FACE IDENTIFICATION, GENDER AND AGE GROUPS CLASSIFICATIONS FOR THE SEMANTIC ANNOTATION OF VIDEOS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

GÖKHAN YAPRAKKAYA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
COMPUTER ENGINEERING

DECEMBER 2010

Approval of the thesis:

FACE IDENTIFICATION, GENDER AND AGE GROUPS CLASSIFICATIONS FOR SEMANTIC ANNOTATION OF VIDEOS

Submitted by **GÖKHAN YAPRAKKAYA** in partial fulfillment of the requirements for the degree of **Master of Science in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences** ________________

Prof. Dr. Adnan Yazıcı
Head of Department, **Computer Engineering** ________________

Assoc. Prof. Dr. Nihan Kesim Çiçekli
Supervisor, **Computer Engineering Dept., METU** ________________

Examining Committee Members:

Asst. Prof. Dr. İlkay Ulusoy
Electrical and Electronics Eng. Dept., METU ________________

Assoc. Prof. Dr. Nihan Kesim Çiçekli
Computer Engineering Dept., METU ________________

Asst. Prof. Dr Tolga Can
Computer Engineering Dept., METU ________________

Asst. Prof. Dr. Alptekin Temizel
Informatics Institute, METU ________________

Umut Tosun
Software Engineer, ASELSAN ________________

Date: 17.12.2010

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Gökhan Yaprakkaya

Signature :

ABSTRACT

FACE IDENTIFICATION, GENDER AND AGE GROUPS CLASSIFICATIONS FOR SEMANTIC ANNOTATION OF VIDEOS

Yaprakkaya, Gökhan
M.S., Department of Computer Engineering
Supervisor: Assoc. Prof. Dr. Nihan Kesim Çiçekli

December 2010, 73 pages

This thesis presents a robust face recognition method and a combination of methods for gender identification and age group classification for semantic annotation of videos. Local binary pattern histogram which has 256 bins and pixel intensity differences are used as extracted facial features for gender classification. DCT Mod2 features and edge detection results around facial landmarks are used as extracted facial features for age group classification. In gender classification module, a Random Trees classifier is trained with LBP features and an adaboost classifier is trained with pixel intensity differences. DCT Mod2 features are used for training of a Random Trees classifier and LBP features around facial landmark points are used for training another Random Trees classifier in age group classification module. DCT Mod2 features of the detected faces morped by two dimensional face morphing method based on Active Appearance Model and Barycentric Coordinates are used as the inputs of the nearest neighbor classifier with weights obtained from the trained Random Forest classifier in face identification module. Different feature extraction methods are tried and compared and the best achievements in the face recognition module to be used in the method chosen. We compared our classification results with some successful earlier works results in our experiments performed with same datasets and got satisfactory results.

Keywords: Face Detection, Face Recognition, Automatic Age Group Classification, Automatic Gender Classification, Semantic Annotation.

ÖZ

VİDEOLARA ANLAMSAL AÇIKLAMA EKLEMEK İÇİN YÜZ TANIMA, CİNSİYET VE YAŞ GRUBU SINIFLANDIRMASI

Yaprakkaya, Gökhan
Yüksek Lisans, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi: Doç. Dr. Nihan Kesim Çiçekli

Aralık2010, 73 sayfa

Bu tezde videolara anlamsal açıklama eklemek için gürbüz bir yüz tanıma metodu, cinsiyet ve yaş grubu sınıflandırma yöntemleri bir arada sunulmaktadır. Cinsiyet sınıflama metodunda, yüzün çıkarılan özellikleri olarak 256 bölümlü yerel ikili örüntü histogram ve piksel yoğunluğu farkı kullanılmıştır. Yaş grubu sınıflama metodunda, yüzün çıkarılan özellikleri olarak DCT Mod2 özellikleri ve yüzdeki önemli noktaların civarına uygulanan kenar bulma metodu sonuçları kullanılmıştır. Cinsiyet sınıflandırma modülünde, 256 bölümlü yerel ikili örüntü histogramı verileriyle eğitilmiş Random Trees sınıflandırıcısı ve piksel yoğunluğu farklılıkları ile eğitilmiş bir adaboost sınıflandırıcısı kullanılmıştır. Yaş grubu sınıflandırma modülünde, DCT Mod2 özellikleriyle eğitilmiş Random Trees sınıflandırıcısı ve yüzün önemli noktalarına uygulanan 256 bölümlü yerel ikili örüntü histogramı verileriyle eğitilmiş bir diğer Random Trees sınıflandırıcısı kullanılmıştır. Yüz tanıma modülünde Aktif Görünüm Modeline ve Barycentric Koordinat Sistemine dayalı iki boyutlu yüz biçimlendirmesi ile biçimlendirilmiş yüzlerden elde edilen DCT Mod2 özellikleri ile eğitilen Random Forest sınıflayıcı kullanılmıştır. Farklı özellik çıkarma yöntemleri denenmiş ve yüzleri tanımadaki başarılarıyla kıyaslanmış ve en iyi bulunan yöntem modülümüzde kullanılmak üzere seçilmiştir. Sınıflandırma metodlarımızı literatürde bulunan bazı başarılı çalışmaların aynı veriler üzerindeki sonuçlarıyla karşılaştırdık ve tatmin edici sonuçlar elde ettik.

Anahtar Kelimeler: Yüz Bulma, Yüz Tanıma, Otomatik Yaş Grubu Sınıflandırması, Otomatik Cinsiyet Sınıflandırması, Anlambilimsel Etiketleme

To my family

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and appreciation to my supervisor Assoc. Prof. Dr. Nihan Kesim Çiçekli for her endless encouragement and support throughout this study. I am very lucky that I have such a friendly, intellectual, patient and benignant supervisor.

I would like to thank Asst. Prof. Dr. İlkay Ulusoy who gave me invaluable suggestions about my thesis. I also want to thank Mehmet Kuzu who helped me a lot for overcoming difficulties.

I am deeply grateful to my parents who devoted their lives to their children, and my sister Gamze Yaprakkaya for their love and support.

I am deeply indebted to my friends, Serkan Demirören, Hasan Yıldız and all the other colleagues at SGK whose suggestions and encouragement helped me a lot during my study.

I would like to thank The Scientific and Technological Research Council of Turkey (TÜBİTAK) for providing the financial means throughout this study.

This work is partially supported by The Scientific and Technological Research Council of Turkey Grant ''TUBITAK EEEAG-107E234''.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

AAM	Active appearance model
DCT	Discrete cosine transform
LBP	Local Binary Pattern
EHMM	Embedded hidden Markov Model
HMM	Hidden Markov model
GMM	Gaussian mixture model
HSV	Hue-saturation-value color space
PCA	Principal Component Analysis
RGB	Red-green-blue color space
SIFT	Scale-invariant feature transform
SVM	Support vector machine
RTREES	Random Forest Classifier
ASM	Active Shape Model
LPP	Locality Preserving Projections
LOPO	Leave-one-person-out cross validation

CHAPTER 1

INTRODUCTION

With the development of the internet connection speed, visual media has become the focus of attention. People have begun to access huge amounts of multimedia content. Video clips, movies and various sorts of video content can be acquired easily. These developments have led to the requirement of an effective search method to access the desired videos. Therefore, video annotation has been a popular research area in recent years. In this thesis the focus is on the facial features extracted from videos. Facial features give lots of clues about the identity, gender, age, ethnicity of that people. If the videos are successfully annotated, they can be searched more easily, thanks to search engines that can be developed, and precise information can be obtained through data retrieval techniques.

Constructing automatic annotation tools to index huge multimedia databases is the basis for video annotation. Since a vast majority of videos contain humans, the extraction of personal data from video images and indexing videos with extracted data become a necessity. Identity, age and gender are valuable features for annotating of people. Human faces provide lots of information about the identity, gender and age of people. The relative positions of the facial landmarks, the intensity differences between facial regions can give the possibility of predicting the identity of a person. Facial landmarks, wrinkles, eyebrows, hair, lips can provide information about the gender and age of a person. Processing the captured image from video enables us to extract those features. Prediction of gender and age group of a person requires detection of frontal faces of people, extraction of facial features and training classifiers with those features, and use of the trained classifiers for prediction. The prediction of the identity of a person requires the detection of frontal faces of people, normalizing and morphing faces, extraction of facial features and training classifiers with those features, and use of the trained classifiers. Facial classification systems such as facial identification, gender classification, race classification, age classification have become requirements for successful annotation tools.

In this study we aim to describe a robust face identification method, a robust method to identify gender from human face, and determine the age group of that face. We work on classification systems separately. We can divide this study into three main modules. 'Gender Classification' is the first module, 'Age Group Classification' is the second module and the third module is 'Face Identification'. The extraction of LBP histogram and the calculation of intensity differences between pixels have been studied for gender classification; the extraction of DCT Mod2 features and LBP values of locations around eyes, mouth, cheek, and nose have been studied for age group classification. Active Appearance Model based face morphing and extraction of DCT Mod2 features are issued in face identification. After the extraction phase, 'Random Trees' and/or Adaboost classifiers are trained with the extracted data. The trained classifiers are used as the decision-maker of the system.

The specific contributions of this thesis are:

- We used multiple classifiers that use LBP and pixel difference data of the face image to classify the gender. Our algorithm combines the results of two classifiers and increases the success rate of the classification.

- Our age group classification module normalizes the face image in pose and expression. Two distinct classifiers that use LBP values around some selected regions on the face and DCT Mod2 values of the face are used in this module. Our algorithm combines the results of two classifiers and increases the success rate of the classification.

- Our face identification module detects the feature points of the face and morphs the face using these feature points. Morphing process is based on point interpolation using triangles. Morphing process normalizes the face image in pose and expression. We compared the results of there feature extraction methods and selected the methods with the best accuracy.

Part of this work was accepted as a short paper in ISCIS 2010 [69].

This thesis is structured as follows: The related work on gender and age classification and facial identification are presented in Chapter 2. The overall architecture of the system is explained in Chapter 3. Chapter 4 describes the details of the preliminary methods used in this thesis. Gender classification module, random trees and Adaboost, age group classification methods are described in the Chapter 5. Chapter 6 includes the description of the algorithms used for face identification and comparative performance evaluation for each method is presented with respect to identification accuracies. Some explanations about the potential usage of our algorithms in "Ontological Semantics" field are given in Chapter 7. The experimental results are presented in Chapter 8. Finally, conclusions and possible extensions to this study are discussed in Chapter 9.

CHAPTER 2

RELATED WORK

In this chapter, the related work about gender classification from faces, age group classification and face identification in images and videos are described. Section 2.1 gives information about the existing work on gender classification. Section 2.2 gives information about the existing work on age group classification. Section 2.3 includes the existing work on face identification techniques.

2.1 Gender Classification

Various studies on gender classification are reported. Some of the first studies reported on gender classification use auto associative networks and perceptrons [1,2]. Similar to these works, there is a considerable amount of work on gender classification [3-18]. Finding the most successful machine learning classifier is the focus of research in gender classification studies. SVMs are used as the classifier with the best success rates. Several methods like; HyperBF networks, elastic graph matching, RBF network, LDA, Adaboost, and etc. have also been suggested. Researchers in [11] found that Bayesian kernel classifiers outperformed SVM. The determination of the hyper parameters based on Bayesian model selection criterion is the main advantage of their classifier. In [62], researchers compared the success rates of SVM and Random Forest classifiers on gender classification. They used SIFT features as feature extraction method. Random Forest outperformed SVM in their experiments.

Image quality, lighting issues, and the extraction method of the faces are the most effective parameters for the accuracy rate. In most of the previous works, high quality images were used and these images were manually extracted. Therefore, rates over %90 were achieved in those studies. However in [14], [20], and [21] classifiers achieved around 80% success rate with lower quality images. In these experiments a set of images of human faces was collected from the World Wide Web. The face detection algorithm which is based on Haar was applied on the images and the detections were manually labeled. False positive detections and faces more than 30° off frontal orientation, as well as those for which it was impossible to establish the ground truth regarding the gender and ethnicity of the subjects are removed from the dataset. Finding the best image resolution for the most successful gender classification study is one of the research aims. In [5], [11], and [18] they measured the success rates of their classifiers with various face image resolutions. They all agreed on the negative effect of decreasing image resolution.

The selection of facial features plays an important role in classification. The selection of facial features has been issued in earlier works. Face color and wrinkles on face are used in [22]. These features are used for age and gender classification in [22]. In [84], various face regions are used as the feature input for their classifier and the performances are measured. In [23], experiments are conducted with extracted facial features in uncontrolled environments. In [12], researchers used facial intensity and range data and got better results than either alone. All the studies agreed on the importance on success rates of the selection of best features. Using whole face image rather than a part of the face improves the success rate of the classifier.

There are also various other concepts to improve the success rate of gender classification systems. In [84], the effect of using clothing features on the performance of gender classification was inspected. Hair, face and clothing are used as features for classifiers and the experiments showed the negative effect of the clothing under uncontrolled environments. In [85], the effect of facial expressions on gender classification was studied. They compared the success rates of the gender classification using raw faces and unexpressioned faces. Preprocessing face images before feature extraction phase for the normalization of face in expressions made the classifier more successful.

In [88], Random Forest and PCA is used for gender classification. The accuracy of the classifier is measured in four distinct test data in their experiments. These groups are four distinct age groups. Young aged, adults, seniors and all ages. 74% accuracy rate is achieved on “all ages” group.

In this thesis, based on the related work, we have selected Random Forest classifier and Adaboost classifier as the machine learning methods, LBP features as feature extraction method, and whole face rather than a facial region as the data to be processed. An automatic face extraction and face normalization method is used. We have worked on 20x20 pixels image resolution in our first classifier and 40x40 pixels resolution in our second classifier. We have compared our results with the study [88].

2.2 Age Classification

There has been much computer vision research on age classification with faces. In [24], researchers worked on the ratios between facial landmarks and on wrinkles that appear in the face. In different ages, location of eyes can be in different positions. Mouth and nose of a person come closer and more wrinkles appear as the person gets older.

Predicting the exact age of a person from the face image is impossible. Therefore, in most of the studies, age is classified into some categories. Age groups were determined as babies, young adults and seniors in [24]. The classifier is tested with 15 images and all the images were classified successfully. However, performance was not measured for a sufficient number of images. In [61], face images were classified into five groups. They used facial landmark locations and facial texture attributes for classification. They also used Gabor features, skin color and hair for classification. An SVM classifier was trained with those features. Shape and texture of the face image is modeled using active appearance model in [59]. They recorded their own database which contains 70 people with 7 images each, with neutral facial expression, different illuminations, and small deviations in head orientation. In their experiments, they achieved around 45% success rate. Earlier works showed that classifying age into fewer categories makes the age classification more successful.

In [25], 2D-LDA and LDA based classifiers are used for age classification. This classification was robust under various lighting conditions. For age ranges of 5, 10 and 15 years accuracies were 46.3%, 67.8% and 78.1%. In [26], researchers worked on the estimation of the exact age of a human. 5 year mean error is achieved when the whole face, hair, lower face including mouth, eyes are used. Quadratic classifier, shortest distance classifier, multi-layer perceptron and self-organizing map classifier were experimented in [26]. Quadratic classifier gave the best result with 3.82 years mean error.

In this thesis, we focused on facial landmark locations and wrinkles for age classification. We tried to compensate for the variations because of facial expressions using face morphing approach based on active appearance model. We used LBP features to extract the wrinkles of the face and DCT Mod2 features to obtain the ratios between facial landmarks. We used Random Forest classifier as the machine learning method. We decided to classify the age groups into three categories like most of the related works such as young, middle aged and old aged. We discuss the experiments and the results in Chapter 8.

2.3 Face Identification

Face identification has received significant attention during the past few years. In this section some related work on the major techniques on face identification are described. First of all we describe some of the two dimensional methods used frequently in the previous studies. After discussing 3D methods used in face identification research, we describe the related work on face recognition in videos.

