Thesis for degree of Master of Engineering

Face Detection and Recognition using Deep Learning Method

by

Adyan Marendra Ramadhani

Department of Information System
(Interdisciplinary Program)

The Graduate School

Pukyong National University

February 2017

Face Detection and Recognition using Deep Learning Method

(심층학습 기능을 사용한 얼굴 탐지 및 인식)

Advisor: Prof. Bong-Kee Sin

by

Adyan Marendra Ramadhani

A thesis Submitted in partial fulfillment of the requirements

For the degree of

Master of Engineering

In Department of Information System (Interdisciplinary Program)

The Graduate School

Pukyong National University

February 2017

Face Detection and Recognition using Deep Learning Method

A dissertation

By

Adyan Marendra Ramadhani

Approved by:

(Chairman) ***Man Gon Park***

(Member) ***Carmadi Machbub***

(Member) ***Bong-Kee Sin***

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AFLW	: Annotated Facial Landmarks in the Wild
CNN	: Convolutional Neural Network
DBN	: Deep Belief Network
FDDB	: Face Detection Data Set and Benchmark
LBP	: Local Binary Pattern
HOG	: Histogram of Gradient
ORL	: AT&T *"The Database of Faces"*
SVM	: Support Vector Machine
VGG	: Visual Geometry Group

심층학습 기능을 사용한 얼굴 탐지 및 인식

Adyan Marendra Ramadhani

부경대학교 정보시스템협동과정

국문 요약

컴퓨터 비전에서 사람 얼굴 탐지 및 인식은 고전적인 분류의 문제이다. 최근에는 심층신경망을 사용하여 처리하는 연구가 많이 진행 되고 있다. 또한 얼굴 탐지 및 인식을 하기 위하여 기계학습 알고리즘의 신경망과 랜덤 포레스트가 사용되기도 한다. 일반적인 얼굴 탐지와 인식은 1. 영상 처리, 2. 얼굴 탐지, 3. 얼굴 인식 등의 삼 단계로 구성된다. 본 논문은 각각 별도로 훈련을 한 두 개의 신경망을 결합한 통합적 접근법을 제안한다. 실험을 통하여 심층학습 기반의 CNN(Convolutional Neural Network)과 DBN(Deep Belief Network)을 통한 얼굴인식은 75%, CNN, DBN 과 LBP(Local Binary Pattern)를 혼합한 방식은 98.3%의 정확성을 보였다. 본 연구는 얼굴 탐지 및 인식 문제에 심층신경망의 기능을 학인하고 성능을 평가 및 비교하였다

키워드: 얼굴 탐지, 얼굴 인식, 심층학습

Chapter I.
Introduction

1.1 Background

Face detection and recognition is a research problem that has drawn much attention from a huge number researchers. Most researchers take both detection and recognition with the majority of the effort going to improving the accuracy and the performance of both tasks. There are many algorithms that have been developed for face detection and recognition such as *Adaboost*, neural networks, and support vector machine (SVM)[1].

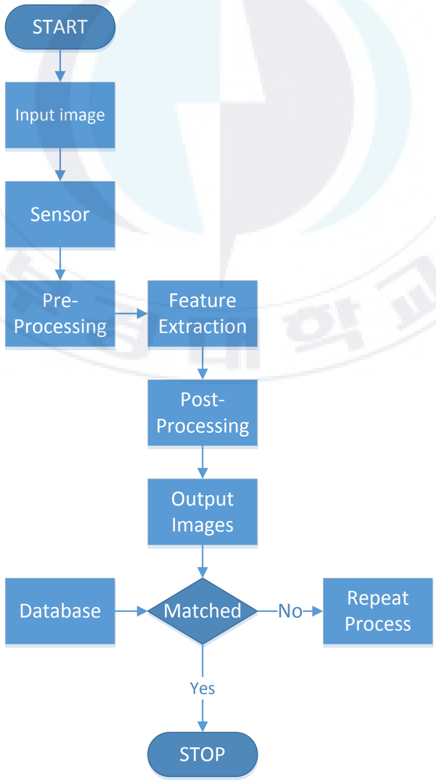

Figure 1.1 Flow Chart of Face Detection & Recognition System

Face detection determines the location of human faces in an input image. This finds many important applications such as video surveillance, human computer interface, video conference and biometric application [2]. Once a human face is detected, it is followed by the face recognition procedure. Face recognition plays an important role in verification and identification tasks [3].

In the verification task, the face of an unknown person along with an authorized face is presented to confirm if the two face images match or not. In the face identification the task is to label the person by comparing the face known individuals in the database [3].

