Face recognition by using hybrid-holistic methods for outdoor surveillance systems

A dissertation submitted to

THE GRADUATE SCHOOL OF ENGINEERING SCIENCE

OSAKA UNIVERSITY

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN ENGINEERING

BY

SADI VURAL

MARCH 2011

OSAKA UNIVERSITY

ABSTRACT

FACE RECOGNITION BY USING HYBRID-HOLISTIC METHODS

FOR OUTDOOR SURVEILLANCE SYSTEMS

BY SADI VURAL

Chairperson of the Supervisory Committee: Professor TATSUO ARAI
Department of Systems Innovation

This thesis introduces a real time many-to-many face recognition approach, which is targeted for recognizing/identifying human faces in unconstraint environments such as in a street. Initially, a brief review of multi-view face detection and eye detection method is presented. In particular, a new filter set of haar-like filters is presented for face detection and six segment filters are presented for eye detection. Following this, an illumination technique is introduced to minimize the illumination effects on a face surface. In addition to these, we introduce a novel face extraction and classification methodology for face recognition in uncontrolled environments. The detected faces in a given image are then processed by the proposed face extraction module and the resulting features are compared with a precompiled face database. If a match is found, system gives an alarm. In an image, more than one face at a time is detected in a crowded environment. All the detected faces are matched with the face database. Database includes one record per person and each record occupies 10KB space.

Illumination normalization technique uses Gaussian decomposition. We propose a six-vector set which we call it as “Ayofa-filter”. This is used as a preprocessing step for the face recognition in our work.

Ayofa-filter is a new approach that consists of Gabor decomposition and albedo estimation on face normal where the direction of the light source is unknown. This approach effectively finds illumination directions and recovers the illumination. The approach computes unknown albedo directions by using spatial frequency components on salient regions of a face. It requires only one single image taken under any arbitrary illumination condition where we do not know the light source direction, strength, or the number of light sources. Ayofa-filter takes the nose tip as a reference point before the normalization of the light.

Face recognition is another item we propose in this thesis. After face is located in an image, we refine to find eye locations, nose tip location, and mouth corners. After that, face is cut, geometrically normalized and then processed by the illumination normalization proposed in chapter 3. The face features are extracted and compared with the face database to confirm if the person has a record. The process that starts from the feature extraction is the part of the face recognition.

To perform face recognition, we developed a hybrid face feature extraction methodology, which uses spatial face filter (SFF) for extracting invariant face features to obtain high accuracy. Spatial face filters are implemented by spatial gaussian bessel mixtures (SGBM). SGBM is the combination of gabor filters and bessel function. Gabor filters are not sufficient to extract face features. Therefore, we combined the gabor filters and bessel function to generate more complicated filters. We applied the SGBM to the whole face surface, extracted the face features and analyzed them by using hierarchical nonlinear PCA (H-NPCA). H-NPCA extracts not only the most distinguished information of the face but also it removes the image noise by approximating the missing data. This method analyzes the face features and each feature set is independently computed by H-NPCA. After extracting the face features, we obtain high dimensional feature arrays. It is important to do dimension reduction without losing the discriminative face features. We achieve this by using H-NPCA.

We tested the overall approach by using still images and video data by using major face recognition databases. The images from these databases are carefully selected so that the testing images represent the outdoor illumination conditions. The efficiency of the methods proposed here

was tested by standard face recognition algorithms, namely principal component analysis (PCA), linear discriminant analysis (LDA), Gabor wavelets and active appearance model (AAM) methods across the multiple face databases. The evaluation results show that our novel method significantly improves the recognition ratio with these recognition methods.

Finally, an automatic face recognition system which consists of detection, recognition and enrollment modules has been developed. The system can effectively detect faces from IP camera video streams, recognize them and retrieve corresponding person information from the application database if a match is found.

Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

AAM	:	Active Appearance Model
AF	:	Ayofa-Filter
AHF	:	Adaptive Histogram Fitting
aIT	:	Anterior Inferior Temporal Cortex
ANN	:	Artificial Neural Networks
CNN	:	Convolutional Neural Networks
CPU	:	Central Processing Unit
EB	:	Extended Bessel
EER	:	Equal Error Rate
FAR	:	False Acceptance Rate
FFA	:	Fusiform Face Area
FFC	:	Face Feature Comparison
FERET	:	Face Recognition Technology
FRGC	:	Face Recognition Grand Challenge
FRR	:	False Rejection Rate
GA	:	Generic Algorithm
GPU	:	Graphics Processing Unit
HD	:	High Definition
HE	:	Histogram Equalization
H-NPCA	:	Hierarchical Nonlinear Principal Component Analysis
ICA	:	Independent Component Analysis
IP	:	Internet Protocol
LAN	:	Local Area network
LDA	:	Linear Discriminant Analysis
LTV	:	Logarithmic Total Variation
MSE	:	Mean Square Error
MSR	:	Multi-Scale Retinex
NICA	:	Nonlinear Independent Component Analysis

NIR	:	Nonlinear Intensity Filtering
NLPCA	:	Nonlinear Principal Component Analysis
N:N	:	Many-to-Many
OP	:	Optimization Problem
QP	:	Quadratic Problem
PCA	:	Principal Component Analysis
POE	:	Power-on-Ethernet
RAM	:	Random Access Memory
ROI	:	Region of Interest
SFS	:	Spatial Face Filter
SGBM	:	Spatial Gaussian Bessel Mixture
SFF	:	Spatial Face Filters
SQI	:	Self-Quotient Image
SQL	:	Structured Query Language
SSF	:	Six Segment Filters
STS	:	Superior Temporal Sulcus
SVM	:	Support Vector Machine

ACKNOWLEDGEMENTS

First of all, I would like to start by giving my sincere gratitude to my supervisor Prof. Tatsuo Arai for his patience, wise advices, orientation, ideas and suggestions to complete this thesis. He accepted me to study in his lab and encouraged me to complete this thesis. He has helped me to the utmost of his ability and understanding the situations all the time. I am very proud of working under his supervision. I also strongly thank to Prof. Hiroshi Ishiguro, Prof. Kosuke Sato, and Assoc. Prof. Yoshio Iwai for their critical reviewing, encouragement and valuable suggestions during my presentations, reviewing of this thesis. I feel genuinely lucky to have such great professors around me.

The work presented in this thesis has been realized with great help of number of people, whom I gratefully would like to acknowledge: Assoc. Prof. Yasushi Mae, Dr. Huseyin Uvet and Yuriy Chesnokov who helped me during paper reviews, corrections and the intensive review of my thesis

I also would like to thank to Prof. Bulent Sankur and M.D. Assoc. Prof. Tahsin Beyzadeoglu for technical assistance, useful discussions and ideas. Bulent Sankur gave me many opportunities and chances all the time. Tahsin Beyzadeoglu gave me many ideas during our talks on micro surgery. I had new concepts and new paths to try. He also supported and encouraged me in various ways.

I would like to dedicate this dissertation to my lovely father, Mustafa Vural and my mother, Rahime Vural, my lovely sister Rabia Cetin and her husband Tarik Cetin and my beloved wife, Kayo Vural, my little baby, Reina Vural and my parent-in-law Yoshio and Fumie Hayashi who supported and encouraged me to continue in my most difficult times. I would not be able to complete this thesis without their support. They have always been with me. Their helps cannot be expressed by words.

Finally, I would like to thank to Ugur Aksoy and Bulent Kilic for their greatest help and motivation. Whenever I needed them, they were one-phone-call distance at any time. They worried a lot about me, the continuously kept asking my situation. They have become my best friends and will stay to be. I will never forget their support throughout my life. I have been so lucky to have such great friends in my life.

CHAPTER 1: INTRODUCTION

The major concern of this thesis is to develop an automatic face recognition system, which is robust against to variance in illumination. At the same time, the system has to take the computational cost into consideration for real time applications. This chapter will give a brief introduction to the reason and background of this research, past studies, organization of my published papers. Following this, the description on challenges behind the research and the organization of the thesis will be introduced.

1.1 Why face recognition?

What motivates us to conduct this research is the increasing demand on security in public locations. Automatic identification of a wanted/criminal person in public locations has become a very important topic for government organizations. In addition to this, accurate automatic identification is now needed in a wide range of civilian applications involving the use of passports, automatic teller machines and driver license. Traditional methods rely on ID Cards and physical inspections and there is no method to find a person in public. Therefore, there is a strong need for a new technology that can identify people in public locations.

1.2 Background and scope of the thesis

Facial recognition is an important capability of human beings in daily life. We can remember hundreds or even millions of faces in our whole life and identify a face in different perspective variations, illuminations, ages, etc. Under very poor illumination conditions, a face can still be recognized, as the position of the different facial features and the face contours are usually sufficient for recognition. This suggests an approach to face recognition, whereby the

geometrical positions of the different facial features are measured first, and then the details of each feature are used for further matching. This is also a common approach for object recognition. Recognition requires distinguishing between millions of faces we keep in our brain from birth to present time. This computational feat involves a network of brain magic. The detection of the faces is done in the fusiform face area (FFA) [1],[2],[3] and recognition event happens inside the Anterior inferotemporal cortex (aIT) [4].

The FFA responds whenever a face is visible. FFA distinguishes faces from any other objects. FFA does facial detection [5], [6]. After FFA detects faces, it feeds the face information to anterior inferotemporal cortex (aIT). For aIT, human lesion and neuroimaging studies suggest a role in face identification. Neuroimaging studies [4],[7],[8],[9],[10] found anterior temporal activation during face recognition with the activity predictive of performance[7]. The studies [11],[12],[13],[14] predict that the right anterior inferotemporal cortex is involved in face recognition.

Kanwisher [1] observed that the cortical region of the lower bank of the superior temporal sulcus (STS) process the faces in holistic manner. Gauthier et al. [16] claims that aIT cannot resolve the high spatial frequencies that make up the fine details of faces by observing the activation regions of the fusiform face area. From this fact, it can be concluded that STS uses holistic methods to recognize the face. Based on these facts in neuroscience, several approaches have been proposed for face recognition by computer scientists [15],[17],[18],[19],[20] .

There are also face recognition algorithms which are based on local face features such as Gabor wavelets [21], holistic methods such as principal component analysis(PCA)[18],[22],[23], linear discriminant analysis (LDA)[24] and canonical correlation analysis[55]. A more comprehensive survey on other algorithms is done by Zhao et al. [25]. Ekenel [26] proposed a face recognition method by using DCT (discrete cosine transform). He took the image features by DCT and applied to KLT (Karhunen-loeve transform). He says DCT is powerful tool to extract face features but he did not mention about face pose and face illumination. DCT is weak against to illumination direction and face pose. The features change in small changes of face appearance. Savvides et al. [27] introduced correlation filters which they claim correlation filters are robust to illumination variances and face shifting. They compute the minimum correlation

energy from image data. However, their method performs poorly when face data appears with distortions (e.g., rotations, scale changes, noises). Wang et al. [28] addresses this challenge with the introduction of a combination of bayesian probabilistic model and Gabor wavelets. He extracted the face features by Gabor wavelets and classified by a bayesian algorithm. He solved the rotation and scaling problems by using Gabor wavelets. Gabor wavelet is strong to such parameters according to the paper.

However, Gabor wavelets are not specially designed for face recognition. Gabor features do not contain face specific features. For example, it is unknown that which scale and frequency components are the most important and discriminative for face recognition. Furthermore, Gabor wavelets are intolerant to illumination changes. Therefore, they have some limitations for using widely in uncontrolled environments.

Wright et al. [29] introduced a sparse classification algorithm in his paper. He developed a sparse representation of each individual and applied the sparse representation to a general classification algorithm. Although sparse representation partly solves the illumination and occlusion problems, the face expression and pose images cause high misrecognition due to the linear feature analysis. To overcome the misrecognition caused by linear feature analysis, Yang [30] compared kernel PCA and kernel LDA to analyze the features by kernel-based algorithms. Kramer [31] proposed nonlinear PCA (NLPCA) with auto-associative neural networks. He generated the NLPCA from PCA components and used standard neural network topology to generate nonlinear features. He used three-hidden-layer architecture to acquire the feature correlations. However, data discrepancy problems occur if there is a data out of cycle. There are more factors to cause discrepancy in face recognition than any chemical data analysis, which Kramer focuses on. In addition to this, the neural network structure has some capacity limitations to extract the nonlinear features from the given data. Tan et al. [32] used a different version of NLPCA for data compression to overcome the limitations. He trained the demapping subnet only instead of training the complete hidden layers as Kramer did. He computed the PCA components to train demapping subnet. Although the proposed method is good for small number of data analysis, it is not appropriate for large number of face data and the demapping subnet gives high training error since face features include huge variations and many factors make the features change. Chen et al. [75] offered a texture feature descriptor by using weber law. He analyzed the

selective features to get high classification accuracy. Zhang et al. [76] introduced a face recognition technique by using SVM and an inductive learning method which is derived from bayes decision theory. Their purpose is to process faces in real-time. Although their method is promising in speed, the learning performance of the proposed method is not robust as multi pose faces are considered. Lee et al. [77] offered a tensor-based face recognition that recognizes under various poses of faces effectively as the paper says.

Both NLPCA offered by Kramer [31] and Tan et al. [32] cannot model the overlapped data. In addition to this, both of them cannot model the parameterizations that have discontinuity region. Kramer method needs continuous function, which can be only achieved by a feedforward neural network. Malthouse [33] introduced a straightforward technique to cover the discontinuity regions. He used a nonlinear optimization routine by filling each neural network by eight hidden nodes. He finally reduced the possibility of incorrect projection. Although he filled up some gaps of Kramer, the NLPCA gives incorrect projection if there is testing data close to ambiguity points. Our proposed H-NPCA method gives solution to both missing data (discontinuity) and overlapping data.

In this thesis, we introduce a face recognition method, which is based on a combination of holistic methods and local feature approaches. The thesis explains both proposed holistic method and local-feature approach. Finally, this thesis gives a system infrastructure for practical use. Apart from this, the effect of some preprocessing techniques, namely illumination normalization, is given in this thesis. Finally, the results are compared with respect to the effects of changes in illumination.

1.3 Technical challenge

Although there is an increasing demand to face recognition, there is no technology that satisfies the industry needs. There are several problems to prevent the industry from using this technology. Although the problems are difficult to classify, we listed some of such challenges which are the most significant problems.

- Illumination challenge

Although the performance of face recognition systems in indoor platforms has reached a certain level, face recognition in outdoor platforms still remains as a challenging topic. Face recognition performance is significantly affected by the problems caused by variations in illumination, face pose, expression, aging and etc. Especially, the effect of variation in the illumination conditions, which causes dramatic changes in the face appearance, is one of the most challenging problems that a practical face recognition system needs to achieve. More specifically, the variations between the images of the same face due to illumination and viewing direction are almost always larger than image variations due to change in face identity.
We introduced an illumination invariant face recognition approach to overcome illumination challenge. We introduced an albedo prediction without any need to light source. The proposed method is explained at chapter 2 in detail.

- Face pose

In a surveillance system, the camera is mostly mounted to a location where the people cannot reach to the camera. Mounting a camera a high location, the faces are viewed by some angle degree. This is the simplest case in city surveillance applications. The next and the most difficult case is that people naturally pass through the camera view. They do not even look at the camera lens. Authorities cannot restrict people behaviors in public places. Recognition in such cases must be done in an accurate way. However, even state-of-the-art-techniques have 10 or 15degree angle limitation to recognize a face. Recognizing faces from more angles is another challenge. The most significant face features are lost after an angle of 25 degree or 30 degree. Hence, the system reliability decreases exponentially. The techniques proposed until now did not give good results in actual work conditions since there are many other factors that are added to face pose problem in outdoor environments.

We researched a face rotation methodology to compensate the angle. Our face rotation method is motivated from active appearance model (AAM). We remodel the frontal face during training. We enroll multiple pose images of each individual from different angles. During the

recognition process, we search the in-class of the angular face. This part is not included in this thesis.

- Face expression

Face expression is less significant issue compare with angle and illumination but it affects the face recognition results. Although a close eye or smiling face does affect the recognition rate by 1% to 10 percent, a face with large laugh has an influence as more as 30% since a laughing face changes the face appearance and distorts the correlation of eyes, mouth and nose. Hence, the features are grouped as a different class. This suddenly increases the false alarms. Many research papers focus on small changes on the face surface. However, huge changes in expression are still an unsolved problem.

During our research work, we understood that reconstructing the face components will solve this challenging problem. Our latest research which remodels the face by using texture, shape and spatial frequency decomposition should overcome this challenge.

- Race difference

Face recognition performance varies with races. Face recognition algorithms for western people cannot give good performance on faces of Asian people or Latin people, African people. The face structure differs by race. In Asian people, the eye bones have no depth. However, in African people, eye bones are opposite to Asian faces. In Latino people, cheek bones are point of interest. Such race differences affect face recognition. State-of-the-art techniques focus on westerns since there are many dataset to test the methodologies. However, there is no large race database. Therefore, each region optimizes face recognition algorithms to their region. Face recognition algorithms must be race-free. However, there is no method for this and this remains a challenge to solve. To overcome this, a race identification method must be done. Race difference is a kind of multi-pose face detection methods. Defining different races in different classes and separation of classes by bayes-based techniques or SVM is needed. However, cross-race people are another problem in grouping the classes.

- Face aging

Face recognition algorithms are using either geometrical techniques or feature-based approaches or holistic methods. All of them do not solve the aging problem. Almost all of them give an age tolerance as long as 20 years after the training. Faces between 1 year and 15 years cannot be recognized since face appearance changes fast. Face appearance becomes stable after teenage years. A recognition algorithm that can recognize faces for all ages does not exist.
We overcome this problem by our proposed hybrid-holistic methods. In hybrid holistic approach, the invariant face features are extracted and each class of feature set is processed separately and the features are classified by hierarchical nonlinear PCA.

- Occlusions

Occlusions cause erroneous facial feature localization. It has been shown that by solving the misalignment problem, very high correct recognition rates can be achieved with a generic local appearance-based face recognition algorithm. In the case of a lower face occlusion, only a slight decrease in the performance is observed, when a local appearance-based face representation approach is used. This indicates the importance of local processing when dealing with partial face occlusion. Moreover, improved alignment increases the correct recognition rate also in the experiments against the lower face occlusion, which shows that face registration plays a key role on face recognition performance. The challenge is on how to obtain the alignment points. Normally, eyes points are used for face alignment. However, a sunglass or occlusion due to the illumination prevents the detection of eye coordinates. There are some studies to use more points to compensate points that are not detected. They need many points from different parts of a face. However, detecting many points from a face is not realistic in real-world conditions. We overcome this problem by our proposed six-segment-based eye detection as well as nose and mouth point detection. We make an alignment by using four points. The details of this method are explained in chapter 3.

- Speed

Recognition speed is very important. In a surveillance system, it is important that megapixel images should be used for the quality of the face recognition scores. Increasing the image size slows down the detection of the face, facial feature extraction. In addition to this, the matching speed becomes slow if the number of enrolled people increases to several millions of records. There are some accurate 3D-based face recognition methods but they take more than 5second per face. However, in a crowded environment, people movement is so fast that face recognition in a 100ms is necessary which includes the DB matching, image acquisition, face detection, image normalization and alarm reporting. We put our focus on this point during this thesis.

In this section, we reported the technical challenges. Since it is impossible to cover all of these in this thesis, we focused on illumination and speed issue during our work. Face expression and Face aging issues are minor problems and we have confirmed that these do not affect the overall performance. Face poses and face occlusions change the performance significantly. We will further continue our research to have a robust facial recognition to pose and occlusions. However, in this thesis, we ignore these two parameters and instead we concentrate illumination and speed.

1.4 Organization of the entire thesis

Chapter 1 is the introduction given about face recognition and the purpose of the thesis. Chapter 1 also gives explanation on background and the research studies in past.

Chapter 2 is about the face detection and eye detection. Proposed enhanced haar filters used for face detection and SSR filters are used for eye detection. The details of adaboost approach for face training are also given in this section.

Chapter 3 is about the illumination normalization technique. In this chapter, we explain the illumination normalization method we developed as a preprocessing step for the face recognition. In this section, adaptive histogram fitting technique, normalization by using gaussian filter and reflection direction estimation techniques are explained in detail.

Chapter 4 gives information on feature extraction. SGBM filters, feature selection, optimizations are all given here. In addition to this, feature analysis technique details are explained. Finally, feature classification technique is explained in this section.

Chapter 5 explains the overall architecture. We use the components of chapter 2, chapter 3 and chapter 4 to build a system which is appropriate for a surveillance system. It explains how the system is used and how effective the proposed model is.

Chapter 6 gives the testing results by using major face databases.
Most of the major databases are used to measure the efficiency and robustness of the methods. In chapter 6, the comparison with other systems is also provided with graphs and tables.

Chapter 7 includes the conclusion of the paper and the summary of the overall work and the extension of the current work program.

CHAPTER 2: MULTI-VIEW FACE DETECTION BY ENRICHED HAAR-LIKE FILTERS AND SIX-SEGMENT FILTERS

This chapter explains the methodology of the face detection and eye detection. Face detection is done by using specific haar-like filters and eye detection is done by using six-segment filters. We introduce 18 new haar-like filters and generate 108 filters by rotating each filter by 30 degree in clockwise. After giving the details of the haar-like filters and their combination, we train different face images by using the proposed filters and then we combine them to achieve multi-view face detection. The details are explained in section 2.2. After this, the principals of SSR computation and SSR theory are given. The last section gives the experimental studies by using indoor as well as outdoor images. The speed, error rates, detection rates are all given in section 2.7. Finally, we summarize and conclude the chapter.

2.1 Introduction

Given an arbitrary image, the goal of face detection is to determine whether or not there are any faces in the images and, if present, return the image location and extent of each face. To build fully automated systems that analyze the information contained in face images, robust and efficient face detection algorithms are required.

Face detection is a challenging problem because faces are nonrigid and have a high degree of variability in size, shape, color and texture. The challenges associated with face detection can be attributed to the following factors: pose (frontal, rotated), presence or absence of structural components (bear, no-bear), facial expression, occlusion, and lighting. These problems also arise

in automatic land-marking so we further discuss about them in section 2.2.

Numerous techniques have been developed to detect faces in a single image [79], [80] until now. A large majority of such studies reports that haar-like filtering methodology is the most powerful and fast method in object detection. Therefore, many subsequent works have been refined and extended the haar-like filter-based approaches [81], [82], [83], [84]. Most of these methods construct a weak classifier by selecting one feature from the given feature set. Schneiderman et al. [85] proposed an impressive object detection method by using the statistics of appearance attributes to construct the likelihood classifiers. The computation of likelihood classifiers is slower than haar-like classifiers.

Here, a face detection method which uses enriched haar-like filters and six-segment filters (SSF) is introduced. P. Viola and M. Jones [58] proposed a multi-stage object classification procedure that reduces the processing time substantially. They developed a new face detection method. Their method uses some face-weighted rectangles and weak cascade classifiers. They applied rectangle black and white shapes to find the face in the image. All rectangles have different weights and scales to reflect the characteristics of a human face. Lienhart et al. [59] introduced a rapid face detection method by using haar-like features to detect a face in real time by improving the P. Viola and M. Jones algorithm [58]. Maydt et al. [62] conducted research on face detection using a support vector machine to increase the performance of face finding. On the other hand, Kawato [60] offered first to find a face among face candidates with low processing time. His basic idea is derived from simple rectangular cell logics. His method relies on two eyes of the face. Precondition is that face is not occluded and eyes are open. Kawato applied SSF to the face image to find eyes on the image. He needs at least 80x80 pixel-sized SSF on 320x240 pixel images. The filter types and details are given in this section.

2.2 Enriched haar-like filters and application to face detection

Haar boosting is a fast face finding method. It uses boosted cascade of simple features. The structure of Haar-like filters is first introduced by Viola et al. [58] and it is improved by Lienhart [59]. Viola et al. [61] offered a complimentary haar features to Lienhart's work in his

paper. Shihavuddin et al. [74] developed a face detection system by haar-like features and adaboost algorithm. He used a different set of haar-like features to process large number of features in real time. Although their approach is good, the method is weak to outdoor illumination. The selected features do not simulate the outdoor environment sufficiently. Therefore, false rejection rates and false acceptance rates are very high. We introduce special haar features which are robust enough to run accurately in outdoor environments. We further make speed optimization for candidate selection among many objects. The resulting graph is as below.

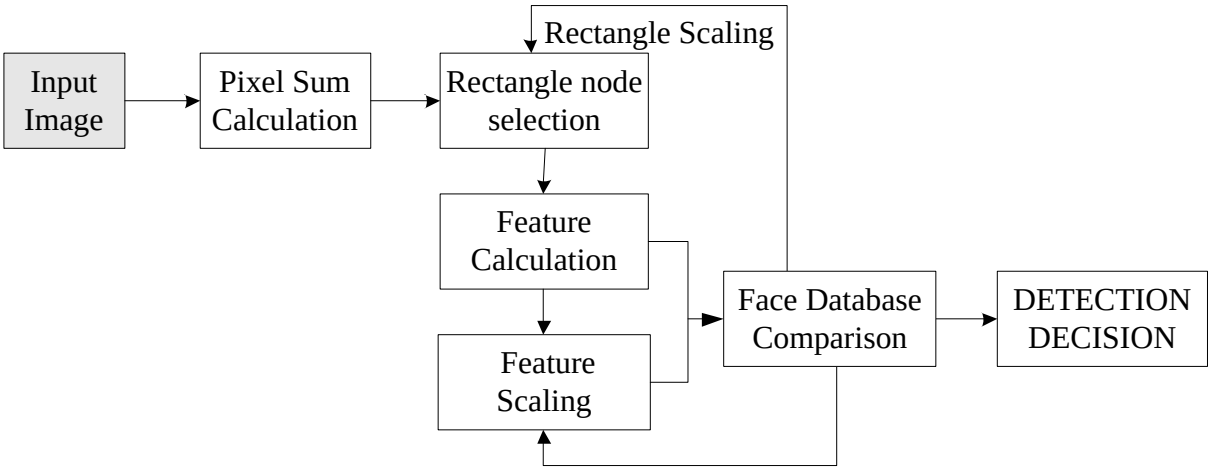
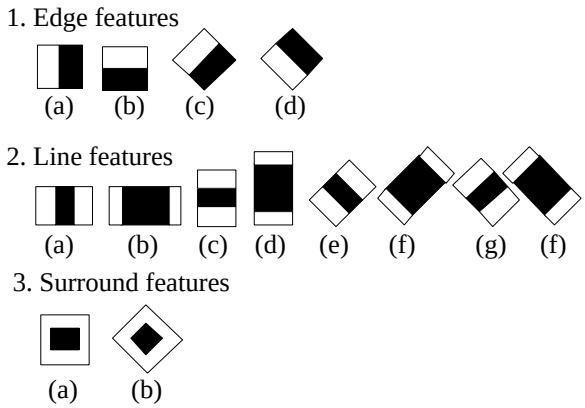


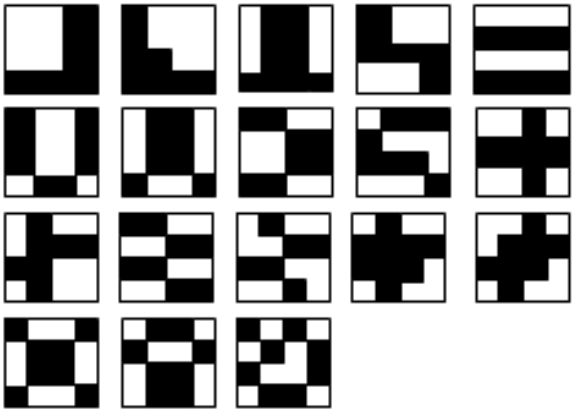
Fig. : Haar-like process flow

As seen in fig. 1, the sum of pixels is computed by integral image and nodes of rectangles are obtained. Using the nodes, haar-like features are calculated by using the filters that we propose and the result is compared with a precompiled face database cascade. The haar-like features that are used in our proposed algorithm are shown in fig. 2.

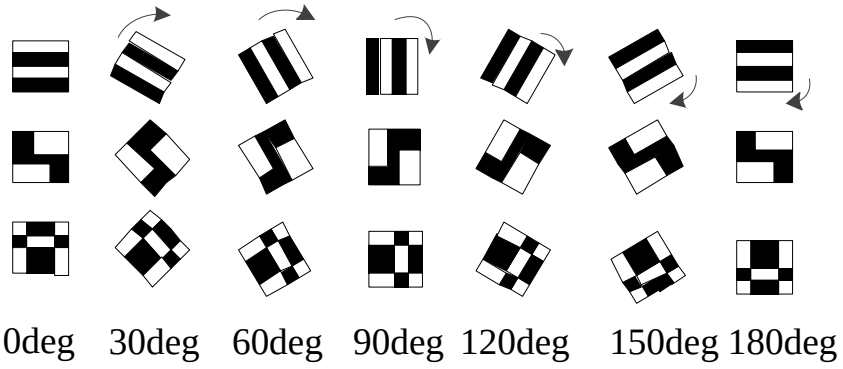
Feature scaling is done by applying an image pyramid. Face database comparison is done for each trained cascade group. We trained the faces from frontal, left, right, up, down and profile. The comparison is done for each one separately. The decision is done by a threshold value. However, there is a possibility to find the same face by two cascades. To prevent such cases, minimum neighborhood is computed and if the two rectangles are very near, one is omitted. If two rectangles do not overlap each other, both rectangle areas pass the face detection process.



(a) Original haar-filters



(b) Proposed extended haar-filters



(c) Rotated filters in six different angles in clockwise

Fig. : Original haars, the proposed haar filters and their rotated filter set

Feature calculation as seen in fig. 1 is done by using the features given in fig. 2. To find a face, several rectangles are used to apply to the images. There are eighteen different filter sets and each filter has six different rotation rotated by 30 degree. Fig. 2(a) shows the extended haar like filters proposed by Lienhart. Fig. 2(b) is our proposed haar-like filters and Fig. 2(c) is some of the rotated filters generated from the filters shown in fig. 2(b). In original haar-like filters, there are more than 5000 features, which include the derived filters from edge features, line-features, and surround-features. In our proposed method, only 108 features are used as rotated filter and 18 original filters, which yield 126 filters. 126 filters are satisfactory to obtain all the necessary features from a face. Using them in a form of parallel cascading further improves the performance by making the proposed method more robust and efficient.

A classifier, which is simply a cascade group of boosted classifiers, is trained with a few thousands of samples of a face. The various face images are named as positive samples and the arbitrary images, which do not consist of the target object we try to detect, are named as negative samples. After a classifier is trained, it can be applied to a region of interest (of the same size as used during the training) in an input image. The classifier outputs a "1" if the region is likely to show the face and "0" if it is not like a face. To search for the face in the whole image one can move the search window across the image and check every location using the classifier. The classifier is designed so that it can be easily "resized" in order to be able to find the objects of interest at different sizes, which is more efficient than resizing the image itself. Therefore, to find an object of an unknown size in the image the scan procedure should be done several times at different scales. Hence, usually an image pyramid is constructed and the same process is repeated for all different scales of the image.

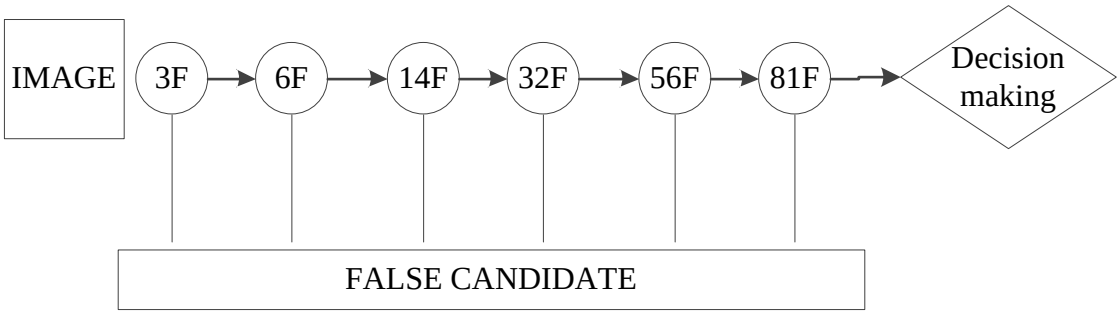


Fig. : Face decision by a series of classifiers (Detector mechanism)

In fig. 3, the in-circle denotation is given by a number and character. Number means the number of features used in the stage. For instance, 3F means 3 features are used. Input image is evaluated on the first classifier of the cascade (3F) and if that classifier returns false, then computation on that stage ends up and classifier returns false. If the classifier returns true, then the image is passed to the next classifier in the cascade (6F). The next classifier also evaluates the image as the same manner of 14F, 32F. If the image passes through all the classifiers successfully, then the candidate is returned as face. The more it looks like a face, the more classifiers are used to evaluate and the longer it takes. However, 80% of the face candidates are discarded as non-face at the first stages of the evaluation.

The evaluation process explained above uses gentle adaboost algorithm to decide if the candidate is a face or not after each classifier.

A detailed description of adaboost is as follows.