2.3.1 2D methods for Face identification

Eigenface method is one of the mostly used methods in face identification. Eigenface method is also called Principle component analysis (PCA). Eigenface method is used in [27-31] for face recognition. In [27], [28] and [29], researchers assigned weights for all faces and they reconstructed face images using their assigned weights. In these studies illumination variation is problematic in pure eigenface method. In [30], the covariance matrix is computed to compensate the negative effects of illumination variation. Three images for each person are taken in different illumination conditions. The success rate in [30] was greater than the experiments stated in [27-29]. Changes in the appearance and in facial expressions are also problematic in eigenface method. In [31] they used eigenfeatures such as; eigeneyes, eigennose, and eigenmouth to compensate the negative effects of changing facial expressions and appearance. They achieved better results for a test set that includes face samples with changing facial expressions than the pure eigenface method in [28]. Despite some extentions on eigenface method in several works, its success rate in different lighting conditions and different facial expressions is not satisfactory and it is not applicable for video based face identification systems. The best success rate achieved in the existing studies was around 70%. Therefore, eigenface method is not applicable for a successful face identification system.

Some other techniques like graph matching are also used in face identification researches. Elastic graph matching technique is one of the graph matching techniques. Researchers used an extension to classical artificial neural networks for distortion invariant object recognition based on elastic graph matching to find the closest stored graph in [32]. This method achieved good accuracy results but its computational cost was huge. In [32], comparing a face with 87 stored objects on a parallel machine with 23 transputers took about 25 seconds. Their recognition rate for their image galleries which include 14 face images was around 80%-85%. This method is extended in [33]. In [33], distortion due to rotation in different facial expressions and depth made the method more successful. Around 68% to 87% of success rates were obtained for a face database which includes 111 different people with neutral frontal (with rotation degrees between -15˚ and +15˚) views. But this change also added some computational costs and it made these techniques not applicable for a video based face identification systems.

Some techniques like Hidden Markov Models - HMMs have been used in facial identification methods. HMM was used in [34] and the implementation was tested with 200 neutral frontal faces with high resolution. Eyes, mouth, nose, forehead are represented as regions in this algorithm. These regions are combined with the states of HMM in [34]. 180 of 200 tested faces are recognized successfully in this study. HMM uses one dimensional sequences which can be temporal or spatial, therefore, a technique which converts images to one dimensional sequence is used as helper. In [35] band sampling technique is used to convert images to one dimensional observation sequence. In [35], 87% success rate is achieved in a test with 100 selected 15˚ rotated faces from ORL face database [40], and 95% success rate is achieved with 2D Hidden Markov Model for neutral frontal views. However the classification time was very expensive (10 seconds for a face) and therefore it is not an applicable method for video based facial identification systems.

A widely used technique which uses geometrical features of the face is geometrical feature matching. In this approach, the main facial landmarks such as; eyes, eyebrows, mouth, nose, and the overall shape of the face are used for identification. These landmarks are represented as vectors. In [36], they developed an automated face identification system with this approach. 75% success rate was achieved in their tests. If features are extracted manually, face identification could achieve satisfactory results as in [37]. In [38], they extract some facial features such as nose width, length, mouth and chin, and construct a feature vector. Bayes classifier is used as the classifier in that system. They used a database of 47 people for testing and training and 90% accuracy is achieved. This approach is extended in [39] by adding mixture-distance technique and Gabor wavelets decomposition. In this study face identification methods achieved 95% accuracy rate on a face gallery which includes 685 individuals. In this work faces were represented as 30 manually extracted distances, however in our case, in a video based automated face identification system, extraction of features manually is not acceptable. In general, geometrical feature method based identification systems include a matching procedure which utilizes a topological graphic representation of the feature points. Accuracy of the feature locators is the most important parameter in this technique. This approach is applicable for matches in a large database. Therefore, in this thesis we focused on some feature extraction methods and geometrical feature matching.

Template matching is another used technique for face identification. Template matching is based on comparing faces with a suitable metric. A set of four feature templates is selected in [38]. These feature templates are the eyes, nose, mouth, and the whole face, for all of the available faces. They also compared the success rates of their template matching and geometrical feature methods and in their experiments their template matching algorithm performed with higher accuracy rates. However, comparing the template of an example face with the faces in the training set requires much more time than geometrical feature methods. Therefore, template matching approach is not suitable for the face identification system in videos.

2.3.2 3D methods for Face identification

Some 3D methods have been also used in face identification studies. 3D morphable model is an example for this approach. A vector space representation of faces is constructed such that any convex combination of shape and texture vectors of a set of examples describes a realistic human face in [41]. They proposed a new method to generate three dimensional face images from two dimensional face images. In this approach, a generated model is fitted to images and then the coefficients of the model are used to identify faces. In their method, the required facial landmarks need to be marked manually. The studies [42] and [43] extend this approach by reconstructing a synthetic view from probe images. After the reconstruction, the view is transferred to a pose-dependent identification method. Some sophisticated extensions like combining deformable 3D models with a computer graphics simulation of projection and illumination are studied in [44]. The morphable model represents shapes and textures of faces as vectors in a high-dimensional face space. The algorithm presented in [44] estimates all 3D scene parameters automatically, including head position and orientation, focal length of the camera, and illumination direction. The success rates of the discussed methods were between 90% and 95% on FERET face database. However, 3D morphable models require high resolution face images and 3D scans of the faces. Therefore, this approach is not applicable for our case.

2.3.3 Methods for Face Recognition

Line edge maps (LEMs) is a recent approach for face recognition. Line edge maps use edge information and they are insensitive to illumination changes. Edge maps have been used in various object recognition applications. It requires lower memory requirement than the approaches such as template matching and geometrical feature matching. Line edge maps (LEMs) are insensitive to illumination changes. 92% accuracy rate in face recognition task is achieved in [45]. In their experiments they used FERET face database. They selected four images of one hundred individuals for training and tested their method with 50 faces. In [47], edge map is thinned and line edge map of the face is obtained by a polygonal line fitting process. Line edge map is more successful than edge map under controlled conditions. Their methods identified 100% and 96.43% of the frontal faces correctly in experiments. In all the studies discussed, edge map of a face is usually constructed incorrectly for a low resolution image. Therefore, this approach is not applicable for our case. In our case, not all videos are high resolution videos.

Recently, some methods used with SVMs are popular in face recognition literature. A binary tree recognition strategy is applied with SVMs in [48]. In this work, features are extracted and then the discrimination functions between each pair are learned by SVMs and the disjoint test set enters the system for recognition. The experiment on ORL face database and another face database showed around 91% of success rate in face recognition task. A component based method for face identification is presented in [49-50]. In this approach, ten components are extracted from face and a vector is constructed by these features. Linear SVMs are used and in testing phase the component based method outperformed the global systems even though a more robust classifier is used. 3D morphable model and component based face recognition is combined in [51]. After taking various face images of people with various poses and in changing illuminated environments, classifiers are trained. Around 98% success rate is achieved. Requirement of a large number of training images taken with various poses and in changing illuminated environments is a drawback of this approach. An array of K optimal pairwise coupling classifier (O-PWC) is constructed in [52]. Combining the results of these K O-PWC results give the final result of the classification. In their experiment they used ORL face database and get satisfactory results around 96% of success. SVM and Independent Component Analysis (ICA) are combined in [53]. ICA is a generalization of the eigenface method. The Yale Database which contains 165 face images, and selected 300 face images from AR Face Database [54] are used in the experiments of [53]. ICA/SVM approach outperformed the eigenface method and their recognition accuracy was around 90%. However, in their experiment the used neutralized and illumination normalized face images. They tested their approach in controlled environments.

Multi classifier systems have begun to be used in face identification researches. In [55], a parameter based combined classifiers improved the generalization capability. Three classifiers trained on different parameters LVQ neural networks are combined and successful results were obtained. In their experiments, they used three faces in training and the accuracy retes in tests were around 100%. However, their experiments were not sufficient. They worked only on 3 people and used one face image of each in testing. A face recognition committee machine (FRCM), which assembles the outputs of various face recognition algorithms (such as Eigenface, Fisherface, Elastic Graph Matching (EGM), SVM and neural network) to obtain a unified decision with improved accuracy is discussed in [57]. 86.1% on Yale face database and 98.8% on ORL face database is achieved.

Face identification in videos is more challenging than face recognition in images. There are some related works on face recognition - identification in videos. Some semi automatic annotation works based on face recognition are introduced in some researches like [60]. The aim of the work was to develop a face identification tool to be used in TV series. Viola-Jones face detector and particle filter face tracker are used in this study. Skin colored region detection was used to help tracking and face detection methods. Local appearance features are extracted from the detected face image. They worked on 3different scenarios such as closed-set identification, automatic retrieval and interactive retrieval. They used DCT features of the face for face recognition. In this thesis, we have compared our results with the results of [60]. We have used the videos from TV Episodes “Coupling” as in [60]. SIFT features in [63] are extracted as the descriptions of the classifiers. Discrete Cosine Transform (DCT) coefficients are extracted from the faces and nearest neighbor classifier is used for automatic labeling. In [63] and [64], the researchers used some episodes as training episodes and some episodes as testing episodes which are manually annotated. They focused on the effects of the aging in face identification. They prepared a dataset consists of 611,770 face images. However, the success rate of their methods was around 60% in their TV series.

To summarize, geometrical feature matching - local appearance features are the most suitable approach for our case. This approach is not a time consuming approach and its success rate was very satisfactory in earlier works. The most used feature extraction methods in earlier works such as; LBP, DCT and HOG feature extraction methods are used in this thesis. We also selected Random forest as the classifier of our system. Random forest is a generic method that classifies the images successfully even though partial occlusions exist [87]. Critical components can be executed completely in parallel in this algorithm. In our case, quick training is necessary, since our classifiers are trained during face detection and tracking and our implementation aims to complete the processing a frame in real time. High accuracy of generalization can be achieved by this machine learning method. The nature of random forests presents a uniform strategy for accomplishing many image classification tasks [87]. Therefore, we selected Random Forest classifier to build a fast and successful face identification system for annotation. In the most of the related work, the identification of an unknown faces have not been considered clearly. In pure Random Forest classifier searches for the likelihood of an input face identified as a face which exists in training set. However, determination of the unknown faces is a necessity. We utilized Nearest Neighbor algorithm to detect unknown faces. We have performed experiments similar to the experiments in [60]. We have compared our results with the results of [60] and we evaluated our methods and classifier. In Chapter 8, we presented comparative evaluations of our approach and plotted the graphs of the results.

CHAPTER 3

THE OVERALL ARCHITECTURE OF THE SYSTEM

The system consists of two main modules. The first module is the Gender and Age Group Classification module. The second module is Face Identification module.

3.1 Gender and Age Group Classification Module

The module mainly consists of two running modes. The first mode is the *training mode* and the second mode is the *classification mode*. As a first step, video frames are captured from the selected video in the training mode. Skin colored regions are obtained from the captured frames. The face detection algorithm is executed on the skin colored regions to extract frontal faces. Then, eyes, mouth and nose locations are determined. In the next step, the detected human faces are cropped, rotated to equalize y coordinates of two eyes and then histogram equalization of the rotated face image is applied. After these normalization steps, LBP feature extraction, pixel intensity comparisons for gender classification and DCT Mod2 feature extraction, edge detection around eyes, mouth, nose and cheeks for age classification are executed. The extracted features are written to files and prepared to train the classifiers. A Random Trees classifier is trained with the data which reside in LBP feature file. An Adaboost classifier is trained with the data which reside in pixel comparison file.

In the classification mode, video frames are captured from the selected video. Skin colored regions are obtained from the captured frame. The face detection algorithm is executed on these locations and frontal faces are obtained. Then, eyes, mouth and nose locations of the faces are determined. In the next step, detected human faces are cropped, rotated to equalize y coordinates of two eyes and then histogram equalization of rotated face image is done. After these normalization steps, LBP feature extraction, pixel intensity comparisons for gender classification and DCT

Mod2 feature extraction, edge detection around eyes, mouth, nose and cheeks for age classification are executed. The extracted features are used as parameters to predict the gender and age group of the detected faces. Figure 3-1 shows the flowchart of the overall architecture.

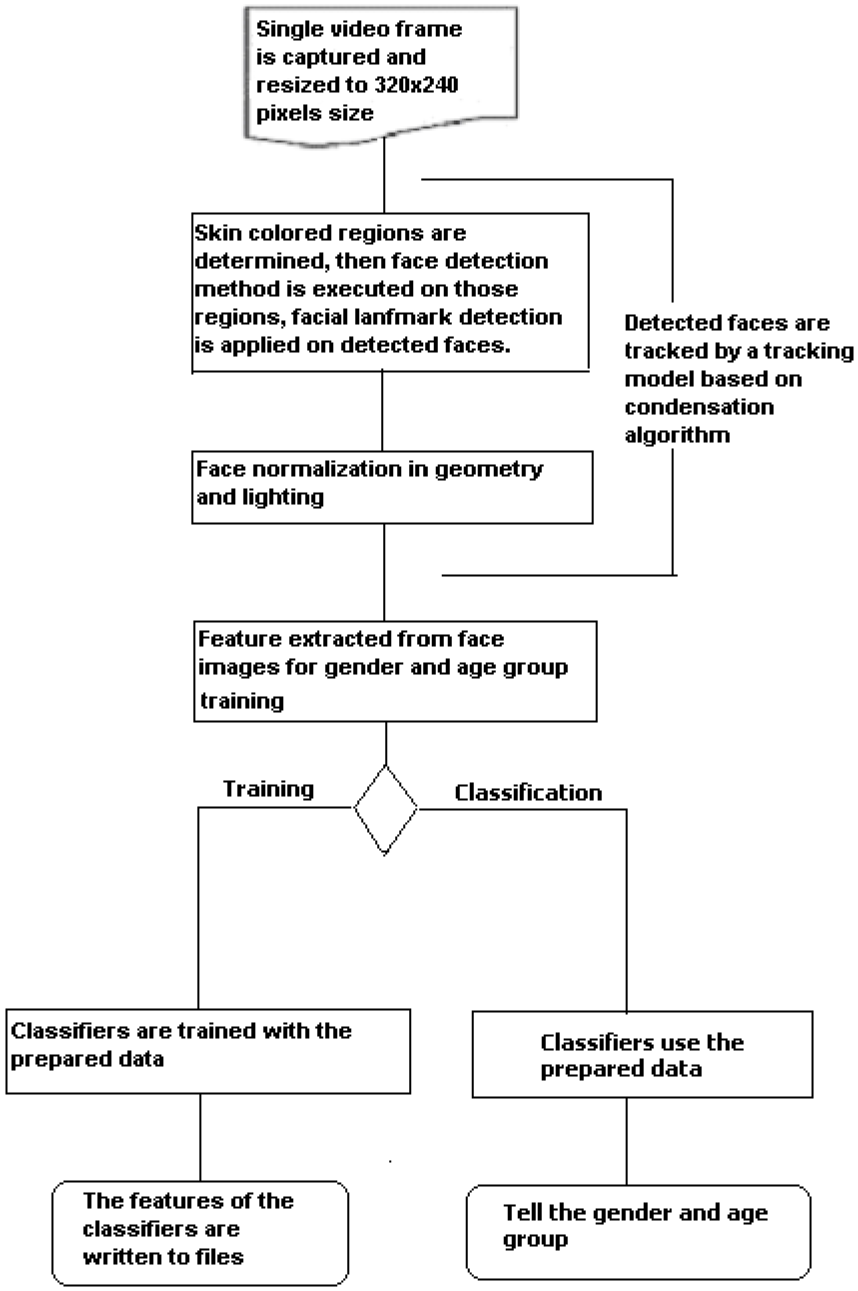


Figure 3-1 Overall architecture of Gender and Age Group Classification

3.2 Face Identification Module

The face identification module also consists of two running modes: the *training mode* and the *classification mode*. First, video frames are captured from the selected video in the training mode. Skin colored regions are obtained from the captured frame. The face detection algorithm is executed on the skin colored regions to obtain the frontal faces. Then, facial landmarks are obtained using Active Appearance Model method and two dimensional faces morphing using facial landmark phase is executed. Next, morphed human faces are normalized for identification. After these normalization steps, feature extraction methods are executed. The extracted features are written to files and prepared to train the classifiers. A Random Trees classifier is trained with the data which reside in the feature file.