The success of the face detection and recognition system relies on the methods and models used to overcome variations in pose, expression and illumination. That's why need a robust and efficient method to improve the detection and recognition.

Recently, deep learning has become a popular mode across a range of image recognition task [4]. With an extensive number of layers and free parameters this model provides enough capacity to represent the variations of complex face images.

Inspired by previous research, this thesis presents an integrated face detection and recognition system architecture using deep learning methods.

1.2 Objective and research method

The objective of this thesis is to present the optimum the performance of detection rate and recognition rate for face detection and recognition system. Although some previous researches have already studied about face detection and recognition optimization, but the previous methods can be used to create the parametric and comparison analysis.

At this research, dependent and independent variable is used. Deep learning is applied to reach the best performance of face detection and recognition

1.3 Outline and contribution of thesis

This section describes the contents of the thesis and their contributions briefly. The contents consist of four sections, i.e. introduction, Understanding Deep Learning Methods, Face Detection and recognition, Experimental description, simulation and results and conclusions as follows:

Chapter 1: Introduction

In this chapter, background and motivation of this thesis are presented. The objective and research method of this thesis are then described. Finally, the outline and contribution of this thesis is explained.

Chapter 2: Fundamental of Theory

This chapter describes the theoretical background used to create Deep Learning (CNN and DBN) and about face detection and recognition as well.

Chapter 3: Face Detection and Recognition System Using CNN (Convolutional Neural Network)

This chapter describe the using of CNN (Convolutional Neural Network) for face detection and recognition system using AFLW, FDDB and VGG Face Dataset.

Chapter 4: Face Recognition using Deep Belief Network and LBP (Local Binary Pattern)

This chapter describe the using of deep belief network, LBP (Local Binary pattern) and HOG for face recognition problem using ORL dataset.

Chapter 5: Conclusions

Conclusions from this research and several suggestions for future works are presented.

Chapter II.
Understanding Deep Learning Methods, Face Detection and Recognition

2.1.Deep Learning Method

Deep-learning is a software that attempts to mimic the activity in layers of neurons in the neocortex, the wrinkly 80 percent of the brain where thinking occurs. The software learns, in a very real sense, to recognize patterns in digital representations of sounds, images, and other data [5]. Deep learning solves this central problem in representation learning by introducing representations that are expressed in terms of other, simpler representations. Deep learning allows the computer to build complex concepts out of simpler concepts [6].

Machine-learning systems are used to identify objects in images, transcribe speech into text, match news items, posts or products with users' interests, and select relevant results of search. Increasingly, these applications make use of a class of techniques called deep learning [7].

Deep learning methods leads at learning feature hierarchies with features from higher levels of the hierarchy formed by the composition of lower level features, automatically learning features at multiple levels of abstraction allows a system to learn complex functions mapping the input to the output directly from data, without depending completely on human-crafted features [8]

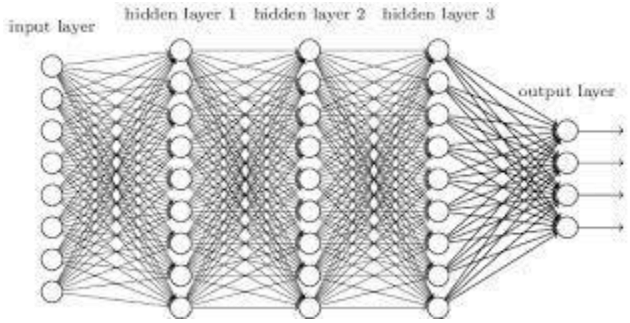

Figure 2.1 Deep Learning Architecture on Neural Network

2.2.Face Detection

Face detection is one of the challenging problems in computer vision. The goal of face detection is to determine whether or not there is any face in the image and the location of each face if there is. This topic attracts popular attention because of its wide range of applications. For example, intelligent video surveillance systems and human-computer interface have been closely linked to the industrial applications where face detection plays a key role [8].