1. Given a training set T={xi, yi}, i=1,….,N, where N is the number of training images, $y_i \in \{+1, -1\}$ is the patterns of xi . +1 represents positive samples and -1 represents the negative images.
2. Given a cost parameter C>0, define a separate Ci for each training cascade

$$C_i = \begin{cases} \frac{2C}{C+1}, & \text{if } y_i = +1 \\ \frac{2}{C+1}, & \text{if } y_i = -1 \end{cases} \quad ()$$

3. Initialize the training set weighting

$$D_1(i) = \frac{C_i}{\sum_j C_j} \quad ()$$

4. For each t=1,…..,T, train the base classifier ht by using D_1.
5. Compute the following model error of the adaboost

$$\varepsilon_t = \sum_i D_1(i) \frac{1 - h_1(x_i) y_i}{2} \quad ()$$

6. And choose the adaboost parameter

$$\alpha_t = \frac{1}{2} \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right) \quad ()$$

7. Update the error

$$D_{t+1}(i) = \begin{cases} \frac{D_t(i) e^{-\alpha_t (2 - C_i}}{Z_t}, & \text{if } y_i h_t(x) = 1 \\ \frac{D_t(i) e^{-\alpha_t (2 - C_i}}{Z_t}, & \text{if } y_i h_t(x) = 1 \end{cases} \quad ()$$

where Z_t is the normalization factor. D_t is the probability distribution.

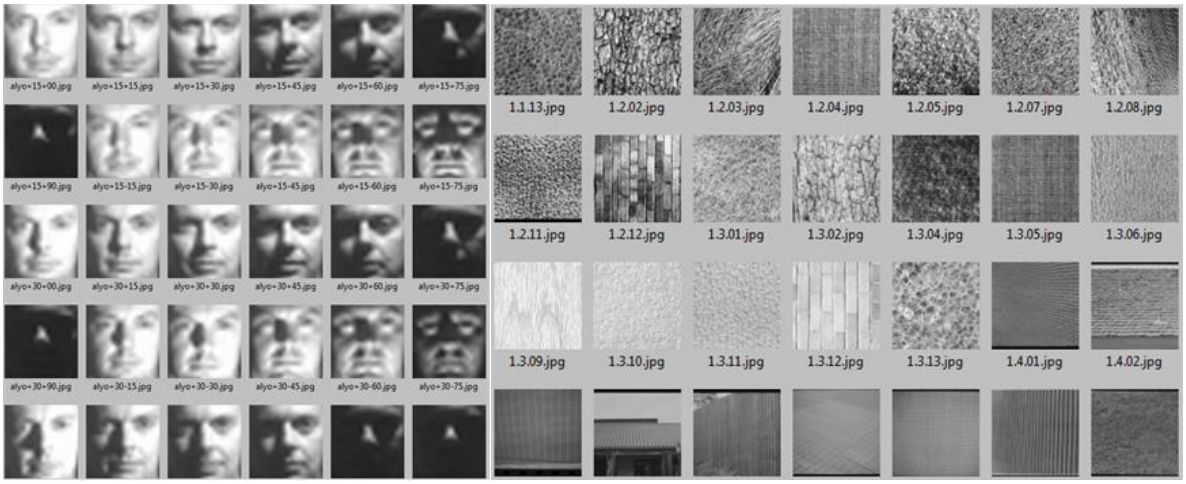
8. Finally, output the final hypothesis

$$H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right) \quad ()$$

In our proposed method, extended haar features are calculated very fast independently of image resolution and they are robust to the environmental noise, change in illumination and the object position. Face detector is trained per object separately by stagewise selection of weak classifiers based on the extended haar filters by using adaboost algorithm. The adaboost is better suited for the classification, besides that it produces perfect separation boundary in large dataset and provides an analysis of margin distribution [86], [87].

2.3 Cascade training

Training is the most exhausting part of the haar-like approach. It takes for long time to learn. It is due to the nature of the haar-like approaches. Haar-like approaches use a few thousands of features. This makes the training step very slow. In addition to this, thousands of training images (positive and negative images) are needed for a proper learning. Haar like approach cannot learn from a few images. Therefore, it is necessary to collect many images. Preparing them, cutting and alignment of them horizontally are all necessary steps before starting training. In our proposed filter set, we use only 126 filters which are all face-specific filters designed to capture the most significant features of a face. By using the proposed filter set, not only training time is short, but also there is no need to use many images. Sample training images (positive images) and negative images are given in fig. 4(a) and fig. 4(b).



(a) Positive images (b) Negative images

Fig. : Training set

These are taken under various illumination conditions. Fig. 4(b) is the negative images. These are random images collected from internet randomly. The only point is that negative images should not include any face image. Negative images can be nature images, building, street, figure, text, logo, car etc.

2.4 Combination of the training cascades

In above section 2.2 and 2.3, we explained the proposed haar filters and the training and detector mechanism. In this section, we explain the how to do the training and the combination of cascades.

Basically, Viola&Jones proposed one single cascade training and trained the images as one binary which was used for detection of the faces. Although the approach is simple and easy to implement, the detection rate is low and false alarms are high. In addition to this, the training takes long time. The reason for low detection rate and high errors is due to their feature selection. They offered very simple haar-like filters which are not adequate to represent the face features. Also, they considered only frontal faces in indoor environments. However, in a surveillance application, the environment is outdoor and faces have large pose variations. Therefore, their initial approaches become insufficient. To fill this gap, we established a framework where multiple cascades are trained and connected in parallel by using the proposed haar-like filters in fig. 2. Its structure is shown in fig. 5.

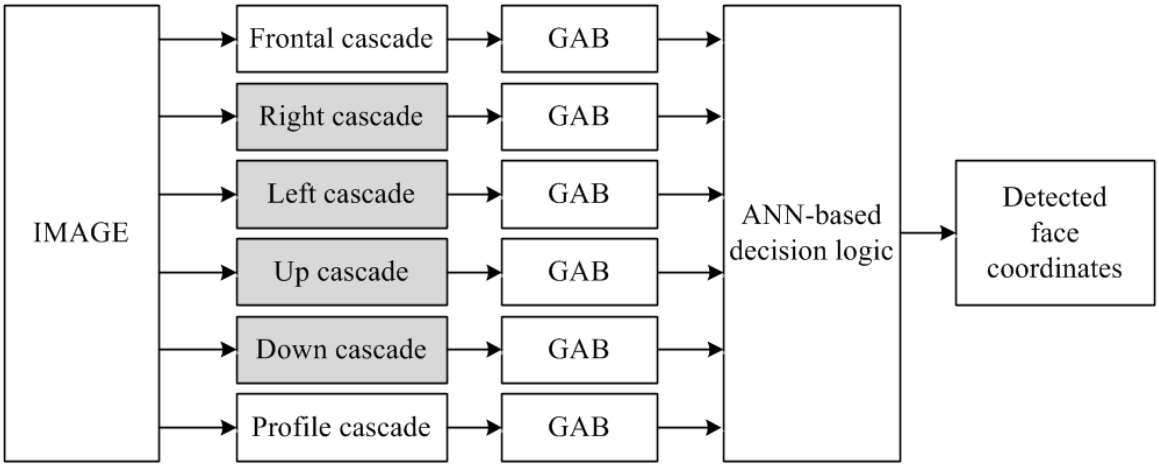
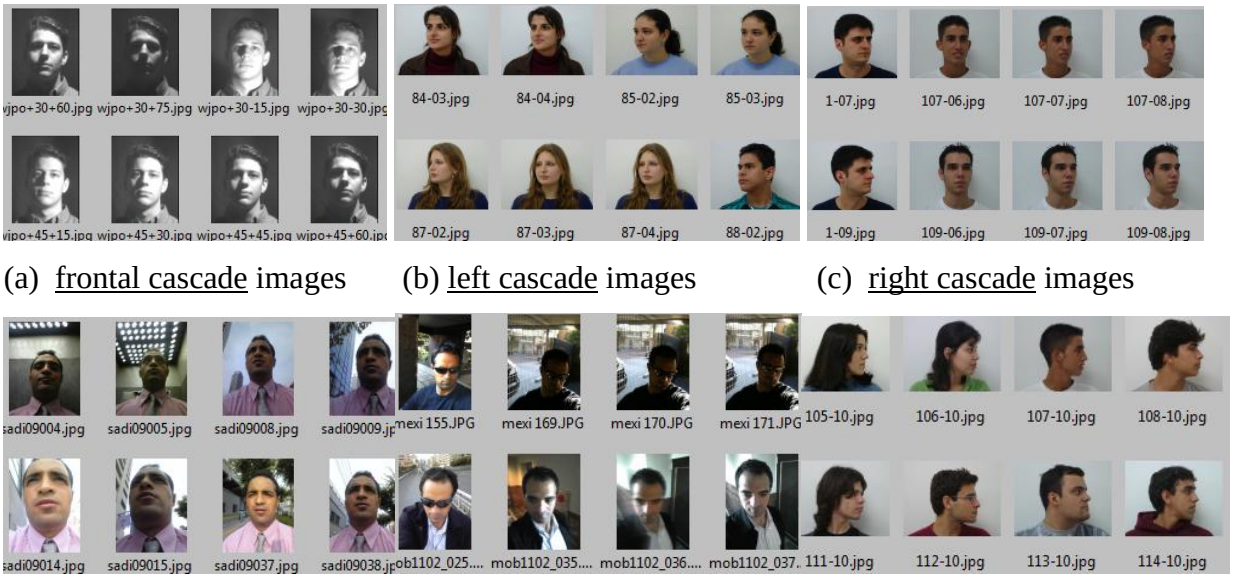


Fig. : Combination of cascades for multi-view

In fig. 5, the image is processed by all cascades at the same time. Each cascade provides a result. The results are evaluated by decision logic and the coordinates of the faces are given.

Decision logic uses neural network topology to decide if the candidate is an actual face or not. GAB is gentle adaboost. It evaluates the output of each cascade. The frontal cascade is trained binary data which is obtained by training frontal ± 5 degree faces. Right cascade contains right faces between 5degree to 45degree. Left cascade contains left faces with 5degree to 45 degree faces. Up cascade is generated by training 5degree to 30 degree up-looking faces. Down cascade is generated by training 5degree to 30 degree down-looking faces. The sample images that are used for training in each cascade are shown in fig. 6. Profile cascade depicts the profile face either from right or left. Profile images contain faces with an angle between 67.5 to 90 degree.



(a) frontal cascade images (b) left cascade images (c) right cascade images

(d) up cascade images (e) down cascade images (f) profile cascade images

Fig. : Cascade training images

The images of each cascade are selected from various databases, television broadcasting and internet. The image selection is done carefully. For example, face with large illumination variation, face with occlusion and large variation, a face with sunglass, mask, faces of different races, sex and ages are preferable images for all cascade trainings. In profile cascade, we train only right profile and invert the features of right profile to the left. Therefore, there is no need to

use any cascade training for the left profile. Each cascade contains indoor and outdoor images. The images have complex background and are taken under the various illumination conditions. The camera position is ignored. However, in most of images, the camera is either located in the front (straight to the face) except for the up and down face images. The further details of each cascade are given in table 1.

Table : Multiple-cascade details

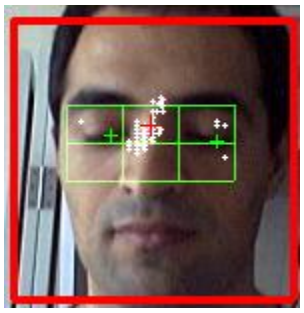
Cascade	# of feature	# of positive images	Object size	Training time	Speed
Frontal	126	100	10x10	10 hours	40ms
Right		50		6 hours	30ms
Left		50		5 hours	30ms
Up		50		6 hours	27ms
Down		50		6 hours	30ms
Profile		50		6 hours	26ms

As seen in table 1, a number of 100 images for frontal cascade was used and training time needed to train 6 cascades by using 126 features was 10 hours. The detection time that is necessary to detect an object of 10x10 pixel size is 40ms. For the right, left, up and down cascade training, 50 images per each category were used and the training of each cascade took approximately 5 hours to 6 hours.

2.5 Face feature extraction by six-segment filters

After finding the face location by using the approach in section 2.4, eye detection is needed for the facial recognition that is used for various purposes such as face normalization. SSR is a filter structure which Kawato [10] discovered to find a face among many candidates with low processing time. The basic idea comes from simple rectangular cell logic. We assume that the person to be recognized has two clear eyes, not hidden. Using this assumption, we apply our SSR

filter to the face image to find whether or not there are eyes on the images and to find where the eyes are in.



Eul	**Nu**	**Eur**
Edl	**Nd**	**Edr**

Fig. : SSR filter application to the face

Fig. 7 shows the filter application to the face surface and the eye and nose bridge candidates. "+" mark shows the right candidate and white circles are other candidates.

To compute the SSR, we assume that nose area is the brightest area. Eyes are the darkest locations on a face. Let N_T be the summation of the nose bridge and nose wall which connects nose bridge to the nose tip, El_T be the summation of the left eye and the lower eye part on the left, Er_T be the summation of the right eye and the lower eye part on the right, Eu_T be the total value of the upper rectangle(left eye area, nose bridge and right eye area) and Ed_T be the lower part of the Eu_T, then these can be computed by $N_T = Nu + Nd$, $El_T = Eul + Edl$, $Er_T = Eur + Edr$, $Eu_T = Eul + Nu + Eur$ and $Ed_T = Edl + Nd + Edr$

where Nu is the upper nose area, Nd is the lower nose area, Eul is upper left eye, Eur is upper right eye, Edl is lower left eye, Edr is lower right eye. Then,

$$N_T > El_T \quad ()$$

$$N_T > Er_T \quad ()$$

$$Eu_T < Ed_T \quad ()$$

$$El_T \cong Er_T \quad ()$$

As seen in the equations, N_T must be bigger than both the left side of the rectangle (left eye region) and the right side of the rectangle (right eye region). In other words, nose area is always whiter than eye areas. Furthermore, Eu_T must have lower value than the Ed_T since the lower part of the eyes has always brighter than the eyes. From this logic, El_T must be almost equal to the Er_T because of the symmetrical properties of the left side and right side of the eye areas.

When expression (7), (8), (9) and (10) are all satisfied, it can be a candidate for eyes.

The denotations given in fig. 7 show the rectangle subscript and their brightness relations are investigated to determine the eye locations.

The SSR filters are computed by using intermediate representation for image called integral image. For the original image i(x, y), the integral image is defined as

$$\text{ii}(x,y) = \sum_{x<x'} \sum_{y<y'} i(x,y) \qquad ()$$

where, i(x,y) is the image to be processed.

The integral image can be computed in one pass over the original image by the following pair of recurrences.

$$\text{es}(x,y) = \text{es}(x,y-1) + \text{i}(x,y) \qquad ()$$

$$\text{ii}(x,y) = \text{ii}(x-1,y) + \text{es}(x,y) \qquad ()$$

where, es(x,y) is the cumulative row sum and es(x,-1)=0 and ii(-1,y)=0.

Then, final equation is formed as below.

$$E_T = (ii(x,y) + ii(x-W, y-H) - (ii(x-W,y) + ii(x,y-H)) \qquad ()$$

where W is the pixel width and H is the pixel height.

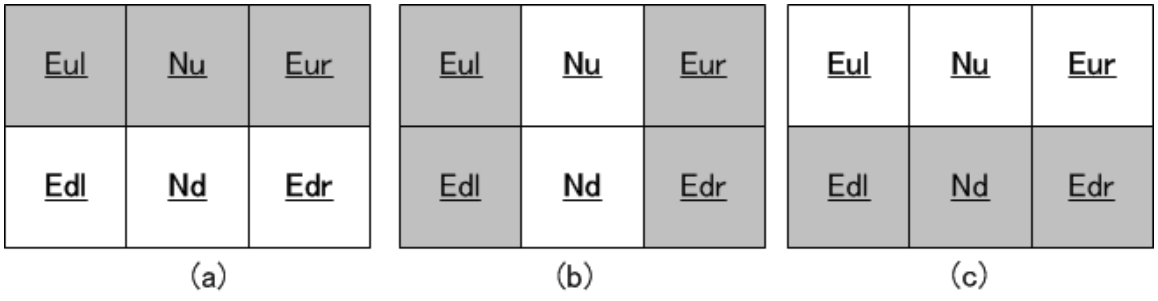


Fig. : SSR filter template patterns

As shown in fig. 8, SSR filter is evaluated in four filter patterns.

Fig. 8(a) depicts upper eye locations. Fig. 8(b) gives four local minimum. Fig. 8(c) is lower eye locations.

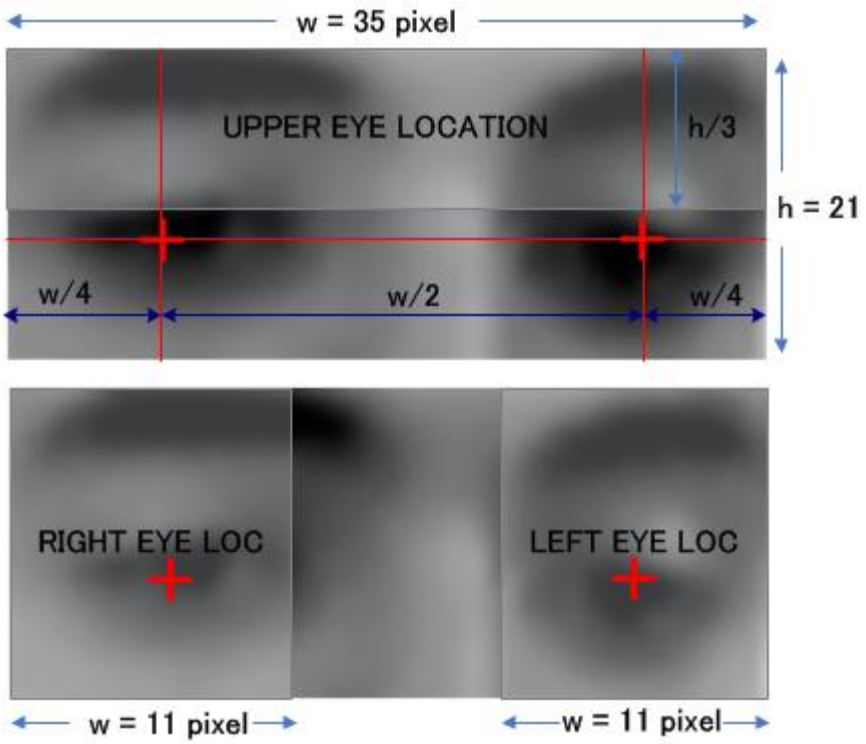


Fig. : SSR filtering for precious eye location determination

To determine the center of the iris, we focus an area between w/4 and h/3. The values of w and h change with the face size but the ratios do not change.

Here, we use 50x50 pixel sizes in fig. 9. First, we decide face region-of-interest (ROI) at a distance of quarter size of total eye area size. The distance of 35pixels between-the-eyes is equal to 100cm distance from the camera if the image size is QVGA (320x240). We scale the eye area size automatically between 20x20 and 180x180 so that it is adaptable to real world conditions.

When we apply SSR filter, we get four local minimums: left eye, right eye, left eyebrow, right eyebrow. Using those four minimum, getting distribution of the four points, we calculate the eye center points. Since we want to find different sizes of faces in real time, we change SSR rectangle sizes with a $1/\sqrt{2}$ ratio. We can find 20x20~180x180 pixel sizes of a face. Calculation of the eye is done as shown in fig. 10.

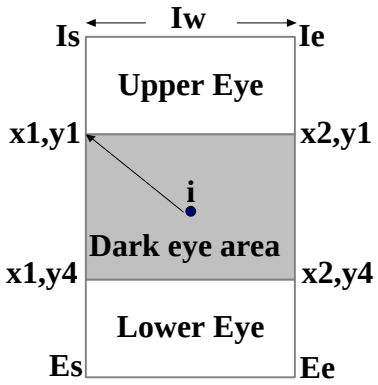


Fig. : Decision of the center of eye-pupil

Based on fig. 10, exact location for eye-pupil is searched on the face image. This process is called eye fine search (EFS). The center of iris is tones of dark color. The upper and the lower side of iris area are examined. When examining the eye location, the following relation must be satisfied.

$$(y1 - Is) = (y4 - y1)/2 \quad ()$$

$$(Es - y4) = y4 - y1) \quad ()$$

The area calculated by eq. 15 is white, while the iris area is dark. The upper side of iris area whose width is (y1 – Is) is half size of iris height and is white. That information is mostly being enough to find the iris location. After iris location is approximately determined, the coordinates of center of it is given in the below equation

$$i(\frac{x4-x1}{2}, \frac{y4-y1}{2}) \quad ()$$

The sum of pixels within rectangle of iris is calculated by

$$Ir(x, y) = i(x4, y4) + i(x1, y1) - i(x1, y4) - i(x2, y1) \quad ()$$

where $i(-1, y) = 0, i(x, -1) = 0$

This is an integral image calculation. We use this logic for a fast computation. Especially during the first stage of image scanning, the SSR finds many candidates. Since more than 90% of them are non-faces, those must be eliminated fast. Knowing eq. 10 and eq. 11 and calculating eq. 12, non-face candidates are easily eliminated.

2.6 Eye training by using SVM for fast detection

Support vector machine (SVM) is used for face training and face decision. SVM is a powerful learning kernel machine and several forms of it is used commonly by many researchers [63],[64]. For example, SVM light is used by Joachims [63].

Here, we use SVM decomposition method. After we find our face candidates by using SSR filtering shown in fig. 7, we do fast histogram equalization in gray scaled images and then we do normalization by scaling and correcting tilt angle.

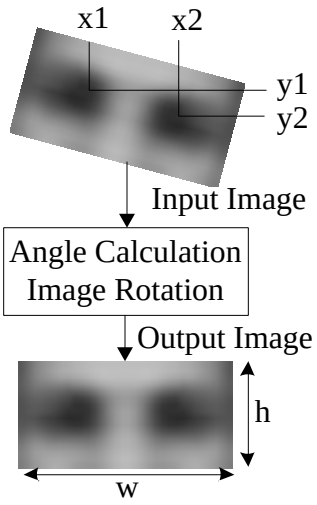


Fig. : Eye image pattern angle correction before SVM

As seen in fig. 11, using x1, x2, y1, y2, the eye angle is computed. Since the only the eyes are rotated rather than the whole image, rotation process is done fast. For SVM training, we use 30x20 pixel size eye area images. They contain two eyes and nose area. The training images are scaled, rotated, translated and finally geometrically normalized before training process.

1000 images for positive and 700 images as negatives are selected for training.

$$W(\propto) = -\sum_{i=1}^{l} \propto_i + \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j \propto_i \propto_j k(x_{i,} x_j) \qquad ()$$

where

$$\sum_{i=1}^{l} y_i \propto_i = 0,\ 0 \leq \propto_i,\ i = 1, \ldots\ldots, N \qquad ()$$

Here, l is the number of the training samples $\propto_i, \propto_j$ is the scalar form of training samples and $k(x_i, x_j)$ is a dot product. $\propto$ is also known as lagrance multiplier.

We put our face candidate to an array and then train the samples and compare the results. If it is less than certain threshold, we assume that it is not a face.

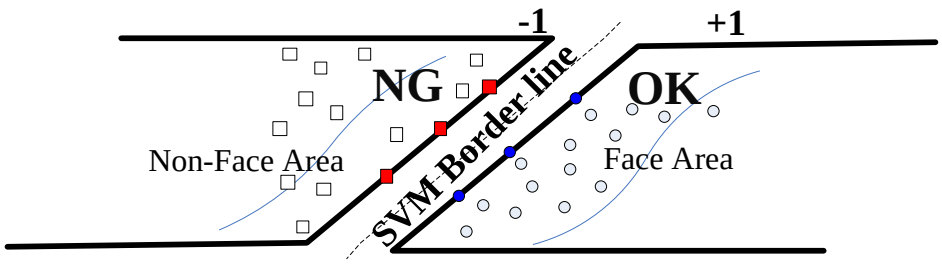


Fig. : SVM face decision border

In fig. 12, rectangles represent the non-face, circles represent the face output.
If the result is +1, it is a face. If the result gives -1, then it means that the candidate is not a face. Hence, it can be dropped. However, to get a better performance, we assign a threshold to filter out the SVM results. If the SVM result is +1 and more than threshold value, we take it as a face. Because some quadratic problems occur for test images to test in real time, we use a decomposition method to implement SVM by separating jobs into the small tasks.

2.7 Evaluation of face detection

Some experimental results are done by using FRGC-DB images and the results are given in fig. 13. FRGC-DB has version 1 and version 2. Version 1.0 has indoor images and version 2.0 includes outdoor and 3D face images. We used a mixture of images from both FRGC-DB ver.1.0 and ver2.0 in our experiments.
During the evaluation of face recognition, we created a benchmark by using FRGC-DB. The results of the benchmark outputs correctly recognized faces, false detected faces and falsely rejected faces. We evaluated not only face detection accuracy, but also we tested the eye detection accuracy. Eye detection depends on the race since the eyebrow thickness is different from race to race. In our approach, we use knowledge-based approach. It is important to include images from different races. FRGC-DB contains western faces and very few African faces and few Asians and latin faces. During evaluation, we also checked outdoor performance of the face detection. We used Sanyo IP camera connected via TCP/IP network for testing faces in outdoor

environments. The images are transferred to the computer in real time in network and the software processes the frames and provides the detected faces and their detection ratios.

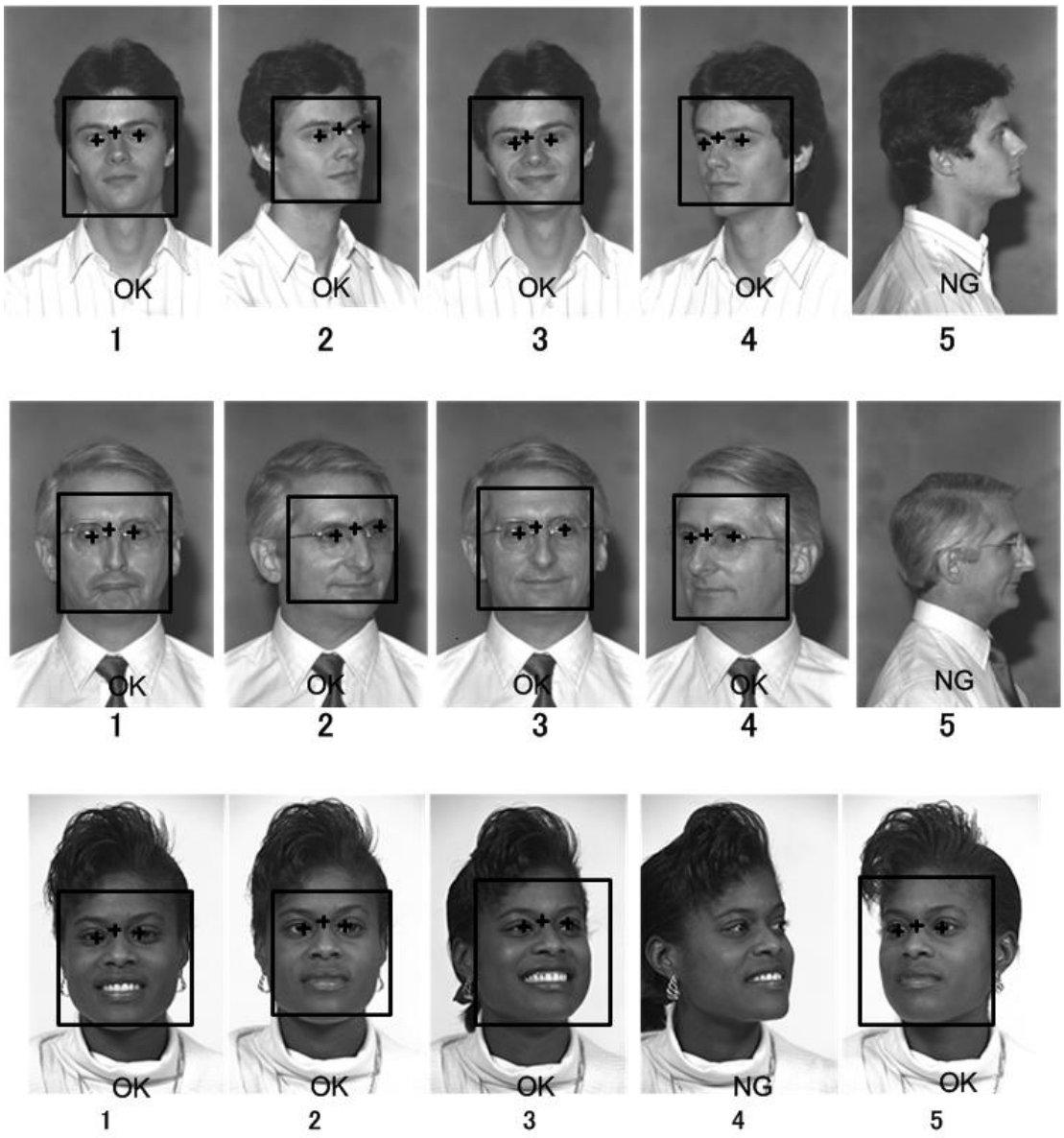


Fig. : Face detection evaluation by using Feret database

Rectangle is drawn to the face area after adaboost and eye crosses are drawn after SVM classification. More results are given in fig. 18.

Fig. 14 shows the detection results in a group, in a street and in a crowd.

Fig. : Face detection by using group images

(a) Detection in a street (b) Detection in a surveillance system

Fig. : Face detection in the street

In fig. 15, detection results taken in uncontrolled environment are seen. Fig. 15(a) gives the face detection results. One face is found by frontal cascade and another face is found by the right-cascade. Face of the middle person between two detected faces could not be detected here. In Fig. 15(b), many-to-many face detection is shown. This frame is obtained from IP camera. Majority of the faces are detected correctly.

We also tested our algorithm by using major databases and obtained the following detection rates. For the different angles and backgrounds, we used different face databases and listed up them in table 2.

Table : DB results (indoor)

DB	# of Images	Detection rate
Feret-DB	5100	96%
CMU-PIE	650	94%
VALIDDB	550	95%
Orl-DB	400	100%
BioID	800	95%
YaleB-DB	1200	93%
UPS-ES	800	98%
Original	20,000	97%

We used color Feret-DB for testing small face changes. Feret-DB has various faces taken under controlled environments. Each face has small variations in angle and in expression. We used fa, fb data set from the Feret-DB images. CMU-PIE has various illumination and various poses. The images are taken in controlled environments. Faces are artificially illuminated by a light source. We selected 650 images from the overall CMU-PIE database. UPS-ES database has faces images with various angles. We selected images within 30degree. We selected 20 images from each individual. Some of occluded faces were also included to the test set. YaleB-DB has various illumination faces. Although face angle is limited to frontal faces, the illumination variations are significant.

The original database has many images that are taken in indoor as well as in outdoor environments. Partially illuminated face images as well as background illumination, faces with significant face expressions were also included to the test set. Although number of individuals is few compared to Feret-DB, there are many variations of face images for each individual (approx. 500 images/person).

Based on the results, the Orl-DB provided the best results. It has frontal faces with small angles and with no occlusion. Original database gave 97% correct detection. Feret-DB gave 96% correct

detection. UPS-ES database gave 98%. Occluded images were also tested during testing and we conformed that the occluded faces were also correctly detected. However, some of angular faces could not be detected. Valid-DB has various illuminated images. Severe illuminated faces failed to detect. Especially significant partial-lighting caused failure.

Table : Detection performance in outdoor

Database	Object type	Proposed method		Original haar filters	
		Detection rate	Error rate	Detection rate	Error rate
Kodak Face-DB	Face	98.2%	1.2%	87.3%	7.9%
FRGC-DB(outdoor)	Face	98.6 %	0.4%	89.1%	5.4%
IP camera(outdoor)	Face	93.3%	4.4%	83.8%	12.1%

Table 3 shows the test results when using outdoor images. Outdoor images are taken by using Sanyo model network IP camera (HD-4600C).

Table 3 also compares the proposed method with the original haar like filters. The detection rate and error rates are significantly different from the original haar like filters.

Table : Detection rate and speed for various environments

Environment	Detection rate	Speed
simple background	96.3%	60ms
complex background	92.1%	
dark environment	94.8%	
bright environment	95.0%	

Table 4 shows the detection rates in different environments. Simple background performs better detection. In complex environments, the detection rate becomes worse due to the object variations. For instance, a tree, corner of house roof, tile, any color text, white board, road, vehicle etc. Dark environments are where faces seem dark such as a place where light is less than 10 lux. In dark environment, the noise is higher than the illuminated environment. Therefore,

noise causes misdetection. Bright environment is where light reflects from the front. In bright environments, the detection rate drops since the features used for detection provide lower scores than the scores of an illuminated face.

The evaluation was done in different databases to make sure of the efficiency of the method. The face detection was tested by using Sanyo megapixel IP camera by running the system one week. During image collection, 120000 face images were collected and used during evaluation. The images collected for evaluation were not included to the training set. Face detection overperformed the original haar like filters. Original haar filters are used in opencv. We used Opencv 2.1 version for the comparison with our proposed method.

Table 5 shows the detection rates, error rates and speeds of other detection algorithms and our proposed method. An implementation of Viola&Jones algorithm is used in Opencv.

Table : Comparison with other detection algorithms

Algorithm	Detection rate	False detections	False rejections	Speed
Proposed method	98.2%	160	22	60ms
Viola&Jones	87.4%	240	152	320ms
Rowley	83.2%	420	203	360ms
Schneiderman	90.5%	80	116	65ms

We used the frontal cascade which comes with the package by default. Rowley implementation was used in matlab.