In the classification mode, video frames are captured from the selected video. Skin colored regions are obtained from the captured frame. The face detection algorithm is executed on those locations and frontal faces are obtained. Then, facial landmarks are obtained using Active Appearance Model method. Then, two dimensional faces morphing using facial landmark phase is executed. Next, morphed human faces are normalized for identification. After these normalization steps, feature extraction algorithms are executed. The obtained feature vectors are used by the classifiers. Classifiers give the identification of the faces and the system shows the results to the user via graphical user interface. The time points representing the appearance and disappearance of faces are output by the system. Figure 3-2 shows the architecture of the face identification system.

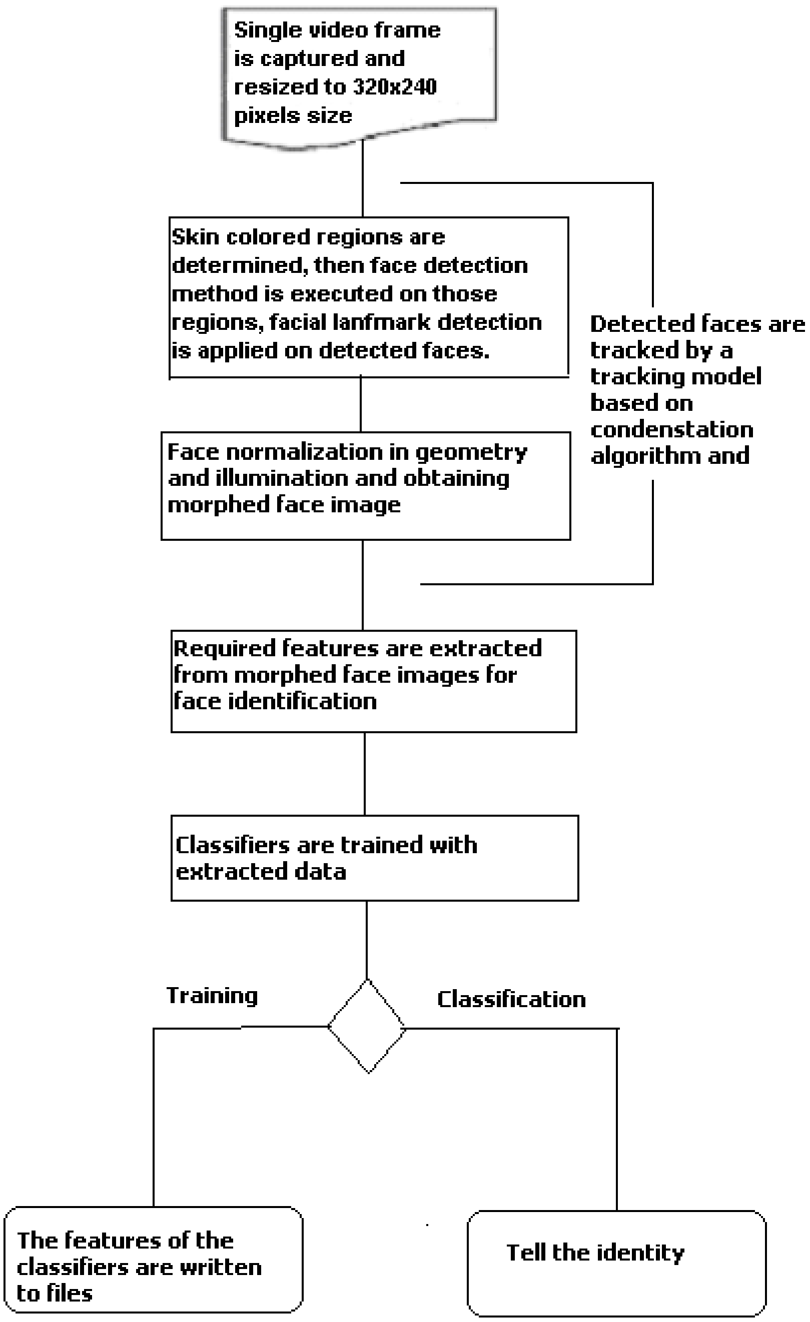


Figure 3-2 The architecture of Face Identification

CHAPTER 4

PRELIMINARY METHODS

4.1 Overview

In both modules of the proposed system (i.e. the Gender & Age Group Classification module and the Face Identification module) video frames are captured and skin-colored regions are identified. The main steps are shown in Figure 4-1.

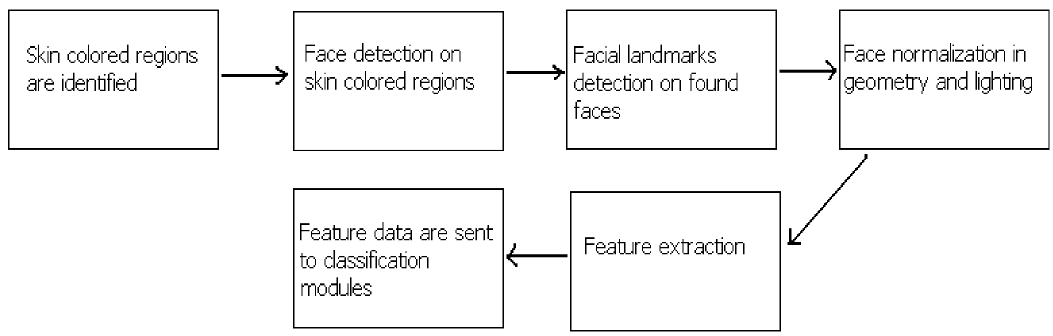


Figure 4-1 Preliminary Methods

The robustness of the skin colored region detection influences the overall success of the system. In both modules, Haar classifier based face detection algorithm is run on the detected skin colored regions. The detected faces are kept in a list. The facial landmarks are detected using Haar classifiers on the detected faces in the gender & age group classification module. In the face identification module, facial landmarks are found by using a model based on the Active Appearance Model (AAM). 67 feature points are determined by this model, and face morphing is applied by using these 67 feature points. 114 triangles are constructed with these points, and a standardized face is obtained by the projection of every point to a template face.

This projection is applied with the help of barycentric coordinate transformation. Then, in both modules, some face normalization algorithms are applied. Feature extraction phase is common in both modules.

4.2 Skin Colored Region Detection

A robust face detector should be executed in order to detect faces on video frames. In our system, the face detection algorithm is run on skin colored regions. The assumption is that, whatever the ethnic group is, the skin color is localized in a precise subset of the chrominance space [66,67]. Therefore, a skin-color probability model is constructed in the form of a bi-dimensional Gaussian function as in [66,67]. The parameters of this function are determined on a learning database. The main aim is to reach a Boolean skin or non-skin decision for each pixel. This can be achieved by setting a threshold on the probability values as in [66,67]. The learned mean vector and the covariance matrix of the bi-dimensional Gaussian model is as follows:

$$\begin{pmatrix} \mu_{Cb} \\ \mu_{Cr} \end{pmatrix} = \begin{pmatrix} -17.0 \\ 24.06 \end{pmatrix}$$

$$\begin{pmatrix} \sigma_{CbCb} & \sigma_{CbCr} \\ \sigma_{CrCb} & \sigma_{CrCr} \end{pmatrix} = \begin{pmatrix} 93.02 & -77.91 \\ -77.91 & 118.02 \end{pmatrix}$$

The resulting skin-colored pixels are compiled with a contouring algorithm. The contours are combined with convex hull construction. Finally, the bounding rectangles of the regions are determined and the coordinates of the rectangles are given to the face detection module. This reduces the required time for face detection. The result is an image of the frame whose non-skin colored regions are darkened, and the skin-colored regions are kept in their original color. An example output of our skin color detection algorithm is illustrated in Figure 4-2.

Figure 4-2 Execution of skin-colored region detection

4.3 Face Detection

A LUT-type boosted cascade classifier based on the same concept of Viola and Jones [65] is used as the face detector. This face detection method uses Haar-like features. These features encode the existence of oriented contrasts between regions in the image and contrasts exhibited by a human face are encoded by using a set of these features [65]. Frontal faces are detected on the captured video frame. There are several robust face detection algorithms for frontal faces in the literature. We used OPENCV [56] face detection implementation on the detected skin colored regions. In the face detector of OpenCV, a classifier is trained with a set of face and non-face images. In this set, face images are named as positive examples, and non-face images as named as negative examples. After the training phase, the classifier is ready to be applied to the captured image. "1" is output if a region is estimated as a face, "0" is output if a region is estimated as non-face. Faces are searched in the whole image. The classifier checks all locations in the image by using a search window. Figure 4-3 illustrates the detection of faces.

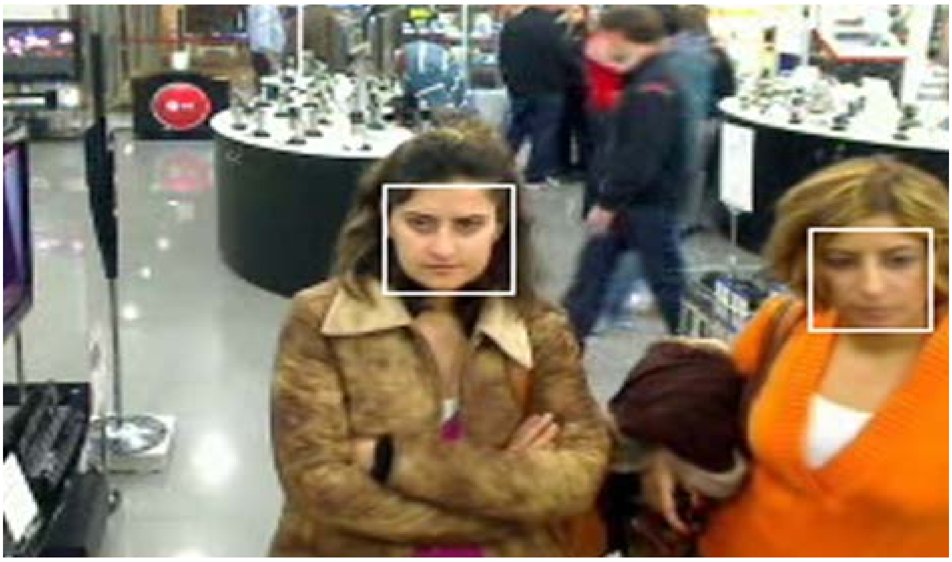


Figure 4-3 Example of face detection

4.4 Facial Landmark Detection

Eyes, mouth and nose are searched on the resized face images. Eyes, mouth and nose detectors are boosted cascade classifiers similar to the face detector. The boosted cascade classifiers are trained in the samples of OPENCV [56]. The coordinates of the rectangles including the mouth, eyes and nose are sent to the face normalization module. The detected eye coordinates in the facial landmark detection module are used to rotate the face image to equalize both eyes' y-coordinates. Figure 4-4 illustrates the detection of facial landmarks.

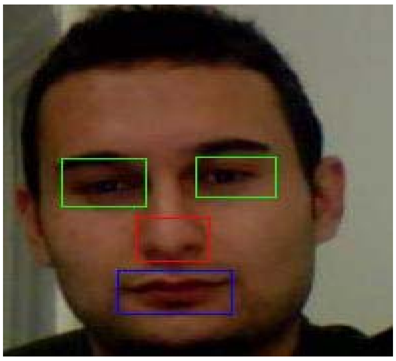


Figure 4-4 Example of facial landmarks detection

4.5 Face Normalization in Geometry and Lighting

Since most of the personal videos are captured in uncontrolled illuminated environments, the illumination compensation is an essential step to deal with the diversity of illumination conditions. Faces are cropped from the original frame and resized to 80x80 pixels size. The histogram equalization method is executed on the face images to equalize the brightness distribution of the image and thereby increase the contrast of the image. After rotation and histogram equalization, we create three sets of images to extract features which will be used by the classifiers. The first set is the histogram equalized images resized to 40x40 pixels. The second set is the histogram equalized images resized to 20x20 pixels. Some transformations like rotation and scaling are applied to map left and right eye locations to (1/4 of image width, 1/4 of image height) and (3/4 of image width, 1/4 of image height), respectively. The third set is the histogram equalized images resized to 40x40 pixels and a rigid transform is computed to map the eyes to (1/4of image width, 1/4 of image height) and (3/4 of image width, 1/4 of image height). As a result images become ready to extract features.

4.6 Face Tracking

The detected faces are tracked by an algorithm based on condensation and template matching. Every detected face is tracked as a human and we collect four different face images of that human. The propagation and factored sampling are iteratively used in condensation algorithm. The first step of the condensation algorithm is to create a set of potential states. The next step is the propagation of samples using a state transition equation. The third step of the algorithm is computing likelihoods from the observation. As a final step, resampling using factoring sampling is performed. The template matching algorithm is used as the confirmatory part of the face tracking module. We assume that the tracking method tracks the faces successfully if the similarity of two phases measured by applying template matching on the existing face and the found face is higher than a fixed threshold. Figure 4-5 shows an example execution of the tracking. Tracking the non-frontal face with the help of template matching can be seen in Figure 4-6.

Figure 4-5 Example execution of face tracking method

Figure 4-6 Tracking with template matching

4.7 Feature Extraction Methods

In this thesis, we have applied some feature extraction methods on the preprocessed face images and used the extracted features as input to classifiers. In our modules, we have not processed the direct gray level values of the cropped faces, since direct pixel values are sensitive to noise and localization errors. In this section, all feature extraction methods used in our study are discussed. Local binary patterns (LBP), discrete cosine transform (DCT Mod2) and histograms of oriented gradients (HOG) methods are explained.

4.7.1 Local Binary Pattern Histogram

Local binary pattern (LBP) is described as an illumination invariant descriptor of the local structure in a given image neighborhood [58]. It is necessary to divide a face image into 8x8 pixels blocks to calculate the LBP feature vector of the image,. The next step is the comparison of each pixel in a block to each of its 8 neighbors. This comparison phase follows the pixels along a circle, clockwise or counter-clockwise. If the center pixel's value is greater than the neighbor, "1" is written, otherwise, "0" is written. This gives an 8-digit binary number which is then converted to decimal format. A 256 bin histogram, over the cell, whose bins are based on the frequency of each number, is computed. The values of the bins of histogram are concatenated to construct the local binary pattern feature vector of the face. The feature vector can then be processed using Random Forest (Random Trees), to produce our first classifier. In Figure 4-7, 8-bit strings for LBP for a pixel are illustrated.

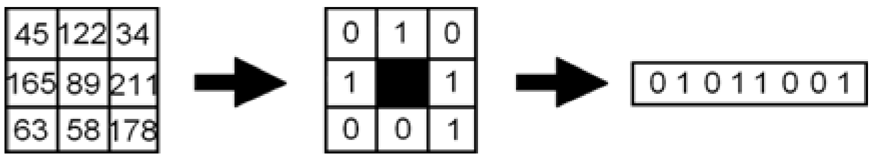


Figure 4-7 8-bit strings for Local Binary Patterns

4.7.2 DCT Mod2

In this method, the set of face images which are 40x40 pixel size, are given to Direct Cosine Transform Mod 2 (DCT Mod2) feature extraction method. The image is divided into (NxN) blocks in this method. An image block is decomposed in terms of orthogonal DCT Mod2 functions. The result is an NxN matrix $C(v, u)$ containing DCT coefficients. Given an image block f(y, x), where y, x = 0, 1, . . . , N-1, th image block is decomposed in terms of orthogonal 2D DCT basis functions [73]:

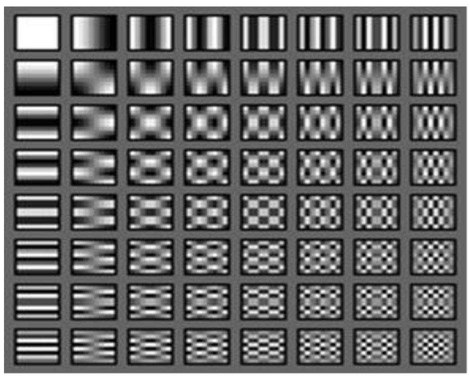
$$C(v, u) = \alpha(v)\alpha(u)\sum_{y=0 \text{ to } N-1} \sum_{x=0 \text{ to } N-1} f(y, x)\beta(y, x, v, u)$$

for $v, u = 0, 1, 2, \ldots, N\text{-}1$, where

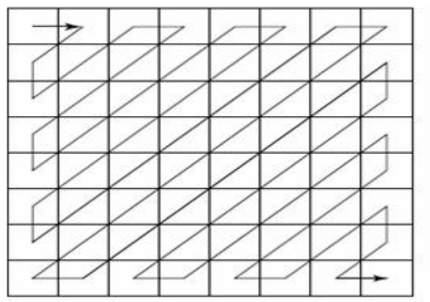
$$\beta(y, x, v, u) = \cos[((2y + 1)v\pi) / 2N] \cos[((2x + 1)u\pi) / 2N]$$

$\alpha(v) = \sqrt{(1 / N)}$ for $v = 0$, and $\alpha(v) = \sqrt{(2 / N)}$ for $v = 1, 2, \ldots, N - 1$

For block located at (b, a), the DCT feature vector is composed of the obtained DCT coefficients. DCT coefficients are ordered in a zig-zag scanning as shown in Figure 4-8. The changes in the illumination direction do not affect the success of the DCT-mod2 approach. In our implementation, we use 48x36 pixels size image and our block size is 8x8 pixels. During the training phase, DCT Mod2 features of all training images are extracted and these feature vectors are written to a file. After this phase, Random Forest classifier is trained with these features. In Figure 4-8, the zig-zag scanning of DCT coefficients and 2D DCT basis functions are shown.



2D DCT Basis Functions



Zig-Zag Scanning of DCT Coefficients

Figure 4-8 DCT Mod2 basis functions and zig-zag scanning

4.7.3 Histogram of Oriented Gradients (HOG)

Our third feature extraction method is Histogram of oriented gradients (HOG). It is usually used for human detection and pedestrian detection. In this method 8x8 pixels blocks are used. In our implementation, morphed faces are resized to 48x36 pixels.

Our morphing method constructs the face with the "width/height" ratio 4/3. In earlier studies, the best success rates on face identification are achieved with the resolutions around 40x40 pixels. 48x36 pixels of size are applicable for our implementation. The algorithm is as follows [81]: The face image is divided into blocks, and horizontal and vertical gradients are extracted. These gradients are obtained by one dimensional derivative. Using these gradients, the angles and magnitude of gradients are calculated. This calculation is performed for every pixel in the blocks. A histogram is calculated for each block. The histogram is named as orientation histogram. The magnitudes of the gradients are scaled and the weights are calculated by these magnitudes. In our implementation, HOG method constructs nine bins for $-\infty$ to $+\infty$. We used 2x2 pixels block which are accumulated to yield a block of 12x12 pixels, and the weights are multiplied within the block so that weights are scaled. The descriptor of a block is obtained by appending four scaled histograms within that block. Appending obtained feature descriptors yields to the overall feature vector. This vector is used as the input of classifiers.

CHAPTER 5

GENDER AND AGE CLASSIFICATION

In this chapter, the methods used in the gender and age classification module are described. Namely, Random Trees classifier, pixel LBP values comparison and adaboost algorithm are explained. Then, the use of these methods is summarized for each classification module separately.

5.1 Random Forest – Random Trees

Random Forest is defined as a set of tree estimators whose classification process starts with taking the input vector and classifying it with every tree in the forest [68, 71]. In our implementation, we defined our Random Forest classifier with 100 trees. The input vector is classified by all trees in the random forest and the output is the label which received the majority of votes. It is suggested that there is no need for any accuracy estimation procedures, such as cross-validation or bootstrap in random trees [68]. In random forest, the error is estimated internally during training. It is assumed that when the training set for the current tree is drawn by sampling with replacement, some vectors are left out which are called as out-of-bag data [68]. In [71], one third of the number of vectors is assigned as out-of-bag data. In [71], the classification error is estimated by using this out-of-bag data by getting estimations for all vectors. All trees are trained in the forest. Non out-of-bag vectors are given as input to these trees and the class that received the majority of votes where the vector was out-of-bag is determined. If this class matches to the ground truth response, it is determined as “correctly classified”; “misclassified” otherwise. Then, the number of misclassified out-of-bag vectors is compared to the number of vectors in the original data to obtain a ratio. This ratio is the classification error estimate. In the case of regression the out-of-bag error is computed as the squared error for out-of-bag vectors difference divided by the total number of vectors [71].

5.2 Pixels LBP Values and Adaboost

Our second classifier is an Adaboost classifier which is trained with the data extracted from comparisons of LBP values of pixels on 20x20 pixels face image. Face images are histogram equalized images resized to 20x20 pixels and a transform is applied which maps the eyes to defined positions. We used ten types of pixel comparison operators:

- LBP value of pixel1 > LBP value of pixel2
- LBP value of pixel1 < 5 * LBP value of pixel2
- LBP value of pixel1 < 10 * LBP value of pixel2
- LBP value of pixel1 < 25 * LBP value of pixel2
- LBP value of pixel1 < 50 * LBP value of pixel2
- LBP value of pixel2 > LBP value of pixel1
- LBP value of pixel2 < 5 * LBP value of pixel1
- LBP value of pixel2 < 10 * LBP value of pixel1
- LBP value of pixel2 < 25 * LBP value of pixel1
- LBP value of pixel2 < 50 * LBP value of pixel1

This algorithm is the same algorithm as the "Boosting Sex Identification Performance" algorithm [19] except for the comparison operators. In this algorithm, several binary features are produced by pixel comparison operations. We can use these binary features as weak classifiers whose accuracy rates are slightly better than random chance. The output of the classifier is the value of the binary feature. If the value of any binary feature is true or "1", the output is male, otherwise female. In this method, we used 20x20 pixels face image and this yields to 10x400x399 = 1596000 distinct weak classifiers. Here Adaboost algorithm is used to combine these weak classifiers together. Its primary goal is to form a single strong classifier with

1200 selected weak classifiers. The adaboost algorithm can be described as follows [72]:

Given $(x_1, y_1), \ldots, (x_m, y_m)$ where $x_i \in X$, $y_i \in Y = \{-1, +1\}$

Initialize $D_1(i) = 1/m$, $i = 1, \ldots, m$.

For $t = 1, \ldots, T$:

- Find the classifier $h_t : X \rightarrow \{-1, +1\}$ that minimizes the error with respect to the distribution D_t:

  $h_t = \arg\min \varepsilon_j$, where $\varepsilon_j = \sum_{i=1 \text{ to } m} D_t(i) \, [\, y_i \neq h_j(x_i)]$

- If $h_t >= 0.5$ then stop.
- Choose $\alpha_t \in \mathbf{R}$, typically $\alpha_t = ½ \ln((1 - \varepsilon_t) / (\varepsilon_t))$ where εt is the weighted error rate of classifier h_t.
- Update $D_{t+1}(i) = (D_t(i) \exp(-\alpha_t . y_i . h_t(x_i))) / Z_t$ where Z_t is a normalization factor.

Output the final classifier:

$H(x) = \text{sign}(\sum \alpha_t h_t(x))$

The distribution D_t is updated by:

$$e^{-\alpha_t y_i h_t(x_i)} \begin{cases} < 1, & y(i) = h_t(x_i) \\ > 1, & y(i) \neq h_t(x_i) \end{cases}$$

The accuracies of all weak classifiers are computed in all iterations. This is a time consuming process. However, this process affects only the training time, it does not have any effect on the classification time. To reduce some of the processing time, the system selects weak classifiers randomly in all iterations. This reduction does not have any negative effect on the success rate of the final classifier. 15% of all weak classifiers are selected randomly in our algorithm. We select the best 1200 weak classifiers to construct a strong classifier. Here comes a question like, “How are the weak classifiers sorted?” In other words, “what is the determiner of the best weak classifiers?” We sort weak classifiers by their accuracy on the training images. If a training image is a male face image, and the weak classifier gives the correct output for that image, the weak classifier’s point is incremented by one. All randomly selected weak classifiers are graded in this way, and finally they are sorted by their total point as in [19]. Our algorithm selects the best 1200 weak classifiers, and writes their pixel coordinates and their operators to a file which will be used to build the strong classifier.

5.3 Summary of Gender Identification

The gender classification module consists of two classifiers. This module takes two face image sets with sizes 40x40 pixels and 20x20 pixels which are prepared by the "face, facial landmark detection and face normalization module". In the classification mode, the first classifier, namely 'Random Forest Classifier' takes the 40x40 pixels image set as input and it extracts the LBP feature vector. Then it predicts the genders of the face images separately with the trained classifier. If the classifier prediction on face images for a human contains more "male" results, the decision "male" is output for this classifier. The second classifier, namely "Adaboost Classifier" takes the 20x20 pixels image set as input and calculates the 1200 selected (weak classifiers) binary operation results, and calculates the genders of the faces. Finally, if the results of these two classifiers are same, the gender of the face is determined as the result of the first classifier directly. If the results of two classifiers are not same, we use all output values of the two classifiers to give the final decision. In our implementation, Random Forest classifier outputs values for every class.The class label which has a greater value is selected as the final decision.

5.4 Age Group Classifiers

In age classification we have four distinct classes. We separate human ages into three classes: young aged, middle aged and old aged. We labeled these classes as A, B and C. In order to classify the age group of faces, we developed two distinct age classifiers in the age classification module.

Our first classifier is based on DCT features of the extracted face and the Random Forest classifier. The ratios between facial landmarks and the wrinkles on the face is the main focus of our first classifier. We can get the information about the relations between facial feature points and the wrinkles by extracting the DCT features of the face. Our Random Forest classifier uses these DCT features as the input. During the training phase, DCT feature vectors of all training images are extracted and these feature vectors are written to a file. After this phase, Random Forest classifier is trained with these features.

Our second classifier is based on LBP features of the selected face regions and the Random Forest classifier again. In this classification, LBP features of some selected regions of the face are computed as a first step. Wrinkle features around some regions on a face can give some clues about the age of a person. Wrinkle structures around the eyes, cheeks and foreheads have different characteristics which can differ by age.

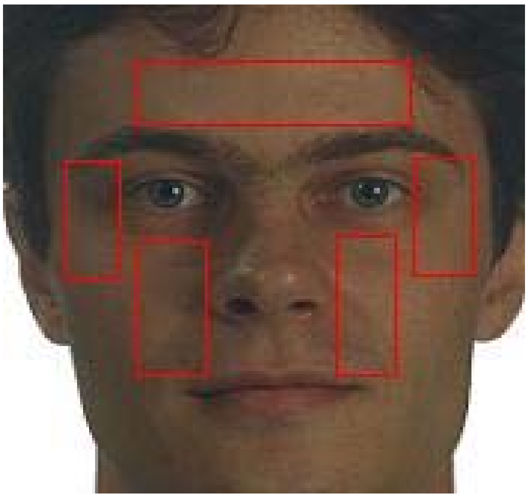


Figure 5-1 Selected regions for age classification

The calculation of LBP features yields to a 256 bin histogram. The vector with the values of the bins can be used as a feature vector. During the training phase, LBP feature vectors of all training images are extracted and these feature vectors are written to a file. After this phase, the Random Forest classifier is trained with these features.

5.5 Summary of Age Group Identification

The age group classification module consists of two classifiers. In the classification mode, the first classifier, namely ‘DCT Mod2 Features Random Forest Classifier’ takes the 40x40 pixels image set as input and it extracts DCT Mod2 feature vectors, then it predicts the age group of the face images with the trained classifier. The output of this classifier is the average of the results of all predictions on the given face image set of a person. The second classifier, namely “Region’s LBP Features Random Forest Classifier” takes the 80x80 pixels image set as input and it extracts LBP feature vectors, then it predicts the age group of the face images with the trained classifier. In our implementation, we give four face images of the person to the first and the second classifier. Our classifiers output eight results for the given face image set of a person. The final output is determined by averaging the results of both classifiers and taking the floor of the average value.

CHAPTER 6

FACE IDENTIFICATION

6.1 Overview

In this chapter, the details of the methods used in face identification are described. All methods are explained in detail and the success rates of facial identification with these methods are presented. The experiments for the evaluation of the performance and execution times of the methods are summarized. The reasoning behind the final choice of the methods is discussed. In face identification module, feature extraction methods like LBP, DCT Mod2, and HOG are used and compared against each other. A nearest neighbor (NN) classifier is implemented whose feature weights are provided by a Random Forest classifier. All feature extraction methods are tested with the nearest neighbor classifier and their execution times and recognition precisions are compared. The execution times and success rates of all methods are plotted and presented in Section 6.6. To identify a person, first of all an illumination compensation method is executed on the detected faces. Then 67 face feature points are found using the active appearance model. Next, a morphing procedure is applied to compensate the facial expression variations. After morphing the face, the feature extraction method is applied and the extracted features are used as input to classifiers.

In this chapter, the illumination compensation algorithms are described first. After that, facial landmarks detection using the active appearance model is described and face morphing is explained. Then the feature extraction and classification methods are explained. Finally the execution times and recognition accuracies of the methods for different conditions are presented.

6.2 Illumination Compensation

The illumination compensation removes the illumination variations so that the virtual variations on faces caused by the illumination changes can be discarded. Personal videos are usually captured in uncontrolled environments; therefore illumination compensation is an essential step to deal with the negative effects of illumination conditions. Illumination compensation is performed in order to increase the recognition performance. The PCA is adopted by discarding the first few principal components and a better performance is achieved for images under different lighting conditions in [75] and [76]. In [77] LBP features with Gabor wavelets are used. In [78], researchers used an extension of LBP for illumination compensation.

In our system we adopted the approach of [20] for illumination normalization. In this approach, face images can be represented by their intensities at every pixel (x, y), therefore face images can be determined as a 2D function I(x, y) where x is the location of the point on x-axis and y represents the location of the point on y-axis. The intensity of a sample image can be represented as the product of source illumination incident on the frame and the amount of reflection in the frame. We can describe this situation with the equation

$$I(x, y) = R(x, y) * L(x, y)$$

where R(x, y) represents the reflectance equation at point (x, y) and L(x, y) represents the illumination incident at point (x, y). This equation is the basis for the illumination-reflectance model to normalize an image captured under poor illumination conditions. Figure 6-1 illustrates the result of our illumination normalization method.

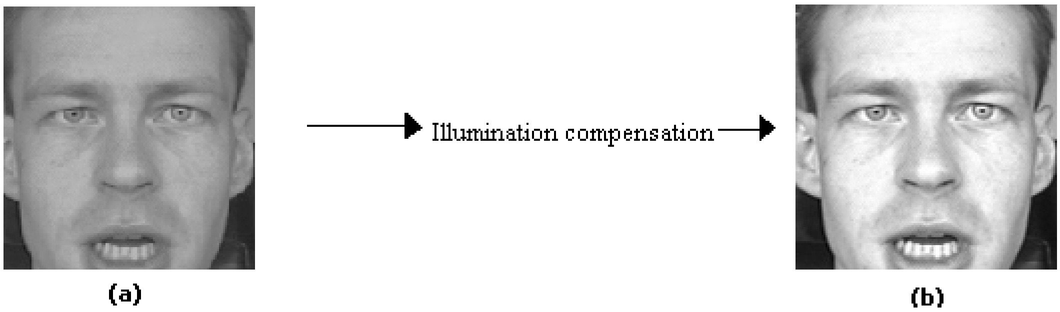


Figure 6-1 Illumination normalization

6.3 Facial Landmarks Detection for Face Identification

Pose and expression are the other most important factors for a successful face recognition system. There are three types of methods to compensate the negative effects of variations in pose and expression: multiple images approach, hybrid approach and single image based approach [79]. In our case, we will focus on single image based approaches. New AAM methods have been proposed to handle both varying pose and expression [79]. In our study, we have implemented an AAM based face morphing approach for varying pose and expressions. For the success of the face recognition module, 67 face feature points on the detected face are found using the active appearance model; then a morphing procedure is applied to compensate the facial expression variations. After morphing the face, the feature extraction method is applied and the extracted features are used as input to classifiers.

In AAM, constructing shape normalized face is an essential step. Warping the texture of the face to a constructed shape normalizes the shape of the face. In active appearance model, the mean texture can be obtained by the shape-normalized faces [79]. Texture variations can be modeled by a set of principle component analysis. Active appearance model can detect the feature points of the face. In our implementation, we first collected a face image gallery which includes 250 unexpressioned face images. We simply marked the feature points of all face images manually and prepared for training. A statistical model of shape and texture variation is generated [79]. Let s is a shape vector of an object and t is a texture vector of the object, and C is the parameter controlling the texture and shape, we obtain the equation as in [79]:

$$s = Q_s\, C + s'$$

$$t = Q_t\, C + t'$$

where t' is the mean shape, s' is the mean texture and Q_s, Q_t are matrices describing the modes of variation derived from the training set. We trained two different models with the gallery images. A texture representation is generated from the vector **t**. In the above equation, the control points are described by t. The texture representation is warped using these control points. After the training phase, a model which enables us to detect 67 feature points on a face is created. Figure 6-2 shows the detection of 67 landmark points on some example faces.

Figure 6-2 Detected facial landmarks by AAM

After the detection of facial landmark points, our system fits a generic mask to reference faces and compensate the variety of expressions and pose. As stated in [80] densely placed vertexes cause a registration difficulty. Triangulated masks eliminate the registration difficulty. In this thesis, a triangulated generic mask is used. Figure 6-3 shows our generic mask.

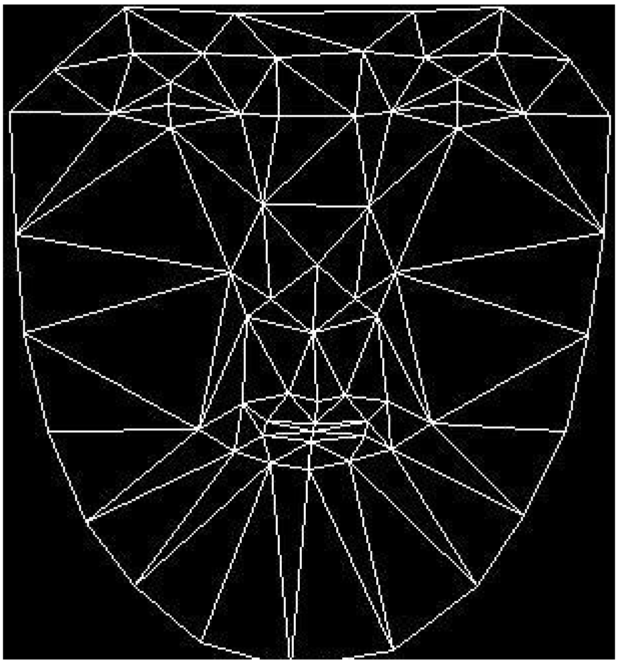


Figure 6-3 Generic face mask

There are 67 vertices and 113 triangles in our generic mask. We use these triangles to morph the given image to a generic mask. The aim of morphing is to find an average between faces. Here, the "average" means, obtaining a frontal and unexpressioned face. There are some methods to obtain the average between faces like cross-dissolve and global alignment [80]. However, in our case cross-dissolve and global alignment do not work. Feature matching is a solution to this problem. Nose to nose, eyes to eyes, mouth to mouth, cheek to cheek matching is the basis for feature matching. This is a local, non-parametric warp.

In this work, we have created a triangulation for non-parametric warping. In [82], triangulation is defined as a set of points in a plane. A convex hull can represent the triangulation of points in a plane. The vertices of the convex hull are the points of the triangulation. 67 feature points found by AAM is the set of points in the plane. Then one-to-one interpolation phase comes. Every pixel in the source image is interpolated to the corresponding point in the generic face mask. This process is done by interpolation using triangles. In Figure 6-3, the triangles can be seen.

The interpolation using triangles is based on Barycentric Coordinates [10]. In this coordinate system, any point in a triangle can be determined by an equation based on the weights of the vertices of the triangle. Figure 6-4 and Figure 6-5 shows the equation and point representations in Barycentric Coordinates.

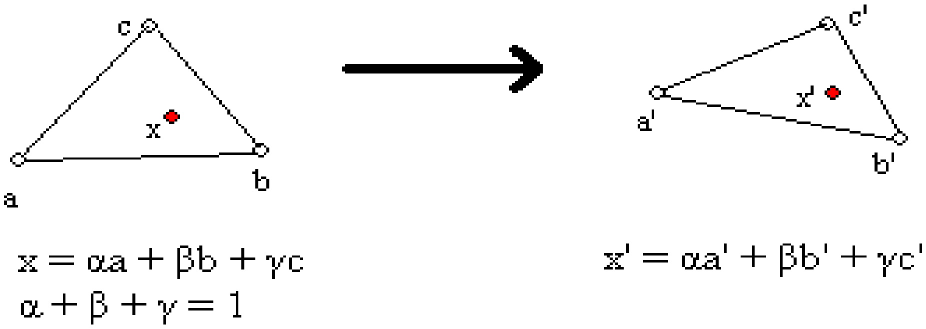


Figure 6-4 Barycentric Coordinates representation of points

There is a linear equation in interpolation using triangles by barycentric coordinates shown in Figure 6-5.

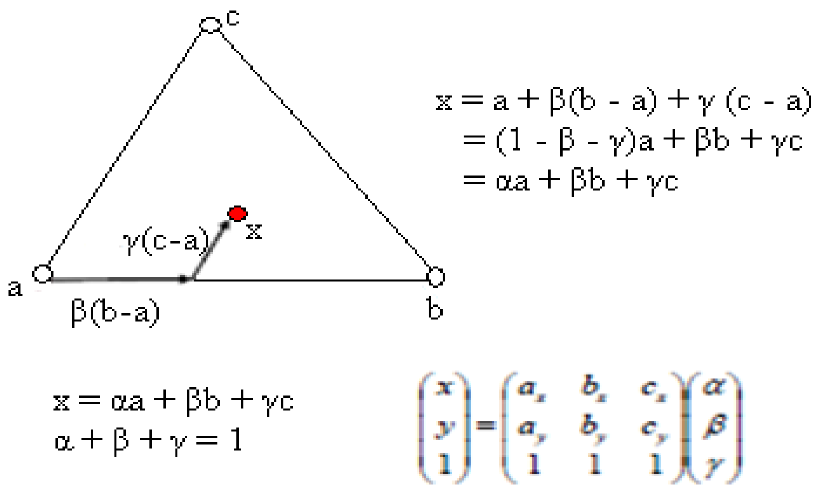


Figure 6-5 Barycentric Coordinates equations

The first step in this interpolation is finding out where each pixel in the new image comes from in the old image. It is first necessary to determine in which triangle the pixel in the raw image is, in order to determine the new position of the pixel on the new image.

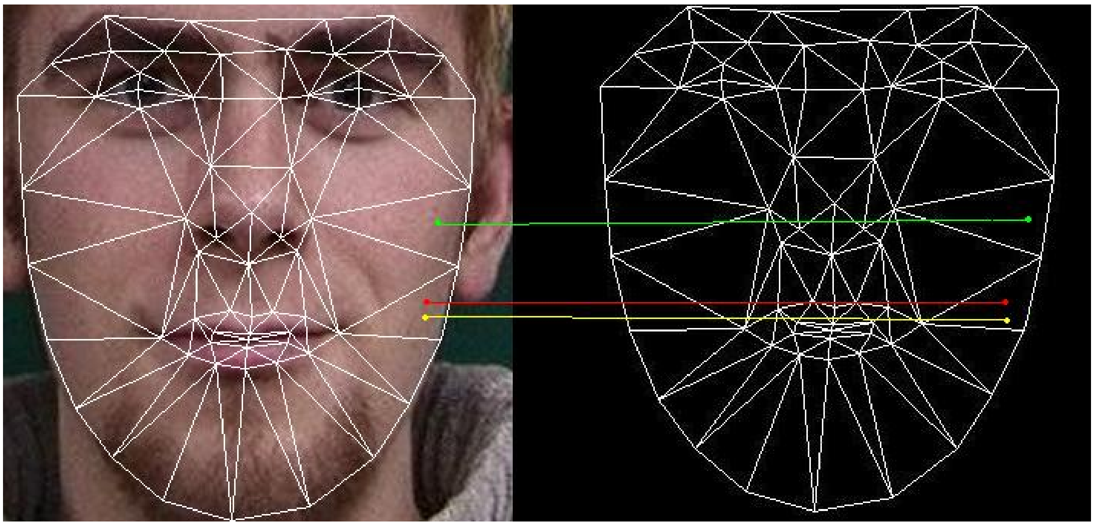


Figure 6-6 Interpolation of pixels

After the interpolation phase, unexpressioned and pure frontal face image is obtained. This image will be used in the feature extraction phase and then in the classification phase.

6.4 Classifier

The popular classifiers such as SVMs, Neural Networks and Random Forests do not consider the classes that are not included in the training set while classification. In our case, modeling the identification of an unknown identity is necessary. Our system should be able to discriminate the unknown faces from the known faces. Therefore we decided to implement a nearest neighbor classifier to correctly determine whether the face is identified and its identity is defined in the training set or not. Our classifier searches for the likelihood of an input face recognized as a known identity or not. In this thesis, we have used a nearest neighbor classifier whose weights are determined by a random forest classifier trained with the same data. A brief explanation of random forests – random trees is given in Section 5.2. In this section, we give a brief explanation of our nearest neighbor classifier.

In [83], nearest neighbor is defined as a method for classifying objects based on the closest training feature in the feature space and in this classification algorithm an object is classified by the majority votes of its neighbors, i.e. the object is assigned to the most common class amongst its nearest neighbors. The training data is used to train a random forest classifier to obtain weights of the features. The obtained weights are used in the voting phase of the nearest neighbor algorithm. If the training vectors include n features, a vector which consists of n weights will be obtained. In our implementation, after obtaining the weights of all features, we calculate the average values of the features for all classes of the training set. The results of these calculations constitute the basis of the classifier. The average values and the weight are used in the classification phase of the classifier. The distance to these values of the input feature vector is calculated by using the Euclidean distance.

The main drawback of the nearest neighbor approach is the domination of the class with more frequent examples in basic majority voting classification [83]. Our approach to solve this problem is to use the calculated weights obtained from the random forest classifier. The training time is the other problem for this type of classifier. The increase in the identification time as the database grows is another problem for the nearest neighbor classification. However, in our case, the identification time is not a big problem as demostrated in our experiments.

6.5 Evaluation of the Feature Extraction Methods

6.5.1 Overview

In our experiments, we measured the processing times and the performance of the algorithms. The processing times and the identification success rates are plotted against different testing samples. In these experiments, the effects of changing the illumination and varying pose and the expression of faces are tested. Therefore, testing is done with various videos which include different environmental situations.

After the training session, several videos are used for the identification testing session. One main assumption is that there are no sudden variations around the scenes. Our tracking algorithm works very well with this assumption. We trained our classifier with a video in which all people we want to search were annotated and we tested the classifier with other videos in which all people we annoted appeared. We used 10-fold cross validation technique. The video database is divided into 10 groups. We used one group for training and the remaining nine groups for testing. The average of the ten test results gives the final precision value. In our experiments we used the "Coupling" (TV Series) episodes as the training and testing videos.

6.5.2 Processing Times

The processing time of an image processing algorithm is important, since the image processing algorithms are process heavy algorithms. In this thesis, the running times of all feature extraction algorithms with our classifier are measured and compared against each other. All of the experiments are performed on a personal notebook whose processor is Intel Core 2 Duo 2.17 GHz. All the methods are implemented in C++. In Figure 6-7, the processing time plots can be seen.

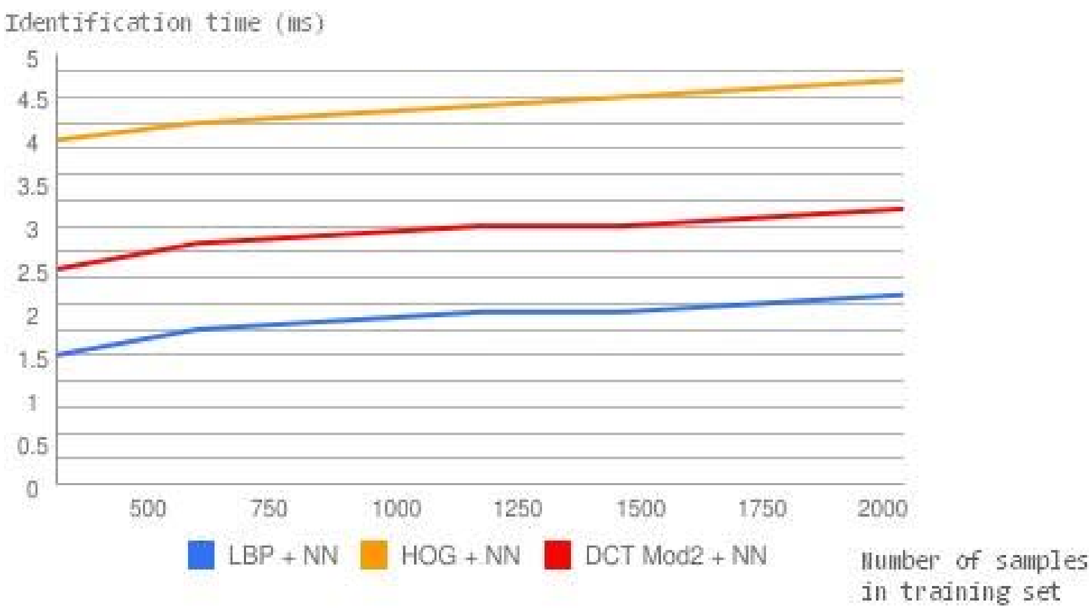


Figure 6-7 Identification time versus the number of samples in the training set

The processing times of the feature extraction methods do not differ in big scales. All the results are reasonable. The processing times increase linearly with the increasing number of training samples with a small bias value. The identification times in the above graphs are measured for the captured frames in which four faces are detected.

6.5.3 Identification Success Rates

The most important part of our experiments is measuring the success rates of the extraction methods. In our first experiment, we work on seven different target classes and the number of training samples was different for all classes. Figure 6.8 shows the results of this experiment. In this experiment, the classification accuracy is compared with the number of face samples for each class for each of the feature extraction methods. In this experiment, there were five classes in the training set.

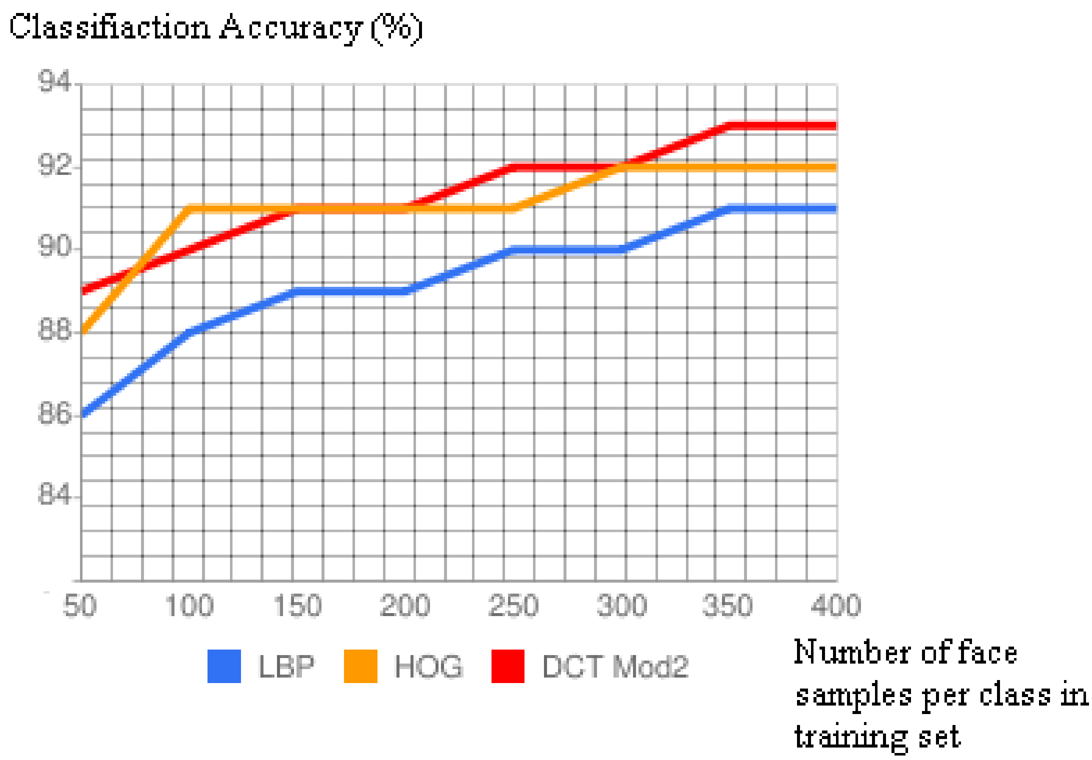


Figure 6-8 Accuracy rate vs. number of samples for each class in training set

The trend of increase in success rates in this graph is due to the enrichment of the classifiers knowledgebase by the increasing number of face samples. To summarize, if the number of face samples for each class increases in training, the success rate of our classifier also increases. The results of our first experiment showed that DCT Mod2 feature extraction method gives the best accuracy values. HOG – Histogram of Oriented Gradients feature exraction method gives the second best classification success rates. LBP histogram feature exteaciton performance is the worst of the three inspected methods.

In our second experiment, we measured the success rates of our feature extraction methods against the increasing class counts. In this experiment we kept the training sample counts constant. In our training set, there were 100 samples for each class and the success rates were measured while the class count in the training set increased. Figure 6-9 shows the results of this experiment.

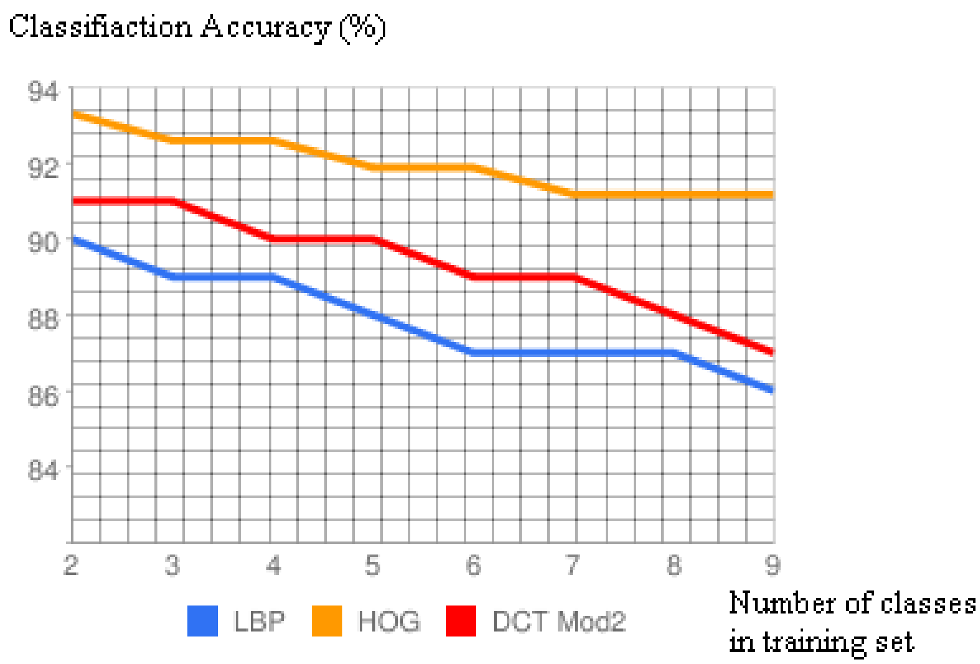


Figure 6-9 Identification accuracy rate versus the number of classes

The trend of decrease in the success rates in this graph shows the fact that while the number of classes increases, the diversity of the model also increases and the decision to choose one of the classes becomes more complex. To summarize, if the number of classes increases while keeping the sample count constant, the success rate of our classifier decreases. The results of our second experiment have showed that HOG – Histogram of Oriented Gradients feature extraction method gives the best accuracy values. DCT Mod2 feature extraction method gives the second best classification success rates. LBP histogram feature extraction performance is the worst of the three inspected methods.

In our third experiment, we compared the training time of our classifier against the increasing number of training samples. We have measured the training times for six classes with 50, 100, 200, 300 and 400 training samples for each class. Figure 6-10 shows the results of this experiment. The training time of the system increases while the number of training samples increases. This is an expected result and these training times are not problematic since the training procedure is applied once in our case and our classifier is then used to identify the faces.

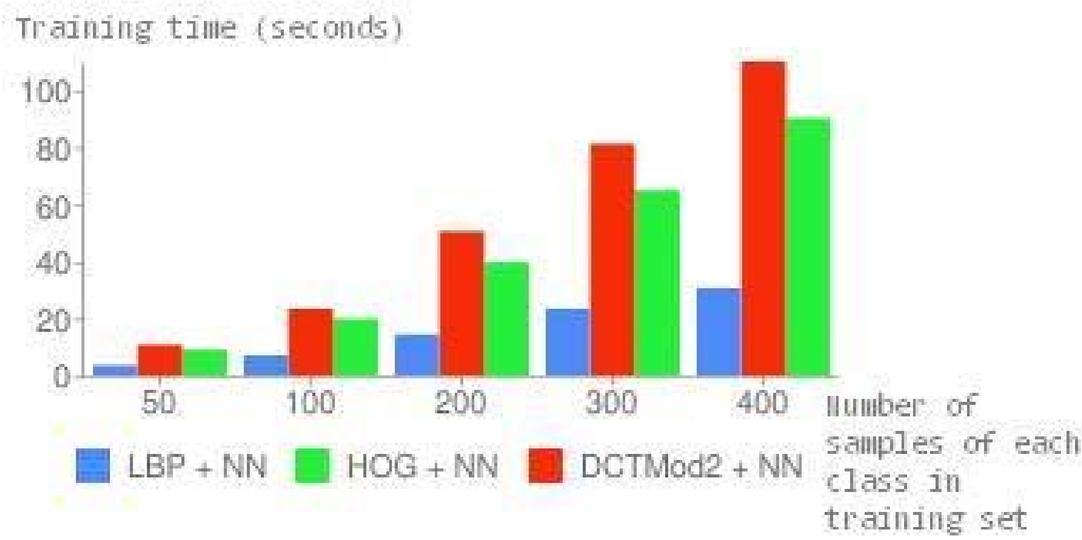


Figure 6-10 Graph of training times

6.5.4 Observations

As discussed before, DCT Mod2 feature extraction method gives the best accuracy results for a limited class count and varying training sample count. Histogram of Oriented Gradients feature extraction method gives the best accuracy values for unlimited class count and constant training sample count. In real world applications, the class count cannot be limited. Therefore, the case in our second experiment is applicable for a real world implementation. The execution times of the inspected feature extraction methods do not differ sharply. We can ignore the results of the execution time experiments to select the feature extraction method for our final implementation. Therefore we have decided to use HOG feature extraction method with our nearest neighbor classifier.

CHAPTER 7

CASE STUDY: ONTOLOGICAL ANNOTATION

Nowadays, as the connection speed of the internet is faster than ever before, people can access video clips, movies and all types of video content. The popularity of personal cameras has increased the amount of personal videos. Because of this situation, effective search tools have become a requirement to access the desired content. Here, semantic annotation of videos becomes a necessity. Video contents can be grouped and labeled by semantic annotation tools. If videos can be annotated successfully, users can reach the desired content. Search queries could be much more specific and target oriented.

We can define ontological annotation as a comprehensive approach to the treatment of multimedia data. Semantic video annotation can be used as an indexing method for multimedia databases or personal videos. This concept is based on automatically assigning metadata in the form of captioning or keywords or the image processing results to a video. Computer vision algorithms can be used to extract valuable information from video contents. The people who appeared in a video, the objects like cars, balls, etc., personal information about the people such as gender and age information can be used as metadata to annotate a video. Thus, our methods can be used as a personal information extractor for a semantic video annotation framework.

7.1 Case Study: Annotation Tool

In this thesis, an example semantic video annotation tool is constructed. In this tool, the video to be processed is selected with a file dialog opened by pressing the “Browse” button. In the training mode, when a new person is detected, the video is paused and the face of the detected person is illustrated in snapview panel of the main window. The user is prompted to add the person to the training set or discard him/her. If the user wants to add the person to the training set, the user should give a name to the person. After adding or discarding, the video is resumed. Then, the tool automatically identifies those people in other videos. Similarly, in this tool, some gender and age data are given and the classifiers are trained with those data. This tool predicts the age group and gender information of the all the detected faces in the videos.

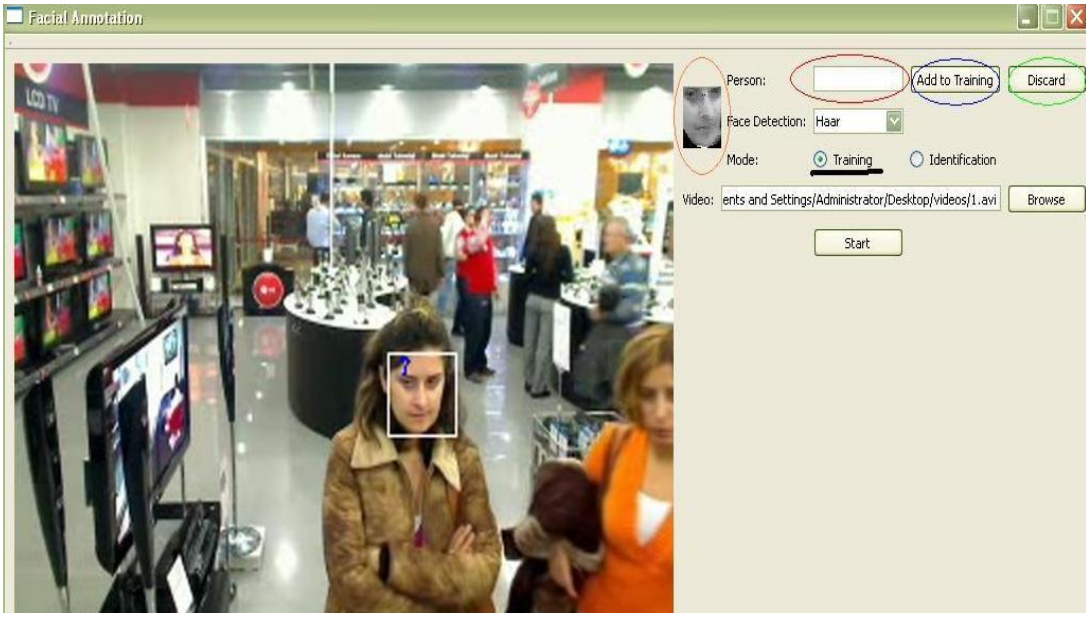


Figure 7-1 Training mode of the face identification tool

Figure 7-1 shows the execution in the training mode. In the training mode, faces detected on the frame are shown as a snapshot and the user is able to insert the annotation for the face and add the face features to the training set. If the detected face is not suitable for training, the user can discard it. If a person is added to the training set or is discarded, he/she will be tracked with the condensation based face tracker discussed in section 4-6 and all faces found for him/her is added to the training set or discarded by the system.

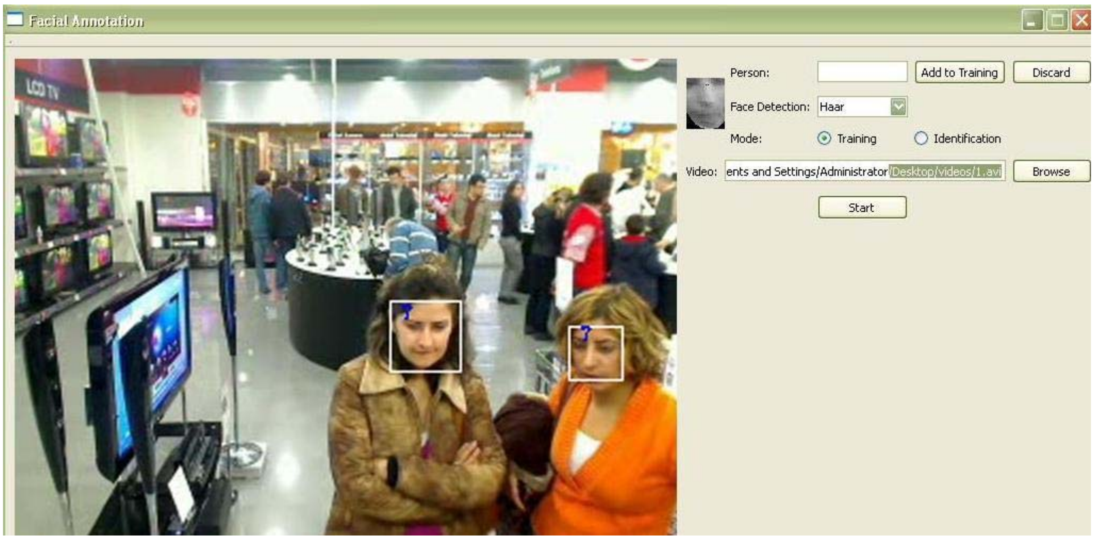


Figure 7-2 Detected faces in the training mode

Figure 7-2 shows the training mode of the system. The detected faces are marked by the system and the user can add the detected faces to the training set.

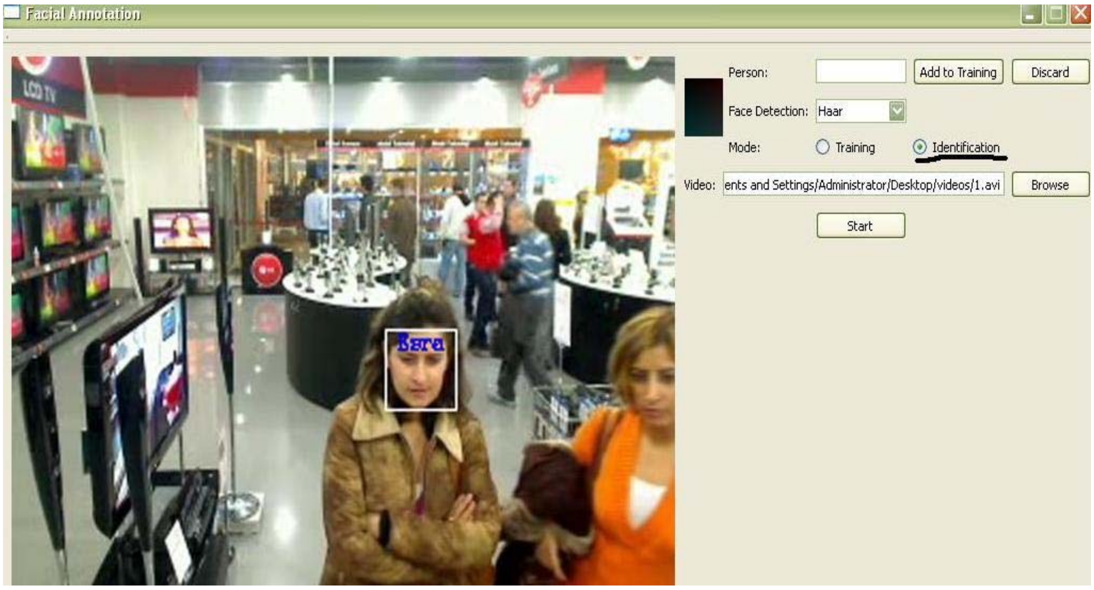


Figure 7-3 Classification mode of the face identification

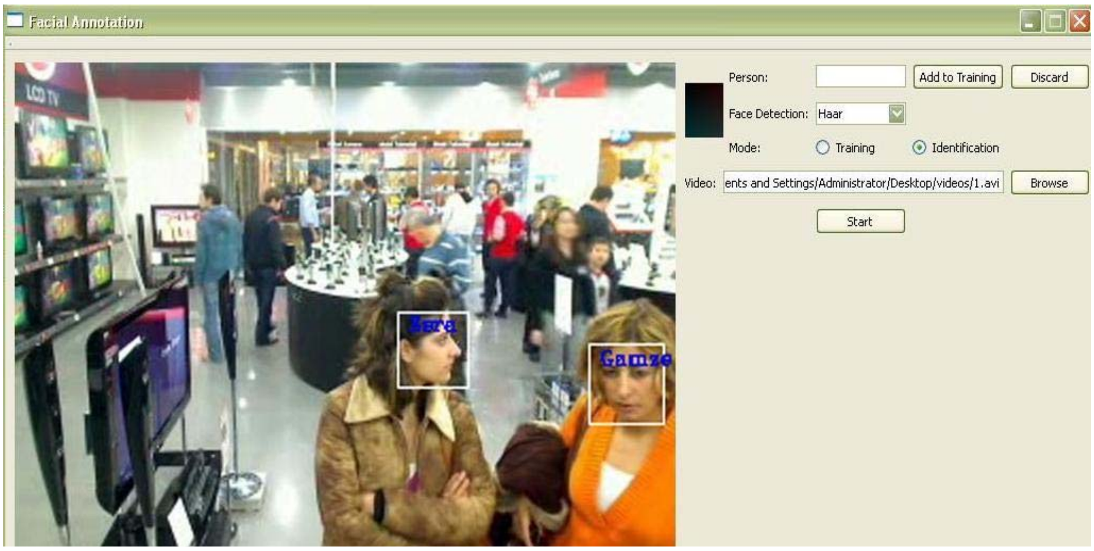


Figure 7-4 Identified people in the classification mode

Figures 7-3 and 7-4 show the identification mode of the system. In the identification mode, the tool finds all faces and determines the identity of the person by the classifiers. The identity is shown by a graphical user interface, and the times (frame numbers) of the appearance of the found people are saved to a file. If a person cannot be identified (this means that he/she is not identified in training set), '?' character is shown on that person and the appearance time of that person is not saved.

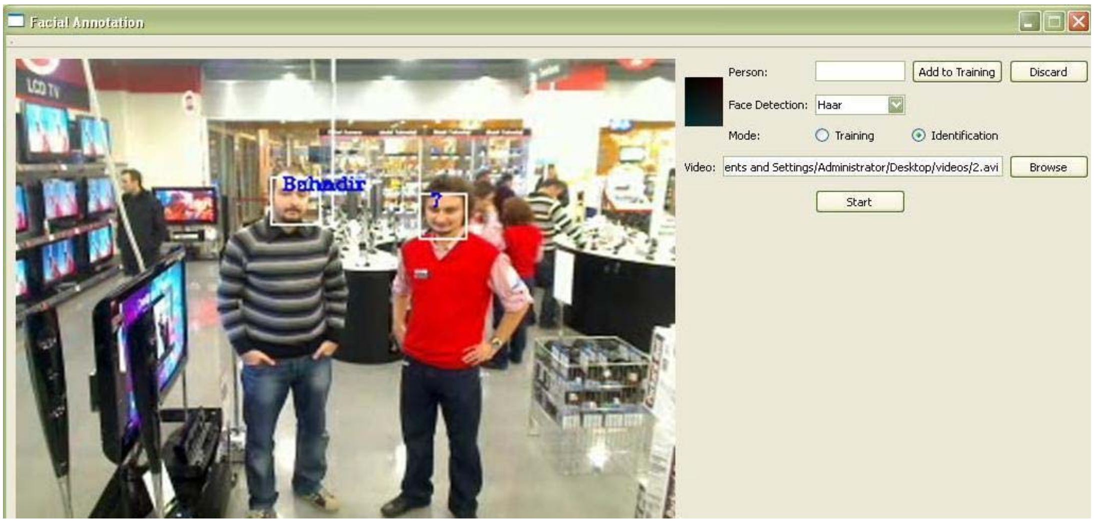


Figure 7-5 Identified and not-identified people in the classification mode

The identification module constructs an annotation file which includes the appearance times of the identified people.

This tool also includes the gender and age group classification module. Some gender and age data are given and the classifiers are trained with these data. This tool predicts the age group and gender classification results of the detected faces in all videos. Figure 7-6 shows an example execution of our gender and age group classifier.

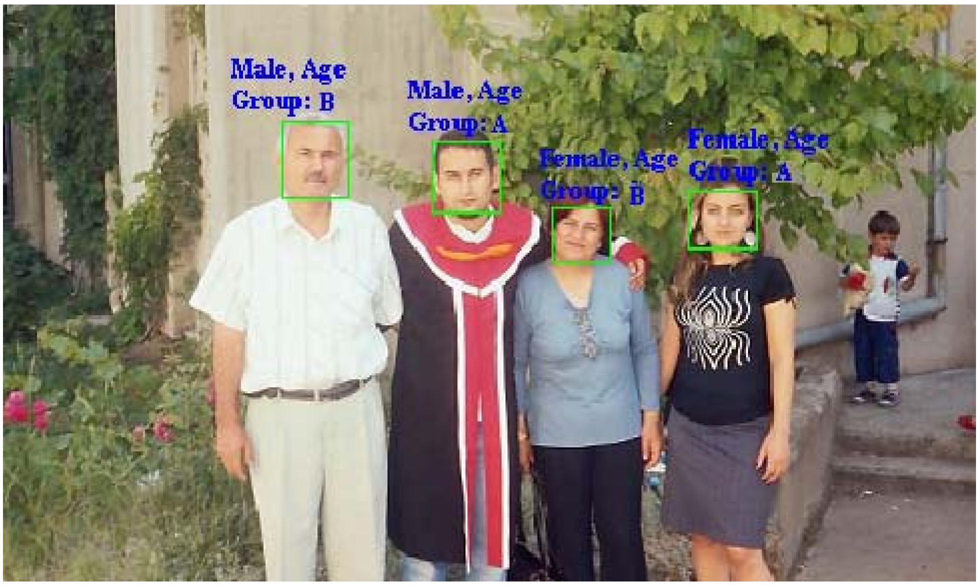


Figure 7-6 A snapshot from the gender and age group classification module

CHAPTER 8

EVALUATION OF THE SYSTEM MODULES

8.1 The Experiments related to Gender Classification

In this section we discuss the results of our experiments on gender classification and compare our results with the results of the study [88]. In our experiments we used FG-NET longitudinal face database [89]. 1002 color and gray face images of 82 people are contained in this database. The faces are grouped by their ages. The ages of the people range from 0 to 70. In [88], this database is used for the experiments of their gender classification algorithm. Table 8-1 shows the distribution of the face images into the age groups in this database.

	Male	**Female**	**Both**
Young (0-18)	389 (38.82%)	298 (29.74%)	687 (68.56%)
Adult (19-55)	161 (16.07%)	145 (14.47%)	306 (30.54%)
Senior (56-70)	5 (0.50%)	4 (0.40%)	9 (0.90%)
All Ages (0-70)	555 (55.39%)	447 (44.61%)	1002 (100%)

Table 8-1 Distribution of the face images into ages in FG-NET database

In our experiments, face detection and facial landmark detection methods were executed on the images. The detected faces were stored with the coordinates of their eyes, nose and mouth. Then face normalization methods were executed. Feature extraction methods were applied and feature vectors were stored in related files.

Classifiers were trained with the stored values and classifier data were stored in xml files for Random Forest and text file for Adaboost. We adopt the Leave-one-person-out (LOPO) cross validation and the 5-fold cross validation respectively to evaluate the classification rates for the proposed approaches as in [88]. The experimental results of the study [88] are illustrated in Table 8-2 and our results are shown in Table 8-3.

Gender classification of different algorithms on Leave-one-person-out CV					
Range	PCA+sequential selection	PCA+Random Forest	PCA+LPP+sequential selection	PCA+all variable	PCA+LPP+all variables
Success Rate	73.35%	**73.45%**	72.26%	55.59%	55.39%
Standard Error	23.54%	23.03%	23.96%	46.38%	49.77%
Gender classification of different algorithms on 5-fold CV					
Range	PCA+sequential selection	PCA+Random Forest	PCA+LPP+sequential selection	PCA+all variable	PCA+LPP+all variables
Success Rate	**84.33%**	82.73%	83.03%	63.77%	55.39%
Standard Error	1.16%	1.49%	2.92%	4.57%	3.16%

Table 8-2 Gender classification of the study [88]

Gender classification of different algorithms on Leave-one-person-out CV			
Range	LBP+ Random Forest	Pixel comparison + Adaboost	**Combination of two classifiers**
Success Rate	89.33%	90.65%	**93.36%**
Gender classification of different algorithms on 5-fold CV			
Range	LBP+ Random Forest	Pixel comparison + Adaboost	**Combination of two classifiers**
Success Rate	92.33%	94.03%	**95.83%**

Table 8-3 Gender classification of our study

Two dimensional reduction methods are used in [88]. These methods are PCA and LPP. The study [88] used the low dimensional data of the extracted features and the best features of the face are tried to be extracted for efficiency. Random Forest classifier, SVM classifier and Sequential Selection classifier are trained with the extracted data. They obtained the accuracy rates for the combination of feature extraction methods and machine learning methods. The Random Forest classifier which uses PCA features achieved the best results with overall classification rate 73.45% in gender classification of different algorithms on Leave-one-person-out cross validation in [88]. The best classification rate 84.33% is achieved by the Principle Component Analysis (PCA) and sequential selection scheme in gender classification of different algorithms on 5-fold cross validation in [88].

We tested our three classifiers separately with the same face images. The Random Forest classifier which uses LBP features achieved the classification rate 89.33% in gender classification on Leave-one-person-out Cross Validation. Our Adaboost classifier that uses the pixel comparisons achieved the classification rate 90.65%. Our final classifier, i.e. the combination of the two classifiers, achieved 93.36% success rate in gender classification on Leave-one-person-out Cross Validation. The

Random Forest classifier which uses LBP features achieved the classification rate 92.33% in gender classification on 5-fold Cross Validation. Our Adaboost classifier that uses the pixel comparisons achieved the classification rate 94.03%. Our final classifier which is the combination of the two classifiers achieved 95.83% success rate in gender classification on 5-fold Validation.

In our experiments, we tested our gender classifiers and compared the results with the success rates obtained in study [88]. The experimental results show that our gender classifiers outperformed the investigated methods used in [88] on both Leave-one-person-out (LOPO) cross validation and 5-fold cross validation. 95.83% success rate is achieved with our combination classifier.

8.2 The Experiments related to Age Group Classification

In this section we discuss the results of our experiments on age group classification and compare our results with the results of the study in [90]. In [90], texture and contour features are used in age classification. They extended local binary pattern method to get pure texture information. Classifiers were trained with the extracted data from the training images. SVM with probabilistic output is used as the machine learning method. A combination mechanism based on fuzzy integral to merge the output of different classifiers is used to get final result. In our experiments, we used BCMI-Omron age database [70] as in [90]. The ages of the people range from 0 to 80 in this database. Table 8-4 shows the distribution of face images into the age groups in the experiments.

Training data		Test data	
Age Group	Number of images	Age Group	Number of images
< 30	455	< 30	151
30 ~ 60	220	30 ~ 60	74
> 60	129	> 60	43
Total	804	Total	268

Table 8-4 Experimental data in age group classification

Researchers in [90] also asked three participants to classify the data to compare the results with human perception. The participants did the test twice. At the first time, they directly classify the image according to their life experience. Two younger participants (22 years old both) did not do well (with precision 67.16% and 70.52%), while the elder (50 years old) reached 76.12%. So they were asked to do the test again. This time, they saw the training data before the classification and reached 81.72% on average. Table 8-5 shows the results of the study [90].

Methods	**Accuracy**
Contour Features + LBP_6	77.61%
Contour + Located Local Binary Patterns	78.73%
LBP_6 + Located Local Binary Patterns	79.48%
LBP_7 + Located Local Binary Patterns	79.48%
Contour+LBP_6 + Located Local Binary Patterns	**80.23%**
Human	81.72%

Table 8-5 Accuracy of different feature extraction method in [90]

A combination of classifiers which use LBP and Located Local Binary Patterns (LLBP) information achieves 80.23% success rate only a bit lower than human's decision in [90].

In our experiments, we have worked on the same training and testing data whose distribution into the age groups is shown in Table 8-4. We tested our age group classifiers separately and then we tested our final (combination) classifier. Our results can be seen in Table 8-6. Our first classifier is a Random Forest classifier which uses DCT Mod2 features of the face image. Our second classifier is a Random Forest classifier which uses LBP features of the selected regions on the face image. Our system combines the results of the classifiers and predicts the age group of the face image. Our two classifiers are Random Forest classifiers which have 100 trees. Therefore, in the classification phase, all 200 trees classify the face images. Our system uses 200 results of all trees as votes. The final decision is the class which gets the majority of votes.

Methods	**Accuracy**
DCT Mod2 + Random Forest	82.46%
Selected Region LBP + Random Forest	83.95%
Combination of two classifiers (Combination Classifier)	85.19%

Table 8-6 Accuracy of our age group classifiers

We tested our three classifiers separately with the same face images in our experiments. The Random Forest classifier which uses DCT Mod2 features achieved the classification rate 82.46% in experiments. Our second classifier that uses the LBP features around the selected regions on the face achieved the classification accuracy rate 83.95%. Our final classifier, which is the combination of the two classifiers achieved 86.19% success rate in age group classification experiments.

In our experiments, we tested our age group classifiers and compared the results with the success rates obtained in study [90]. The experimental results show that our age group classifiers outperformed the investigated methods used in [90]. 86.19% success rate is achieved with our combination classifier.

8.3 The Experiments related to Face Identification in Videos

In this section we discuss the results of our experiments on face identification in videos and compare our results with the results of the study in [60]. After our experiments on feature extraction methods for face identification, we decided to use HOG feature extraction method in our final implementation. A graphical user interface as a case study is designed for the system to train and test our final implementation. In our experiments, two application scenarios are explored and these application scenarios work on the TV Episode *Coupling*:

8.3.1 Closed-set identification

In this scenario, only the main characters of an episode of a TV series are considered. A part of the episode is used to train a classifier, while the rest of the episode is used to test the classifier. More specifically, one fourth of the Coupling video is used for training and the rest is used for testing. Six main characters are selected in this scenario. The face images that do not belong to one of the six main characters are discarded. The results of the study [60] for this scenario are shown in Table 8-7.

Algorithm	**Correct classification rate**
DCT	**70.5%**
EHMM	67.9%
Fisherfaces	63.2%
Eigenfaces	50.4%

Table 8-7 Results of closed-set identification in [60]

Occurrences of the main characters in the training and testing sets are shown in Table 8-8.

Character	Training images	Testing images
Steve	316	1292
Jane	118	673
Susan	198	1255
Sally	772	353
Jeff	501	2224
Patrick	6	423

Table 8-8 Occurrences of the main characters in [60]

Occurrences of the main characters in the training and testing sets of our experiments are shown in Table 8-9.

Character	Training images	Testing images
Steve	300	1200
Jane	142	600
Susan	180	1255
Sally	700	548
Jeff	570	2297
Patrick	9	320

Table 8-9 Occurrences of the main characters in the training and testing sets

After our experiments on feature extraction methods for face identification, we decided to use HOG feature extraction method in our final implementation. Our face identification system achieved **88.24%** accuracy rate. In [60], the best identification rate is achieved (70.5%) with DCT method. The accuracy rates of our face identification methods outperformed the results of the study in [60]. The main reason for this is that the eye locators in [60] are not reliable. Another reason for the low results is that even though only faces, where the confidence for a frontal view is very high, were selected, still a significant number of faces are non-frontal. Since their Euclidean alignment cannot compensate out-of-plane rotations, this also affects the performance negatively. In our study, we used face morphing method to align and normalize the faces in geometry, pose and expression.

8.3.2 Automatic retrieval

In this scenario, all persons of the episode are considered. Again a part of the episode is used to build query sets for the main characters, while the rest is used as a database from which images that depict the queried person, have to be retrieved. In this scenario, the persons different from the six main characters were not removed from the dataset. A part of the episode is used to train a classifier, while the rest of the episode is used to test the classifier. Again one fourth of the Coupling video was used for training and the rest were used for testing. The threshold of the distance between a probe image and the query images was varied to give a plot of precision and recall for the automatic retrieval. Precision and recall graph of the study [60] for the automatic retrieval can be seen in Figure 8-10.

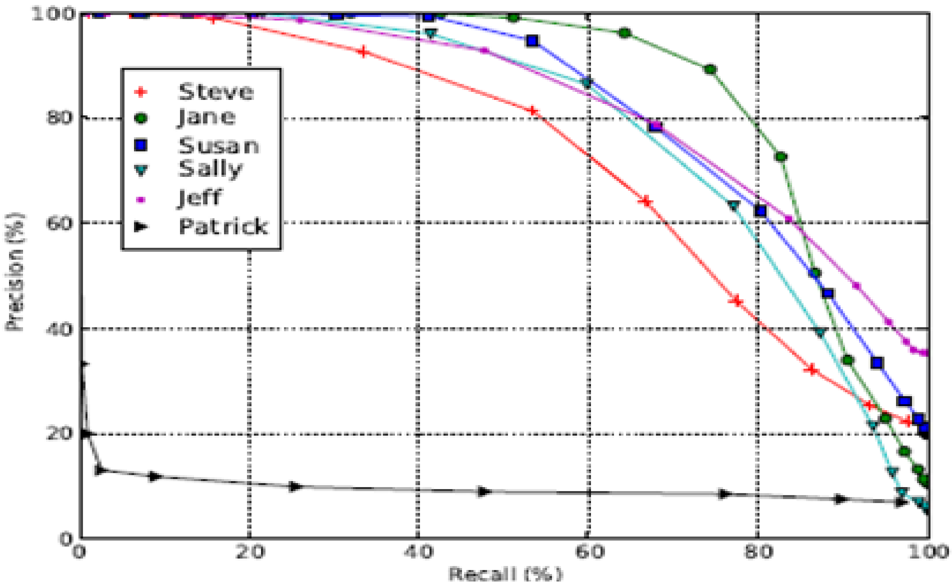


Figure 8-10 Results of automatic retrieval in [60]

Precision and recall graph of our study for the automatic retrieval can be seen in Figure 8-11.

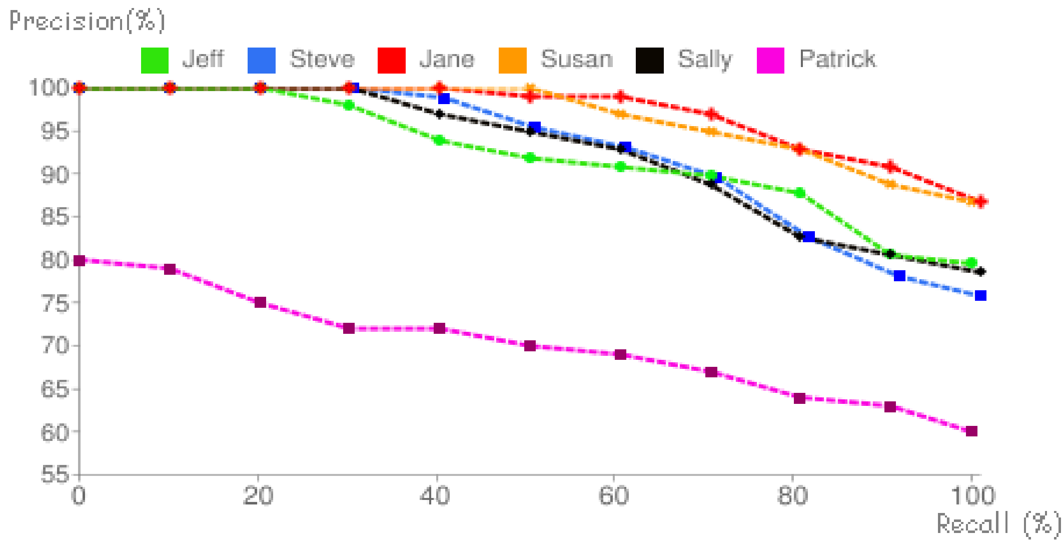


Figure 8-2 Results of automatic retrieval in our experiments

The results are encouraging because the performance can yield a useful semantic annotation system.

8.3.3 Adding Gender and Age Group Classification to the Face Identification Module

In this thesis, we added gender and age group classification methods to our face identification module to improve the success rate of the face identification method. We tested our system with this addition. In this implementation, the age group and gender of the detected face is determined first, and then, our face identification classifier search for the identity of the face amongst the identities of the same age group and gender. However, our experiment showed that the addition of other classifiers for indexing does not improve the success rate of the face identification process. The false classification rates of the gender and age group classifiers are the main cause of this situation. A wrong classification of gender and age group classifiers brings the false identification of the face. This problem even outs the advantages of indexing by gender and age group classifications. As a result, we decided to present these classifiers as separate modules.

CHAPTER 9

CONCLUSIONS

We have introduced some methods to classify the genders and age groups of human faces. Our Gender Classification module contains two distinct gender classifiers. The first classifier, “LBP and Random Forest”, works with 89% accuracy rate in our experiments on LOPO cross validation. The second classifier, “Pixel LBP values comparison and Adaboost”, works with 90% accuracy rate on LOPO cross validation. Our final method, which is the combination of these two classifiers has 93% accuracy rate on LOPO cross validation. These accuracy rates prove the usability of our method in any application which requires a robust gender classification module.

Our Age Group Classification module contains two distinct age group classifiers. The first classifier, “DCT Mod2 and Random Forest” worked with 82% accuracy rate in our experiments. The second classifier, “Selected Region LBP and Random Forest” has worked with 83% accuracy rate. And finally, our final method, which is the combination of these two classifiers had 85% accuracy rate. 85% is a high success rate for a process of classification of age groups on videos. Therefore we conclude that our method is a successful method for age classification and can be used in applications like semantic video annotation.

In this thesis, we have also introduced a robust face identification method to identify people by their faces in videos. Some face identification algorithms have been evaluated before deciding which methods to be used. The identification performances and execution times of the face identification methods are measured and the best algorithm is chosen for our system. Since this thesis is based on videos, the execution time is also important. Viola-Jones face detector [65] is used as the face detector and a Gaussian-mixture skin color detector used in this work. A condensation algorithm and template matching algorithm based robust face tracker is developed. By this way, face images are collected and used for training or identification. The success of the face tracker is so important because the success of the overall system performance depends on good tracking performance. The pose of the detected face is also important because, face identification methods are very sensitive to pose changes. Therefore frontal or near-frontal faces are the main focus of this thesis. This focus has also some disadvantages. This assumption limited our training and testing samples as faces are encountered under a variety of poses like profile or near-profile faces. Changing illumination conditions and varying facial expressions make face identification more challenging. To neutralize the facial expression, an AAM based facial morphing is used. HOG based feature extraction with nearest neighbor classifier is used in our face identification module. Our methods can be used in real time systems as we showed that its execution time is suitable for real applications. The success rates for face identification in videos are also satisfactory. The results show the applicability of our methods for face identification and annotation in videos.

The presented results do not show the performance of the system for low-resolution videos, therefore some further testing of the framework may still be required for low-resolution videos. However, in particular, the results are quite sufficient for many applications. It is difficult to recognize human faces or determine their genders and ages in videos of low resolution. There are some solutions like Super Resolution (SR) and Multiple Resolution-faces (MRF) approach. Implementing these methods for our system can be a useful extension. This can improve the success rate of the classifications. Using profile faces in the identification process may be another improvement. Non-frontal face detector may be used and the training set can be clustered according to the pose of the detected faces. The identification performance of the tracking method can be improved by working on profile faces. .

REFERENCES

[1] Cottrell, G.W., Metcalfe, J.: Face, emotion, and gender recognition using holons, Proceedings of the 1990 conference on Advances in neural information processing systems 3, p.564-571, Denver, Colorado, USA, 1990.

[2] Golomb, B.A., Lawrence, D.T., Sejnowski, T.J.: A neural network identifies sex from human faces, Proceedings of the 1990 conference on Advances in neural information processing systems 3, p.572-577, Denver, Colorado, USA, 1990.

[3] BenAbdelkader, C., Griffin, P.: A Local Region-based Approach to Gender Classification from Face Images, Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, pp.52-56, 2005.

[4] Brunelli, R., Poggio, T.: HyberBF networks for gender classification. In: Proc. DARPA Image Understanding Workshop, pp. 311-314, 1995.

[5] Buchala, S., Davey, N., Frank, R.J.: Dimensionality reduction of face images for gender classification, Intelligent Systems, Proceedings, 2004 2nd International IEEE Conference, 2004.

[6] Costen, N., Brown, M., Akamatsu, S.: Sparse models for gender classification. Proc. of Intl. Conf. on Automatic Face and Gesture Recognition (FGR'04), pp: 201–206, 2004.

[7] Graf, A., Wichmann, A.: Gender Classification of Human Faces. Proceedings of the Biologically Motivated Computer Vision Springer LNCS 2525, 491-501, 2002.

[8] Gutta, G.: Gender and Ethnic Classification of Face Images; In Proceedings of the IEEE International Automatic Face and Gesture Recognition, pp. 194-199, 1998.

[9] Iga, R., Izumi, K., Hayashi, H., Fukano, G., Ohtani, T.: A gender and age estimation system from face images, SICE Annual Conference, 2003.

[10] Barycentric Coordinates in Wikipedia, http://en.wikipedia.org/wiki/Barycentric_coordinate_system_%28mathematics%29, Last Access: November, 2010.

[11] Kim, H.C., Kim, D., Ghahramani, Z., Bang, S.Y.: Appearance-based gender classification with Gaussian processes, Pattern Recognition Letters, Elsevier, 2006.

[12] Lian, H.C., Lu, B. L.: Multi-view gender classification using local binary patterns and support vector machines, In: Proc. 3rd Intl. Symposium on Neural Networks (ISNN'06), Chengdu, China, vol. 2, pp. 202-209, 2006.

[13] Moghaddam, B., Yang, M.H.: Learning Gender with Support Faces, IEEE Transactions on Pattern Analysis and Machine Intelligence, v.24 n.5, p.707-711, 2002.

[14] Shakhnarovich, G., Viola, P.A., Moghaddam, B.: A unified learning framework for real time face detection and classification, Proc. Intl. Conf. on Automatic Face and Gesture Recognition (FGR'02), IEEE, pp: 14–21, 2002.

[15] Su, E.: Multi-view gender classification using local binary patterns and support vector machines, International Journal of Neural Systems 17(6), 479–487, 2007.

[16] Tivive, F.H.C., Bouzerdoum, A.: A Gender Recognition System using Shunting Inhibitory Convolutional Neural Networks, International Joint Conference on Neural Networks (IJCNN '06), 5336 – 5341, 2006.

[17] Sharma, R., Yeasin, M., Walavalkar, L.: A method for human gender recognition using audio-visual cues, Master's thesis, Sound and Image

Processing Laboratory, School of Electrical Engineering, Royal Institute of Technology, Stockholm, Sweden, 2009.

[18] Wu, B., Ai, H., Huang, C.: LUT-based Adaboost for gender classification. In: Proc. Intl. Conf. on Audio and Video-based Biometric Person Authentication (AVBPA'03), Guildford, United Kingdom, pp. 104-110, 2003.

[19] Baluja, S., Rowley, H.A.: Boosting Sex Identification Performance, International Journal of Computer Vision, v.71 n.1, p.111-119, 2007.

[20] Ruiz-del-Solar, J., Quinteros, J.: Illumination compensation and normalization in eigenspace-based face recognition: A comparative study of different pre-processing approaches. Pattern Recognition Letters 29(14): 1966-1979 (2008).

[21] Castrillón, M., Déniz, O., Hernández, D., Domínguez, A.: Identity and Gender Recognition Using the ENCARA Real-Time Face Detector, Conferencia de la Asociación Española para la Inteligencia Artificial, 2003.

[22] Hayashi, J., Koshimizu, H., Hata, S.: Age and Gender Estimation Based on Facial Image Analysis, Lecture Notes in Computer Science, Volume 2774/2003, 863-869, DOI: 10.1007/978-3-540-45226-3_118, 2003.

[23] Lapedriza, A., Jiménez, M., Vitrià, J: Gender Recognition in Non Controlled Environments, International Association for Pattern Recognition, 2006.

[24] Kwon, Y. H., Vitoria Lobo, N.: Age classification from facial images, Computer Vision and Image Understanding Journal, Vol 74, No. 1, 1999.

[25] Ueki, K., Hayashida, T., Kobayashi, T.: Subspace-based Age-group Classification Using Facial Images under Various Lighting Conditions, In: Proc. of IEEE International Conference on Automatic Face and Gesture Recognition , pp. 43-48, 2006.

[26] Lanitis, A., Draganova, C., Christodoulou, C.: Comparing different classifiers for automatic age estimation, IEEE Transactions on Systems, Man, and Cybernetics, Part B, vol. 34, no. 1, pp. 621-628, 2004.

[27] Sirovich, L., Kirby, M.: Low-Dimensional procedure for the characterisation of human faces, Journal of the Optical Society of America, vol. 4, pp. 519-524, 1987.

[28] Kirby, M., Sirovich, L.: Application of the Karhunen-Loève procedure for the characterisation of human faces, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, pp. 831-835, 1990.

[29] Turk, M., Pentland, A.: Eigenfaces for recognition, Journal of Cognitive Neuroscience, vol. 3, pp. 71-86, 1991.

[30] Zhao, L., Yang, Y.H.: Theoretical analysis of illumination in pca based vision systems, Pattern Recognition, vol. 32, pp. 547-564, 1999.

[31] Pentland, A., Moghaddam, B., Starner, T.: View-Based and modular eigenspaces for face recognition, Proc. IEEE CS Conf. Computer Vision and Pattern Recognition, pp. 84-91, 1994.

[32] Lades, M., Vorbruggen, J.C., Buhmann, J., Lange, J., Von Der Malsburg, C., Wurtz, R.P., Konen, M.: Distortion Invariant object recognition in the dynamic link architecture, IEEE Trans. on Computers, vol. 42, pp. 300-311, 1993.

[33] Wiskott, L., Von der Malsburg, C.: Recognizing faces by dynamic link matching, Neuroimage, vol. 4, pp. 514-518, 1996.

[34] Samaria, F., Fallside, F.: Face recognition and feature extraction using hidden markov models, Image Processing: Theory and Application, G. Vernazza, ed., Elsevier, 1993.

[35] Samaria, F., Harter, A.C.: Parameterisation of a stochastic model for human face identification, Proc. of the Second IEEE Workshop on Applications of Computer Vision, 1994.

[36] Kanade, T.: Picture processing by computer complex and recognition of human faces, technical report, Dept. Information Science, Kyoto Univ., 1973.

[37] Goldstein, A.J., Harmon, L.D., Lesk, A.B.: Identification of human faces, Proceedings of the IEEE, vol. 59, pp. 748, 1971.

[38] Bruneli, R., Poggio, T.: Face recognition: features versus templates, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 15, pp. 1042-1052, 1993.

[39] Manjunath, B.S., Chellappa, R., Von der Malsburg, C.: A Feature based approach to face recognition, Proc. IEEE CS Conf. Computer Vision and Pattern Recognition, pp. 373-378, 1992.

[40] ORL Face Database http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html, Last Access: November, 2010.

[41] Vetter, T., Poggio, T.: Linear object classes and image synthesis from a single example image, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 19, no. 7, pp. 733-742, July 1997.

[42] Beymer, D., Poggio, T.: Face recognition from one model view, Proc. Fifth Intl Conf. Computer Vision, 1995.

[43] Zhao, W., Chellappa, R.: SFS based view synthesis for robust face recognition, Proc. Intl. Conf. Automatic Face and Gesture Recognition, pp. 285-292, 2000.

[44] Blanz, V., Vetter, T.: Face recognition based on fitting a 3D morphable model, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 25, no. 9, 2003.

[45] Takács, T.: Comparing face images using the modified hausdorff distance, Pattern Recognition, vol. 31, pp. 1873-1881, 1998.

[46] Gao, Y., Leung, K.H.: Face recognition using line edge map, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 6, 2002.

[47] Leung, K.H., Yang, Y.H.: Dynamic two-strip algorithm in curve fitting, Pattern Recognition, vol. 23, pp. 69-79, 1990.

[48] Guo, G., Li, S.Z., Chan, K.: Face recognition by support vector machines, In proc. IEEE International Conference on Automatic Face and Gesture Recognition, pp. 196-201, 2000.

[49] Phillips, P.J.: Support vector machines applied to face recognition, processing system, Technical Report NISTR 6241, National Institute of Standards and Technology, Geithesburg, 1998.

[50] Heisele, B., Ho, P., Poggio, T.: Face recognition with support vector machines: Global versus component-based approach, in International Conference on Computer Vision (ICCV'01), 2001.

[51] Huang, J., Blanz, V., Heisele, B.: Face recognition using component-based support vector machine classification and morphable models, SVM '02 Proceedings of the First International Workshop on Pattern Recognition with Support Vector Machines, 2002.

[52] Jonsson, K., