaneously acquired unregistered visible and thermal image pairs per person. The database contains variations in pose, illumination, and expression as challenges of face recognition. Training was performed on randomly selected 70% of images for each subject. Due to the limited size of the database rigorous data augmentation was performed on the training images. The data was augmented using image rescaling, rotation, shifting, and flipping horizontally. The size of the augments data set for training is about 26,000 images in each modality. Reserved 30% of images were used for testing, no data augmentation was performed on the test dataset. The training and validation accuracy and loss are presented in Fig. 2. A recognition accuracy of 98.97% was achieved on the test data using our proposed method.

4. Conclusion and Future Work

Robust facial recognition systems are of increasing importance for security and access control. With the rising challenges of spoofing, facial recognition systems need to adapt stay ahead. Multi-modal facial recognition is promising approach against disguise and replay attack scenarios. In this paper we present a multi-input architecture for face recognition and have proven its effectiveness against pose, illumination, and expression. Our experimental work also shows that the proposed approach of using thermal and visible images for multi-input CNN is promising and merits further work.
Various other modalities such as infrared and depth images



(a)

(Figure 2) Training and Validation results for 14 epochs of training on the IRIS thermal/visible face dataset.

can be combined using the same methodology and improved accuracies achieved. We plan to extend our experimental work to perform face recognition on other challenges of face recognition such as disguise and occlusion.

References

1. O. Nikisins, K. Nasrollahi, M. Greitans and T. B. Moeslund, "RGB-D-T Based Face Recognition," 2014 22nd International Conference on Pattern Recognition, Stockholm, 2014, pp. 1716-1721.
2. T. I. Dhamecha, A. Nigam, R. Singh and M. Vatsa, "Disguise detection and face recognition in visible and thermal spectrums," 2013 International Conference on Biometrics (ICB), Madrid, 2013, pp. 1-8.
3. S. Liu and W. Deng, "Very deep convolutional neural network based image classification using small training sample size," 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), Kuala Lumpur, 2015, pp. 730-734.