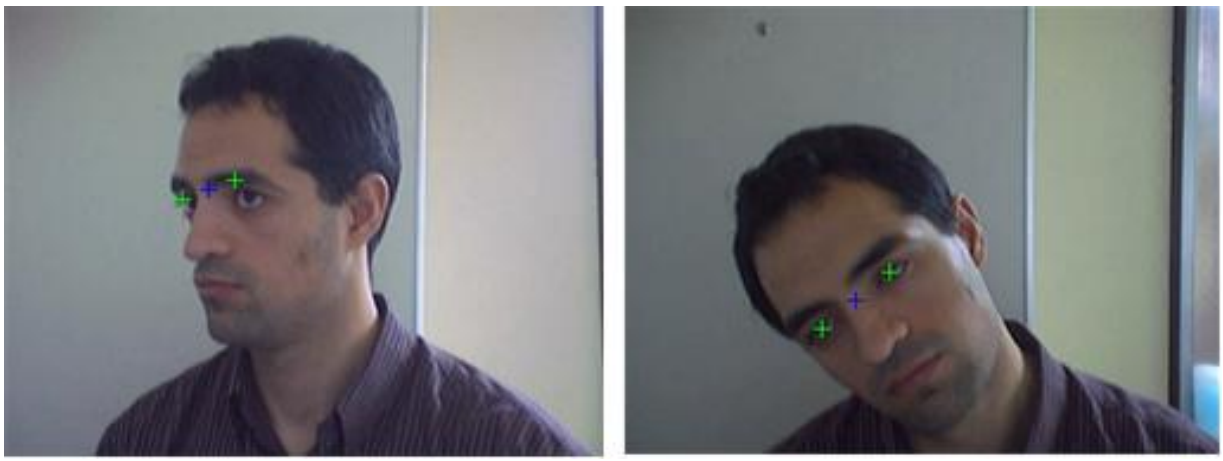
Schneiderman method was commercialized by Pittspat Corporation. We used their implementation for comparison. Both Viola&Jones and Schneiderman methods use the same bootstrap training. The difference is that Viola&Jones uses features that are similar to haar wavelet features and Schneiderman uses wavelet features. Furthermore, Viola&Jones averages the overlapping area and Schneiderman removes such overlapping. Both methods need many negative images during training.

500 images were randomly selected from Kodak-DB images. More than half of the images have more than two faces. Furthermore, most of the image are taken in outside. Hence, the database is

appropriate for surveillance system testing. The images are all high resolution. We applied a downscaling to the images to save the detection time.

Although Viola&Jones method gives better results than Rowley, the method was slow in detection. Schneiderman is better than Viola&Jones in both speed and performance. For the same number of images, Viola&Jones detected 1059 correct faces from 1212 faces in 500 images. There are 240 wrong detections. The method detected the tiles, necklace, part of a face (nose, chin, mouth only, upper eye area) as faces. Schneiderman detected 1096 faces from 1212 faces correctly by yielding 116 non-detected faces and 80 wrong detections. Rowley provided worst results compared with other methods that we used during comparison. Rowley gave 1409 detections. 1009 detections were correct faces and there were 420 wrong detections. Our proposed method detected 1350 locations. 1190 detections of 1300 locations were correct face locations. The speed of our proposed method overperformed Viola&Jones, Rowley and Schneiderman.

In addition to the above testing, we conducted a real-time test to see eye detection performance and the results are given in fig. 16 and fig. 17.



(a) Detection result in left pose (b) Detection in tilt-angle.

Fig. : Face extraction based on SSR with some angles

Above figure shows two implementation results. Fig. 16(a) shows the eye detection from the left pose angle and fig. 16(b) shows the detection of eyes in tilt-angle. As seen in fig. 16(b), the detection of eyes is very accurate. However, when the face is nonfrontal, the error happens as in fig. 16(a). We further investigated for the various face poses and their results are given in fig 17.

Fig. 17(a) is the detection result when the face is looking at the left side with 40 degree. Fig.17(b) shows that face looks at on the left down side.

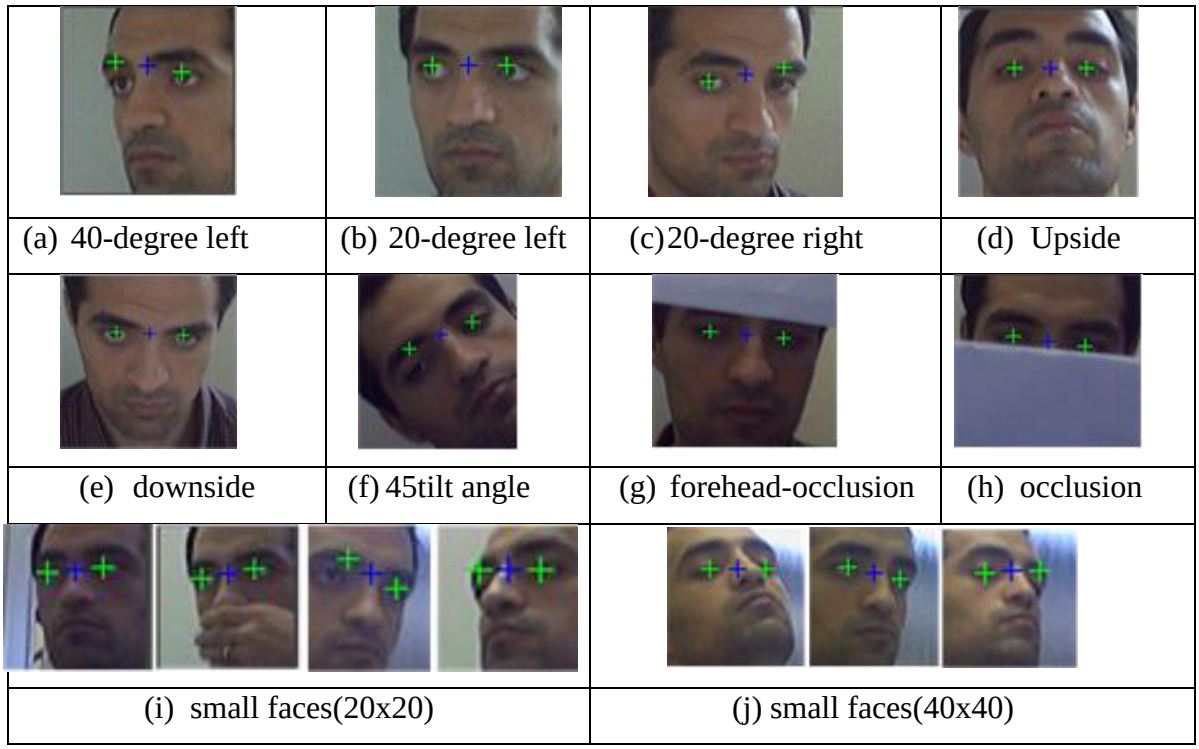
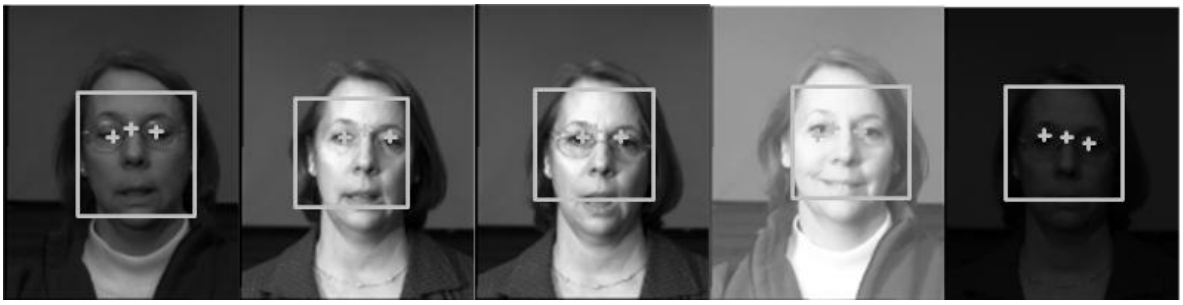


Fig. : Miscellaneous examples for eye detection

We tested our algorithm by using some angles and distances. Distance is directly related to the camera image pixel size.

Fig. 17(a) is the eye detection from a face pose of 40 degree. Fig. 17(b) is the detection of eyes from 20 degree pose. If face angle increases, the eye detection accuracy decreases. Fig. 17(d) and (e) show the eye detection from frontal view. fig. 17(i) shows the faces with 20x20 pixel sizes and fig. 17(j) shows 40x40 pixel faces. In both of them, the eye detection accuracy is the same. In fig. 17(h), we see almost all part of face is hidden. As far as both eyes are visible, our algorithm detects eyes correctly.

More results are given in fig. 18.



(a) Infrared images from IRFace-DB

(b) Images from ARFace-DB

(c) Images from Indian-DB

Fig. : Results of face detection and eye-search

Fig. 18 shows the detected faces and detected eyes in several databases. Fig. 18(a) is the test results on IR images.

2.8 Summary

In this chapter, we gave the principals of the proposed face detection and eye detection. Face detection uses the enriched haar-like filters which contain less number of features than what Viola&Jones proposed in their paper. The enriched features provide increased accuracy, fast detection and fast training. The training by using new features becomes fast since we do not use

thousands of features. The features are computed by integral image and image pyramid architecture is used to speed up further.

In addition to face detection, we introduced eye detection method which uses similar features to haar-like filters. We segment the eye area into 6 segment which includes the below part of eyebrow and nose tip. The six segmented filter is separated into 3 columns and 2 rows. The final classification is done by using support vector machine.

First of all, face is found and eyes are searched inside the face area. The searching is done simultaneously for each detected face.

Finally, the evaluations were done by using default databases such as FRGC-DB as well as network IP camera. The experimental results show that the face detection and eye detection perform well in outdoor environments. The error rates are significantly lower than the original haar-like filters. An improvement more than 65% in equal error rates has been obtained.

CHAPTER 3: ILLUMINATION NORMALIZATION BY USING PHOTOMETRIC & FREQUENCY TRANSFORM

In this chapter, an illumination normalization technique is given in detail. The chapter explains the details of the proposed illumination technique. After giving some past studies, the proposed method is explained in detail. The illumination normalization technique given here is used as a preprocessing step to the face recognition. A comparison with the existing illumination technique is also included.

3.1 Introduction

We present a novel illumination normalization method, which apparently improves the face recognition accuracy in outdoor environments. Different types of illumination normalization techniques such as illumination cone, linear subspace and spherical harmonics cannot properly normalize the illumination in outdoor environment because most of them utilize a linear filter structure. Thus, these approaches are not appropriate to cope with light homogeneity, nonlinearity of sunlight, harmonic effects and cast-shadow effects.

The proposed approach, which is named as "Ayofa-Filter", considers the frequency variability and Albedo direction of local face regions. Ayofa-filter is a new approach that consists of gabor decomposition and albedo estimation on a face normal where the direction of the light source is unknown.

This approach effectively finds illumination directions and recovers the illumination. Whereas,

most of the conventional illumination normalization methods need constant albedo coefficients as well as known illumination source direction to recover the illumination.

Our novel approach computes unknown reflection directions by using spatial frequency components on salient regions of a face. Our method requires only one single image taken under any arbitrary illumination condition where we do not know the light source direction, strength, or the number of light sources. It relies on the spatial frequencies and does not need to use any precompiled face models. Ayofa-filter references the nose tip to evaluate the reflection model.

3.1.1 Past studies

In recent years, face recognition in outdoor environments has a great interest among researches because of the increasing demand on face recognition for public security. So far, many papers have been published and some of the effective algorithms have been commercialized. Although most of them give satisfactory results in indoor environments, they suffer from high misrecognition rates in outdoor environments [34]. Face recognition techniques are not steady in outdoor environments due to the complexity of the real world conditions. Some surveys were conducted to address these problems caused by nonlinearly changing outdoor environments [35],[36]. Some of the problems presented in these reports are illumination, shadow, face pose, face expression, face occlusion, aging, natural changes on the face, race difference, human behavior, distortions during enrollment or recognition and hardware factors. Among them, specifically illumination problem is one of the most important issues.

In the past decades, a significant effort has been devoted to address the illumination issue. Zou and Kittler [37] reported different illumination normalization techniques, which are photometric techniques, subspace techniques and frequency transformation techniques. They also investigated an albedo technique published by Belhumeur et al. [38]. Belhumeur puts a prior condition that the number of reflections should be known initially. This is not possible to determine the number of reflections in real world conditions since the intensity of the light is different in outdoor environments. Furthermore, the direction of light frequently changes by sunlight, by weather conditions, by human factors and by the condition of the environment.

Biswas et al. [39] introduced an estimation method of albedo as an alternative to

Belhumeur's technique. His approach is capable to estimate the illumination direction. However, cast shadows cause the failure of albedo computation. In addition, his method strongly depends on the accuracy of face landmark detection. Thus, a small misdetection of even one landmark point significantly increases the illumination normalization error. Zhang et al. [40] proposed a face recognition, which reports high performance under arbitrary unknown illumination conditions by using spherical harmonics representation. They estimate the spherical harmonic basis images from one frontal image taken under an arbitrary illumination condition. In their method, a face illumination database set by using frontal images was generated and the statistics of spherical images in 2D space were computed. However, the proposed method also relies on the landmark accuracy. Furthermore, their illumination estimation depends on the illumination database. Normally, a database, which covers all the illumination conditions, is not feasible. Lee et al. [41] approached the illumination problem by using linear superposition of images without considering landmark points. The advantage of this method is that neither does it require 3D model of faces as in the spherical harmonics approach [40] nor does it need any training. Nonetheless, if the images are taken under extreme illumination conditions, the approximation error becomes high. Moreover, the problems turn into even more difficult to handle if there are cast shadows and partial occlusions. Chen et al. [42] proposed logarithmic total variation (LTV) to overcome extreme illumination conditions. LTV is used to remove larger scale lighting fields and it keeps the small-scale facial features. However, LTV causes high false acceptances in face recognition by causing deformations under severe illumination conditions.

There are more works such as illumination cone methods [43], [44], spherical harmonic based representations [45],[46],[47]. Georghiades et al. [43] showed that the illumination cones of human faces can be approximated well by low dimensional linear subspaces. Therefore, the set of face images in fixed pose but under different illumination conditions can be efficiently represented using an illumination cone. However, illumination cone approaches are expensive methods in computation and they need more than one image to normalize the illumination. Spherical harmonic based representations are faster than Illumination cone methods but they are not effective techniques under partial lights that frequently happen in outdoor environments. Basri and Jacobs [88] proposed that the set of images of a convex Lambertian object obtained under a variety of lighting conditions could be approximated by using linear subspace that is

formed by harmonic images. However, this method requires knowledge of the object's surface normals and albedos before the harmonic subspace can be computed. To overcome this, Feng et al. [45] proposed an algorithm for estimating the illumination parameters. It includes the direction and strength of a point light source and strength of the ambient illumination for an illumination model consisting of one point light source and ambient illumination. He projects the images into an analytical subspace and estimates the illumination from these projected coefficients using a nonlinear least-squares method. The estimated error of the light direction increases when the ambient lights are active.

Okabe et al. [46] proposed a different approach to recover the illumination under ambient lights and cast shadow. He compared the results of spherical harmonics and haar wavelets to calculate the inverse lighting. However, he did not consider the cast-shadow effects. To recover it, Adini et al. [47] used an illumination database of faces, in which each of the imaging conditions was controlled. The distances between the pairs of images of different people were computed and compared with the distances of images of the same face in a different viewing condition. He extracted the edge map of an image by filtering with 2D Gabor features. All of the above representations were also followed by a log function to generate additional representations. However, the recognition experiment on a face database with lighting variation indicated that none of his representations is sufficient by itself to overcome image variation due to the change of the illumination direction.

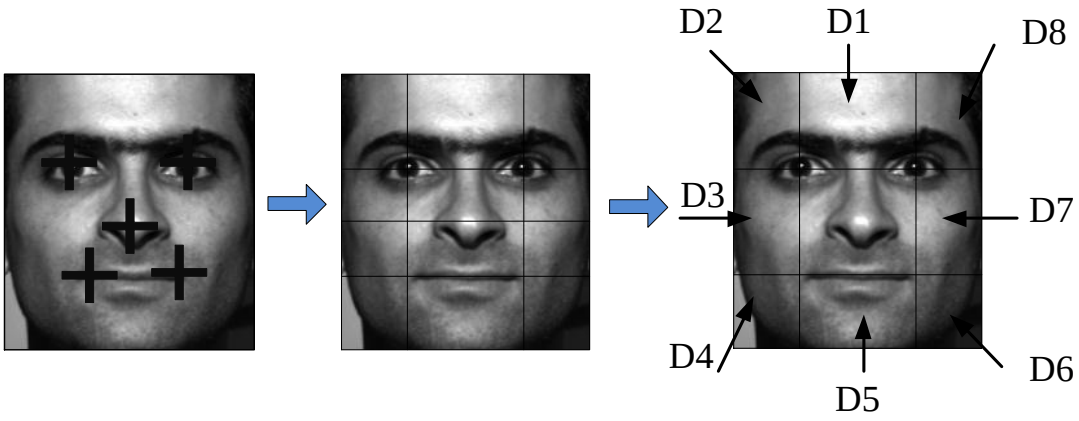
Hadjidemetriou [48] proposed a multiple histogram approach, which does neither require a constant light source nor needs an illumination database. He calculated the image histogram for different resolutions of a face image by using gaussian filters. He used the features of gaussian filters and combined image intensity information with spatial information. He used the features directly to acquire the intensity information. However, intensity information does not always provide accurate results in outdoor environment since the illumination causes different reflections from the part to part. If the light is equally distributed to the face surface and if the whole face image is dark or bright, the method of Hadjidemetriou works well. This is simply not realistic in outside since the light direction cannot be predicted. Xie et al. [49] proposed an effective illumination correction methodology. He assumes that face is a Lambertian surface and

he measures the albedo on each image pixel (x,y). He corrects the illumination by estimating albedo. Albedo estimation is done by computing the light intensity. Although his method is based on simple arithmetic calculations, the method overperforms the algorithm of Hadjidemetriou [48]. However, his method needs fix light source to estimate albedo. In our method, Ayofa-filter does not need any light direction. We extract the features from a face, calculate the albedo by estimating the light direction, and finally reconstruct the face model.

Matsushita et al. [50] focused on removing the cast shadow effect by using the illumination eigenspaces. Eigenspaces contain a previously constructed database set and image sequence information, which have similar structure with Adini et al. [47]. In his method, scene must be constant and camera must be set to a fixed location. Blanz et al. [51] solved these limitations by using a morphable 3D model. The 3D morphable model describes the shape and texture of face separately based on the eigenfaces of the shape and texture obtained from a database of 3D scans. His algorithm works well across pose and illumination but the computational expense is quite high. Gross [52] proposed an anisotropic diffusion based normalization technique. By using a single brightness image, his technique first estimates the illumination field and then compensates the illumination in order to recover the scene reflectance. He applies a smoothness constraint to the image. He claims that his proposed technique does not require any training steps, knowledge of 3D face models or reflective surface models. Although his method removes the significant illumination changes, significant amount of shadow remains. In addition to this, the method does not deal with cast-shadow. Therefore, it is not efficient to use in outdoor illumination. There are also other methodologies which are based on wavelet transformation [95], homomorphic filtering [94] and non-local means based approaches [96]. The wavelet technique applies the discrete wavelet transform to an image and then processes the obtained sub-bands. It emphasizes the matrices of detailed coefficient and applies histogram equalization to the approximate coefficients of the transform. After the manipulation of the individual sub-band the normalized image is reconstructed using the inverse wavelet transform. In homomorphic normalization approach, the high frequency components are emphasized and the low-frequency components are reduced. As a final step the image is transformed back into the spatial domain by applying the inverse Fourier transform and taking the exponential of the result. Non-local mean normalization technique computes the luminance function to estimate the reflectance. However, none of them is adequate to fully normalize the illumination.

In our illumination technique, we developed a novel feature based illumination normalization approach by using Ayofa-filters. Estimation of Albedo is frequently used with shape from image shading (SFS) [89],[90]. We propose a different approach to find the amount of reflection to remove the illumination effects on a face surface. We segment the face area into nine parts before the feature extraction. The segmentation is important since the illumination in small image regions is more homogeneous. If we divide the face image into smaller regions and use a different set of filter parameters for each region, it is expected that the overall estimation error to be smaller than processing the whole face. However, there are other problems, which arise in this case. One of them is the interpolation problem and another problem is the pixel saturation problem. If the majority of the pixels in an arbitrary region has saturated due to the heavy illumination such as direct sunlight or strong light from background, the features obtained from the region are distorted heavily. To avoid such problems, we first process the face by using nonlinear intensity filtering and then extract the face features by using Gabor filters. Based on the features, we estimate the reflection direction by measuring the amplitude of Ayofa vectors. We interpolate the face after correcting the illumination. The efficiency of the method mostly depends on the Ayofa-filter parameter selection. Serrano et al. [91] introduced a similar method to Ayofa-filter by using Gabor jets. However, he used Gabor features only for face recognition. We use Gabor decomposition technique to compute Ayofa-filter.

3.2 Face normalization by five novel points



(a) Landmark points (b) Grid window (c) Light directions

Fig. : Light source direction

A face is automatically marked by five points. These five points contain two eyes, nose tip and the corners of the mouth as seen in fig. 19(a). A pre-trained cascade detects the center of two eyes, nose tip and mouth corners. There is no human action to do any manual input. In fig. 19(b), horizontal lines show the landmark lines and vertical lines show the eye centers. Based on the points, the face is segmented into nine parts inside the inner-face area in fig. 19(c).

There are some recent research studies, which use some different methods by dividing the face area into 8 parts or 16 parts or 64 parts to estimate the albedo. The more face area is divided into parts, the more processing power is needed. Our purpose is to introduce a method, which runs fast and still provides high accuracy when using megapixel images. The nine-part division is decided empirically. More divisions increase the normalization performance slightly but the speed becomes slow. For example, performance is 2% better with 20% speed degradation if 16-division is used and performance is 5% better with 40% speed degradation if the number of division is 64. More detailed indication is given in table 6.

Table : Face partitioning and trade-offs

Resolution	Partition	Speed	Accuracy
1280x960	1x1	27ms	91.86%
	2x2	28ms	93.54%
	3x3	30ms	94.21%
	4x4	36ms	95.88%
	5x5	38ms	96.82%
	8x8	42ms	98.70%
	9x9	46ms	99.64%
	12x12	60ms	96.35%

In table 6, different face partitions are investigated and their speed-accuracy relations are given. ARFace-DB images are used during the accuracy measurements.

We found that nine-part division (3x3) is optimal in terms of speed and accuracy. The only disadvantage of nine-part division is that we cannot perfectly eliminate the background effects during the illumination normalization. If a strong light comes from the back, the face boundary is

distorted. The distortion causes misrecognition. We prevent such misrecognitions by cutting the face area smaller. However, the smaller face cut leads to performance degradation. Partitioning the face smaller than the nine parts degrades the speed but increases the accuracy. Partitioning the face into 81 equal parts gives the best accuracy with 65% speed degradation. Further partitioning decreases the performance. Since the proposed method is for outdoor use, the speed is important.

We decided that nine-part division (3x3) is optimal in terms of speed and accuracy trade-off. The only disadvantage of nine-part division is that we cannot perfectly eliminate the background effects during the illumination normalization. If a strong light comes from the back, the face boundary is distorted. The distortion sometimes causes misrecognition. We prevent such misrecognitions by cutting the face area smaller. However, the smaller face cut leads to performance degradation. Partitioning the face smaller than the nine parts degrades the speed but increases the accuracy. Partitioning the face into 81 equal parts gives the best accuracy with 65% speed degradation. Further partitioning decreases the performance. Since the proposed method is for outdoor use, the speed is important. We selected 3x3 partitioning as we considered the best speed and accuracy trade-off. Rather than using a fixed partitioning, a new scheme in which face image is divided into 3x3, 5x5 and 9x9 blocks and "vote" on the best overall interpretation can be employed. Although this increases the accuracy, the processing time increases. Downscaling the face image by keeping 3x3 partitioning is another scheme to try. Downscaling the face image does not affect the speed. However, it destroys some of the important face features and it leads higher equal error rate.

The estimation of reflection requires that face be not affected by a spotlight from the middle of the face. Another assumption is that the light comes from any of the eight directions as shown in Fig.19(c). In addition to these D1~D8, there are also harmonic light effects. Harmonic lights are the second and third reflections from face normal. The harmonic illumination effects are taken into account in this thesis.

Before the Gabor decomposition, five points from a face are automatically detected. These points are used to normalize the face to 100x100 pixels. Five-point normalization is important for accurate normalization since face shape should not be distorted when face is normalized. Furthermore, nose point gives us information about the pose orientation of the face. Pose orientation is used for the decision of the Ayofa-filter parameters. Based on the five points, face

is divided into the nine parts. Ayofa-filters are applied to each segmentation area in clockwise direction and are implemented in three parts. The first part does the contrast filtering to remove the DC components of the illumination. The second part of the filter extracts the face features. The last part does the normalization of the illumination and finally, face is reconstructed by a linear interpolation technique.

Let $I(x, y)$ be the normalized face image. Let i_w be the face image width, i_h be the face image height, r_w be the rectangle width of each segmentation area, r_h be the rectangle height of each segmentation area. Let Er_x be right eye x coordinate and El_x be the left eye x coordinate, ml_y be the left mouth y coordinate.

$$\begin{aligned} x_1 &= El_x\ ,\ y_1 = El_y \\ x_2 &= Er_x - El_x\ , y_2 = ml_y - El_y \\ x_3 &= i_w - (Er_x + El_x),\ \ y_3 = i_h - ml_y \end{aligned} \qquad ()$$

where x_1 is the width of the rectangle area from the top left of the face, x_2 is the distance between eyes which is also width of the middle rectangle and x_3 is the width of the rectangle area from the top right of the face. y_1 is the y coordinate of x_1, y_2 is the height of x_2 and y_3 is the height of x_3.

3.3 Adaptive histogram fitting (AHF)

Histogram equalization can be computed easily and efficiently. They are robust to noise and local image transformations [53]. For many applications, however, the histogram is not adequate, since it does not capture spatial image information. The proposed adaptive histogram fitting method not only combines intensity with spatial information, but it also preserves the efficiency, simplicity, and robustness of the plain histogram. In AHF, a mean face is used. A mean face is a single face image as given in fig. 20.

Fig. : Mean face samples used in AHF

As seen in fig. 20, any frontal and well-illuminated without any shadow, no-occlusion and neutral expression can be used as mean face.

A similar technique to AHF was used by Phillips [54]. However, he used a random sample face, which made their technique dependent to the certain environments. Another artifact of the histogram proposed by Phillips [54] is the discontinuities in sudden changes of illumination such as dark right side and white left side. Some techniques solved this by averaging the left and right side of the face. However, averaging introduces noises and some deformations if the face is not frontal. Noise and such deformations become significant with proportional to the face angle. As a solution to prevent deformations, Hadjidemetriou [48] introduced multi-resolution histogram technique. He used histogram differences between various locations of the image data. He used gaussian filters to compute the histogram. Although the method is promising, it is restricted to indoor illumination. He did not consider cast-shadow and partial illumination effects.

As an alternative solution to deal with partial illumination, we first apply AHF as a preprocessing step. There are some state-of-the-art techniques, which use a mean face database to normalize the illumination [56], [57]. Shan et al. [56] used gamma intensity correction and histogram equalization to recover the side-lighting effects. Furthermore, they re-synthesized the face for relighting. Although the method works well for ambient light, it fails to recover the partial lights. To overcome partial lighting problem, we divide the face area into 9 parts (three from left and three from right) and each part is processed separately by AHF. After processing all the pre-segmented parts, a linear interpolation is used to reconstruct the image.

Let Hi(i) be the histogram function of an image, and let Ho(i) be the desired histogram to be mapped via a transfer function $f_H \rightarrow Ho(i)$. Let $f_{Hi} \rightarrow T(i)$ and $f_{Ho} \rightarrow T(i)$ be the transfer functions for Hi(i) and Ho(i) respectively. $f_{Hi} \rightarrow T(i)$ and $f_{Ho} \rightarrow T(i)$ are given as below.

$$f_{Hi} \rightarrow T(i) = \frac{\sum_{j=0}^{i} Hi(i)}{\sum_{j=0}^{n-1} Hi(i)} \quad ()$$

$$f_{Ho} \rightarrow T(i) = \frac{\sum_{j=0}^{i} Ho(i)}{\sum_{j=0}^{n-1} Ho(i)} \quad ()$$

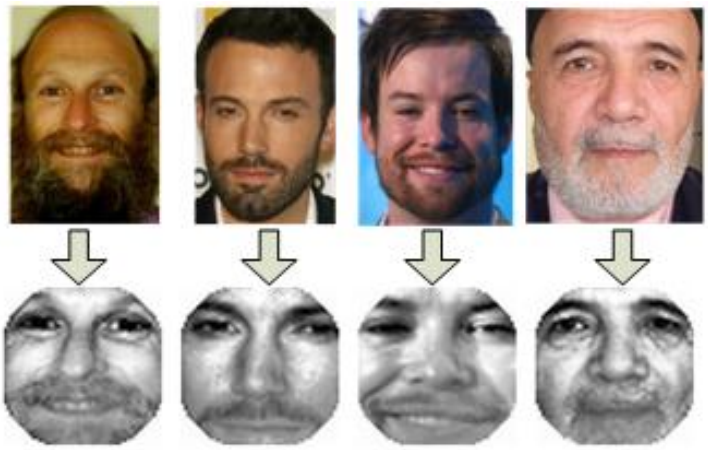
To calculate the $f_{Hi} \rightarrow Ho(i)$, it is needed to invert $f_{Ho} \rightarrow T(i)$ to get $f_T \rightarrow Ho(i)$.
$f_T \rightarrow Ho(i)$ transforms the histogram equalization of each pixel found in histogram Ho(i). This gives the definition of the $f_{Hi} \rightarrow T(i)$ as below:

$$f_{Hi} \rightarrow Ho(i) = f_T \rightarrow Ho[f_{Ho} \rightarrow T(i)] \quad ()$$

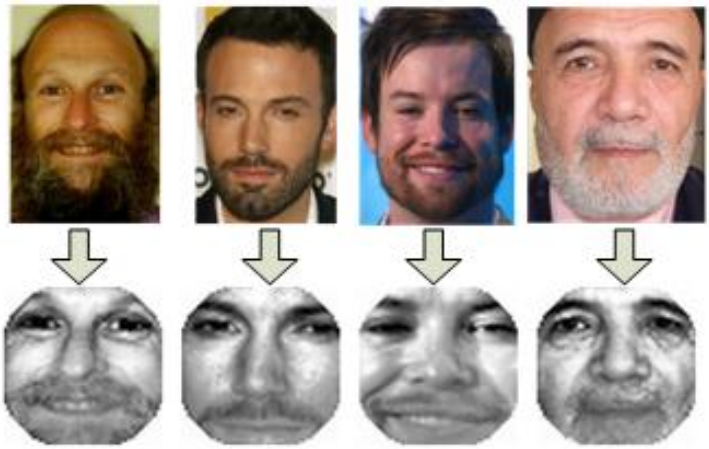
To correct the illumination accurately, each part of the face divided is processed independently. Then, each part contains a transfer function $f_{Hi_n} \rightarrow Ho_n(i)$.

$$f_{Htotal} \rightarrow Ho_n(i) = \sum_{i=1}^{N} (f_{T_i} \rightarrow Hi_i) \times \frac{I(x,y)}{2^i} \quad ()$$

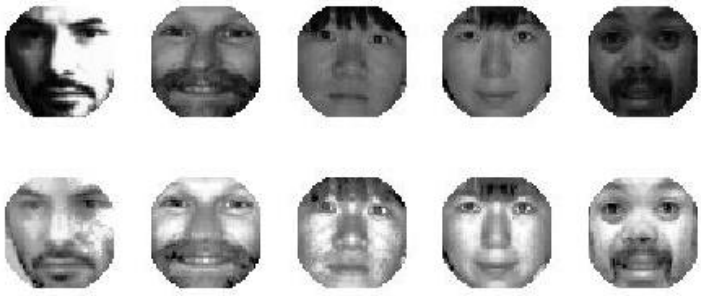
where I(x,y) is the segmented each area. $(f_{T_i} \rightarrow Hi_i)$ is the transfer function of the image. N is set to 9, which is the maximum number of image segments.
The resulting images are shown in fig. 21. The images are normalized by using AHF method. The method can be applied to angular faces.



(a) The method increases the contrast of dark areas and reduces the light level on bright areas. AHF processes the faces by diving it to 9 parts.



(b) Beards are also illuminated. Beards areas are whitened.

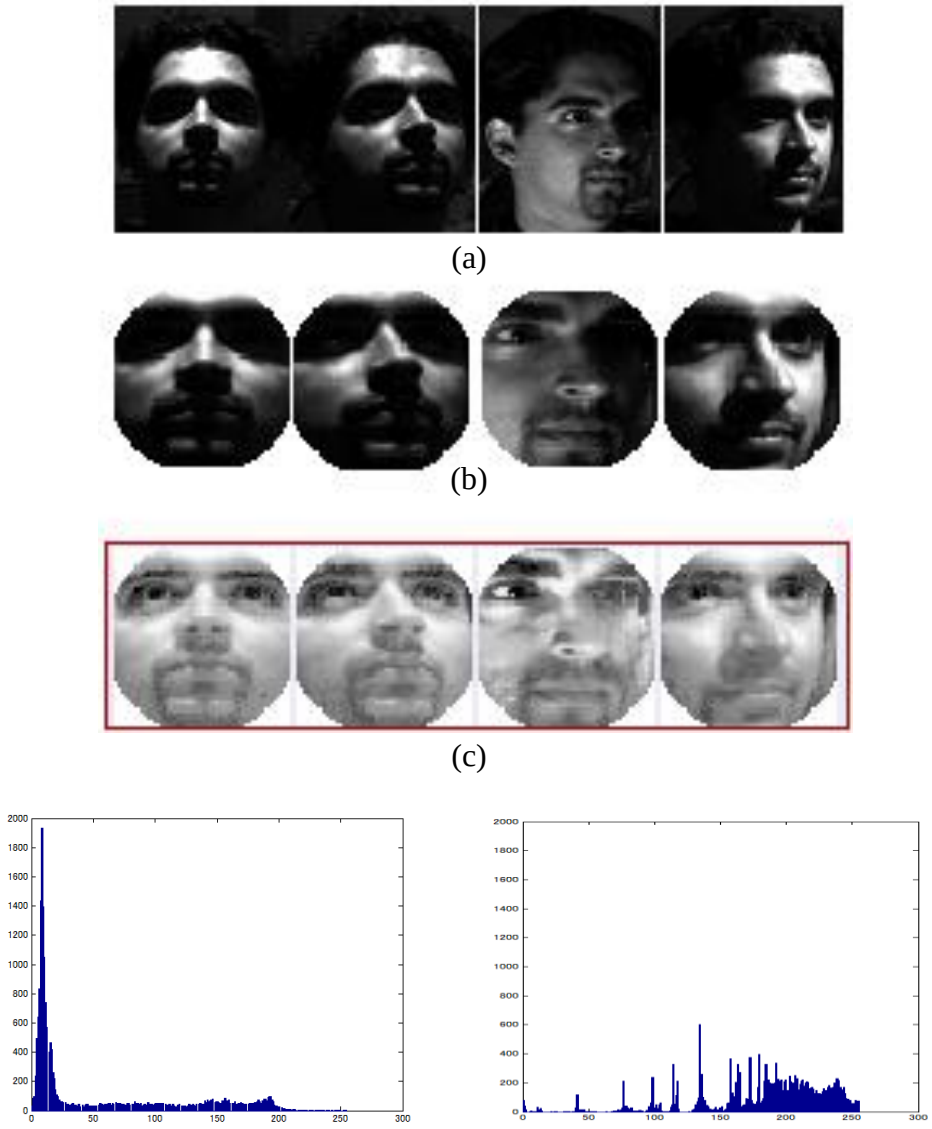


.

(c) Moustache and beards are also whitened as other parts of the face. The whitening decreases the effect of moustache and beards that significantly improves FAR rates.

Fig. : Adaptive histogram fitting

In Fig. 22, (a) shows the illumination images. (b) shows the beard images and (c) is various images.



(a)

(b)

(c)

(d)

Fig. : Test results and image histogram after AHF

Fig. 22(a) shows the original images from the YaleE-DB. Light angle of 45-90 degrees and frontal face to 30-degree face pose images are chosen during the testing. Fig. 22(b) gives the face crop area. The original face images are cropped by 64x64 sizes. Fig. 22(c) shows the illuminated face images. The images are equally illuminated. There are distortions for completely dark side

of the faces. Such deformations do not significantly affect H-NPCA performance since those components are insignificant during the recognition. Fig. 22(d) is the histogram chart of one image selected from YaleE-DB. Left histogram is before AHF, Right histogram is after AHF. After the AHF, the histogram is equally distributed.

The third image from the left side in fig.22 is heavily distorted after the AHF. It is because the texture information of the image right side is lost due to the darkness. However, the distortion does not affect the face recognition rates since the distortion is around 10% of the overall face area and the 90% variances of the face features can be still used in H-NPCA.

3.4 Ayofa-filter design

An overall illumination normalization flow is given in fig.23 and the area inside dot rectangle shows the Ayofa-filter structure.

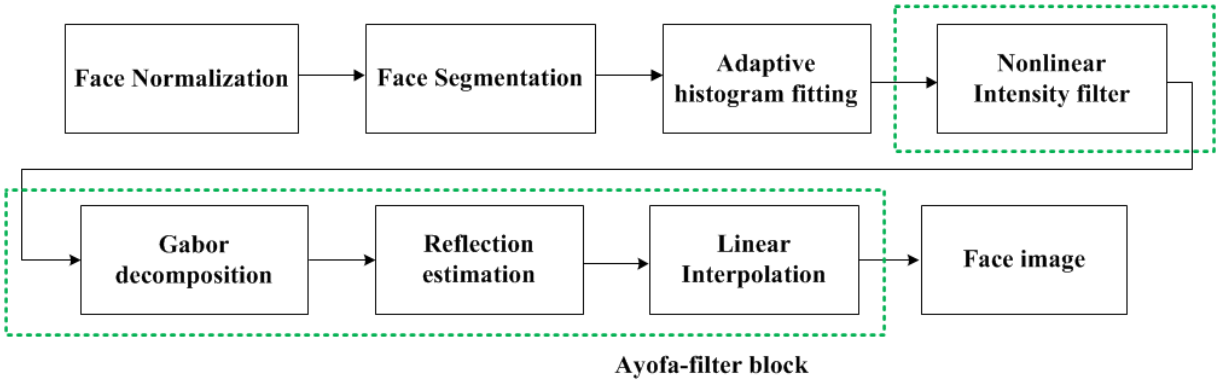


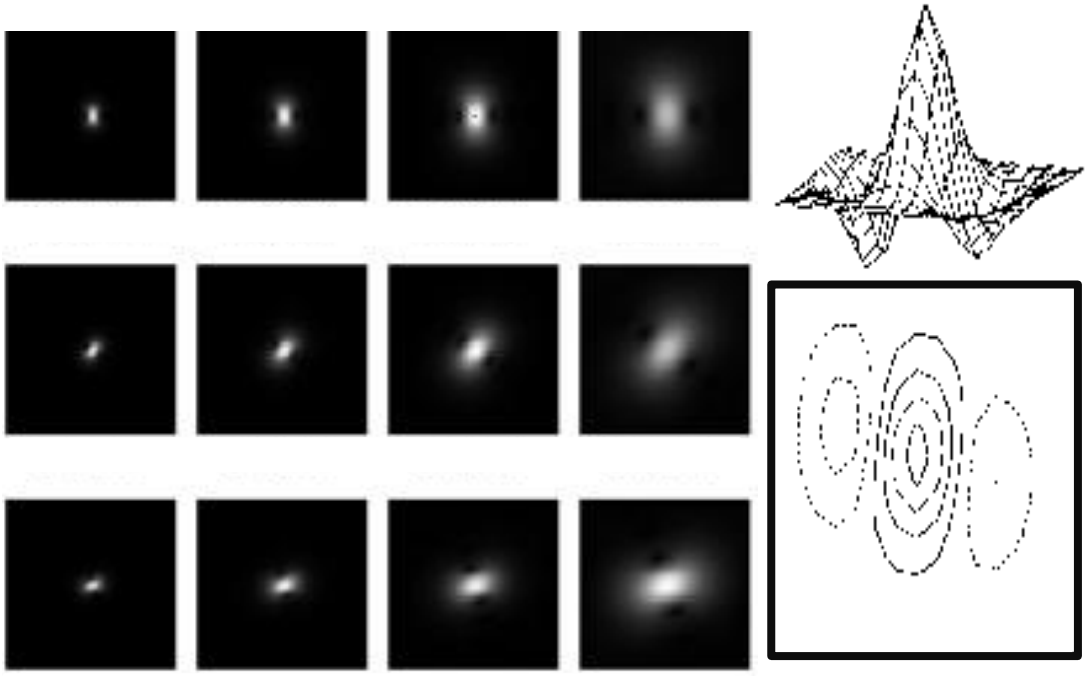
Fig. : Ayofa filter schematic diagram

In fig. 23, Face normalization process contains face gray-scaling, resizing, cutting and tilt-angle correction.

Each segment of the face is first processed by using AHF and then processed by nonlinear intensity filter (NIF) as seen in fig. 23. AHF shifts the histogram within the dynamic range and NIF removes the high frequency components of the raw image data. NIF is the part of the Ayofa filtering and it is computed by the following equation.

$$Ni(x,y) = \sum_{i=0,j=0}^{2\pi} \frac{\varphi_j}{\vartheta_j} \times I(x,y) \qquad ()$$

where I(x,y) is the raw input image data, φ_j is a bandpass filter and ϑ_j is the low pass filter. j shows the sampling frequency which changes between $0 \sim 2\pi$. Maximum frequency is the frequency value which both bandpass and low pass filters are limited by maximum frequency. The resulting data is processed by Gabor decomposition.

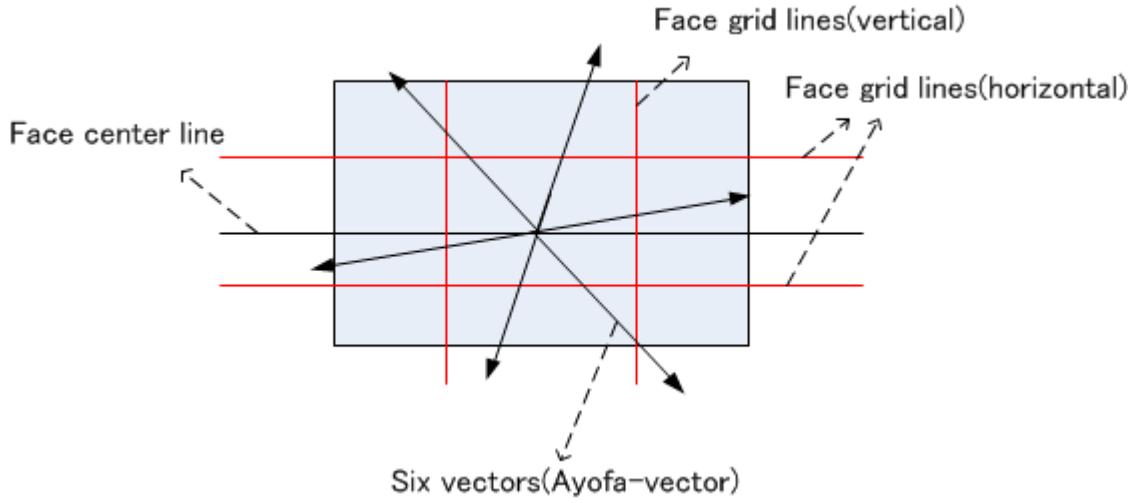


(a) Gabor filters (4scales x 3orientations) (b) Gabor waveform

Fig. : Gabor decomposition and its waveforms

Fig. 24(a) shows some Gabor filters that are used for Gabor decomposition. Fig. 24(b) is the Gabor waveform. The middle part is weighted. Filters extract the global texture features from the target face object. The selection of the scale and orientation is done based on Ayofa-vector logic. The spatial features are used to compute the albedo. It is important what parameters should be used and where the filters must be applied on the face area. To find answer to these, the light directions are computed for each direction (D1~D8). Fig. 25 depicts the relation of source and reflection. As an example, we assume that light comes from the direction of D7. Light is reflected to the six different directions by scattering on the face surface. We name these six different

directional vectors as Ayofa-vector. Magnitude of each scattered reflection vector is computed by integral image. The computation is repeated for all directions and the estimation is done for each direction separately.



(a) Ayofa-filter vectors

(b) Light effects from different directions

(c) Ayofa vectors and light source

Fig. : Illumination source

Fig. 25(a) depicts the six illumination vector which is called here as "Ayofa-vector, fig. 25(b) shows the various images that are affected by the sunlight and fig. 25 (c) shows the Ayofa vectors on actual images. Once the light reaches to the face surface, we assume that light is reflected into six different directions.In each direction, we compute the Gabor decomposition along with the vector length. Vector length is computed by (27). The nose tip is taken as a reference point of the Ayofa vectors. This is because of three facts: nose is the highest point on a face, nose is the middle of the face and the nose is always visible.

Let i_w be the face width, i_h be the face height. The length of each Ayofa-vector is computed by

$$|Al| = \sqrt{(\frac{i_w}{2} cos\theta)^2 + (\frac{i_h}{2} sin\theta)^2} \quad ()$$

In above equation, θ is 60 degree, which is the angle of each Ayofa-vector.

Xiao et al. [44] made the normalization of the illumination with the assumption that light source is known. Their method does not work in uncontrolled environment. The illumination direction is not predictable in uncontrolled environments since the sunlight direction changes with time, weather conditions and other factors such as occlusions by a vehicle and so on. Ayofa-filter does not depend on sunlight direction. The theory of our filter is discussed in detail below.

The energy spectral density of the albedo is computed by the following equation.

$$\rho(\omega) = \varepsilon + \left| \frac{A_0}{\sqrt{2\pi}} \sum_{n=1}^{M} Al_n \times \rho(n) e^{-i\omega n} \right|^2 \quad ()$$

where ω is the angular frequency and ε is error factor. ρ(ω) is the energy spectral density of the image lambertian surface ρ of Ni(x,y) image. A_0 is the coefficient value. M is the maximum number of reflections. This is typically set to six in this paper. This number should be set to higher values in practical use because there are more reflections in real world conditions. We consider six reflection directions for each light source. The angle between each reflection direction is calculated by

$$\theta_{av} = 2\pi/N \quad ()$$

where θ_{av} is the angle between Ayofa-vectors, N is the number of reflections.
If N is increased, the computation time decreases while the error factor of the albedo estimation ε increases. To compute the albedo, we do the following steps:

Step 1:

After face is segmented into parts, Ayofa-filters are applied onto each part of the face. Ayofa-filters are expressed by a complex exponential function and its expression is given below.

$$\varphi_{\mu,\gamma}(\vec{z}) = \frac{\|k_{\mu,\gamma}\|^2}{\sigma^2} e^{(-\frac{\|k_{\mu,\gamma}\|^2 \|\vec{z}\|^2}{2\sigma^2} + \frac{\theta_{av}}{\sigma^2})} \times \left[e^{-ik_{\mu,\gamma}\cdot z} - e^{\frac{\sigma^2}{2}}\right] \quad ()$$

where each $\varphi_{\mu,\gamma}$ is a plane wave characterized by the vector $k_{\mu,\gamma}$ enveloped by a Gaussian function, where σ is standard deviation of the gaussian. The center frequency of the filter at μ,γ is given by the characteristic wave vector, $k_\gamma cos\theta_\mu$, $k_\gamma sin\theta_\mu$ having a scale and orientation given by k_γ and θ_μ. Here, scale factor γ is 4, orientation factor μ is set to 3. The first term inside the brackets in eq.6 gives the oscillatory part of the kernel. The second part compensates the DC value of the kernel.
The expression $\exp(\sigma^2/2)$ is the cut-off frequency and it is subtracted to remove the DC components of the Gaussian octaves.
The convolution of an intensity-corrected image Ni and a Gabor wavelet $\varphi_{\mu,\gamma}$ can be defined as follows:

$$G(\vec{z}) = Ni(\vec{z}) * \varphi_{\mu,\gamma}(\vec{z}), \vec{z} \quad ()$$

$G(\vec{z})$ is the convolution output with a wavelet at a position $\vec{z}$. μ denotes the scale, γ denotes the orientation.
$\varphi_{\mu,\gamma}$ is tuned to 4 scales and 3 orientations and they are used for extraction of the illumination strength.

The convolution results at a pixel position $\vec{z}$ consist of important local information, and can be concatenated to form a discriminative local feature.

Step 2:

Let $Ni_{x,y}$ be the nonlinear image intensity function, which is calculated by (2). The definition of a surface can be explained in terms of depth and albedo by giving $S_{x,y} = S_{-x,y}$ and $\rho_{x,y} = \rho_{-x,y}$ due to the symmetry relation of the face.

A standard Lambertian reflectance function has the following form:

$$Ni_{x,y} = \rho_{x,y}\frac{(1+p_{x,y}P_s+(q_{x,y}Q_s))^2}{(1+p_{x,y}^2+q_{x,y}^2)(1+P_s^2+Q_s^2)},\ Ni_{-x,y} = \rho_{-x,y}\frac{(1-p_{x,y}P_s+(q_{x,y}Q_s))^2}{(1+p_{x,y}^2+q_{x,y}^2)(1+P_s^2+Q_s^2)} \quad ()$$

where $P_s = \frac{\cos(\tau)\sin\sigma}{\cos\sigma}, Q_s = \frac{\sin(\tau)\sin\sigma}{\cos\sigma}$, τ is the tilt and σ is the slant of the illumination. $\rho_{x,y}$ is the albedo of the face normal. $p_{x,y}$ and $q_{x,y}$ are the surface gradients which their gradients are tangent to the surface normal. The ratio of $L_{x,y}$ and $L_{-x,y}$ gives a new reflection which yields

$$r_{x,y} = \frac{Ni_{x,y}-Ni_{-x,y}}{Ni_{x,y}+Ni_{-x,y}} = \frac{p_{x,y}P_s}{1+q_{x,y}Q_s} \quad ()$$

By using discrete approximation for surface gradients, $p_{x,y} = S_{x,y} - S_{x-1,y}$, $q_{x,y} = S_{x,y} - S_{x,y-1}$, the reflection function (33) is modified as

$$f(r_{x,y}, S_{x,y}, S_{x-1,y,S_{x,y-1}}) = r_{x,y} - \frac{P_s(S_{x,y}-S_{x-1,y}}{1+Q_s(S_{x,y}-S_{x,y-1})} = 0 \quad ()$$

To find two unknowns, $S_{x-1,y}$ and $S_{x,y-1}$, we expand (34) into taylor series and use only the first order terms. Finally, we obtain

$$f_n \cong f_{n-1} + \Delta S_1\frac{\partial f_{n-1}}{\partial S_{x,y}^{n-1}} + \Delta S_2\frac{\partial f_{n-1}}{\partial S_{x-1,y}^{n-1}} + \Delta S_3\frac{\partial f_{n-1}}{\partial S_{x,y-1}^{n-1}} = 0 \quad ()$$

where $f_n = f(r_{x,y}, S^n_{x,y}, S^n_{x-1,y}, S^n_{x,y-1})$, $f_{n-1} = f(r_{x,y}, S^{n-1}_{x,y}, S^{n-1}_{x-1,y}, S^{n-1}_{x,y-1})$

For the simplicity of the eq.35, two or more constant values are set to zero ($\Delta S_2 = \Delta S_3 = 0$)

$$S^n_{x,y} = S^{n-1}_{x,y} - \frac{r_{x,y} - \frac{P_s(S^{n-1}_{x-1,y} - S^{n-1}_{x-1,y})}{1+Q_s(S^{n-1}_{x,y} - S^{n-1}_{x,y-1})}}{\frac{\partial f^{n-1}}{\partial S^{n-1}_{x,y}}} \quad ()$$

where $\frac{\partial f^{n-1}}{\partial S^{n-1}_{x,y}}$ is the partial gradient term and is defined as below:

$$\frac{\partial f^{n-1}}{\partial S^{n-1}_{x,y}} = \frac{P_s(1+Q_s(S^{n-1}_{x-1,y} - S^{n-1}_{x,y-1}))}{(1+Q_s(S^{n-1}_{x,y} - S^{n-1}_{x,y-1}))^2} \quad ()$$

Substituting (37) into (36) and also substituting (34) to (36) gives the iterative depth map equation

$$S^n_{x,y} = S^{n-1}_{x,y} + \frac{r_{x,y}W^2_{x,y} - P_s W_{x,y}(S^{n-1}_{x,y} - S^{n-1}_{x-1,y})}{P_s(W_{x,y} - Q_s(S^{n-1}_{x,y} - S^{n-1}_{x-1,y}))} \quad ()$$

where $W_{x,y} = 1 + Q_s(S^{n-1}_{x,y} - S^{n-1}_{x,y-1})$

Substituting (37) into (33) gives us the pixel albedo values.

$$\rho_{x,y} = Ni_{x,y} \times \frac{\sqrt{(1+(S_{x,y} - S_{x-1,y})^2 + (S_{x,y} - S_{x,y-1})^2)}}{1+(S_{x,y} - S_{x-1,y})P_s + (S_{x,y} - S_{x,y-1})Q_s} \left[\sqrt{(1 + P_s^2 + Q_s^2)}\right] \quad ()$$

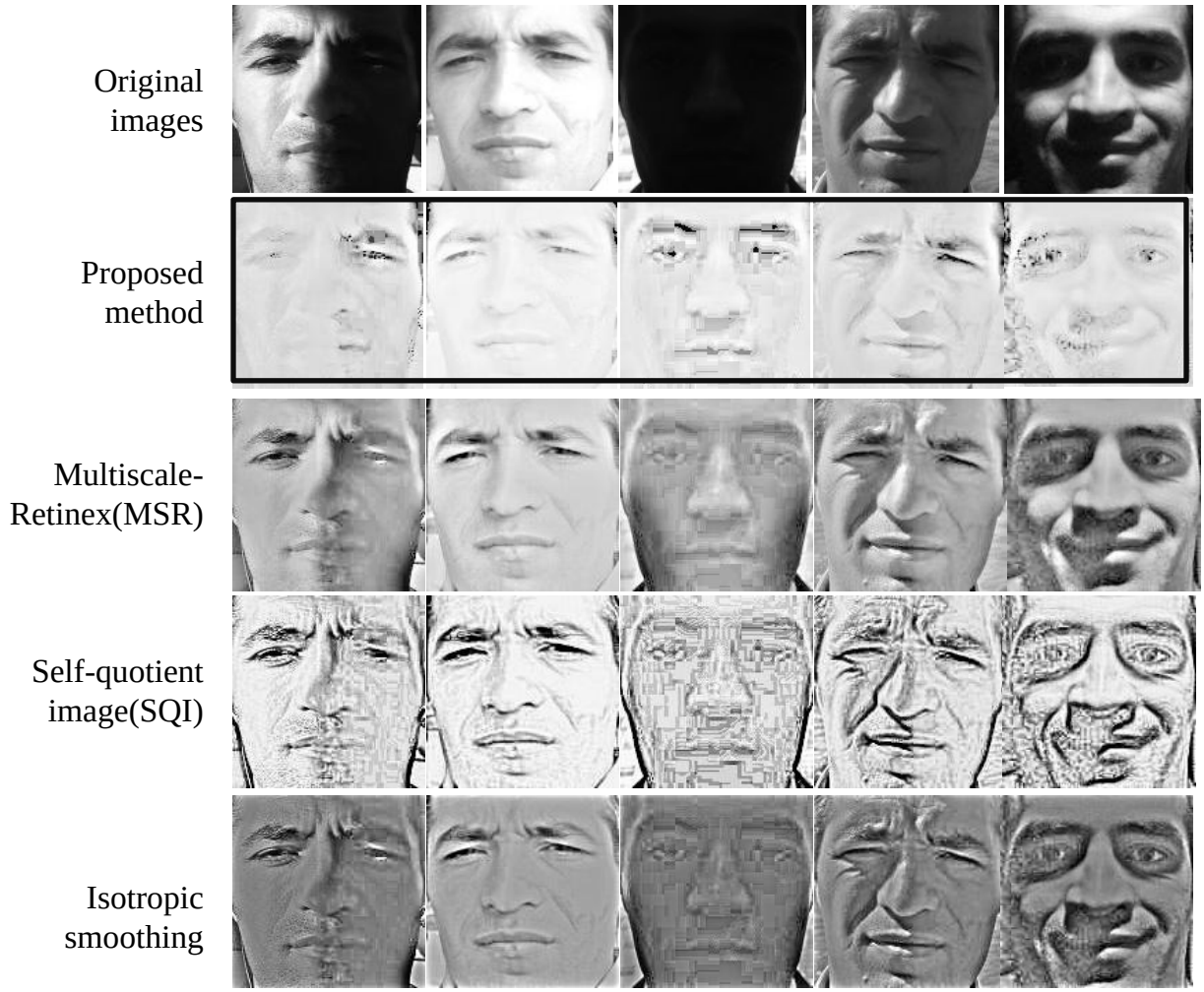
where $Ni_{x,y}$ is the intensity normalized face image. We compute the albedo pixel values on $Ni_{x,y}$. The output of the computation is given in cartesian domain.

Step 3:

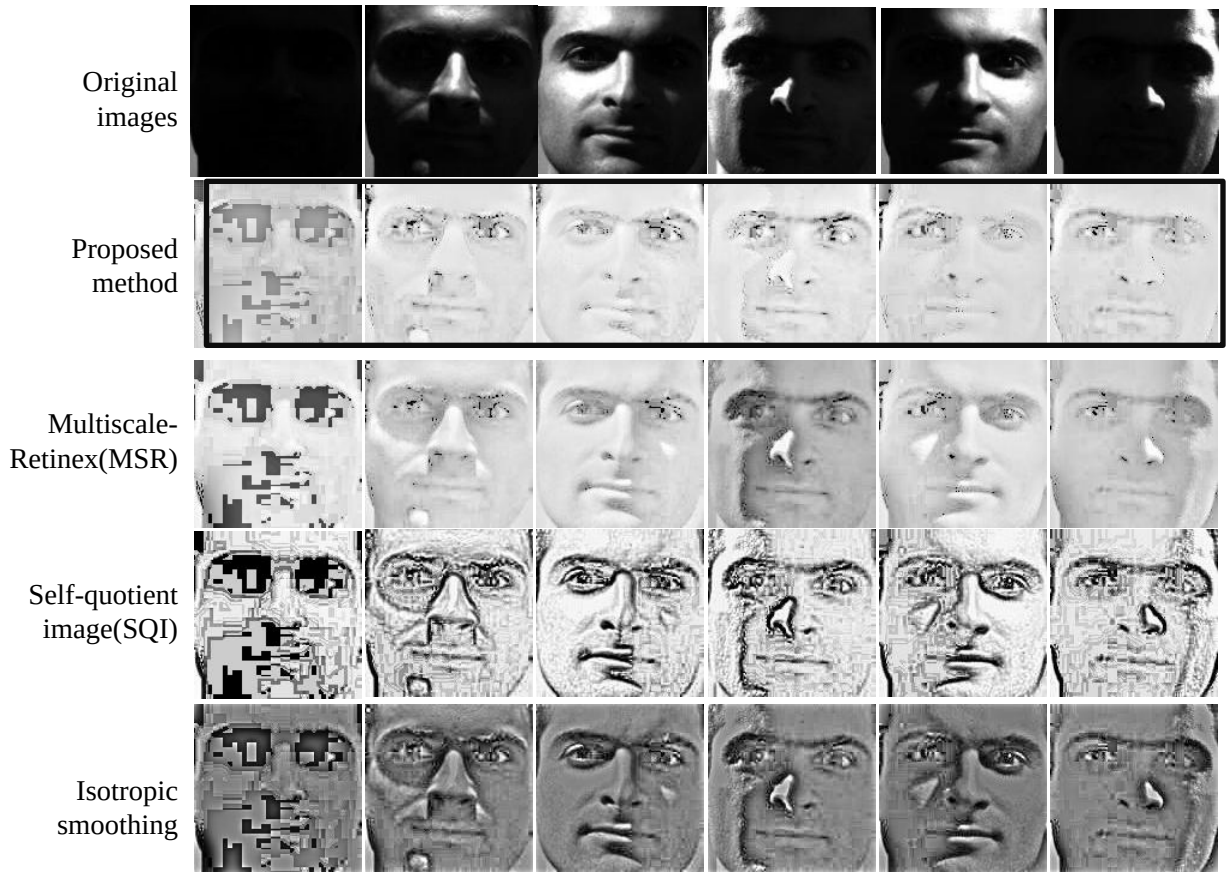
After the intensity at each pixel value is recovered for all areas, the face is reconstructed by linear interpolation.

As a summary, in step 1, Ayofa-filters are computed to extract the face features. In step 2, albedo function ρ and computed albedo value in a given pixel value is calculated. If the neighboring gradient point in similar directions gives near values, the peaks produce an averaging effect among the neighbors. Such information is computed in step 2. In step 3, pixel recovery is done by interpolating the parts.

Fig. 26 (a) depicts the original images and their normalized images, which are taken under different illumination environments.



(a) Outdoor images (IPCAM). First row: the input images from outdoor. The third, fourth and fifth rows: the results from different illumination techniques. Shadow and any illumination effect due to the unknown light directions are removed if only one input image is available.



(b) Yale database images. First row: the input images extended Yale database. The second row: the results of the proposed method.

Fig. : Face illumination normalization results.

In fig. 26 (b), the images from extended Yale database (YaleE-DB) are processed. The proposed method extracts the features, albedo estimation is done, and the image is reconstructed by linear interpolation technique to obtain the illumination-normalized face image.

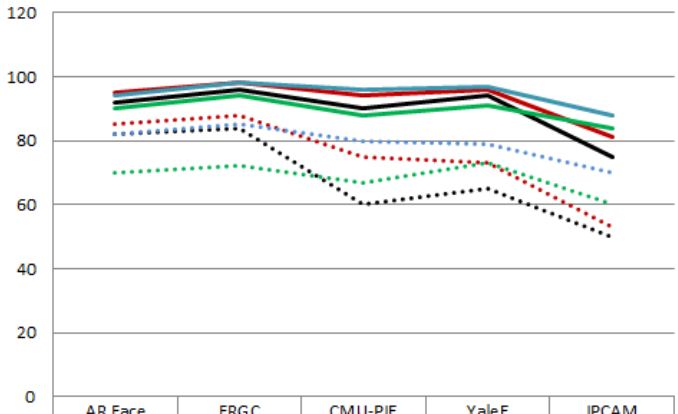
The proposed method is compared with multi-scale retinex (MSR) [92], self-quotient image (SQI) [93] and Isotropic diffusion based normalization technique [52]. Multi-scale retinex (MSR) processing has been shown to be an effective way to enhance image contrast, but it often has an undesirable desaturation effect which can be seen in fig. 26 (the third row). The SQI technique exhibits similarities to the MSR, but unlike the MSR, it uses an anisotropic filter for the smoothing operation. Isotropic technique uses a smoothing of the image to estimate the

luminance. It represents a simpler variant of the anisotropic diffusion based normalization technique proposed by Gross et al.[52]. Compared with the other techniques, our proposed technique provides the same level of illuminated images. For example, for the third column image, other techniques provide dark and distorted face compared with other images although the proposed technique gives very similar illumination normalization for all environments that we experienced.

3.5 Evaluation results of the proposed method

The proposed illumination normalization technique has been tested in different aspects. The first testing was done by using FRGC-DB and YaleE, CMU-PIE databases and the second testing was done by outdoor images directly from the IP camera which was mounted in outside. During all the evaluations, face is detected first. After that, two eyes are detected. Face detector and eye detector details are given in chapter 3. Eye detection has an average error rate of three pixels in outdoor and this error becomes significantly high on different poses of a face. The accuracy difference between the images with the ground-truth data and the images with three-pixel errors was approximately 0.1%. The method of Biswas et al. [39] caused inaccuracy around 5%. Our proposed method compensates the pixel errors up to 10 pixels. Errors more than 10 pixels caused significant performance drop. The effect of the proposed illumination technique on face recognition is given in fig. 27.

To measure the face recognition performance, we selected the basic face recognition algorithms which are PCA, LDA, Gabor and AAM. Each method is weak against to illumination. Gabor is easily affected by small variations of illumination. AAM modeling fails if the illumination is not stable. PCA and LDA subspace representations change with small illumination effects. The proposed method is applied before faces are processed by PCA, LDA Gabor or AAM techniques.



	AR Face	FRGC	CMU-PIE	YaleE	IPCAM
PCA with AF	92	96	90	94	75
PCA without AF	82	84	60	65	50
LDA with AF	95	98	94	96	81
LDA without AF	85	88	75	73	53
Gabor with AF	94	98	96	97	88
Gabor without AF	82	85	80	79	70
AAM with AF	90	94	88	91	84
AAM without AF	70	72	67	73	60

Fig. : Comparison with/without Ayofa-filter

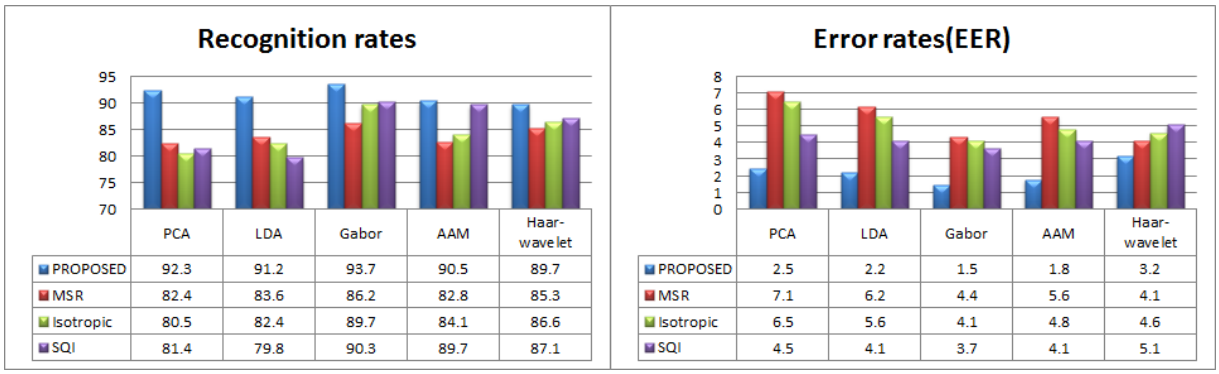
PCA and LDA are holistic approaches. Gabor is feature-based approach and AAM is hybrid approach, which is a mixture of feature-based approaches and holistic approaches. By choosing four different methodologies, we can make sure of the performance of the proposed method. We evaluated the proposed method with these methodologies and then evaluated them without our method to see the effectiveness of our method. LDA performed better than PCA in all databases and Gabor overperformed PCA, LDA and AAM both in the first stage and the second stage.

During the evaluation, we ignored the first three components of the PCA to eliminate the illumination effects. Furthermore, we performed 10 times the same test for each database, enabling us to construct an average result with a 90% confidence interval for both PCA and LDA. The illumination affects the performance significantly as seen in fig. 27. The proposed method significantly improves the face recognition performance.

FRGC-DB gave the best results among the databases we used. ARFace and CMU-PIE gave similar results. IPCAM results are worse than the ARFace, FRGC, CMU-PIE and YaleE. There are several reasons to cause it such as pose changes, landmark detection errors, blurring on the face, face expression, camera focus etc. As a result, we confirmed that our proposed method

improved the recognition rates significantly.

We conducted some more tests by using other illumination techniques. We compared our method with multi-scale retinex algorithm, self-quotient image and isotropic smoothing methods. Their results are seen in fig. 28.



	PCA	LDA	Gabor	AAM	Haar-wavelet
PROPOSED	92.3	91.2	93.7	90.5	89.7
MSR	82.4	83.6	86.2	82.8	85.3
Isotropic	80.5	82.4	89.7	84.1	86.6
SQI	81.4	79.8	90.3	89.7	87.1

(a) Recognition rate

	PCA	LDA	Gabor	AAM	Haar-wavelet
PROPOSED	2.5	2.2	1.5	1.8	3.2
MSR	7.1	6.2	4.4	5.6	4.1
Isotropic	6.5	5.6	4.1	4.8	4.6
SQI	4.5	4.1	3.7	4.1	5.1

(b) Error rate

Fig. : Comparison of the proposed technique with other illumination techniques

We compared the proposed method by using MSR, Isotropic and SQI methodologies. MSR and SQI are similar methods. The difference of MSR is that it uses isotropic smoothing while SQI uses anisotropic smoothing. We used multiple scales of Gaussian filters during the computation of SQI. We used four scales and four orientations. In MSR, we used three different filter scales. As it can be observed in fig. 26, neither MSR, SQI nor Isotropic method is enough by itself. They need additional methodologies to provide promising results. These methods remove the some part of high frequency components of the given face. This is logical since illumination variation highly affects the high frequency face components. On the other hand, removing high energy distribution removes some discriminative parameters of the face, as well. Therefore, although illumination is corrected virtually, the recognition rates do not increase. In most cases, error rates increase significantly. Our proposed method provides the stable illumination normalization and the error rates do not increase after the illumination. There is a possibility to have some deformations after the illumination normalization such as fig. 26(b) top left image. Although the deformation is significant, the error rate increment is ignorable in most of cases.

To see the speed performance of the proposed method, we compared the proposed method with MSR, Isotropic and SQI and the results are given in table 7.

Table : Speed comparison

	Illumination methods				
Speed(ms)	Image resolution	**Proposed**	MSR	Isotropic	SQI
	640x480	**15**	20	25	27
	1280x960	**30**	40	42	55
	1920x1080	**50**	65	65	72

The highest resolution (1920x1080) was used during evaluation in outdoor (the second stage). The other resolutions (1280x960 and 640x480) were used in the first stage.

To measure the speed, we used Intel pentium Core2Duo, 1.8GHZ with 1GB RAM. We used single CPU, single thread to measure the speed. Our proposed method computation was faster than MSR, Isotropic, and SQI computation. In all methods, resolution and computation time are directly related. As the resolution of the image increases, the computation time also increases.

3.6 Summary

We proposed the Ayofa-filter for albedo estimation and illumination normalization for accurate face recognition in uncontrolled environments. We used one single image and normalized it by five points from the eye, the nose, and the corners of the mouth. Our Ayofa-filter extracts the face features by using Ayofa-vectors and estimates the albedo. Instead of processing a face as a whole, we divided a face into small subsets (nine equal pieces) and estimated the reflection function parameters by Ayofa-filter for each face part. More partitioning of the face was also investigated in this paper to make sure nine-equal division is appropriate. We removed the DC components of the face and used a set of filters for extracting the most meaningful face features. We generated specific filters with various orientation and frequencies with a kernel frequency. This improved the computation speed of the Ayofa-filter by making it advantageous compared with the traditional methods such as decomposition techniques, illumination cones, frequency analysis techniques etc.

The experimental results showed that the approach is promising and it increases the performance of conventional face recognition algorithms. We compared the efficiency of the proposed method by using MSR, Isotropic and SQI image techniques. We confirmed that the proposed

method overperforms all these illumination technique. Furthermore, we tested the proposed method and MSR, Isotropic and SQI techniques by using PCA, LDA, Gabor and AAM face recognition algorithms to see the effects of our proposed illumination technique on these algorithms. The results before and after Ayofa-filter were quite different. Before Ayofa-filter, the experimental results produced poor recognition. After the Ayofa-filter, we observed a significant improvement on recognition rates. The best results were obtained on FRGC-DB. We used randomly selected five frontal images per person during enrollment. The enrollment images were chosen carefully not to have any angle or occlusions on faces.

CHAPTER 4: HYBRID HOLISTIC-BASED FACE RECOGNITION BY NONLINEAR FEATURE EXTRACTION

In this chapter, we explain the feature extraction and nonlinear feature analysis methods. After shortly giving introduction in the next section, the details of the local feature extraction method is given. The feature extraction is a combination of Gabor filters and Bessel functions. The feature extraction follows with nonlinear feature analysis, namely H-NPCA. The last section explains the feature classification. It gives the details of SVM. Finally, section 4.7 concludes the section with a short summary of the works.

4.1 Introduction

Face recognition algorithms are separated into three parts: feature-based approaches, holistic-based approaches and hybrid-based approaches. General approaches classify the faces as either holistic based, where faces are recognized using global features from faces, or, featured based, where faces are recognized using local features from faces. However, the features used in holistic and feature-based approaches are fundamentally different. Features found from the holistic approaches represent the optimal variances of pixel data in face images that are used to uniquely identify one individual to another. Alternatively, the features found from feature-based approaches represent face features like the eyes, noses and mouth, where these features are used to uniquely identify individuals. Holistic methods rely on face appearance and feature-based approaches rely on distances between eyes, nose and mouth.

In this chapter, we introduce a face extraction method, which extracts the local features of faces,

and find the optimal variances of the extracted features. The method is called hybrid-holistic based method since local features are extracted by applying specific Gabor wavelets into the specific points of the face and the low-level features are processed in a holistic way. A detailed block diagram shows it as below.

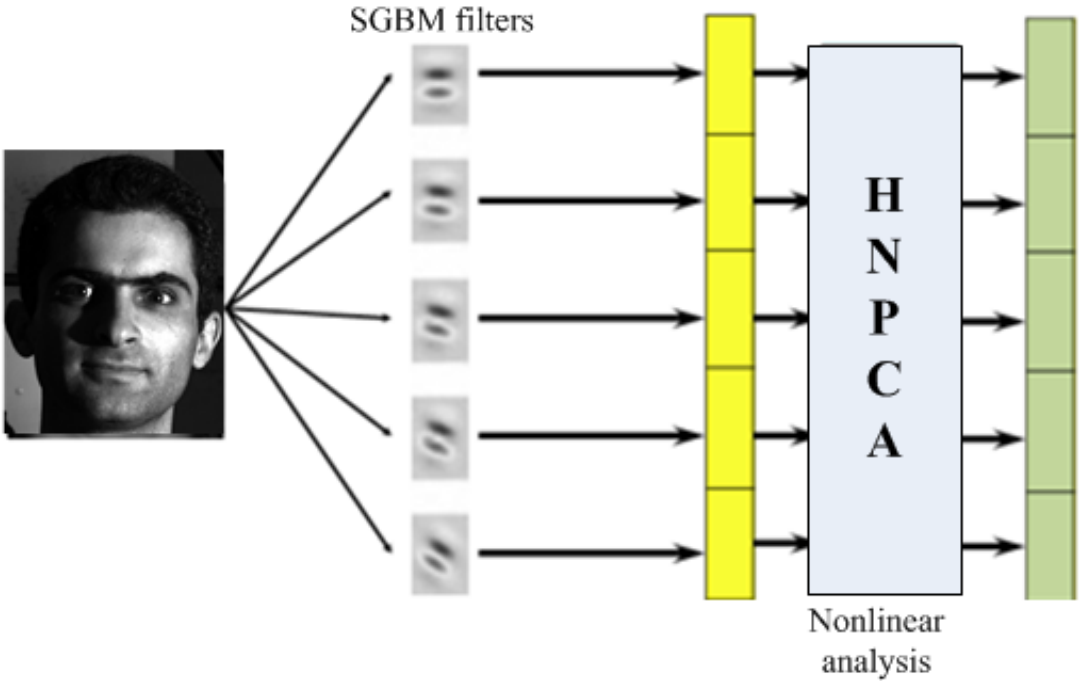


Fig. : Hybrid holistic based face recognition

Face is processed by Gabor filters. Gabor filters are constructed as 8 orientations and 5 scales. However, in this research study, we used an optimal orientation and scales. Optimal filters give both speed and performance. The features after the Gabor filters are grouped into subparts. The extracted features are displayed as yellow line in fig. 29. Each feature set is processed by nonlinear analysis method (H-NPCA part). H-NPCA finds the optimal variance of the features. In this chapter, first of all, we explain the SGBM filters in detail and we further give how we choose the best feature points from a face. Next, we explain the H-NPCA analysis. H-NPCA processes the features of SGBM. The relation of both parts, namely SGBM and H-NPCA is given in detail after that. After H-NPCA, the interrelation of both methods and their combination are given in short. The theory of support vector machine, which is a linear feature classification method, concludes the chapter 4.

4.2 Local feature extraction by using SGBM

To model SGBM filters, we used a method which is based on a gaussian approximation.

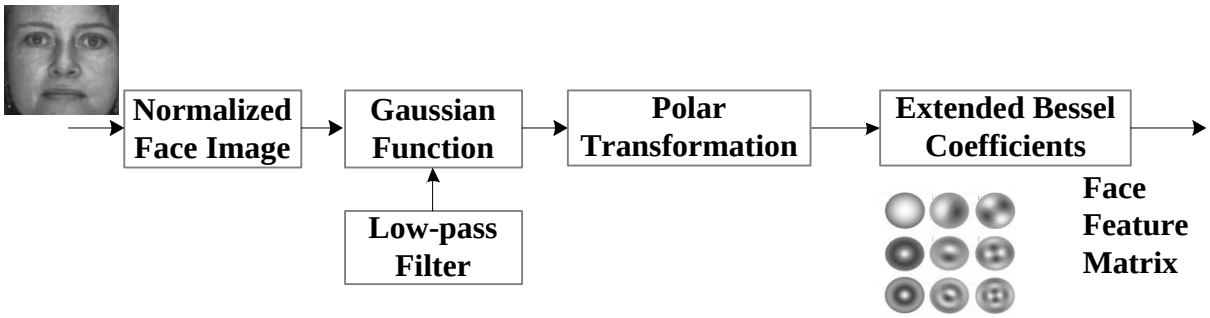


Fig. : SGBM filter flowchart

Filter implementation order is given above. A normalized face image is first processed by a two-dimensional gaussian function. The specified high frequency components are cut off by a low-pass filter. Proper selection of low-pass filter parameters is important. Probability density function of gaussians is low for $\pi/2$ and $2\pi/3$. Only $3\pi/4$ is primarily important since it is the optimal size as we consider a face size of 50x50 pixel. To avoid the low energy parts from being processed, low-pass filter tuning must be done properly. We choose $3\pi/4$ to be optimal for 50x50 pixel size face image.

The resulting filter is converted to polar coordinates and then extended bessel coefficients are applied.

Let S_f be spatial frequency, fmax be maximum spatial frequency. SGBM filter is composed of 7 orientations and 3 scales. Orientation is decided by frequency and scale is decided by a scale factor δ.

$$S_f(x,y) = e^{-2\pi^2\sigma_x^2(x^2+y^2)} \quad ()$$

Let $G_a(x,y)$ be two-dimensional gaussian function.

$$G_a(x,y) = \frac{1}{2\pi\sigma_x^2} e^{-\frac{(x^2+y^2)}{2\sigma_x^2}} \quad ()$$

$\sigma_x$is the scale of the gaussian envelope.

The steps to get SGBM are given below:

1. Apply cut-off frequency into gaussian function to drop some parts of the high frequency components
2. Apply the output of the resulting gaussian to low-pass filter.
3. Represent it in terms of polar frequency.
4. Combine the result with extended bessel (EB) coefficients.
5. Take the convolution of the input image with the result.

Step1:

Applying the cut of frequency $\frac{f}{\sigma}$ yields the following equation.

$$G_a(x,y) = \frac{1}{2\pi\sigma_x^2} e^{-\frac{(x^2+y^2)}{2\sigma_x^2}-\frac{\pi}{\sigma_x}} \quad ()$$

where $\frac{\pi}{\sigma_x}$ is the cut-off frequency.

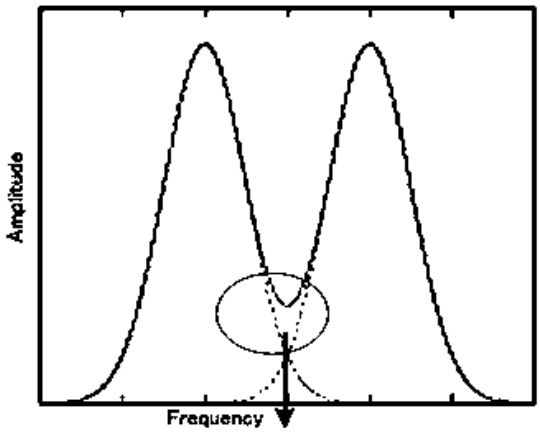


Fig. : Zeroing the cut-off frequency

Step 2:

Let L(x,y) be the low pass filter.

$$GL(x,y) = L(x,y)G_a(x,y) \quad ()$$

Step 3:

Now, let $P(r,\theta)$ be the polar frequency representation for GL(x,y).

$$P(r,\theta) = P(GL(r,\theta)) \quad ()$$

Combining PG with EB yields the final result.

Step 4:

Let $fb(r,\theta)$ represent the expanded Bessel (EB) function.

$$\theta = \frac{\tan^{-1}(y-y_0)}{(x-x_0)},\ r = \sqrt{(x-x_0)^2+(y-y_0)^2} \quad ()$$

$$fb(r,\theta) = \sum_{i=1}^{i=\infty}\sum_{n=0}^{n=\infty} A_{n,i}Jn(A_n, i^r)\cos(n\theta) + B_{n,i}Jn(A_n, i^r)\sin(n\theta) \quad ()$$

Jn is the bessel function of order n and $A_{n,i}$ is the ith root of the Jn function.

The orthogonal coefficients $A_{n,i}$ and $B_{n,i}$ are given by the following equation:

$$A_{0,i} = \frac{1}{\pi R^2 J_1^2(A_{n,i})}\int_0^{2\pi}\int_0^R f(r,\theta) r J_n\left(\frac{A_{n,i}}{R}r\right) dx d\theta \quad ()$$

$$\begin{bmatrix}A_{n,i}\\B_{n,i}\end{bmatrix} = \frac{2}{\pi R^2 J_n^2(A_{n,1})}\int_0^{2\pi}\int_0^R f(r,\theta) r J_n\left(\frac{A_{n,i}}{R}r\right)\begin{bmatrix}\cos(n\theta)\\\sin(n\theta)\end{bmatrix} dr d\theta \quad ()$$

where, n>0, $B_{0,i} = 0$

We transform the normalized face images up to the 20th bessel order with angular resolution of 20. In the polar frequency domain, the bessel root is related to the radial frequency (number of cycles along the image radius) while the bessel order is related to the angular frequency (number of cycles around the center of the image).

Step 5:

The output is then the convolution of the input face image with the product of $(fb(r,\theta)PG(r,\theta))$.

$$SFF(r,\theta) = I(r,\theta) \| (fb(r,\theta)PG(r,\theta)) \| \quad ()$$

4.3 Selection of the most discriminate SGBM features

Despite the advantages of SGBM filters in recognizing face images with different illumination, pose and expression, they require high computational efforts. Even when a multi-core CPU is used, the convolution of a 50×50 pixel image takes about few seconds. It is because there are many features introduced as a result of filter convolutions. Therefore, it is necessary to eliminate

the features that are ignorable. There are several parameters to optimize the features, which lead more robust to noises. For this point, a few works have been introduced until now. Wang et al. [34] introduced an optimization algorithm to select the Gabor features. They located 34 points manually on each image to represent faces. A set of Gabor wavelets with 4 scales and 6 orientations is then designed as candidates in his paper. In a similar way, we applied a generic algorithm (GA). From the GA algorithm, feature points are determined by using

$$R_j(x_0, y_0) = \max_{(x,y)} R_j(x, y) \quad ()$$

$$R_j(x_0, y_0) > \frac{1}{N_1 N_2} \sum_{x=1}^{N_1} \sum_{y=1}^{N_2} R_j(x, y) \quad ()$$

where R_j is the response of face poins, N_1, N_2 is the size of input image. As a result of GA, a 9x9 window is used to search feature points. 20 filters($\mu = 5, v = 4$) is used here.

4.4 Feature analysis based on hierarchical nonlinear PCA (H-NPCA)

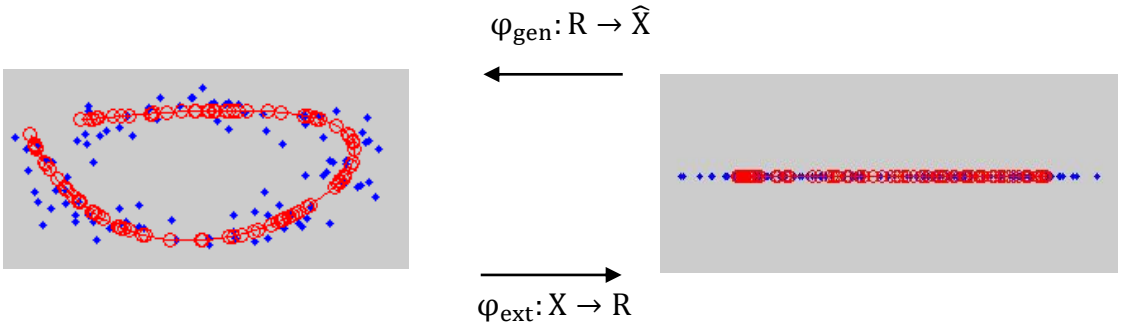
PCA is a well-established and frequently used unsupervised statistical dimension reduction technique. PCA aims to represent the high dimensional vectors with low-dimensional vectors. However, PCA works on linear data and data representation is tangled with each other as the database increases. In addition to this, there are more reasons to use nonlinear approaches:

1. PCA does not satisfy the performance because of the ignorance of the subsequent analysis.
2. Approximation of the missing data is not easy by linear analysis methods.

Face representation is not perfect by linear approaches since there are many parameters to change the face appearance in uncontrolled environments. This is a theory from brain studies. The brain neurons generally act differently over the time [78]. Neuron structures do not give linear results. Brain does not serve as general analyzer but serve as a flexible way to efficiently extract important information from stimuli using much more specialized structural analyses. These are specific stimuli that we need to remember and recognize (e.g one's car, friends' faces). By designating neural pathways to important stimuli, individual IT cells projecting to areas such as the frontal cortex, medial temporal lobe, amygdala, and striatum will facilitate recognition,

appropriate affective responses, and the correct associations.

With specialized visual experience, neural responses in IT will change their responses. Assuming the above studies are correct, we behave the observed face data as nonlinear structure. Therefore, we propose H-NPCA as a nonlinear dimension reduction technique. Fig. 32 shows a nonlinear transformation. The right graph (fig. 32 (a)) shows the input data that is the features of the face to train and it is obtained by drawing the face features in the space. The left graph (fig. 32 (b)) gives their transformed feature space which is mapped onto one dimensional-space. Input data is obtained from face images after processing them by gaussian filters. Face features are distributed into the feature space after they are processed. Some noises are introduced during feature extraction. The introduced noises are mainly caused by environmental factors. Some of environmental factors are background, low contrast, camera lens noise, weather conditions such as rain, wind etc.



(a) Original features (b) Mapped feature space

Fig. : Nonlinear dimensionality reduction.
The transformation is given by φ_{gen} and φ_{ext}.

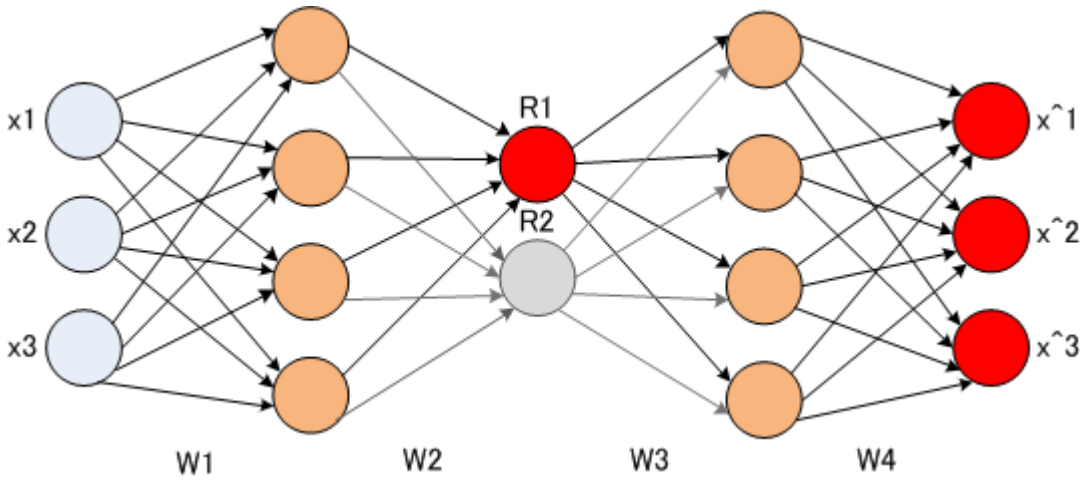
In fig. 32, input features and resulting feature space are shown. Before transformation, data has a circular shape on a multi-dimensional space and data is located on one-dimensional subspace after transformation. The blue dots are noises. Circles are the feature values. φ_{gen} is the generation function which changes the dimension into its original space and φ_{ext} is the extraction function which transforms the original features onto one-dimensional space.

We prove that we can represent the multi-dimensional face features in one dimension by using H-NPCA. Expressing the features in one dimension gives us several benefits such as simplified computation, faster calculation time etc.

There are various types of implementation for nonlinear PCA [65],[66],[67]. Among them, H-NPCA is the most stable and fully scalable. H-NPCA is based on artificial neural network [68],[69]. By using H-NPCA, both nonlinear subspace spanned by the optimal set of the components and the order where the components appear are primary important.

Let $X = \{x_1, x_2, x_3, \dots, x_n\}^T$ be the high-dimensional image data, which is called as observed data. Here, $X = \{x_{i1}, x_{i2}, x_{i3}, \dots, x_{in}\}^T \in R^n$ where R is the subspace of the X. The superscript T is the transposition.

H-NPCA targets to find both the subspace R and the mapping between X and R. Mapping is given by nonlinear functions φ_{ext} and φ_{gen}. $\varphi_{ext}: X \to R$ transforms the coordinates $X = \{x_1, x_2, x_3, \dots, x_n\}^T$ of the high dimensional data space X into the corresponding coordinates $R = \{r_1, r_2, r_3, \dots, r_k\}^T$ on R. R is the low dimensional data. $\varphi_{gen}: R \to \hat{X}$ is the inverse mapping which is used to reconstruct the data from the low dimensional data. Hence, φ_{gen} calculates the assumed data generation process.



$\varphi_{ext}: X \to R$ $\varphi_{gen}: R \to \hat{X}$

Fig. : H-NPCA neural network

In fig. 33, a network topology is shown. The artificial neural network (ANN) is hierarchically extended to perform H-NPCA. Whole network structure is [3-4-2-4-3] network and there is a [3-4-1-4-3] sub-network to consider. The layer in the middle has one, two or more nodes to represent the first, second and higher number of components.

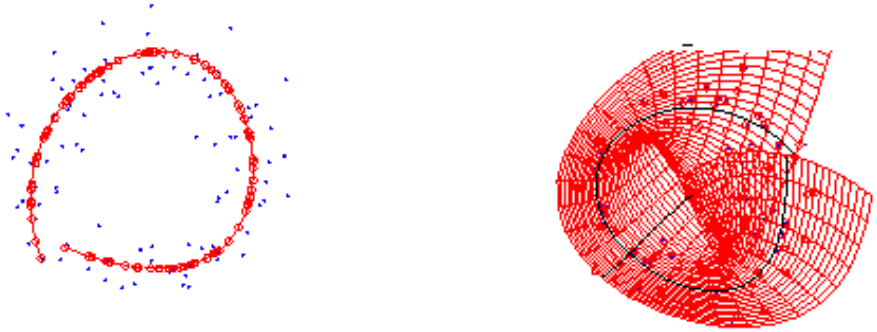
The ANN can be considered as two parts: the first part represents the original data function $\varphi_{ext}: X \rightarrow R$ and the second part is the reconstruction function $\varphi_{gen}: R \rightarrow \hat{X}$. Hidden layers of both φ_{ext} and φ_{gen} enable the network to perform nonlinear projection functions. R2 network is a hidden network. $\hat{X}$ is generated from the R which is seen in the middle of the fig. 33. $\hat{X}$ is noise free representation of X. During reconstruction, we remove the complex nonlinear correlations between data. Some of nonlinear correlations are environmental noises, background effects etc. Removing them reduces the noise and provides useful and meaningful components. Removing complex nonlinear correlations also leads nonlinear whitening transformation. However, removing noises may remove face discriminative features.

The error coefficient (E) of the sub-network with one component and the error of the total network with two components are estimated separately per iteration. The network weights are then adapted jointly with regard to the total hierarchic error, $E = E_1 + E_{1,2}$. E_1 and $E_{1,2}$ are the mean reconstruction errors which happen when using only first few components of the feature data. To minimize such errors, $\|x - \hat{x}\|^2$ must be achieved.

Mean square error (MSE) is calculated as below:

$$E = \frac{1}{\partial N} \sum_n^N \sum_k^D (x_k^n - \hat{x}_k^n)^2 \qquad ()$$

where, x_k^n is the original face features and $\hat{x}_k^n$ is the reconstructed data. N is the number of features and d is the dimensionality.



(a) 3-4-1-4-3 network (b) 3-4-2-4-3 network

Fig. : H-NPCA inversing

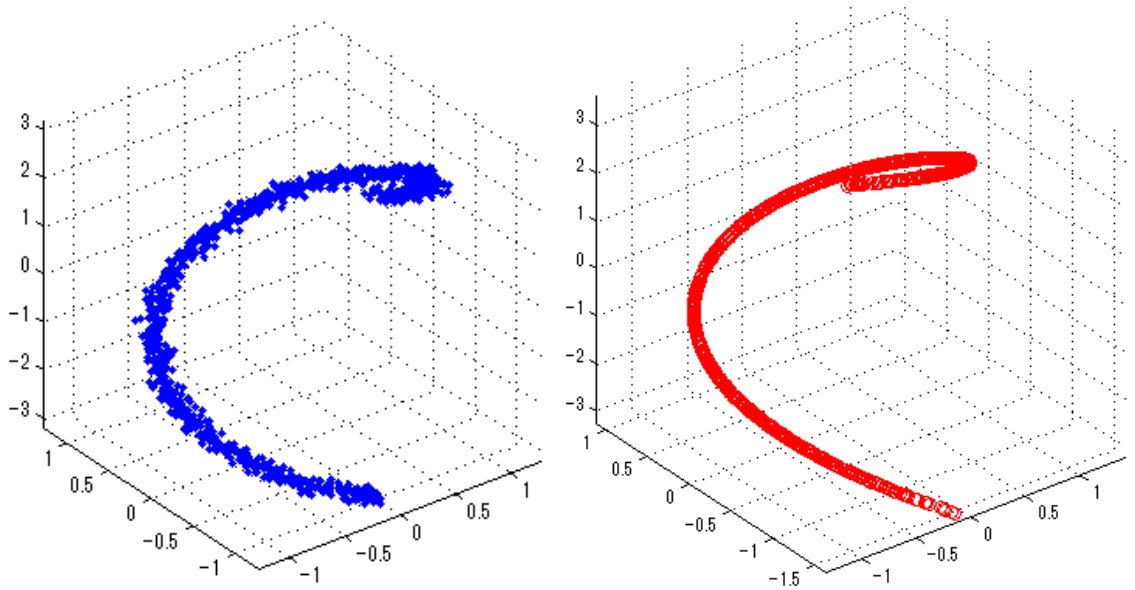
In fig. 34, data inversion is shown. In fig. 34(a), a network topology of 3-4-1-4-3 is used and in

fig. 34 (b), 3-4-2-4-3 is used. The number of features used is 100. The network topology 3-4-1-4-3 and 3-4-2-4-3 are used for the same data. If the number of network nodes is added, the inversing becomes more complicated.

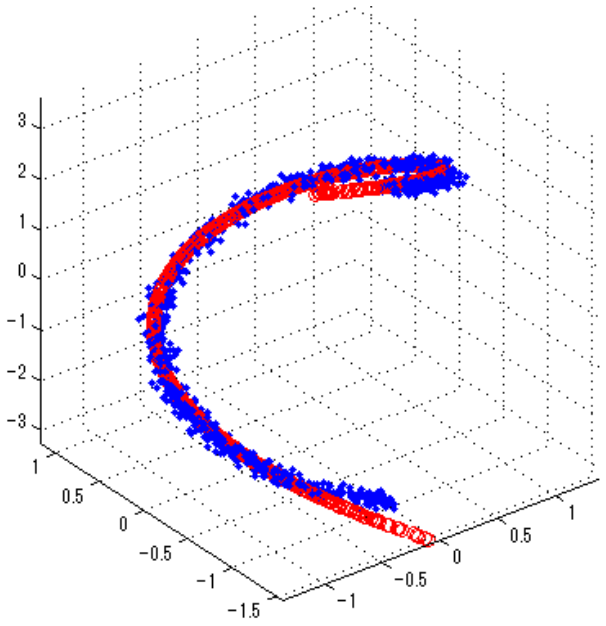
4.4.1 Missing data approximation

One of the main problems in analyzing face feature is the frequent absence of some values in the feature vectors due to measurement failures such as occlusion, environmental effects, face pose, illumination. There are many missing data estimation methods such as blind inverse models where the input and the model are optimized to match a given feature. The output-feature vector does not necessarily have to be complete. One method is to replace the missing values by mean over the available values of corresponding variable. However, this method may lead poor results since it assumes each feature has no relation, which is not the case if the target is a face. There is a maximum likelihood approach, which uses expectation-maximization (EM) algorithm. It estimates the missing values first. However, there is a possibility that it leads problems when distinct or even incompatible assumptions are used. Bishop et al. [70] introduced a bayesian missing value estimation, which uses some linear techniques for estimation. It is good for general object recognition. However, it is not applicable for face since face vectors are not linear.

One approach we used with H-NPCA is a modified inverse model method. Assume that i_{th} element of x_i^n of the nth feature vector xn is missing data. The partial error σ_i^n is taken as zero before back propagating to eliminate the effect of the noise during estimation. All nonlinear features are extracted and original data can be reconstructed including the missing parts. The feature network output $\hat{x}_i^n$ gives the estimation of missing element x_i^n.



(a) Input data (b) Missing data estimation



(c) Input data and estimated missing data

Fig. : Estimation of the missing data.

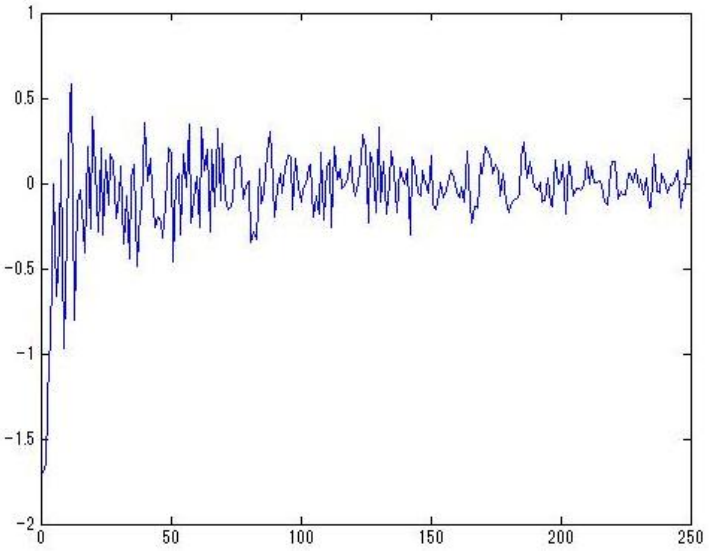
As seen in fig. 35, the lines plotted as dots '·' depicts the known features and circles gives the estimated values. The inverse NLPCA extracts the nonlinear component from this highly incomplete data set, and hence gives a very good estimation of the missing values.

Fig. 35 is obtained by a neuron structure [1-3-2] which is used during the estimation. Fig. 35(b) is

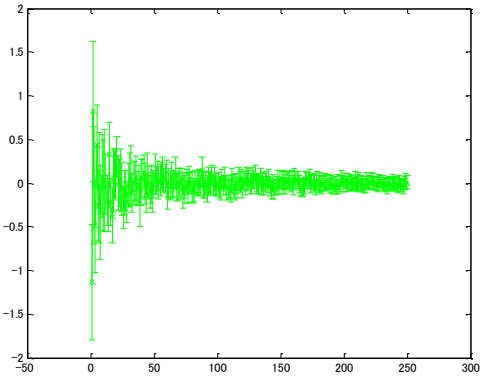
obtained by 3000 iterations by yielding a mean square error (MSE) of 1.5058 which its calculation is given in (52). The structure is not limited to one single component. If the number of components in the input layer is increased, multiple components can be estimated at the same time.

4.4.2 Whitening

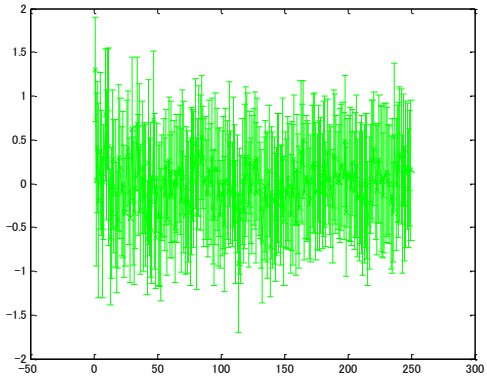
Whitening is a stronger constraint that requires both de-correlation and unit variance. Whitening the observation variables improves H-NPCA performance.



(a) Score deviation before whitening



(b) H-NPCA basis with larger values



(c) Whitened H-NPCA

Fig. : Whitened H-NPCA

Fig. 36 depicts that first projection vectors provide higher multitude scores, e.g. ± 2 and all the next vectors decreases in those projection limits such as ± 1.5, ± 1, ± 0.5. Those first vectors have low frequency and correspond to illumination, face hairs. On holistic face features, rest is the high frequency filters which discriminate the small features in the face. Fig. 36 (b) shows the H-NPCA components before the whitening step and fig. 36 (c) shows the whitening H-NPCA components seen in fig. 36 (b). It is possible to equalize dynamic ranges of H-NPCA projection vectors that all scores are equalized within ± 1.5 values limit. That provides similar significance to all H-NPCA features and prevents the mismatches due to the face hair such as beard, moustache etc. Although the whitening increases the noises in high frequency components, the classification accuracy of features in low frequencies improves significantly.

4.5 Hybrid holistic face feature analysis by using SGBM and HNPCA

SGBM are similar characteristics to human visual system, and it has been found to be particularly appropriate for texture representation. In spatial domain, Gabor filter is a gaussian kernel function modulated by a sinusoidal. Its impulse response is defined by a harmonic function multiplied by a gaussian function. Because of the multiplication-convolution property, the Fourier transform of a Gabor filter's impulse response is the convolution of the Fourier transform of the harmonic function and the Fourier transform of the gaussian function. The filter has a real and an imaginary component representing orthogonal directions. The two components may be formed into a complex number or used individually.

Gabor filters are used for extracting the local face features. The input to the Gabor filters is raw image data and the output is the texture information. Gabor filters are applied to the specific points of a face and most invariant feature points are processed. The points are determined by generic algorithm. The selected features are further processed by H-NPCA. H-NPCA is a nonlinear feature analysis method. It analyzes the global texture information. We use this structure. Because, analyzing the texture information by using H-NPCA minimizes the effects of illumination, image noise and distortions. In addition to this, this hybrid holistic method improves speed. Because, analyzing the entire image is much slower than analyzing a few bits of

features. Furthermore, the Gabor filters are robust to small face pose variances. Because of the circular structure of gabor filters, the face curves can be computed accurate. However, direct process from images is not tolerant to small changes on face angle.

H-NPCA estimates the missing parts by using its neural network architecture. For example, if an occlusion happens to one part of face, the recognition ratio decreases in significant percentage. The detailed method of the H-NPCA is given in section 4.4. Fig. 37 gives more information about the proposed method.

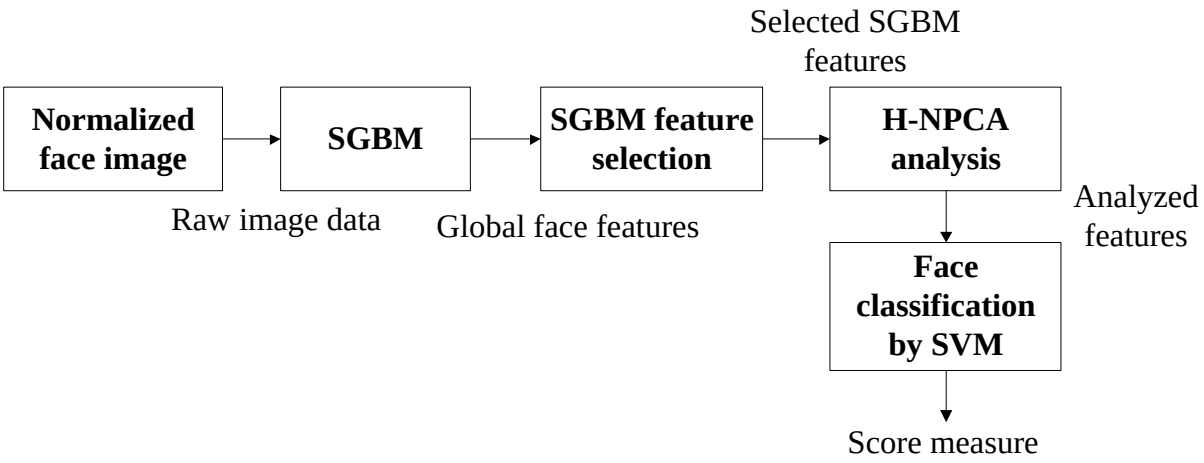


Fig. : Structural diagram of SGBM and H-NPCA

Normalized face image is the face image, which is cut by five points as given in chapter 3. Face object is cut, illumination is corrected, and it is processed by SGBM. The results of SGBM are selected. Some of features are dropped and deleted. The most meaningful features are selected by generic algorithm. The results are analyzed by HNPCA. Finally, the features are classified. The details of the classification method are given in section 4.6.

4.6 Feature classification by support vector machine.

As commonly known, studies on training algorithms for support vector machines (SVM) are important issues in the field of machine learning. It is a task to improve the efficiency of the algorithm without reducing the generalization performance of SVM. To achieve this task, we

introduced a new SVM training algorithm based on the set segmentation and k -means clustering. In this method, we divide all the face features into small subsets by clustering each subset using k -means clustering. The results show that the proposed SVM computation classifies nonlinear features successfully.

Support vector machine performs classification by reconstructing an N-dimensional hyperplane. In SVM, a predictor variable is called an attribute, and a transformed pixel attribute which is used for defining the hyperplane is named as feature. The task of choosing the most suitable representation is known as feature selection. A set of features that describes one case like a row of predictor values is called a vector. Then, the goal of SVM modeling is to find the optimal hyperplane that separates clusters of vector

Linearly separable two classes and one border are shown in fig. 38. The decision boundary should be as far from the data of both classes as possible.

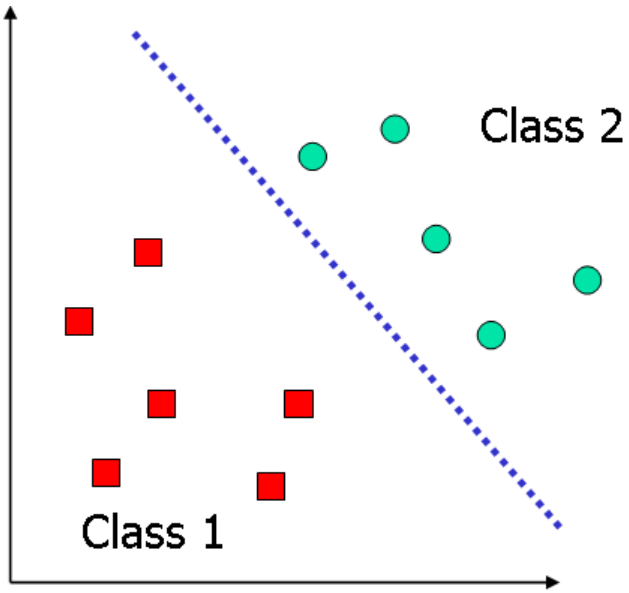


Fig. : SVM decision boundary

Class 1 and class 2 are the two given data needed for training in SVM.

Class 1 is the negative data and class 2 contains the positive data. Class 2 contains a large dataset of face images.

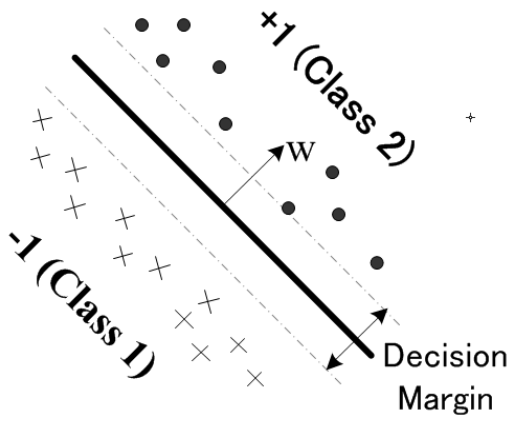


Fig. : SVM class separation

Let $\{x_1, \ldots, x_n\}$ be our data set and let $y_i \in \{1, -1\}$ be the class label of x_i. The decision boundary in fig. 39 must satisfy (53) to classify the class 1 and class 2.

$$y_i(w^T x_i + b) \geq 1 \qquad ()$$

The decision boundary can be found by solving the following constrained optimization problem

$$\text{Min}(\tfrac{1}{2}\|w\|^2) \text{ with subject to } y_i(w^T x_i + b) \geq 1 \qquad ()$$

This is a constrained optimization problem. To solve this problem, a lagrange multiplier for each constraint is used. The lagrangian is

$$\text{L} = \tfrac{1}{2} w^T w + \sum_{i=1}^{n} \propto_i (1 - y_i(w^T x_i + b)) \qquad ()$$

where $\|w\|^2 = w^T w$

By setting the gradient of L and b to zero, we can obtain

$$\text{w} + \sum_{i=1}^{n} \propto_i (-y_i) x_i = 0 \implies \text{w} = \sum_{i=1}^{n} \propto_i y_i x_i \qquad ()$$

and $\sum_{i=1}^{n} \propto_i y_i = 0$

Substituting (56) to (55) leads the following calculation

$$L = \frac{1}{2}\sum_{i=1}^{n} \propto_i y_i {x_i}^T \sum_{j=1}^{n} \propto_j y_j x_j + \sum_{i=1}^{n} \propto_i (1 - y_i(\sum_{j=1}^{n} \propto_j y_j {x_j}^T x_i + b)) =$$

$$-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \propto_i \propto_j y_i y_i {x_i}^T x_i + \sum_{i=1}^{n} \propto_i \quad ()$$

where $\sum_{i=1}^{n} \propto_i y_i$ is assumed to be zero valued.

The new objective function is in terms of $\propto_i$.

It is known as the dual problem: if we know **w**, we know all $\propto_i$ if we know all $\propto_i$ we know **w.** The original problem is known as the primal problem. The objective function of the dual problem needs to be maximized. The dual problem is therefore:

$$\text{Max}(W(\propto)) = \sum_{i=1}^{n} \propto_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \propto_i \propto_j y_i y_j {x_i}^T x_j \quad ()$$

where $\propto_i \geq \mathbf{0}$, $\sum_{i=0}^{n} \propto_i \boldsymbol{y_i} = \mathbf{0}$

$\sum_{i=0}^{n} \propto_i \boldsymbol{y_i} = \mathbf{0}$ is obtained when we differentiate the original lagrangian with respect to b.

w is a linear combination of a small number of data points and $\mathbf{x}_i$ with non-zero $\propto_i$ are called support vectors. The x_i can determine the decision borders. Let t_j (j=1... s) be the indices of the *s* support vectors. Hence, w can be written as below

$$\boldsymbol{w} = \sum_{j=1}^{n} \propto_j \boldsymbol{y_j x_j} \quad ()$$

Note that **w** needs not be formed explicitly. Perceptron concept is given below

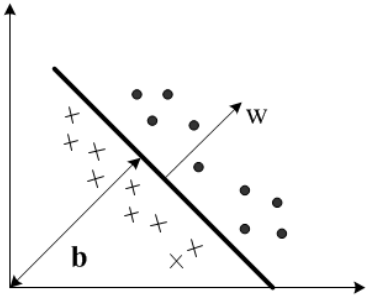


Fig. : Perceptron

Here, the linear separation or border decision classification of input space is calculated by below

$$f(x) = w^T x + b \quad ()$$

$$h(x) = \text{sign}(f(x)) \quad ()$$

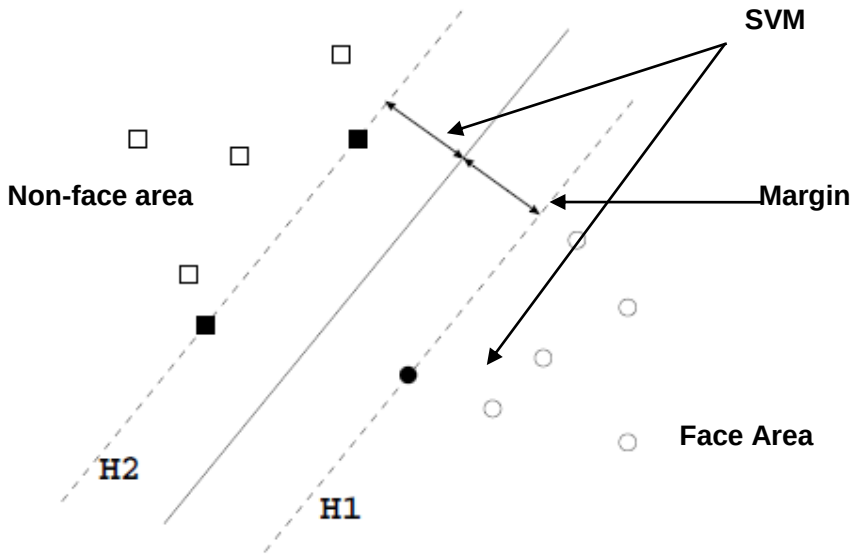


Fig. : SVM border margin

Black rectangles and circles show support vectors and the line between them is support vector margin.

H1 area $w^T x - h = 1$, and H2 area $w^T x - h = -1$ are perfectly separated.

Between H1 and H2, there is no face sample.

If $w^T x - h \geq 1$, then the sample is behaved as face and if $w^T x - h \leq 1$, then it is behaved as non-face.

Separation plane and hyperplane distance is $1/\|w\|$

Then, the calculation of the w and h are done using the following control equation (61) and (62)

$$t_i(w^T x_i - h) \geq 1, (i = 1, \ldots\ldots, N) \quad ()$$

$$\text{L(w)} = \frac{1}{2}\|w\| \quad ()$$

4.6.1 Soft margin hyperplane

If we minimize $\sum \delta_i$, δ_i can be calculated by the following equation

$$\begin{cases} w^T x_i + b \geq 1 - \delta_i \\ w^T x_i + b \leq -1 + \delta \\ \delta_i \geq O_i \end{cases} \quad \begin{cases} y_i = 1 \\ y_i = -1 \\ \forall_i \end{cases} \qquad ()$$

δ_i is slack variable and it must be noted that there is no error for x_i when δ_i becomes zero.

For the minimization of $\sum \delta_i$, δ_i

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \delta_i \qquad ()$$

We obtain the following representation

$$\text{Min}\left[\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \delta_i\right], \quad y_i(w^T x_i + b) \geq 1 - \delta_i \qquad ()$$

where C is the trade-off parameter between error and margin.

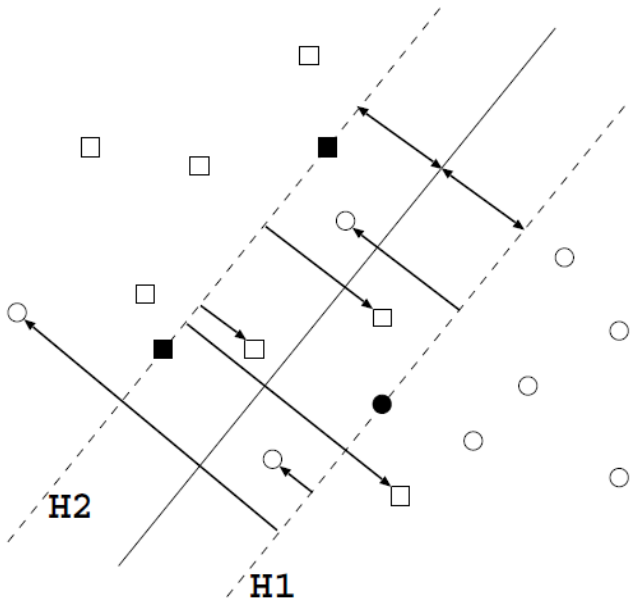


Fig. : SVM soft margin

(○ is class 1 sample,□ is class -1 sample , ■ and ● are support vector)

Some methods are needed for using support vector machine to solve optimization problem (OP) One of those methods is to compute the optimization error. This calculation is known as soft margin. The dual of this new constrained optimization error is done by the following equation

$$\text{Max}(\text{W}(\propto)) = \sum \propto_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \propto_i \propto_j y_i y_j {x_i}^T x_j \quad ()$$

where $\text{C} \geq \propto_i \geq 0, \sum_{i=1}^{n} \propto_i, y_i = 0$

Here, w can be obtained using the following equation

$$w = \sum_{j=1}^{s} \propto t_j \times yt_j \times xt_j \quad ()$$

To make matrix calculation during the optimization problem, we define the matrix Q as

$$\varrho_{i,j} = y_i y_j y x_i x_i \quad ()$$

The size of the optimization problem depends on the number of training examples ℓ

If the size of the matrix Q is ℓ^2 for learning tasks with more than 10,000 training examples it becomes impossible to keep Q in memory. The implementations of QP solvers require explicit storage of Q, which prohibits their application. For example, re-computing Q every time it is needed is one solution. However, this is computationally expensive if Q is needed more often such as face detection for every camera frame. Hence, the calculation of Q every time is not practical. One approach is to decompose the problem into a series of smaller tasks. Our decomposition splits OP1 into an inactive and an active part.

One disadvantage of this method is the longer learning time. For solving the problem, following ideas are used

1. An efficient and effective method for selecting the working set.
2. Successive “shrinking" of the optimization problem. This exploits the property of many support vector learning problems.
3. Computational improvements like caching and incremental updates of the gradient and the termination criteria.

4.6.2 QP problem solving

The algorithm decomposes OP1 and solves the smaller QP-problem. The decomposition assures that this will lead to progress in the objective function W(∝), if the working set B fulfills some requirements of optimization problem solving .With conjunction of this, OP1 is decomposed by separating the variables in the working set B from those which are fixed. Let us assume ∝ and y and Q are well arranged with respect to B and N, so that

$$\propto = \begin{vmatrix} \propto_B \\ \propto_N \end{vmatrix},\ \mathrm{y} = \begin{vmatrix} y_B \\ y_N \end{vmatrix},\ \varrho = \begin{vmatrix} \varrho_{BB} & \varrho_{BN} \\ \varrho_{NB} & \varrho_{NN} \end{vmatrix} \qquad ()$$

Since $\varrho_{N,B}$ is symmetric with respect to $\varrho_{B,N}$, OP2 minimization is calculated as

$$\mathrm{w}(\propto) = -\propto_B^T \left(1 - \varrho_{B,N} \propto_N\right) + \frac{1}{2} \propto_B^T \varrho_{BB} \propto_B + \frac{1}{2} \propto_N^T \varrho_{NN} \propto_N - \propto_N^T 1 \qquad ()$$

where, we assume that

$$\propto_B^T y_B + \propto_N^T y_N = 0 \qquad ()$$

Eq. 71 is obtained by the symmetric optimization.

Since the variables are fixed on N, the terms, $\frac{1}{2} \propto_N^T \varrho_{NN} \propto_N$ and $-\propto_N^T 1$ are constant. The terms can be omitted without using OP2. OP2 is positive semi definitive quadratic problem. To reduce the computational time during the calculation of OP2, we shrink the size of OP1.

During the optimization process, it is clear that certain examples are unlikely to end up as support vectors. By eliminating these variables from OP1, we get a smaller problem OP1' of size ℓ'. From OP1' we can construct the solution of OP1. Let X denote those indices corresponding to support vectors, Y those indexes which correspond to negative support vectors, and Z the indices of non-support vectors. The transformation from OP1 to OP1' can be done by using decomposition. Let us assume that α , y, and Q are properly arranged with respect to x, y and z, then the following decomposition equation is satisfied for ∝.

$$\mathrm{Min}\left(\mathrm{W}(\propto_x)\right) = -\propto_x^T \left(1 - \left(\varrho_{xy} 1\right) \times C\right) + \frac{1}{2} \propto_x^T \varrho_{xy} \propto_x + \frac{1}{2} C 1^T \varrho_{yy} C 1 - |y| \times C \qquad ()$$

Here, $\propto_x, \propto_y, \propto_z$ are the matrix indices of α and those indices are equal to $\propto_x$. Similarly, y_x, y_y, y_z, are matrix indices for y and $\begin{pmatrix} \varrho_{xx} & \varrho_{xy} & \varrho_{xz} \\ \varrho_{yx} & \varrho_{yy} & \varrho_{yz} \\ \varrho_{zx} & \varrho_{zy} & \varrho_{zz} \end{pmatrix}$ are for Q. During the calculation of (72),

$$\frac{1}{2} \propto y^T \varrho_{yy} - |y| \times C = \frac{1}{2} C1^T \varrho_{yy} C1 - |y| \times C \quad ()$$

is taken as zero since it is a constant variable by assuming that $\propto_x^T y_x + C1^T yy = 0$ and $0 \leq \propto_x \leq C1$.

Some of these variables of (73) are eliminated using the decomposition as shown above. In another meaning, these variables are fixed (eq. 73). Neither the gradient, nor the optimality conditions are computed. This leads to a substantial reduction in the number of kernel evaluations.

To better explain, let us suppose we have 5 one-dimensional datasets {x1, x2, x3, x4, x5} = {1, 2, 4, 5, 6} with

$$\begin{bmatrix} class1 \\ class2 \end{bmatrix} = \begin{bmatrix} 1,2,6 \\ 4,5 \end{bmatrix} = (y_1 = y_2 = y_5) = 1, (\, y_3 = y_4) = -1$$

Using a polynomial kernel of degree 2, $K(x, y) = (xy + 1)^2$ and C=100

We evaluate $\propto_i$ for its 1~5 indices using the following

$$\text{Max} \sum_{i=1}^{5} \propto_i - \frac{1}{2} \sum_{i=1}^{5} \sum_{j=1}^{5} \propto_i \propto_j y_i y_j (x_i x_j + 1)^2 \quad ()$$

where $0 \leq \propto_i \leq 100$, $\sum_{i=1}^{5} \propto_i y_i = 0$

By solving QP, we get

$$\begin{bmatrix} \propto_1 \\ \propto_2 \\ \propto_3 \\ \propto_4 \\ \propto_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 2.5 \\ 0 \\ 7.333 \\ 4.833 \end{bmatrix}$$

The support vectors are {x2,x4,x5} = {2,5,6}. The discriminant function based on this,

$f(z)=2.5(1)(2z+1)^2+7.333(-1)(5z+1)^2+4.833(1)(6z+1)^2+b = 0.6667z^2 – 5.333z + b.$

b is obtained by resolving f(2)=1, f(5) = -1, f(6) = 1 as x2 and x5 lie on the +1 border line and x4 lies on -1 border line. Then combining all those together, we get b=9. Replacing b=9 at above calculation, we get $f(z) = 0.6667z^2 – 5.333z + 9$

4.7 Summary

In this chapter, two different algorithms, SGBM algorithm and hierarchical nonlinear PCA (H-NPCA) have been proposed. In addition to this, we gave some explanation on SVM computation. The adaptive Gabor filter algorithm is used to extract the face features. The Gabor filters are applied to the specific locations of the face based on eyes, nose and mouth. The extracted features are further analyzed and very important features are selected by using genetic algorithm. The selected Gabor features are further enhanced in the nonlinear kernel using H-NPCA. H-NPCA is an unsupervised nonlinear data analysis method which consists of neural network topology. Neural network is used for estimating the missing data. Furthermore, SVM is used for data classification. The algorithm shows advantageous in computation cost and efficiency.

CHAPTER 5: THE DEVELOPED PUBLIC SURVEILLANCE SYSTEM

In this chapter, we introduce a surveillance system which works on TCP/IP networking and SQL database. The system runs the proposed methods by using a megapixel IP camera. This chapter is especially prepared to show the efficiency of the proposed methods in real world conditions. The first section gives a brief explanation. The second section explains about the overall architecture of the system with system components. Following this, the camera image acquisition is given in detail and explanation on SQL database is given. Afterwards, an important parameter N:N matching method and its importance in surveillance system are given detail. The explanations on how the records are done, how the network infrastructure works are also given within the chapter. Finally, the parallel processing is shortly given. Parallel processing is designed for Intel CPUs. Its structure and the multicore CPU usage and memory allocations are explained and the explanation of N:N by using parallel processing is also done here. The last section concludes the chapter 7.

5.1 Introduction

Based on the proposed face extraction and recognition algorithms, we created a surveillance system with Sanyo HDD-4600 IP camera. It provides 10fps per second in 1920x1080 full-hd mode. It has 10X optic-zoom in maximum. In surveillance system, the speed and the accuracy are very important as we use the high resolution images like 1920x1080 image pixels. The speed improvement is given in detail at section 5.9.

5.2 System architecture

The system structure is constructed as shown in fig. 43.

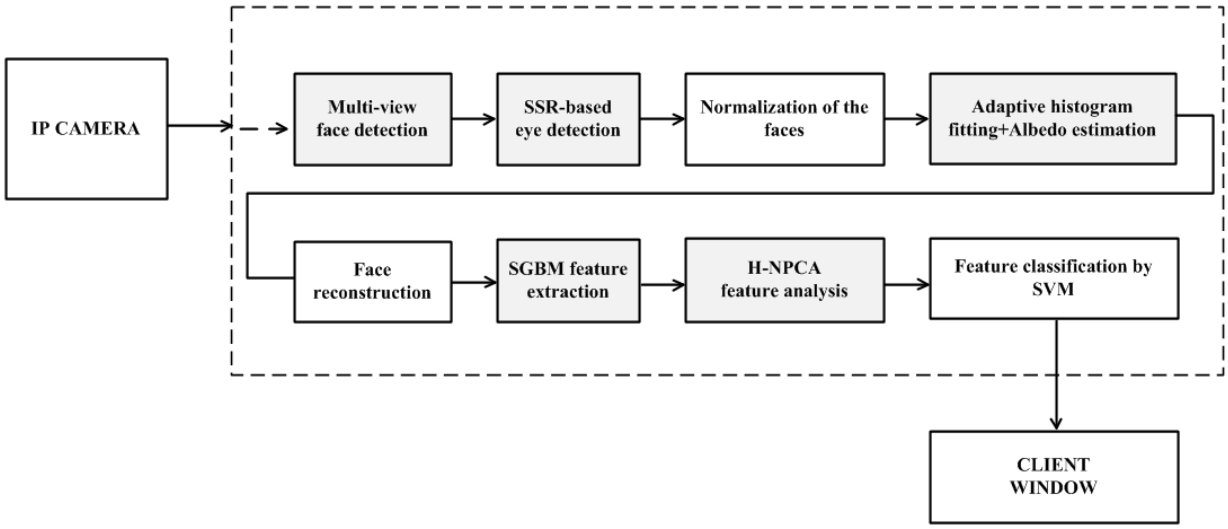


Fig. : Engine structure

- IP camera

The network IP camera provides the continuous frames through TCP/IP network. One jpeg shot is captured from the camera and this input image frame is processed in the system to recognize a face.

Camera used here is Sanyo HD-4600P model which provides Full-HD image. It provides 10 frames per second in full-hd mode.

- Enriched haar-feature-based multi-view face detection

Its details are given in chapter 3. We use enriched haar-based face detection. The original haar features cannot run in real time if full-hd images are considered. We developed enriched haar features which significantly increase the detection performance. Our haar features are implemented in a multi-core platform. It runs with a speed of 10fps in FULL-HD images. The original haar-features run at a speed of 1fps.

- SSR-based eye detection

SSR-based eye detection focuses on the eye region. It uses 6 different black-white filters. The computation is done by integral image which is a high-speed computation method. The eye region is detected and then middle of iris is estimated by several assumption-based calculations.

- Normalization of the faces

After eyes and borders of face are found, the face is normalized by using the eye and face coordinates. The normalization is done by considering a triangle relation of two eyes, nose and mouth. Face alignment is important for the accuracy of the face recognition. The details of the normalization are given in chapter 3 in detail.

- Adaptive histogram fitting and albedo estimation

In outdoor conditions, the illumination changes rapidly and neither its direction nor its strength is controllable. We compute the illumination direction and make some estimation for the strength of the light before extracting the face features. The light direction estimation is done based on eight pre-determined directions. Depending on which direction the light is coming, several solutions are possible. We estimate the albedo by considering the face object as lambertian object. For the estimation, Gabor-based approach is used.
During the light source estimation and albedo estimation, we use several image-processing methods. One is histogram analysis. The face is separated into several parts and each part is processed by fitting the histogram. The fitting model is obtained by a precompiled illumination face database.

- Face reconstruction

After illumination is normalized for each part of the face, the problem is how to reconstruct a face image by keeping the correlation of the parts. Direct linear interpolation is used here.

- SGBM feature extraction

Face is well aligned, illumination is corrected, noise factors are removed, and image is smoothed. Now, face features must be extracted and analyzed. Extraction is done by SGBM. If the filters are applied randomly, some other features are extracted as face features such as part of the background, noise etc. We compute the most effective face points by using a generic algorithm and apply SGBM to those points. This increases our speed and recognition performance. The details are given in chapter 4.

- H-NPCA feature analysis

Now, the features are extracted and it is important to choose the most meaningful features and compensate the missing data. We use H-NPCA feature analysis for this purpuose. It is a nonlinear feature analysis method. It uses auto-associate neural network architecture for missing data estimation. Feature matrices are created and dimension reductions are also done during H-NPCA.

- Feature classification by SVM

SVM is a dynamic feature classifier. It has a hyperplane. Features are grouped on a hyperplane. Features near the line are considered the most similar faces. The decision border is decided by a threshold that is adjustable by software.

- Client window

It is the application window which shows the recognition alerts, detected faces, and enrolled people. The parameters of the algorithms such as recognition threshold, minimum face size, comparison method, minimum face neighborhood, classification matrix selection are also adjusted by the client window.

5.3 Image acquisition from IP surveillance cameras

Camera used here is a megapixel IP camera from Sanyo. Its model is HD-4600P. Its image acquisition and automatic background compensation are available. It has both day and night function. During the night, it provides IR images and during the daytime, it gives RGB color images. The acquisition speed is 10fps. Supported image resolutions are 1280x960, 1600x1200 and 1920x1080. The zoom function supports up to 10X optical zoom. It has also face light compensation. However, this function distorts the face in outdoor. Therefore, we have not used it during our experiments. We used 1920x1080 resolution and a fix iris value of 0.45.

5.4 DB enrollment and feature ID creation

During enrollment, a console is used. Console reads from image folders or from IP camera directly. Enrollment is done by using one image or multiple images. If one image is used for enrollment, internally 10 images are synthesized. Synthesizing is done by resizing the face, shifting by 5%-10%, tilt rotation. The record size is 10K per person. Each record is saved as an independent file.

5.5 Real-time N:N recognition in outdoor

Recognition is done by still images from major face recognition databases or Sanyo IP camera. If the input is set to “still image”, jpg, bmp, gif, tif and png formats are supported. The input image is processed and the extracted faces are shown in real time. If the input is set to “camera”, the frames are processed by using a time interval. Face tracking is not used. However, simple motion analysis and motion-based tracking are used. Once faces are detected, the locations of each face are recorded by the software and each location is separately tracked. If there are N faces, N motion-based tracking process runs in parallel. The output of the matching is ordered from the highest score to lowest score. A threshold is used to filter out the output. If no threshold value is used, the system lists up all matches.

5.6 User record management

SQL 2008 DB is used. DB saves names of enrolled people, record files, and thumbnail display images. In SQL, user enrollment category information is also kept in addition to personal information.

5.7 TCP/IP network infrastructure

The infrastructure is based on full TCP/IP communication. Server side manages the IP camera or image folder reading. Client console displays the images to be processed, alarms and DB management. A token communication is active between server and client. The captured face images are saved in Server side and these are used for enrollment.

5.8 System modules

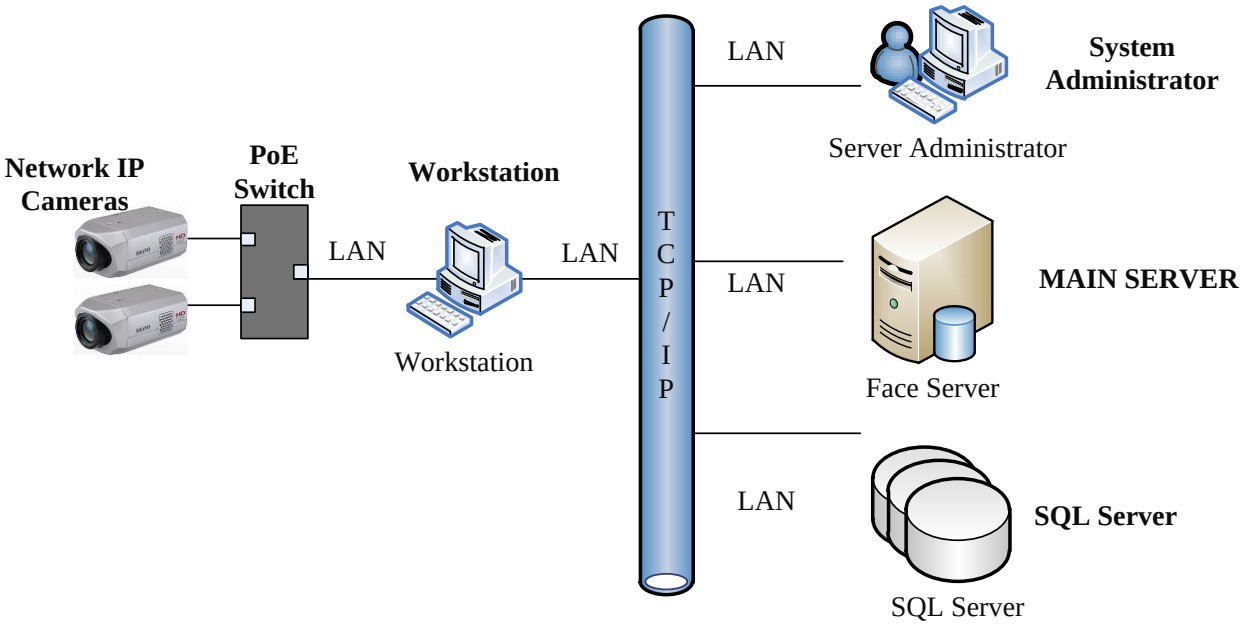


Fig. : Face surveillance system

Fig. 44 shows a face surveillance system we developed based on the method we proposed in this thesis. Power-on-Ethernet (POE) network switch is a switch, which provides 48volts for camera power. Cameras do not need any external power supply. Only LAN connection is done.

Workstation is a high-computing machine, which has Intel i7 processor. It has also nvidia graphic board and 10GB RAM. Since processing Full-HD images from two cameras at the same time needs high computation power, we need a high spec server machine. We use parallel processing. Intel i7 has 8 threads. Face detection and eye detection, tracking are separated into these threads equally. DB comparison is done in GPU. Graphic board provides an independent computation. Increasing the number of records in DB decreases the overall speed. To keep it stable, we pass DB comparison to GPU since GPU runs completely independent from the processor.

Client console manages DB records. It adds new records, modifies existing records, delete the existing records or add records in batch. SQL Server and client console can be installed to the different machines. However, we installed them to the same PC. Server PC and client console can also be installed to the same machine. We separated them during our evaluation to prevent the performance loss.

5.9 Parallel processing for N:N matching

As we propose a real-time surveillance system in outdoor, it is important to process the faces simultaneously. We used the machine specs given in table 8.

Table : Specifications of the PC machine

OS	Windows 7 home premium 64bit
PC Model	HP Pavilion Elite HPE
CPU	Intel Pentium® Core™ i7 930 © 2.8GHZ
RAM	9GB
GPU	Nvidia 320

Furthermore, we set the minimum face size to 50x50 pixels.

Camera resolution is 1920x1080.

In this condition, we did job sharing by doing parallel programming.

Parallel processing is a processor-dependent work. We used the maximum number of cores and threads. In Pentium i7, there are 4 cores and 8 threads and allocated RAM area is 5GB as seen in fig. 45.

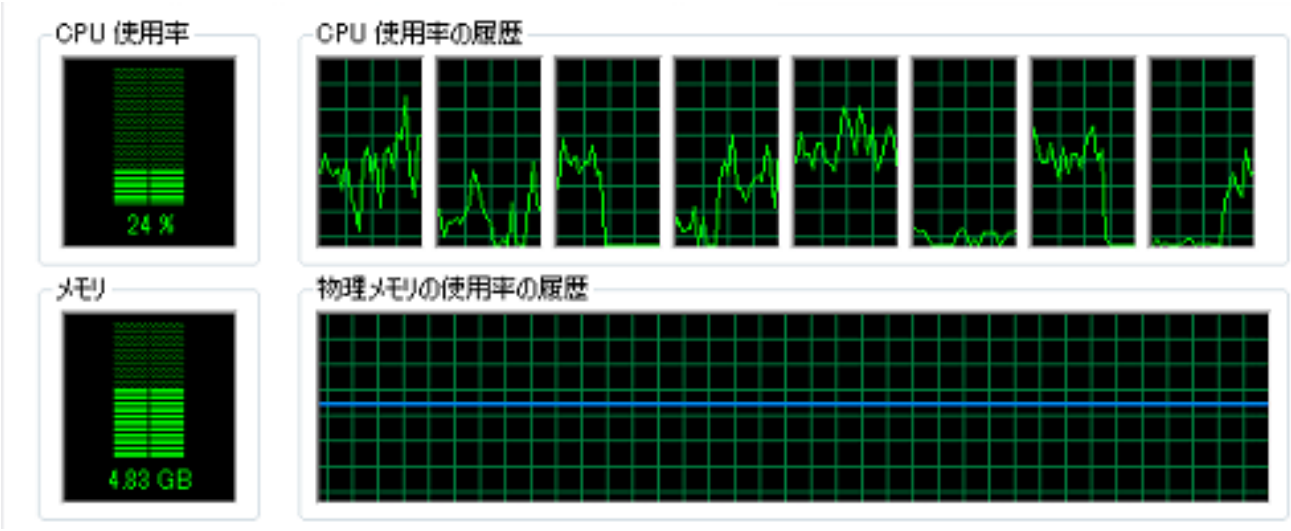


Fig. : Server CPU allocation

We assigned the each weak classifier to separate threads and kept one thread empty to compute the outputs from the other threads. This process made the face detection 10 times faster. Furthermore, we designed the program to equally distribute the number of detected faces to the threads. For example, if there are 8 faces found in the camera image, we processed each face in a separate CPU thread. The feature extraction is called separately by each thread in random order. In ordinary systems, this is done in one thread. Therefore, the program repeats the same process 8 times. In our case, equally distributing the works to the threads makes the overall evaluation finish quickly.

5.10 Summary

We introduced face detection algorithm, eye detection method, feature extraction algorithm and classification algorithms in this thesis and finally we joint up all these components and set up a surveillance application to better see the overall performance in real world conditions.

As the input device, Sanyo model megapixel network IP camera is used during the evaluation. The structure is done in server-client framework. Server captures the camera images, process them and the client browses the images, results and displays the alarms.

The processing server machine contains multi-core CPU. Therefore, we parallelize the work by using hardware power to further speed up the recognition. The recognition is done in N:N mode. The first N means the number of detected faces and the second N means the number of records in the database.

An SQL database is used for saving the records. Each record size is 10KB per person. Network infrastructure uses fully TCP/IP protocols. Hence, all the communications between camera and client, server to client, server to SQL database are done by packet communication.

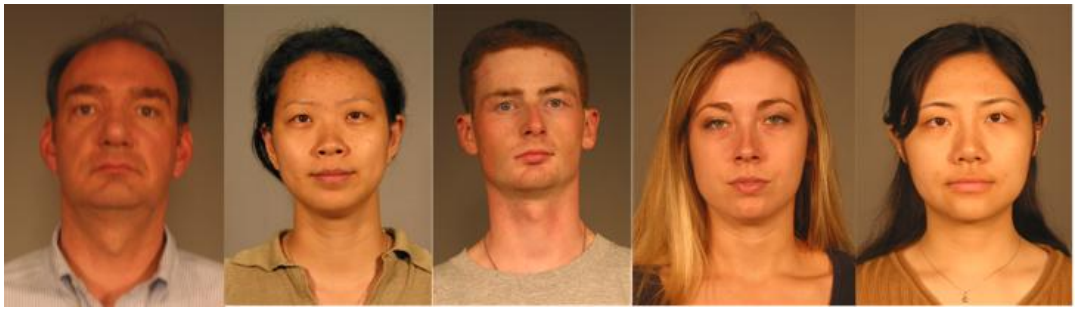
During enrollment, either 1 image or 10 images are used. The performance values depending on the number of images recorded are given in chapter 6 in detail.

CHAPTER 6: PERFORMANCE EVALUATION OF THE PROPOSED METHODS

We conducted the experiments under variations in pose and illumination by using images from different databases as well as real-time images from an IP camera. The primary goal of our experiments is to test the illumination technique introduced in chapter 2. The second goal of the experiment is to test the performance of the proposed nonlinear feature analysis which is introduced in chapter 3.Our proposed technique is compared with several state-of-the-art face recognition algorithms. Each method is trained by ten images per individual and each image has an attribute of frontal, small-illumination variations and small angle variations.

6.1 Environment and image selection

We setup the environment from still images and IP camera. Environment from still images is selected from FRGC-DB [71], CMU-PIE-DB [72] and YaleE-DB [73]. Environment from IP camera consists of live images taken from the outdoor environment. The images from databases are arranged as train and testing folders. Train folders consist of 1 to 10 images per person with a number of 440 people. The specification of these train images is given in table 8. Fig. 46 shows the training images.



(a) FRGC-DB training samples

(b) CMU-PIE-DB training samples

(c) YaleE-DB training samples

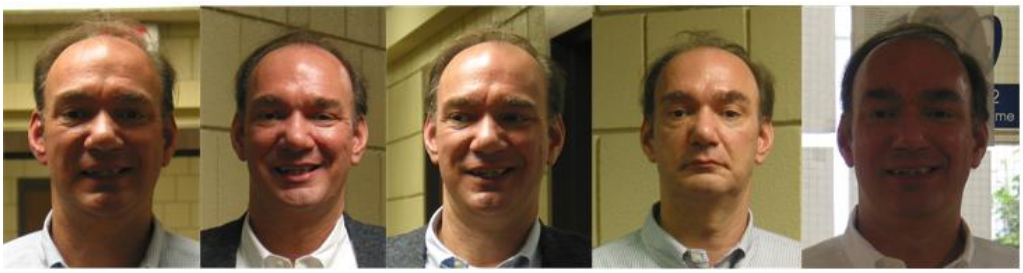
(d) Training samples from live images

Fig. : Train images

Fig. 46(a), (b), (c) shows some samples of train images, fig. 46 (d) shows images from live

images (IPCAM). Fig. 47 (a), (b), (c) shows samples of test images from each database and fig. 47 (d) shows the testing images from live images.

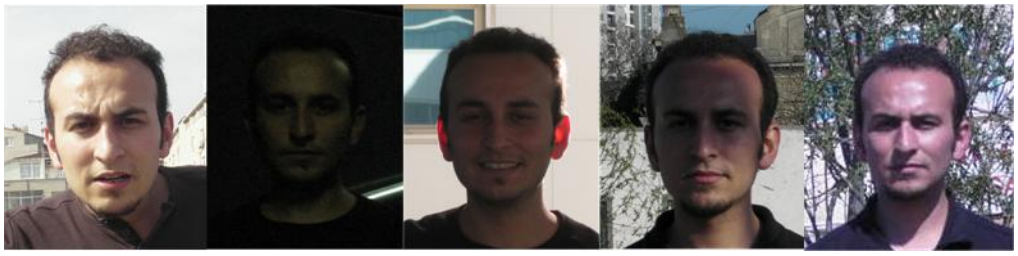
More than 10 images per person are selected for testing. The testing images are selected so that they have different illumination and face expression.



(a) FRGC-DB test images

(b) CMU-PIE-DB test images

(c) YaleE-DB test images



(d) Test samples from live images

Fig. : Test images

In fig. 47, test examples are given. The basic idea is that a person is enrolled in one environment and should be recognized from other environments. Therefore, test images are selected with large variations of illumination.

In case of live images, one image or multiple images of a person is enrolled in one outdoor environment and another camera recognizes the person in a different location in outdoor. In addition to that, the same camera is used to recognize the person in different periods in a sunny day such as morning hours, noon hours, evening hours, and night hours.

Table : Specifications of training and testing images

TRAINING IMAGES					
Frontal	Illumination	Pose	Occlusion	Expression	Age/Gender/Race
Yes	Neutral	± 5	No	Neutral	Any
TESTING IMAGES					
Yes	Severe	± 10	No	Yes	Any

In table 9, we list the specification of training and testing images. Training images must be frontal, well illuminated with no expression. The images can have small angle within five degree from upright direction. Age, gender does not matter. In terms of race, we worked with western and Asian faces. African faces and Latin faces may differ.

6.2 The points for database selection

We used FRGC-DB, CMU-PIE and YaleE-DB. FRGC-DB and Yale-DB has images taken both indoor and outdoor. CMU-PIE DB has indoor images. It has plenty of lighting and small poses changes with small face expressions such as smiling, closed-eyes etc. The databases we are using are the face recognition grand challenge (FRGC) database [71], CMU-PIE [72] and Yale extended database (YaleE-DB) [73].

The FRGC consists of a single controlled still image of a person and each test consists of a single controlled still image. In addition, it studies the effect of using multiple still images of a person on performance. Each biometric sample consists of the four controlled images of a person taken

in a subject session. For example, the train set is composed of four images of each person where all the images are taken in the same subject session.

The FRGC consists of 50,000 recordings divided into training and validation partitions. The data contains high-resolution still images taken under controlled lighting conditions and with unstructured illumination, 3D scans, and contemporaneously collected still images.

CMU-PIE database provides images of both pose and illumination variations. It includes 41368 face images of 68 subjects divided in two major partitions, the first with pose and expression variation only, the second with pose and illumination variation. We use the images from the second partition.

YaleE-DB has various illumination images with various poses. YaleE-DB consists of 16128 images of 40 human subjects under 9 poses and 64 illuminations per pose. We used illumination range of -40 to +90. Numbers "-40, "90" mean the illumination direction.

CMU-PIE-DB and Yale-DB are not properly aligned and the ground truth data is not available. Therefore, we marked all the images from these databases manually before using them for training and testing. This step is necessary and important since specific feature points are necessary for us.

FRGC-DB has also good images, which are illuminated by different light sources from different directions. FRGC-DB has various images taken from different seasons and different environments. However, the conditions are very limited although the images are enough for a controlled environment. Since our objective is to measure the effectiveness of the proposed method, we selected a limited number of images, which are very similar to real world conditions. We used 1 to 10 images per enrollment and used a large number of images per person to check false acceptance rates and false rejection rates. In addition to tests with various databases, we conducted comparison tests with the other state-of-the-art face recognition algorithms.

6.3 Results on various databases

We first conducted a test to check the effect of the illumination. As the title of the paper shows, we needed to prove that the proposed method was illumination-invariant. We selected a number of 25 images from each database. Ten images were used for training and rest of them was used for the testing. Each face image has random illumination. Random illumination means that

illumination is not controlled. The left part of some faces was affected and some of them were affected from the top, down, left-up, right-up, left-down, right-down. We did not use the same illumination conditions during our testing.

6.3.1 The results on FRGC face database

(a) Enrollment image (b) Test image set

Fig. : FRGC image set

As in FRGC-DB in fig. 48, we used 10 images per person during enrollment. We enrolled 100 people. Total 1000 images were enrolled. The testing was done by using different illumination images of the same person.

Table : Recognition rate for FRGC-DB

Database	Number enrollment	# of image/person	Equal error rate
FRGC-DB	10	10	1.6%
	20	10	1.8%
	50	10	2.2%
	100	10	3.8%

Table 10 shows the details of the enrollment and number of images per person. It also shows the equal error rate. Equal error rate is the error rate where false acceptance rate and false rejection

rate becomes equal. If we use 10 images per person and if there is 10 records in the database the error rate is 1.6%. The error rate becomes high if the number of enrolled people increase. For example, if the records increase to 100, the error ratio becomes 3.8%.

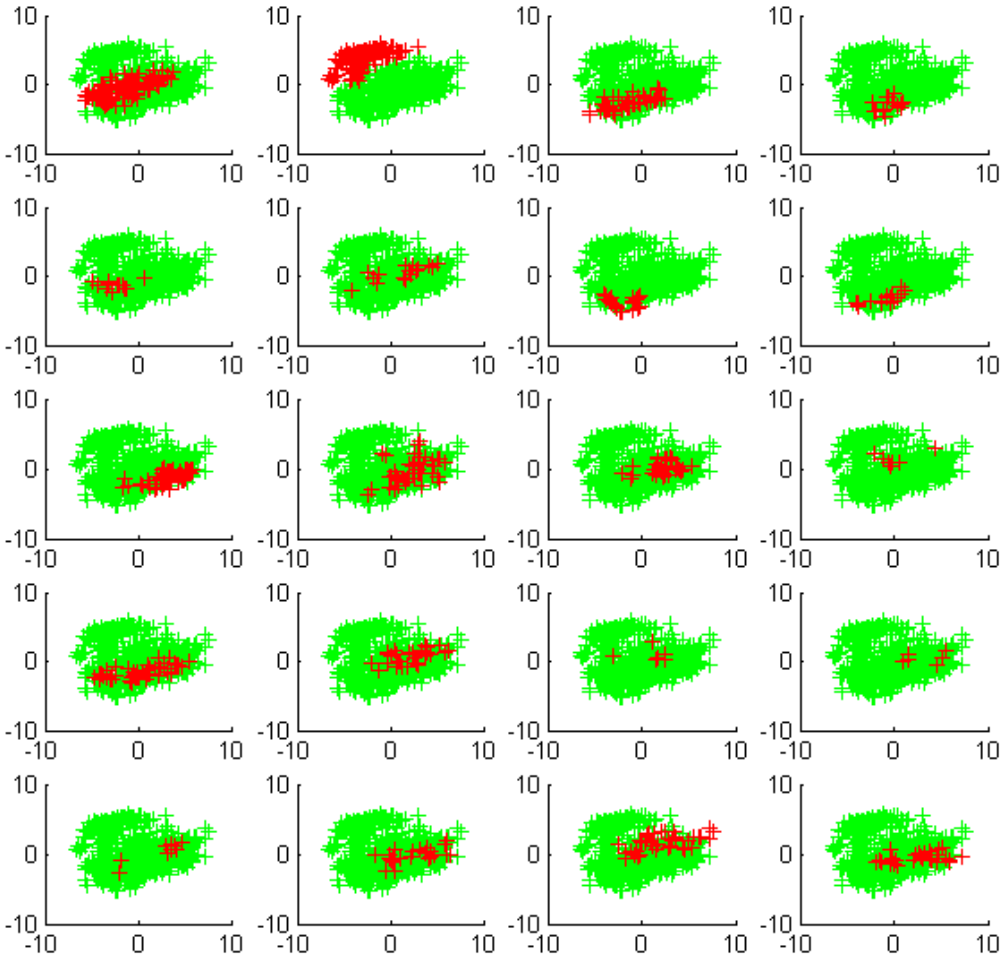


Fig. : Feature group for the first 20 people.

Fig. 49 shows the feature group for 20 people tested in FRGC-DB. The green + marks are unknown people. Red + marks are the recognized person. In addition to performance testing of the methodology, we also conducted a comparison testing with other methodologies. The results were given in fig. 50.

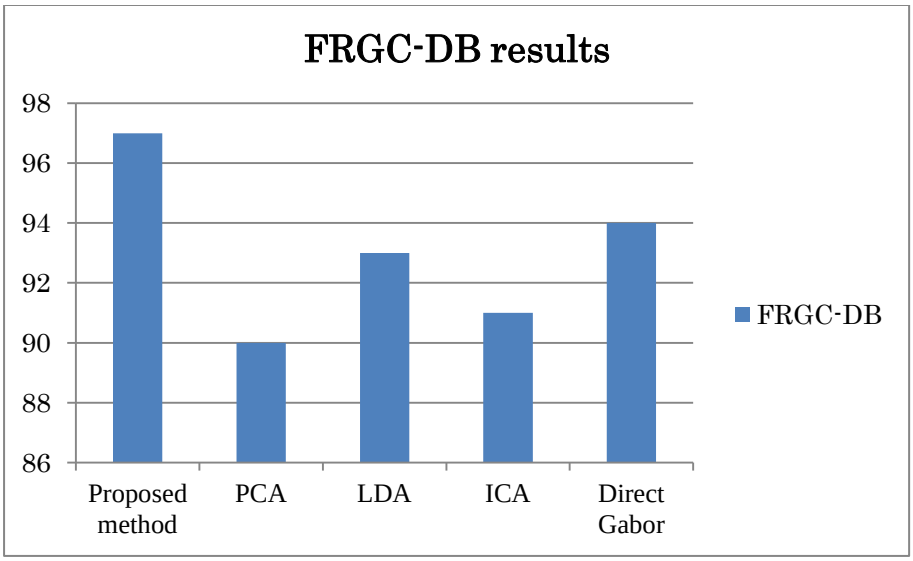


Fig. : Comparison results with other methodologies

6.3.2 The results on CMU-PIE database

(a) Enrollment image (b) Test image set

Fig. : CMU-PIE image set

In fig. 51, there are some images that are used from CMU-PIE. CMU-PIE contains many different illumination images from different directions. For each person, there are 13 different poses, 43 different illuminations, and 5 different expressions. There are 13 cameras used for the creation of database. We used almost frontal images with small angle variation from the upright position.

During training, we enrolled 100 people and we enrolled 1 image per person and 10 images per person. Then we observed the performance.

Table : Recognition rate for CMU-PIE

Database	# of enrollments	# of image/person	Equal error rate
CMU-PIE	10	1	6.0%
	100	1	9.0%
	10	10	2.0%
	100	10	2.8%

As seen from the results of table 11, when we used one image per person, the error rate becomes high. When we enrolled multiple images per person, the error rates significantly improve. As the number of enrollment increases in database, the similarity distance of the individuals becomes closer. This causes error rates an increment. When using one image per person, it is difficult to regenerate other features from that. A small change in face pose or camera position affects the error rate significantly. This change becomes lower as we increase the number of images. If small pose changes are included to the database during enrollment, the error rates further improves.

6.3.3 The results on YALE-DB face database

(a) Enrollment image (b) Test image set

Fig. : YaleE-DB test set for illumination performance.

As seen in fig. 52, YaleE-DB is very challenging since it includes very dark and strong partial lightings. The light is applied to the face from different directions. Cast-shadow also occurs with illumination from left up, left-down, right-up, right-down. We tested the methodology with YaleE-DB by enrolling 40 individuals and the results are given in table 12.

Table : Recognition rate for YaleE-DB

Database	# of enrollments	# of image/person	Equal error rate
YaleE-DB	10	1	10.0%
	40	1	12.0%
	10	10	3.0%
	40	10	3.5%

6.3.4 The results on live images

(a) Enrollment image (b) Test image set

Fig. : Live image test set

Fig. 53 shows the actual camera images. Live images are acquired from IP camera. The testing is done in two ways: enrolling one image / person and 10 images / person. The images for testing contain various poses and illumination. In addition to this, the image quality is different depending on the environment.

6.3.5 More results on surveillance images

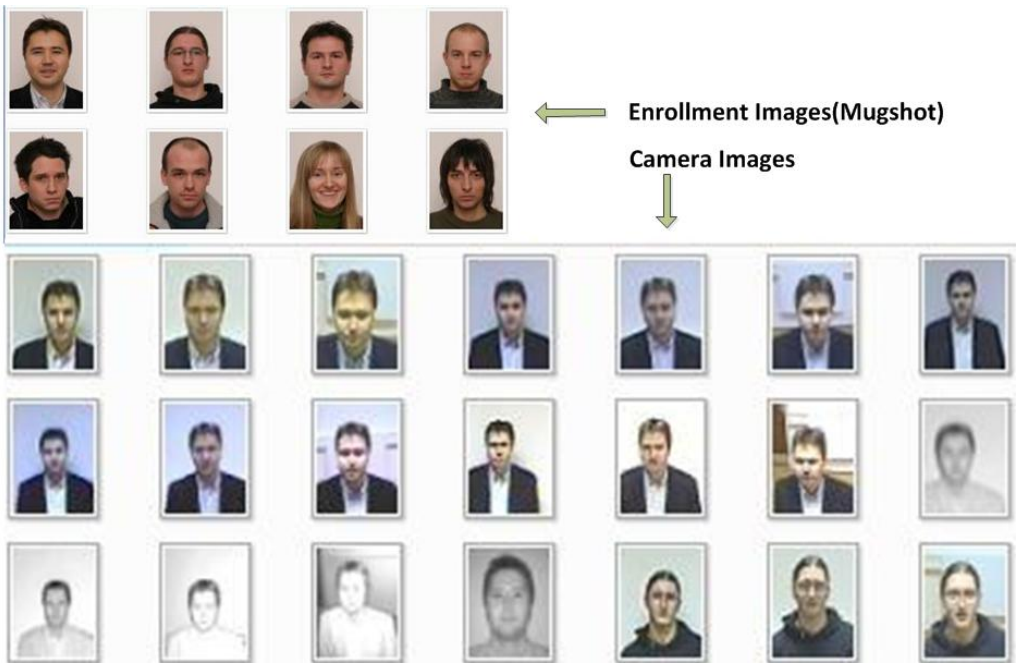


Fig. : Enrollment and recognition image set in a surveillance system

In fig. 54, enrollment is done by using one mug-shot per person and matching is done from IP camera.

Fig. : Image from the proposed surveillance system

In fig.55, recognition results are given. Left screen shows the detected faces. Right screen shows

two matched results (green rectangles). Fig. 56 shows the detected face images.

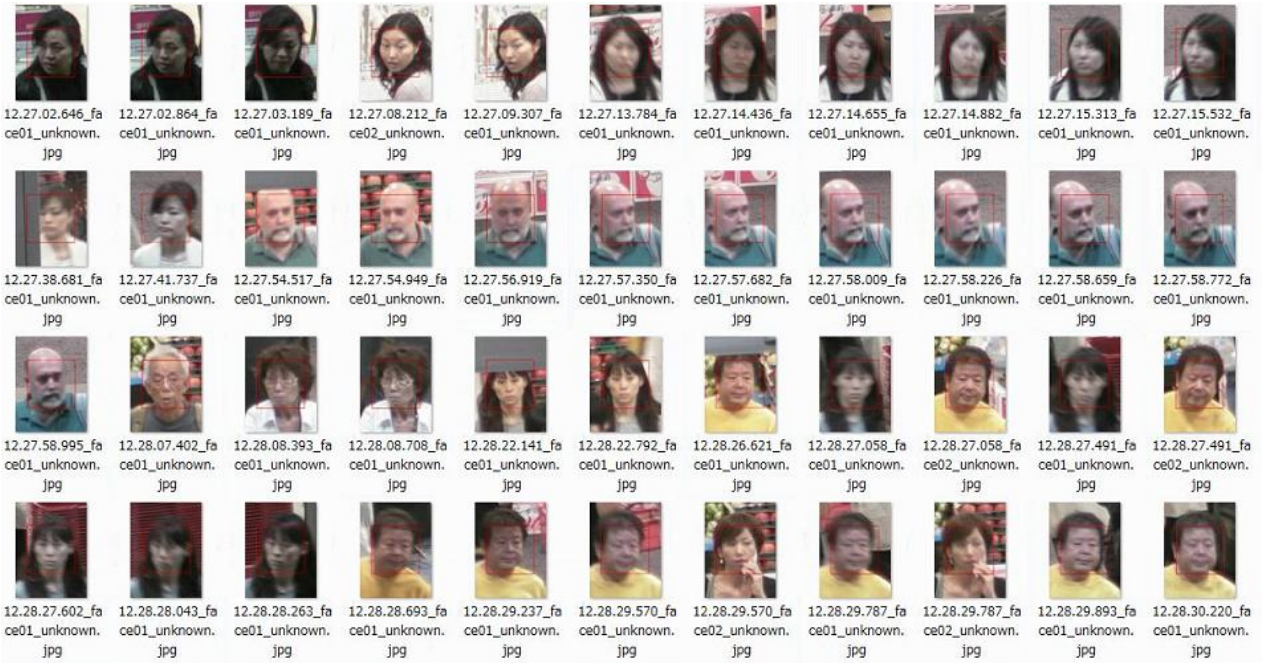


Fig. : Detected faces

The detected faces are all saved automatically. If a face is detected and recognized, the face is marked in green and the person name is shown. Further results and error rates are given in table 13.

Table : Recognition rate for live image

Input	# of enrollments	# of image/person	Equal error rate
Live image from IP camera	10	1	3.0%
	100	1	4.0%
	1000	1	4.5%
	10	10	1.0%
	100	10	1.4%
	1000	10	1.8%

During the daytime, the images are captured in color. When it is night-time, IR images are used. The error rate increases if one image is used per person. If the number of images is more than 1, the error ratio becomes lower. We used 1 image and 10 images per person during the evaluation

6.3.6 The overall results

During the enrollment, we first enrolled one image per person. However, this did not give good performance. We increased the number of images one by one up to 10 images. In different papers, it is reported that the more pose variations of images is used, the more the performance of the algorithm increases. Following results are taken when we used five images:

Table : Recognition rate

Name of database	# of images	# of train	Rec. ratio	Error rate
FRGC-DB	2000	5	96%	2.1%
CMU-PIE	1000	5	94%	4.3%
YaleE-DB	1000	8	89%	15.2%
AVERAGE	N/A	6	94%	7.2%

Error ratio is the equal error rate where false acceptance rate and false rejection rates are equal.

As seen in table 14, FRGC gave 96% recognition rate with 2% error rate. The best values were obtained in Caltech-DB with 97% recognition rate. In fig. 55, there are face recognition sample images from each database.

6.4 Comparison with other methodologies

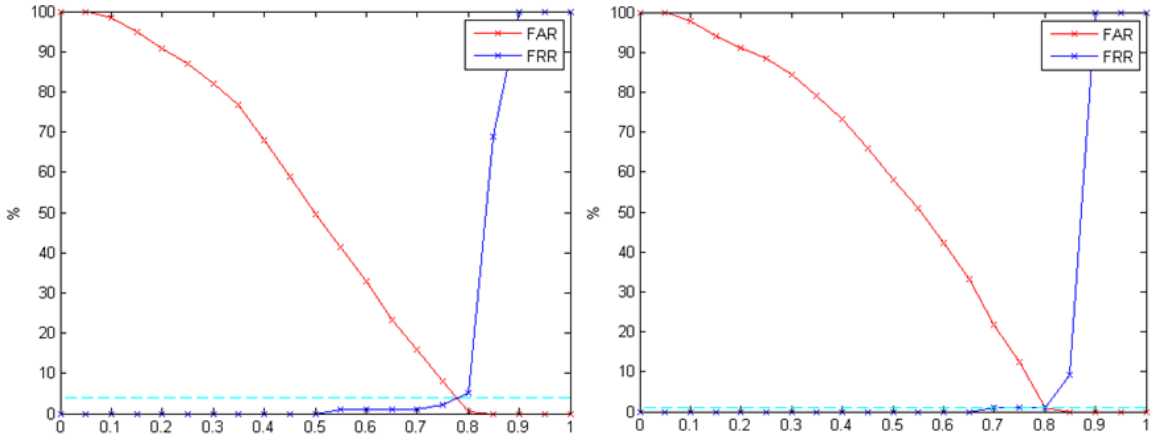
We compared the proposed method by using Gabor wavelet, PCA and ICA algorithms and the results are given in table 15. We used FRGC-DB for the comparison. We used three subset images. Subset 1 consists of well-illuminated frontal faces with neutral expression. Subset 2 images are well-illuminated frontal faces with various face expression. Subset 3 images consist of frontal faces with neutral expression. However, the images are taken in uncontrolled environments. The results of Subset 1 show that the proposed method performs well with the lowest error rate. Gabor wavelet gave good results compared with PCA and ICA. The proposed method gave the worst error rate in subset 3. During the comparison, one to ten images per person are used for enrollment.

Table : Comparison of other face recognition methodologies by using FRGC-DB

Methodology comparison table						
Method	Subset 1		Subset 2		Subset 3	
	Rec. rate	Error rate	Rec. rate	Error rate	Rec. rate	Error rate
Proposed	**96%**	**2.1%**	**94%**	**4.6%**	**95%**	**7.0%**
Direct Gabor	93%	5.4%	90%	8.4%	92%	6.4%
PCA	90%	6.5%	88%	8.1%	88%	10.1%
ICA	93%	4.3%	90%	5.5%	93%	7.8%

6.5 Effect of the number of enrollment images

The number of images affects the recognition rates. When using one image, the EER is higher than using many images per person. When using one image, the EER rate is 3.83% and EER rate is 1.16% when using more than 2 images.



(a) One image enrollment (EER=3.83%) (b) Many image enrollment (EER=1.16%)

Fig. : Effect of the number of enrollment images

Fig. 57 shows the EER rates. We used FRGC-DB indoor images for getting the graph. Frontal images with small illumination variations are used for enrollment.

The results show that the generation of face features is better when using multiple face images.

6.6 Recognition performance on large datasets

We tested the proposed method on a large dataset and the results are shown in fig. 58.

When we enrolled 100,000 people to the database, EER rate was 6.83%. We enrolled 5 images per person during testing. Enrollment images are selected from FRGC-DB, YaleE-DB, ARFace-DB and Network IP camera. The enrollment images are frontal, no-occlusion, small variations on face expression and pose. A threshold value of 0.4 is used. In outdoor environments, a lower threshold value between 0.2 and 0.4 is used to recognize more people. Lowering threshold increases the false alarms and decreases the false rejections. FAR is 10.2% and FRR is 2.2% at 0.2.

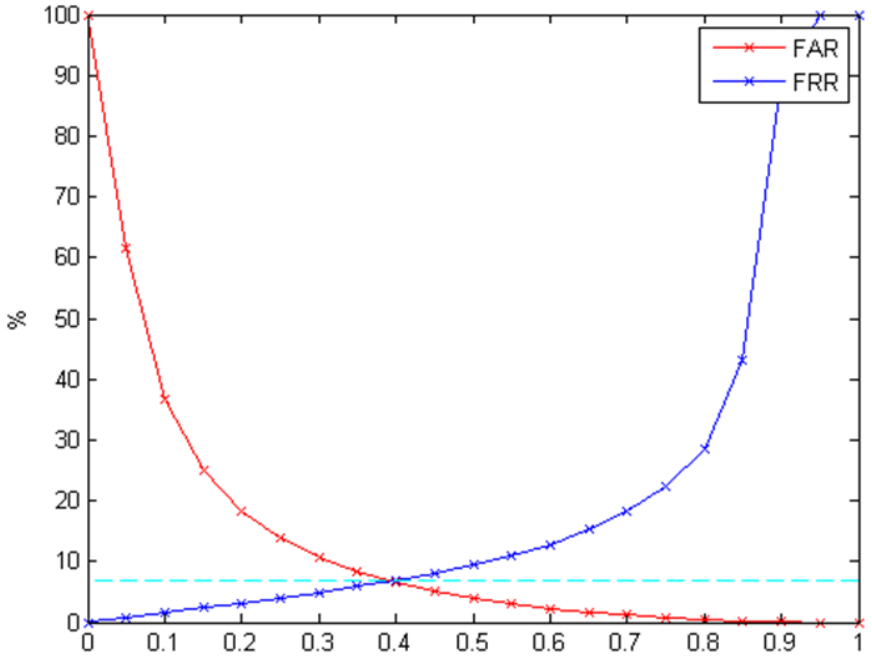


Fig. : Recognition performance on large datasets.

As seen in fig. 58, the best ratio is obtained in the middle of the intersection of FRR and FAR curves. In an uncontrolled environment, the lower then 0.4 threshold value is desirable and in a controlled environment, setting the threshold more than 0.5 is preferred. The recognition rates with respect to threshold value are given in table 16.

Table : FAR-FRR rates with different thresholds on large dataset

Threshold value	FAR	FRR
0.1	36.2%	1.1%
0.2	18.5%	3.6%
0.3	10.0%	5.8%
0.4	6.83%	6.83%
0.5	3.6%	10.3%
0.6	1.3%	13.7%
0.7	0.7%	19.2%

6.7 Summary

In this chapter, we gave the evaluation results of the overall system and algorithms. We used several images from the FRGC, CMU-PIE and Extended Yale-DB databases. In addition to testing from these database images, we also tested the system by using megapixel network Ip camera to see the real world conditions.

We compared our results with state-of-the-art algorithms and confirmed the efficiency of the proposed approached.

The results change when using one image or multiple images for enrollment. If one image is used, the performance is worse than multiple-image enrollments. This is the same as other algorithms.

When increasing the number of people in database, the error rates increase. For example, when we enrolled 100 people, the EER is about 1.16% but when we enrolled 100 thousand people to the database, the EER increased to 6.83%.

CHAPTER 7: CONCLUSIONS AND FUTURE WORKS

In this thesis, we proposed face detection, eye detection, illumination normalization, local feature extraction method and nonlinear feature analysis methods. In addition this, we prepared a network system to test the performance of the methods in actual working conditions. We also tested the methods by using indoor images from major face databases. Finally, we summarize the work and mention about the future studies in this chapter.

7.1 Summary of the works

This thesis presents an approach for face recognition in outdoor platforms. We introduced a complete system of face recognition. First, we introduced face detection by using haar-like features and adaboost learning algorithm. Haar-like features are combined of black and white rectangles.

We introduced a specific haar feature set which significantly increases the performance of the face detection. In our proposed method, it is important to train and test the faces by using the same haar features. After face detection, we introduced an eye detection methodology by using six-segment-filtering (SSR). SSR has six filters. These six filters reflect the eye area characteristics. For example, eyes are the darkest part and nose is the brightest area of a face. By using this assumption, we determine the eyes on a face.

After finding the face on an image, we apply SSR to the face area. The rejection of the wrong eye points are dropped by support vector machine (SVM). To make the classification operation fast,

we introduced a soft-hyperplane in SVM. We used SSR not only for detecting the eyes, but also for the tracking of the face. SSR performs well both in indoor and outdoor if there is no cast-shadow on the region of interest.

Since our purpose is to recognize a face in outdoor, we introduced illumination normalization, face extraction and classification methodologies. The reason to use the illumination normalization is to remove the illumination effects on the face surface and extract the proper features. For the illumination normalization, we introduced a filter "Ayofa-filter".

Ayofa-filter approach computes unknown albedo directions by using spatial frequency components on salient regions of a face. Our method requires only one single image taken under any arbitrary illumination condition where we do not know the light source direction, strength, or the number of light sources. It relies on the spatial frequencies and does not need to use any precompiled face models. Ayofa-filter references the nose tip to evaluate the reflection model that contains six different reflection vectors named as Ayofa-vector.

To further improve the efficiency of Ayofa-filter, we preprocess the face images by a gaussian-based illumination correction method called "adaptive histogram fitting (AHF)". We divide the face into subgroups and each group is independently processed by AHF. Then the face is reconstructed by a linear interpolation technique. The reconstructed face is further processed by Ayofa-filter. Illumination-normalized face images are processed by Gabor wavelets.

We extract the most invariant face features by using SBGM filters. SBGM filters are created by a combination of Gabor filters and extended Bessel functions. SBGM filters are more complex than Gabor filters and they extract the high frequency components of the face by discarding the noise factor during extraction.

After the extraction of the face features, we analyze the features to remove the data reluctance, noises, and irrelevant features by using hierarchical nonlinear principal component analysis (H-NPCA). The missing part of the face due to the occlusions, illuminations, angles and other factors is approximated by using inverse H-NPCA structure. To better separation of the features in subspace, the PCA components are whitened within +1 and -1. The auto-associative network used. Artificial neural networks (ANN), which also helped a lot during the reconstruction of the missing parts from the known components.

Experimental evaluations of the overall method are done by using FRGC-DB, CMU-PIE and YaleE-DB. FRGC-DB has plenty of images per person taken in different periods and different

seasons.

The images are all high resolution. Eye points are colored in some images. These are very useful to test the performance of the eye detection and tracking. However, we did not conduct such tests in this thesis. CMU-PIE has illumination images, different lightened images, face expression images. 13 different cameras take the picture of the individual.

We only used the images, which were captured from an IP camera. YaleE-DB has some pose images under severe illumination. Some images are so heavily affected by illumination that we discarded them from the evaluation. In addition to the testing by using these databases, we also tested the method by using live images from IP camera to see the system performance in real world conditions. In addition to these tests, we compared the method with the other state-of-the-art face recognition techniques to make sure the proposed method outperforms the other face recognition methodologies.

The empirical results showed that our proposed approach is strong enough under uncontrolled environment where the illumination strength and direction are all uncontrolled.

We focused on the face recognition under various illuminations for this thesis. However, there are other factors to be considered in real world conditions such as face pose and face expression. Our feature extraction and analysis method compensate some degree of poses within 10degree from all directions. Since our target is to have strong face recognition under uncontrolled environments where the face illumination cannot be predicted, we did not intensively test the various poses. This remains one item to be done as the next item.

7.2 Extensions of the present works

There are several open problems with respect to both feature extraction theory and face pose.

- SGBM filters should be explored in comparison to new decompositions. Additionally, an appropriate filter design procedure needs to be applied to extract the most invariant features.

- Recognition under arbitrary face pose should be done. To achieve it, there are two ways: either synthesizing the angular faces from frontal pose and using them during enrollment or face pose correction. Many research studies focus on the face pose correction. 3D-based face

recognition technique is one of them. Each approach should be investigated and one method should be chosen for the next work. We are thinking of developing an image descriptor by using three properties of images: texture, shape and spatial frequency.

- Face alignment is done by using eyes, nose and mouth corners. The alignment is done geometrically. Depending on the face size, some faces are shrunk or extended. Any change in face image during alignment changes the discriminative face features. A robust and better alignment technique should be investigated for better recognition rates.

- Live stream has some advantages. Continuous live stream enables us to evaluate a face in several times. Some kind of voting can be used. In addition to this, some general logics should be applied. For example, in the same image, if the system matches the same person to two different people in the image, one needs to be rejected since it is logically an impossible case to give two recognized people to one record since the same person cannot be in two different locations in the image.

REFERENCES

[1] N. Kanwisher and G. Yovel, “The fusiform face area: a cortical region specialized for the perception of faces,” Philosophical Transactions of the Royal Society B, vol. 361, no. 1476, pp. 2109-2128, Nov. 2006.

[2] M.J. Farah, “Is face recognition ‘special’? Evidence from neuropsychology,” Behavioral Brain Research, vol. 76, no. 2, pp. 181–189, Apr. 1996.

[3] I. Gauthier, Tarr MJ, Anderson AW, P. Skudlarski and J.C. Gore, “Activation of the middle fusiform ‘face area’ increases with expertise in recognizing novel objects,” Nature Neuroscience, vol. 2, no. 6, pp. 568–573, Jun. 1999.

[4] G. Isabel and N.K.Logothesis, “Is face recognition not so unique after all?,” Cognitive Neuropsychology, vol. 17, no. 1, pp. 125-142, Feb. 2000.

[5] K.G. Spector, N. Knouf and N. Kanwisher, “The fusiform face area subserves face perception, not generic within-category identification,” Nature Science, vol. 7, no. 5, pp. 555-562, May 2004.

[6] N. Kanwisher, J. McDermott and M.M. Chun, “The fusiform face area: a module in human extrastriate cortex specialized for face perception,” The Journal of Neuroscience, vol. 17, no. 11, pp. 4302-4311, Jun. 1997.

[7] M. Sugiura, R. Kawashima, K. Nakamura, N. Sato, A. Nakamura, T. Kato, K. Hatano, T. Schormann, K. Zilles, K. Sato, K. Ito and H. Fukuda, “Activation reduction in anterior temporal cortices during repeated recognition of faces of personal acquaintances,” NeuroImage, vol. 13, no. 5, pp. 877–890, Mar. 2001.

[8] M.A. Kuskowski and J.V.Pardo, “The role of the fusiform gyrus in successful encoding of face stimuli,” Neuroimage, vol. 9, no. 6, pp. 599–610, Jun. 1999.

[9] T. Tsukiura, H. Mochizuki-Kawai and T. Fujii, “Dissociable roles of the bilateral anterior temporal lobe in face-name associations: an event-related fMRI study,” Neuroimage, vol. 30, no. 2, pp. 617–626, Apr. 2006.

[10] N. Kanwisher, “Domain specificity in face perception,” Nature Neuroscience, vol. 3, no. 8, pp. 759–763, Aug. 2000.

[11] J.J. Evans, A.J. Heggs, N. Antoun and J.R. Hodges, “Progressive prosopagnosia associated

with selective right temporal lobe atrophy: A new syndrome?," Brain, vol. 118, no. 1, pp. 1–13, Feb. 1995.

[12] D. Tranel, H. Damasio and A.R. Damasio, "A neural basis for the retrieval of conceptual knowledge," Neuropsychologia, vol. 35, no. 10, pp. 1319–1327, Oct. 1997.

[13] G. Gainotti, A. Barbier and C. Marra, "Slowly progressive defect in recognition of familiar people in a patient with right anterior temporal atrophy," Brain, vol. 126, no. 4, pp. 792–803, Apr. 2003.

[14] F. Tong, K. Nakayama, M. Moscovitch, O. Weinrib, and N. Kanwisher, "Response properties of the human fusiform face area," Cognitive Neuropsychology, vol. 17, no. 1, pp. 257–279, Feb. 2000.

[15] N.M. Kleinhans, T. Richards, L.C. Johnson, K.E. Weaver, J. Greenshon, G.Dawson and E. Aylward, "fMRI evidence of neural abnormalities in the subcortical face processing system in ASD," Neuroimage, vol. 54, no. 1, pp. 697-704, Jan. 2011.

[16] I. Gauthier, Tarr MJ, Anderson AW, P. Skudlarski and J.C. Gore, "Activation of the middle fusiform face area increases with expertise in recognizing novel objects," Nature Neuroscience, vol. 2, no. 6, pp. 568–580, Jun. 1999.

[17] K. Okajima, "Two-dimensional Gabor-type receptive field as derived by mutual information maximization," Neural Networks, vol. 11, no. 3, pp. 441-447, Apr. 1998.

[18] M. Turk and A. Pentland, "Eigenfaces for recognition," Journal of Cognitive Neurosicence, vol. 3, no. 1, pp. 71-86, Mar. 1991.

[19] J. Lu, K.N. Plataniotis and A.N. Venetsanopoulos, "Face recognition using LDA-based algorithms," IEEE Trans. on Neural Networks, vol. 14, no. 1, pp. 195-200, Jan. 2003.

[20] S. Pittner and S.V. Kamarthi, "Feature extraction from wavelet coefficients for pattern recognition tasks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, no. 1, Jan. 1999.

[21] J. Daugman, "Gabor wavelets and statistical pattern recognition," The Handbook of Brain Theory and Neural Networks, MIT Press, pp. 457-463, 2002.

[22] H. Moon and P.J. Phillips, "Computational and performance aspects of PCA-based face recognition algorithms," Perception, vol. 30, no. 3, pp. 303-321, Jan. 2001.

[23] J. Jiang, "Asymmetric principal component and discriminant analyses for pattern classification," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 5, pp. 931-937, May 2009.

[24] M. Li, and B. Yuan, "2D-LDA: a statistical linear discriminant analysis for image matrix," Pattern Recognition Letters, vol. 26, no. 5, pp. 527-532, Apr. 2005.

[25] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, "Face recognition: a literature survey," ACM Computing Surveys (CSUR), vol. 35, no. 4, pp. 399-458, Dec. 2003.

[26] H. K. Ekenel and R.Stiefelhagen, "Local appearance based face recognition using discrete cosine transform," 13th European Signal Processing Conference (EUSIPCO-2005), Sept. 2005.

[27] M. Savvides, B.V.K. Vijaya Kumar and P.K. Khosla, "Face verification using correlation filters," Proc. Of the Third IEEE Automatic Identification Advanced Technologies, pp.56-61, Tarrytown, NY, Mar. 2002.

[28] X.Wang and X.Tang, "Bayesian face recognition using gabor features," International Multimedia Conference Proceedings of the 2003 ACM SIGMM workshop on Biometrics methods and applications, pp. 70-73, 2003.

[29] J. Wright, A.Y. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, no. 2, Feb. 2009.

[30] M.-H. Yang, "Kernel eigenfaces vs. kernel fisherfaces: face recognition using kernel methods," Proc. of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 215-220, May 2002.

[31] M.A.Kramer, "Nonlinear principal component analysis using autoassociative neural networks," AIChE Journal, vol. 37, no. 2, pp. 233-243, Feb. 1991.

[32] S.Tan and M.Mavrovouniotis, "Reducing data dimensionality through optimizing neural-network inputs," AIChE Journal, vol. 41, no. 6, pp. 1471-1480, Jun. 1995.

[33] E.C.Malthouse, "Limitations of nonlinear PCA as performed with generic neural networks," IEEE Transactions on Neural networks, vol. 9, no.1, pp. 165-173, Jan. 1998.

[34] X. Wang and X. Tan, "Improving indoor and outdoor face recognition using unified subspace analysis and Gabor features," Image Processing, 2004 International Conference, vol. 3, pp. 1983-1986, Oct. 2004.

[35] W.Y. Zhao, R. Chellappa, P.J. Phillips and A. Rosenfeld, "Face recognition: a literature survey," ACM Computing Survey, vol. 35, no. 4, pp. 399-458, Dec. 2003.

[36] G. Shakhnarovich, J.W. Fisher, and T. Darrel, "Face recognition from long-term observations," Proc. of IEEE European Conference on Computer Vision, vol. 3, pp. 851–865, 2002.

[37] X. Zou, J. Kittler and K. Messer, "Illumination invariant face recognition: a survey," First IEEE Conference on Biometrics: Theory, Applications and Systems, pp. 1-8, Sept. 2007.

[38] P.N. Belhumeur and D.J. Kriegman, "What is the set of images of an object under all possible illumination conditions," International Journal of Computer Vision, vol. 28, no. 3, pp. 245–260, Jul. 1998.

[39] S.Biswas, G. Aggarwal and R. Chellappa, "Robust estimation of albedo for illumination-invariant matching and shape recovery," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 5, pp. 884-899, May 2009.

[40] L. Zhang and D. Samaras, "Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 28, no. 3, pp. 351-363, Mar. 2006.

[41] K.Lee, J. Ho, and D.J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 5, pp. 684-698, May 2005.

[42] T. Chen,W. Yin, X. Sean Zhou, D. Comaniciu, and T. S. Huang, "Total variation models for variable lighting face recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 9, Sept. 2006.

[43] A.S. Georghiades, D.J. Kriegman, and P.N. Belhumeur, "Illumination cones for recognition under variable lighting: faces," Proceedings of IEEE Conf. on Comp. Vision and Patt. Recogniton, pp. 52–58, Jun. 1998.

[44] S.S. Xiao and M. Jin, "From 2D to 3D: using illumination cones to build 3d face nodel," Journal of Physics: Conference series, vol. 48, no. 1, pp. 318-323, Aug. 2006.

[45] F. Xiea, L. Taoa, and G. Xu, "Estimating illumination parameters using spherical harmonics coefficients in frequency space," Tsinghua Science & Technology, vol. 12, no. 1, pp. 44-50, Feb. 2007.

[46] T. Okabe, I. Sato, and Y. Sato, "Spherical harmonics vs. Haar wavelets: Basis for

recovering illumination from cast shadows," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 50-57, Jul. 2004.

[47] Y.Adini, Y.Moses and S.Ullman, "Face recognition: the problem of compensating for changes in illumination direction," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, Jul. 1997.

[48] E. Hadjidemetriou, M.D.Grossberg and S.K. Nayar, "Multiresolution histograms and their use for recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 7, Feb. 2004.

[49] X. Xie and Kin-M. Lam, "An efficient illumination normalization method for face recognition," Pattern Recognition Letters, vol. 27, no. 6, pp. 609-617, Apr. 2006.

[50] Y. Matsushita, K. Nishino, K. Ikeuchi, and M. Sakauchi, "Illumination normalization with time-dependent intrinsic images for video surveillance," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 10, pp. 1336-1347, Oct. 2004.

[51] V. Blanz, S. Romdhani and T.Vetter, "Face identification across different poses and illuminations with a 3D morphable model," Proc. of the fifth IEEE International conference on Automatic Face and Gesture Recognition, pp. 202-207, May 2002.

[52] R.Gross and V. Brajovic, "An image preprocessing algorithm for illumination invariant face recognition," 4th Int. Con. on Audio- and Video-Based Biometric Person Authentication (AVBPA), pp. 10-18, Jun. 2003.

[53] E. Hadjidemetriou, "Use of histograms for recognition," Columbia university, ftp://ftp.cs.columbia.edu/pub/CAVE/stathis/thesis/Hadjidemetriou.pdf, 2002.

[54] P.J. Phillips and Y. Vardi, "Data-driven methods in face recognition," International Workshop on Automatic Face- and Gesture-Recognition, pp. 65-70, 1995.

[55] N. Sun, Z-h. Ji, C-r. Zou and L. Zhao, "Two-dimensional canonical correlation analysis and its application in small sample size face recognition," Journal of Neural Computing and Applications, vol. 19, no. 3, pp. 377-382, Jul. 2009.

[56] S. Shan, W. Gao, B. Cao, and D. Zhao, "Illumination normalization for robust face recognition against varying lighting conditions," In Proc. IEEE workshop on AMFG, 2003.

[57] X. Xie and K. Lam, "An efficient illumination normalization method for face recognition," Pattern Recognition Letters, vol. 27, no. 6, pp. 609-617, Apr. 2006.

[58] P.Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features,"

Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2001), vol. 1, pp. 511-518, 2001.

[59] R.Lienhart and J.Maydt, "An extended set of haar-like features for rapid object detection," IEEE ICIP-2002, vol. 1, pp. 900-903, 2002.

[60] S.Kawato, N.Tetsudani, and K. Hosaka, "Scale adaptive face detection and tracking in real time with SSR filter and support vector machine," IEICE Transactions on Information and Systems, vol.88, no. 12, pp. 2857-2863, Dec. 2005.

[61] P.Viola and M. Jones, "Robust real-time object detection," International Journal of Computer Vision, vol. 57, no. 2, pp. 137-154, May 2004.

[62] J.Maydt and R.Lienhart, "A fast method for training support vector machines with very large set of linear features," IEEE ICME2002, vol. 1, pp. 309-312, Nov. 2002.

[63] T. Joachims, "Advances in kernel methods – support vector learning," MIT Press, Cambridge, USA, 1999.

[64] E. Osuna, R. Freund and F. Girosi, "An improved training algorithm for support vector machines," Proc. IEEE NNSP'97, pp. 276-285, Sept. 1997.

[65] B.Scholkopf, A.Smola and K.R.Muller, "Nonlinear component analysis as a kernel eigenvalue problem," Neural Computation, vol.10, no. 5, pp.1299-1319, Jul. 1998.

[66] P.Baldi and K.Hornik, "Neural networks and principal component analysis: learning from examples without local minima," Neural Networks, vol.2, no. 1, pp. 55-58, Aug. 1989.

[67] M. Scholz, "Analysing periodic phenomena by circular PCA," Bioinformatics Research and Development, vol. 4414, pp. 38-47, 2007.

[68] B.B. Nasution and A.I. Khan, "A hierarchical graph neuron scheme for real-time pattern recognition," IEEE Transactions on Neural Networks, vol. 19, no.2, pp. 212-229, Feb. 2008.

[69] M.Egmont-Petersen, D. Ridder and H.Handels, "Image processing with neural networks – a review," Pattern Recognition, vol. 10, no. 35, pp. 2279–2301, Oct. 2002.

[70] C. Bishop, "Variational principal components," Proceedings Ninth International Conference on Artificial Neural Networks, ICANN'99, pp. 509–514, 1999.

[71] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques and W. Jaesik Min Worek, "Overview of the face recognition grand challenge," Proc. Computer Vision and Pattern Recognition (CVPR 2005), vol. 1, pp. 947- 954, Jun. 2005.

[72] T.Sim and S.Baker and M.Bsat, "The CMU pose, illumination, and expression (PIE)

database," Proceedings of the 5th International Conference on Automatic Face and Gesture Recognition, 2002.

[73] A.S. Georghiades and P.N. Belhumeur, "From few to many: illumination cone models for face recognition under variable lighting and pose," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pp. 643-660, Jun. 2001.

[74] A.S.M Shihavuddin, M. Mohammad N. Arefin, M.N. Ambia, S.A. Haque and T. Ahammad, "Development of real time face detection system using Haar-like features and adaboost algorithm," IJCSNS International Journal of Computer Science and Network Security, vol. 10, no. 1, pp. 171-178, Jan. 2010.

[75] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W. Gao, "WLD: a robust local image descriptor," IEEE transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 9, pp. 1705-1719, Sept. 2010.

[76] Y. Zhang and Z.-H. Zhou, "Cost-sensitive face recognition," IEEE transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 10, pp. 1758-1769, Oct. 2010.

[77] H.-S. Lee and D. Kim, "Tensor-based AAM with continuous variation estimation: Application to variation-robust face recognition," IEEE transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 6, pp. 1102-1116, Jun. 2009.

[78] N. Hadjikhani, R.M. Joseph, J. Snyder and H. T.-Flusberg, "Anatomical differences in the mirror neuron system and social cognition network in autism," Cerebral Cortex, vol. 16, no. 9, pp. 1276–1282, Sept. 2006.

[79] A. King, "A survey of methods for face detection," Technical Report. McGill Unv., Mar. 2003.

[80] M. Yang, D.J. Kriegman and N. Ahuja, "Detecting faces in images: a survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 1, pp. 34-58, Jan. 2002.

[81] S. Li and Z. Zhang, "Floatboost learning and statistical face detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 9, pp. 1112-1123, Sept. 2004.

[82] H. Schneiderman, "Feature-centric evaluation for efficient cascaded object detection," IEEE Conf. Computer Vision and Pattern Recognition, vol. 2, pp.29-36, Jun. 2004.

[83] R. Xiao, L. Zhu, and H.-J. Zhang, "Boosting chain learning for object detection," Proc. Of Computer Vision, vol. 1, pp. 709–715, Oct. 2003.

[84] J. Sun, J. Rehg, and A. Bobick, "Automatic cascade training with perturbation bias," Computer vision and Pattern Recognition, vol. 2, pp. 276-283, Jul. 2004.

[85] H. Schneiderman and T. Kanade, "Object detection using the statistics of parts," International Journal of Computer Vision, vol. 56, no. 3, pp. 151-177, Mar. 2004.

[86] Y. Freund and R.E. Schapire, "Experiments with a new boosting algorithm," in Machine learning: Processings of Thirteenth International Conference, pp. 148-156, 1996.

[87] Y. Freund, "Boosting a weak learning algorithm by majority," Information and Computation, vol. 121, no. 2, pp. 256-285, Sept. 1995.

[88] R. Basri and D. Jacobs, "Lambertian reflectance and linear subspaces," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 2, pp. 218-233, Feb. 2003.

[89] W. Zhao and R. Chellappa, "SFS based view synthesis for robust face recognition," Proceedings of 4th IEEE Conferences on Automatic Face Gesture Recognition, pp. 285–292, Mar. 2000.

[90] Q. Zheng and R. Chellappa, "Estimation of illuminant direction, albedo, and shape from shading," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, no. 7, pp. 680–702, Jul. 1991.

[91] A. Serrano, I.M. Diego, C. Conde, E. Cabello, L. Shen, and L. Bai, "Influence of wavelet frequency and orientation in an SVM based parallel Gabor PCA face verification system," IDEAL Conference 2007, Springer-Verlag, vol. 4881, pp. 219–228, Dec. 2007.

[92] D.J. Jobson, Z. Rahman and G.A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observations of scenes," IEEE Transactions on Image Processing, vol. 6, no. 7, pp. 965–976, Jul. 1997.

[93] H. Wang, S.Z. Li, Y. Wang and J. Zhang, "Self quotient image for face recognition," Proceedings of the International Conference on Pattern Recognition, vol. 2, pp. 1397-1400, Oct. 2004.

[94] K. Delac, M. Grgic and T. Kos, "Sub-image homomorphic filtering technique for improving facial identification under difficult illumination conditions," International Conference on Systems, Signals and Image Processing (IWSSIP'06), pp. 95-98, Sept. 2006.

[95] S. Du and R. Ward, "Wavelet-based illumination normalization for face recognition," Proc. of the IEEE International Conference on Image Processing, vol. 2, pp. 954-957, Sept. 2005.

[96] Y-L. Liu, J. Wang, X. Chen, Y-W. Guo and Q-S. Peng, "A robust and fast non-local means

algorithm for image denoising," Journal of Computer science and technology, vol. 23, no. 2, pp. 270-279, Mar. 2008.

AUTHOR BIOGRAPHY

Sadi Vural received his B.S. degree in Electronic engineering department from Istanbul University, Istanbul, Turkey in 1997 and M.S. degree in information science and engineering department from Ritsumeikan University, Shiga, Japan in 2000. He is currently working toward the Ph.D degree in the department of systems innovation, graduate school of Engineering Science at Osaka University. Osaka, Japan.

He worked as a software engineering intern at IBM Corporation, Shiga, Japan in 1999. He joined the development of AIBO robot in Sony Corporation, Tokyo, Japan in 2000 and hardware development department in Rohm Co., Ltd., Kyoto, Japan in 2001. He involved in Mobile phone development projects in Fujitsu Corporation, Akashi city, Hyogo prefecture, Japan between 2003 and 2005. He is currently working at Ayonix Incorporation, Osaka, Japan since of 2007.

His current interests include face detection, face recognition, Object recognition, neuro-brain recognition, gender and age estimation and neuro-imaging. He is presently working on a new facial recognition algorithm, which is appropriate for surveillance systems in public areas.

LIST OF PUBLICATIONS

Journal publications (Face recognition related papers)

Some parts of the work presented in the thesis have been published in the following articles:

1. **S. Vural**, Y. Mae, H. Uvet, and T. Arai, "Multi-view fast object detection by using extended haar filters in uncontrolled environments," in review

2. **S. Vural**, Y. Mae, H. Uvet, and T. Arai, "Illumination normalization for outdoor face recognition by using Ayofa-filters," Journal of Pattern Recognition Research, vol. 6, no. 1, pp. 1-18, Feb. 2011.

3. **S. Vural**, Y. Mae, and T. Arai, "Illumination-invariant face texture analysis by gaussian-histogram equalization and hierarchical nonlinear principal component analysis," Journal of Signal Processing, vol. 14, no. 6, pp. 459-474, Nov. 2010.

4. **S. Vural**, Y. Mae, and T. Arai, "Face perception based on spatial gaussian bessel mixture and nonlinear feature analysis," Journal of Signal Processing, vol. 14, no. 5, pp. 369-379, Sept. 2010.

5. **S. Vural** and H. Yamauchi, "Real-time face detection and tracking using six-segment filters," Journal of Signal Processing, vol. 11, no. 1, pp. 83-91, Jan. 2007.

6. **S. Vural** and H. Yamauchi, "Facial analysis and authentication for high-speed access gates," WSEAS Transactions on Signal Processing, vol. 2, no. 8, pp. 1046-1052, Aug. 2006.

International conference papers (Face recognition related papers)

1. **S. Vural** and H. Yamauchi, "Bio security using face recognition for industrial use," AIC'06 Proceedings of the 6th WSEAS International Conference on Applied Informatics and Communications, Greece, pp.120-124, Aug. 2006.

2. **S. Vural** and H. Yamauchi, "Practical aspects of face recognition," Proceedings of world academy of science, engineering and technology, vol. 2, no. 3, pp. 153-157, Aug. 2006.

3. H. Yamamoto, D. Kanbara, Y. Okazaki, **S. Vural**, and H. Yamauchi, "Biometrics authentication system by referencing human face," Proceedings of the 2006 RISP International Workshop on Nonlinear Circuit Signal Processing, vol. 10, no. 4, pp.295-298, Mar. 2006.

Awarded papers

Best conference paper award NCSP'06 (Proceedings of the 2006 RISP International Workshop on Nonlinear Circuit Signal Processing) (Hawaii) for the paper entitled "Biometrics authentication system by referencing human face".

Other journal publications

1. **S. Vural**, H. Tomii, and H. Yamauchi, "Robust digital cinema watermarking," International Journal of Information technology (IJIT), vol. 1, no. 1, pp. 255-259, Apr. 2005.

2. **S. Vural**, H. Tomii, and H. Yamauchi, "Traceable watermarking system using SoC for digital cinema delivery," International Journal of Information technology (IJIT), vol. 2, no. 1, pp. 91-96, Dec. 2004.

Other international conference papers

1. **S. Vural**, H. Tomii, and H. Yamauchi, "Video watermarking for digital cinema contents," EUSIPCO 2005 (13th European Signal Processing Conference), Sept. 2005.

2. **S. Vural**, H. Tomii, and H. Yamauchi, "Traceable robust watermarking for digital cinema system," 8th International Symposium on Signal Processing and Its Application, vol. 1, pp. 98-102, Aug. 2005.

3. **S. Vural**, H. Tomii, and H. Yamauchi, "Robust video watermarking using additional watermarking techniques," Proceedings of SPIE, vol. 5915, Aug. 2005.

4. **S. Vural**, H. Tomii, and H. Yamauchi, "DTW based robust watermarking embed using CRC-32 techniques," The 3rd World Enformatika Conference Proceedings, vol. 5, pp. 106-109, May 2005.

5. **S. Vural**, H. Tomii, and H. Yamauchi, "Traceable watermarking system using SoC for digital cinema delivery," Proceedings of the International Conference on Signal Processing, pp. 268-271, Dec. 2004.

6. **S. Vural**, T. Tsutsumi, and H. Yamauchi, "Comparison between dispersion model and diffusion model using ray-tracing method for pulse radar simulation," The 1st IEEE International Symposium on Signal Processing and Information Technology, pp. 89-93, Dec. 2001.