These available techniques mainly deal with the challenges that are summarized as follows:

- Lighting condition: Lighting change brings new cast shadow that may lead to extra computation complexity.
- Facial expression: This may result in a completely different appearance of the facial structure.
- Occlusion: Evidence shows that occluded faces are difficult to be detected due to unavailability of facial entities.
- Pose variations: This action may result in 1, 2, and 3

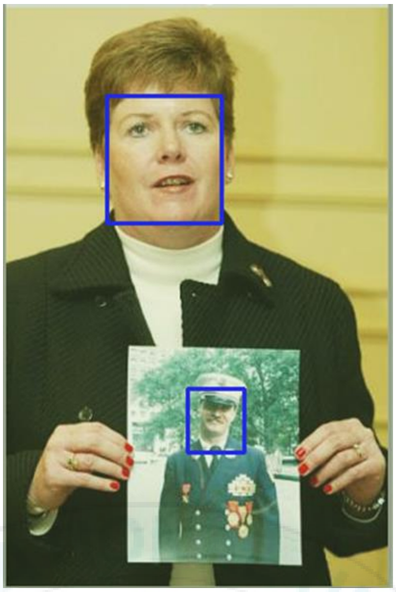

Figure 2.1. Face Detection Example

2.3.Face Recognition

Face recognition is a classical problem in the computer vision. Face recognition is an interdisciplinary field which integrates different techniques such as:

- Image processing,
- Pattern recognition,
- Time attendance system,
- Access control,
- Physiology, and
- Visitor management system [9].

Face recognition algorithms used simple geometric models, but the recognition process has now matured into a science of sophisticated mathematical representations and matching processes. Face recognition can be used for both verification and identification (open-set and closed-set) [10].

2.4.LBP (Local Binary Pattern)

Local binary patterns (LBP) is a type of visual descriptor used for classification in computer vision. LBP is the particular case of the Texture Spectrum model proposed in 1990[13]. The local binary pattern operator is an image operator which transforms an image into an array or image of integer labels describing small-scale appearance of the image.

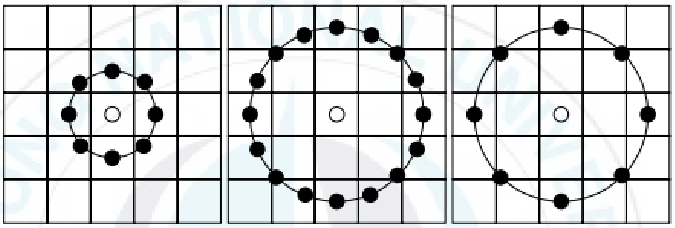

Figure 2.2. The Circular (8,1) (16,2) and (8,2) Neighborhoods on Local Binary Pattern

Figure 2.3. LBP on ORL Sample

2.5.HOG (Histogram of Gradient)

The histogram of oriented gradients (HOG) is a feature descriptor used in computer vision and image processing for the purpose of object detection. The technique counts occurrences of gradient orientation in localized portions of an image. The HOG descriptor has a few key advantages over other descriptors. Since it operates on local cells, it is invariant to geometric and photometric transformations, except for object orientation [14]

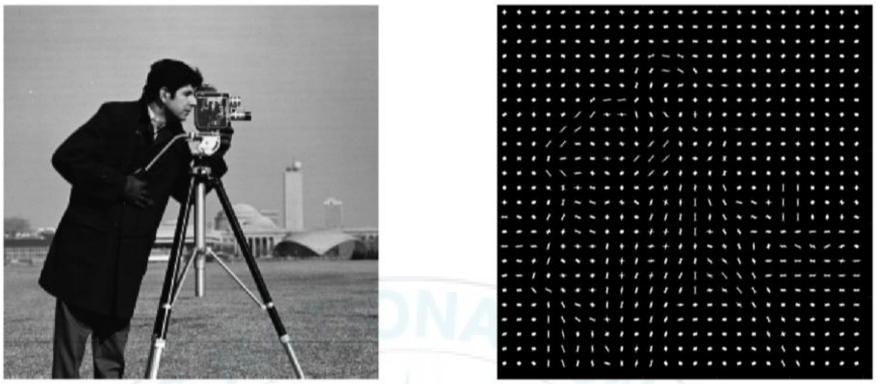

Figure 2.4. HOG Feature Extraction

2.6.CNN (Convolutional Neural Network)

CNNs are biologically-inspired models. The research investigations carried out by D. H. Hubel and T. N. Wiesel in their paper [6] proposed an explanation for the way in which mammals visually perceive the world around them using a layered architecture of neurons in the brain, and this in turn inspired engineers to attempt to develop similar pattern recognition mechanisms in computer vision.

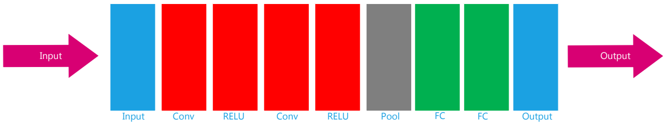

Figure 2.5. CNN Basic Architecture

2.7.DBN (Deep Belief Network)

The deep belief network (DBN) (Hinton et al., 2006) is a multilayer generative model where each layer encodes statistical dependencies among the units in the layer below it; it is trained to (approximately) maximize the likelihood of its training data. DBNs have been successfully used to learn high-level structure in a wide variety of domains, including handwritten digits (Hinton et al., 2006) [11] and human motion capture data (Taylor et al., 2007). [12]

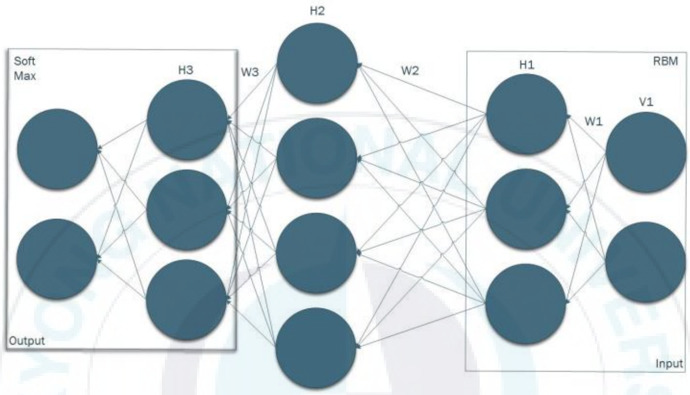

Figure 2.6. Deep Belief Network Basic Architecture

Chapter III.
Face Detection and Recognition Using Convolutional Neural Network

3.1.CNN (Convolutional Neural Network Architecture)

For CNN, There are 3 stages in this experiment, first is the preprocessing and training phase, second is detection and the final phase is recognition. We use AFLW dataset (Annotated Facial Landmarks in the Wild) which is provided by Martin Koestinger, Paul Wohlhart, Peter M. Roth, Horst Bischof Institute for Computer Graphics and Vision, Graz University of Technology for the detection, and for detection test we use Face Detection Data Set and Benchmark (FDDB) and additional dataset develop by our team. For recognition we use VGG-Face dataset provided by Department of Engineering Science, University of Oxford.

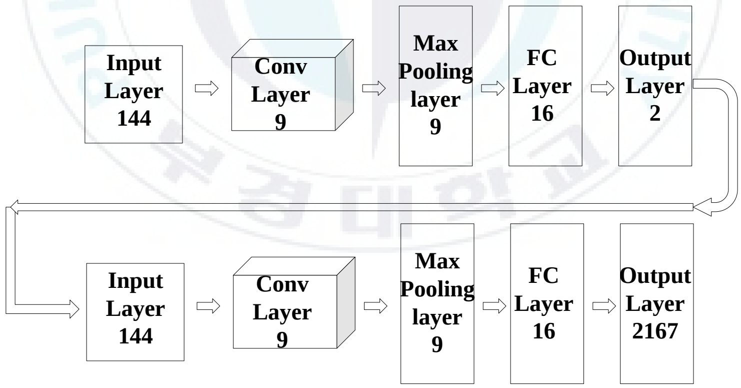

Figure 3.1. Proposed Convolutional Neural Network

Figure show the proposed convolutional neural network for face detection and recognition system. The first CNN is used for detection CNN and the second is used for recognition. The CNN is consist of:

- INPUT [12x12x3] will hold the raw pixel values of the image, in this case an image of width 12, height 12, and with three color channels R, G, B.
- CONV layer will compute the output of neurons that are connected to local regions in the input
- Pooling layer will perform a down-sampling operation along the spatial dimensions (width, height)
- ReLU (Rectified Linear Units) layer will apply an elementwise activation function such as the max (0, x) thresholding at zero.
- FC (Fully Connected) layer will compute the class scores.

3.2.Experiment Result

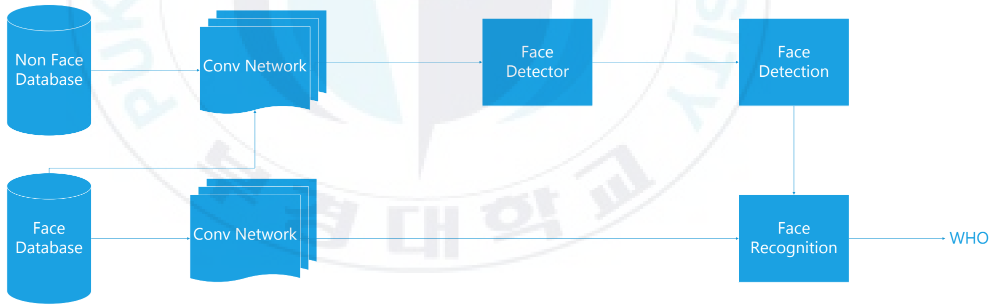

Figure 3.2. Proposed Architecture of Face Detection and Recognition System

For the integration system result for face detection can be input for the face recognition system. For face detection we scan all the areas to detect if there is a face picture or not and test with some of non faces images to measure the accuracy. Using 1200 FDDB dataset for testing, the average time to detect faces in one single images is about ±5.48 Second.

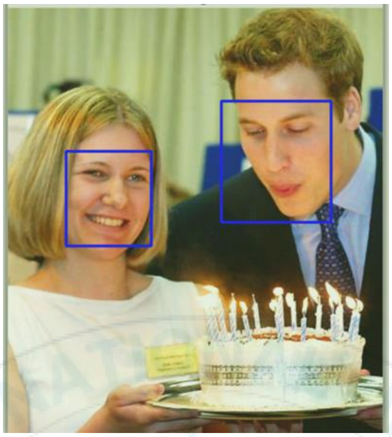

Figure 3.2. Detection Result on FDDB Sample Dataset

Figure show the detection result on FDDB Dataset Sample Figure 4 tells about the learning curve graph on the CNN and it tells that the CNN was trained well in the first stage of training CNN.

Table 3.1. Face Detection Accuracy on CNN (Convolutional Neural Network)

Test Result

Test Set	**Accuracy**
FDDB 1	65%
FDDB 2	62%
FDDB 3	75%

After the CNN is trained until 3 stages then the nets is test using FDDB database. We create 3 or more data test based on FDDB and the accuracy is about ±75%. Figure is show the detection error on false images and positive images, the detection error can be fixed by adding the training set.

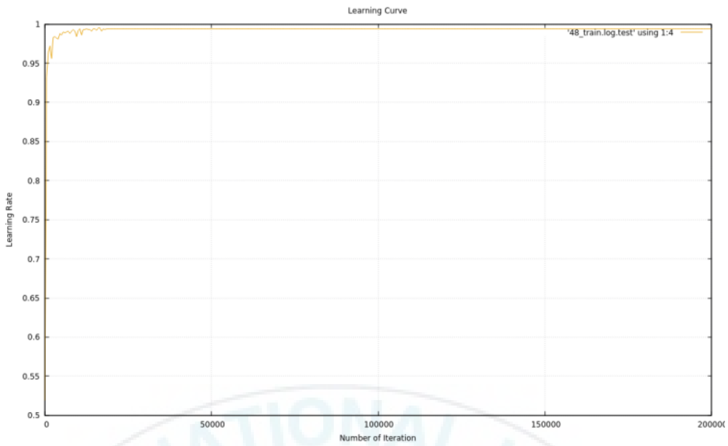

Figure 3.3. Learning Curve on CNN (Convolutional Neural Network)

Figure 3.4. Detection error on Faces Images

For face recognition, the CNN is trained with input 224*244 dimensional size with 2622 outputs. The weights is initialized by random sampling from Gaussian distribution. The learning rate is initially set to 0.01. Then the net is test using test images. For testing, the net is test using Chris Hemsworth and Lea Michele picture. Figure show that the net is correct recognize the picture of Chris Hemsworth and figure the net is incorrect recognize the picture of Lea Michele.

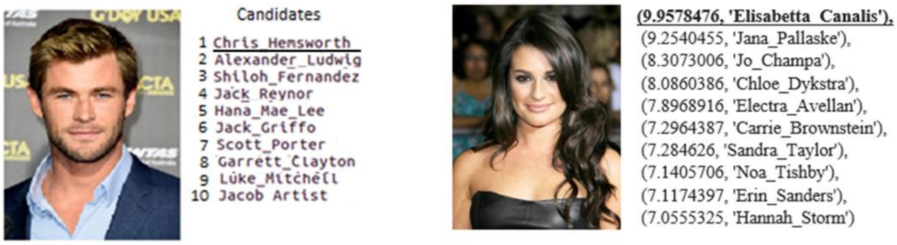

Figure 3.5. Recognition Sample on Chris Hemsworth and Lea Michele

Chapter IV.
Face Recognition Using Deep Belief Network and LBP (Local Binary Pattern)

4.1 DBN (Deep Belief Network) Structure

RBMs can be stacked and trained in a greedy manner to form so-called Deep Belief Networks (DBN). For deep belief network experiment, it's configure based on RBM (Restricted Boltzmann Machine) and it has 4 layers, each layers represent: input, hidden layers and output layer. Each layer has for input layer. For input layer it has 944 visible nodes, and each of hidden layer has 200 hidden nodes and finally the output layer has same value with sum of face labels, for this case the sum of face label is 40. The net is trained with 3000 iteration and 0.001 of learning rate.

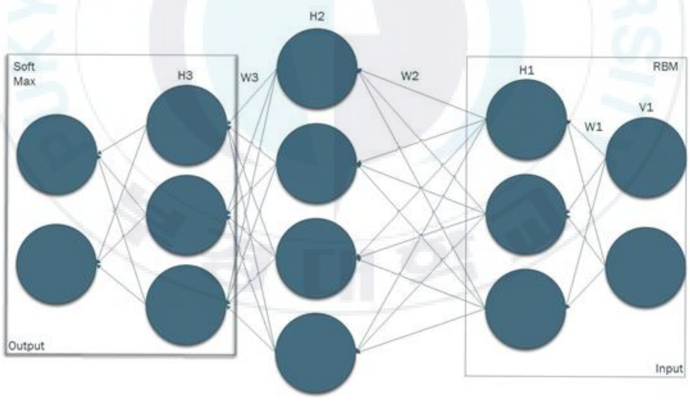

Figure 4.1. Proposed Deep Belief Network

The proposed DBN is consist of 2 hidden layer and DBN is setting and training by:

- Train the first layer as an RBM that models the raw input $x = h^{(0)}$ as its visible layer.

- Use that first layer to obtain a representation of the input that will be used as data for the second layer. Two common solutions exist. This representation can be chosen as being the mean activations $p(h^{(1)} = 1|h^{(0)})$ or samples of $p(h^{(1)}|h^{(0)})$.
- Train the second layer as an RBM, taking the transformed data (samples or mean activations) as training examples (for the visible layer of that RBM).
- Iterate (2 and 3) for the desired number of layers, each time propagating upward either samples or mean values.
- Fine-tune all the parameters of this deep architecture with respect to a proxy for the DBN log- likelihood, or with respect to a supervised training criterion (after adding extra learning machinery to convert the learned representation into supervised predictions, e.g. a linear classifier).

4.2 Local Binary Pattern & HOG (Histogram of Gradient)

For Feature Extraction. The Images is extracted using LBP (Local Binary Pattern) and HOG (Histogram of Gradient). Before the images is extracted. The images is convert into grayscale images. LBP feature vector will returned as a 1-by-N vector of length N representing the number of features. Figure is Represent the LBP Histogram of ORL dataset samples.

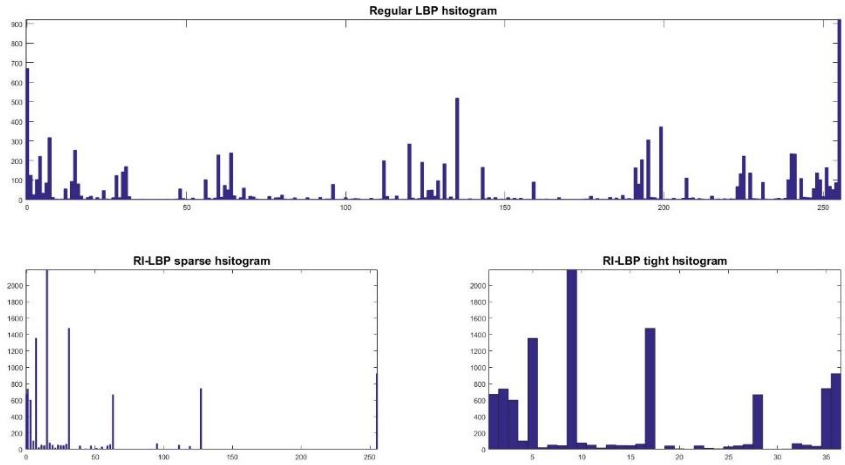

Figure 4.2. Histogram of LBP (Local Binary Pattern) on ORL Sample

4.3 Result

The images is Extract into LBP and HOG. Figure and figure show the LBP and HOG result in image representation. From figure 16 and 17, the HOG show the gradient in the ORL image sample. Figure show the LBP feature extraction using ORL image sample.

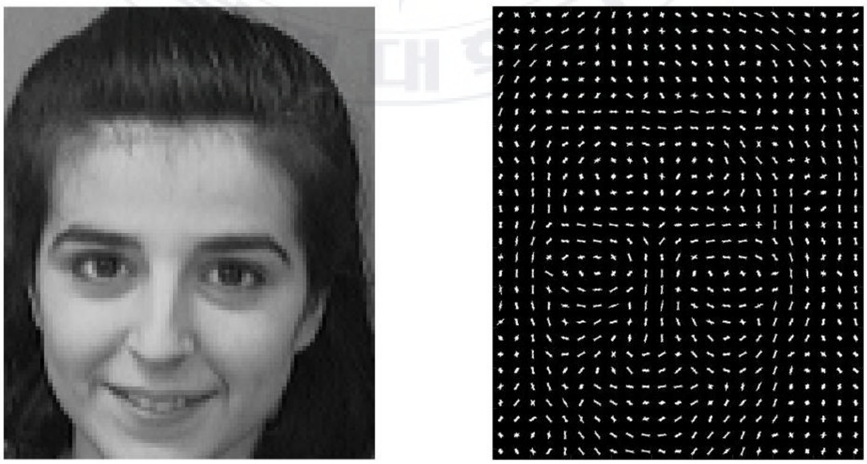

Figure 4.2 HOG on ORL Sample

Figure 4.3. LBP on ORL Sample

From the training, using LBP and DBNs the recognition rate has the accuracy ±98.3%. In other hand using HOG and DBN the recognition rate has the accuracy ±90% and using HOG overlapping and DBN the recognition rate has the accuracy ±90%. Table is show the comparison of the Recognition rate using SVM, RF and DBN, from the table the LBP+DBN has the highest recognition rate among the methods.

Table 4.1. Comparison of Recognition Rate

No	Method	Accuracy
1	HOG+DBN	90%
2	HOG (Overlap) + DBN	90%
3	HOG+RF[4]	95.1%
4	HOG+SVM[4]	94.4%
5	DBN+LBP	98.3%

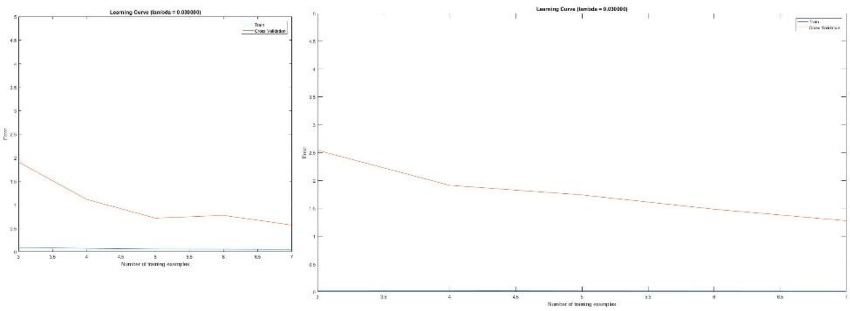

Figure 4.3. Training Curve LBP & HOG with DBN Network

Chapter V.
Conclusions & Future Research

5.1 Conclusions

The result shows that CNN (Convolutional Neural Network) is a good method for detecting and recognizing the human face with the detection rate 75% and recognition rate is 70%. CNN can be trained without using feature extraction. CNN need to be trained with a lot of dataset because is based on deep learning methods.

From the DBN (Deep Belief Network), the result shows that DBN is a useful tools for recognizing the human face with the recognition rate 98.3% using ORL dataset for training and testing using LBP (Local Binary Pattern) and The DBN (Deep Belief Network) using HOG (Histogram of Gradient) has 90 % of recognition rates. The system is still on development to make the reliable for different condition such as illumination and low light. The ORL dataset is not covering the illumination variant and low light condition. DBN need to be tune in a systematic way in the Number of Layers and Number of Units in each layer to make it more optimize.

5.2 Further Studies

For further research, CNN and DBN can be combine into another method like montecarlo searching and etc. after combining the CNN and DBN with another methods, then the CNN and DBN is integrated into one whole system to program more reliable for solve the illumination variant and low light condition in face detection and recognition system.

For feature extractions, the other feature extraction beside HOG and LBP can be used for extract information from the images and combine with the DBN or CNN such as :

- Edge detection
- Corner detection
- Blob detection
- Ridge detection
- Scale-invariant feature transform

The PCA (Principal Component analysis) can be used to optimize the feature extraction and the machine learning as well so they can get more information and training well.

References

[1]. X. Zhang and J. Ding, "An improved Adaboost face detection algorithm based on the different sample weights," 2016 IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Nanchang, China. 2016.

[2]. D. Reney and N. Tripathi, "An Efficient Method to Face and Emotion Detection," *Communication Systems and Network Technologies (CSNT), 2015 Fifth International Conference on, Gwalior*, 2015, pp. 493-497.doi: 10.1109/CSNT.2015.155

[3]. N. Soni, N. Singh, A. Kapoor and E. K. Sharma, "Face recognition using cloud Hopfield neural network," *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, Chennai, India. 2016.

[4]. A. Krizhevsky, I. Sutskever, G. E. Hinton, A. Krizhevsky, I. Sutskever, and G. E. Hinton, *ĀImageNet Classification with Deep Convolutional Neural Networks,ā Advances in Neural Information Processing Systems*, pp. 2012, 2012.

[5]. R. D. Hof, Deep Learning, 2013. [ONLINE] Available at: https://www.technologyreview.com/s/513696/deep-learning/. [Accessed 10 October 2016].

[6]. Hubel, David H., and Torsten N. Wiesel. “Receptive fields and functional architecture of monkey striate cortex.” *The Journal of physiology 195.1* (1968): 215-243.

[7]. LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. “*Deep learning.” Nature* 521, no. 7553 (2015): 436-444.

[8]. Bengio, Yoshua , 2012. Learning Deep Architectures for AI. Learning Deep Architectures for AI, [Online]. Available at: https://www.iro.umontreal.ca/~bengioy/papers/ftml.pdf [Accessed 16 December 2016].

[9]. Yingzi Du, Eliza. Biometrics from Fiction to Practice. Florida: CRC Press, 2012

[10]. Goodfellow. Ian, Bengio. Yoshua, and Courville, Aaron. 2016. Deep Learning Book. [ONLINE] Available at: http://www.deeplearningbook.org/contents/intro.html. [Accessed 10 October 2016].

[11]. Y. Tian, T. Kanade, J. Cohn, S. Z. Li and A. K. Jain *Handbook of Face Recognition* pp. 487-519, 2011, Springer

[12]. Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. Neural Computation, 18, 1527–1554.

[13]. Taylor, G., Hinton, G. E., & Roweis, S. (2007). Modeling human motion using binary latent variables. Adv. in Neural Information Processing Systems.

[14]. DC. He and L. Wang (1990), "*Texture Unit, Texture Spectrum, And Texture Analysis", Geoscience and Remote Sensing, IEEE Transactions on, vol. 28*, pp. 509 - 512

[15]. Wikipedia. 2016. Histogram of oriented gradients - Wikipedia. [ONLINE] Available at: https://en.wikipedia.org/wiki/Histogram_of_oriented_gradients. [Accessed 10 October 2016].

Acknowledgement

First of all, I would like to give my highest praise to The God, Allah Subhanallahu Wa Ta'allaa and my lovely family for all helps and guidance that have been given such that I could finish all my research and thesis in Master program very well. All achievement and success that have been granted to me only can I return to the One who owns them. In this right moment of time, I would like to express my deep gratitude to all people who helped me during my study and meaningful life in Rep. of Korea.

Firstly, I would like to express my gratitude to my beloved father who died because of kidney failure when i was study in Rep. of Korea.

Then, I would like to express my deepest gratitude to my professor, 신봉기 교수님, who has helped me significantly in completing my Master program. I really appreciate his kindness and helps for completing my master program.

Secondly, I would like to express my sincere gratitude to the members of my thesis committee, Professor 박만곤, 김창수, and 이경현 for the guidance for living in South Korea.

Thirdly, I would like to express my profound gratitude to Prof. Charmadi Machbub and Dr. Tech. Ary Setijadi P. for his great help and advice such that I could finish my research project and my master thesis very well.

Furthermore, I would like to express my gratitude to all members of IMedia Laboratory Pamela, 양우희, 임아현, 김정주 for their cooperation, kindness and friendship, especially my Indonesia lab member Jemmy

Kristianto Wiguno who help me a lot for finishing my thesis, paper and help me for living in Rep. of Korea

Also I would like to express my deepest gratitude for my friends in Bandung who help me with my personal matters in Bandung (Khalis Khalifanatama Adiranti and Ikhsan Maulana Malik)

And I would like to express my deepest gratitude to my family in Korea and ITB students Ariana, Aditya, SPL, Annisa, Gilang, Yogha, Vincent, Aldias, Devi Saskia, Zahra Pradanya, Heny, and all my relatives for their love, endless prayers, encouragement and support for me not only in completing Master program but also in the whole of my life.

Busan, February 2017

Adyan Marendra Ramadhani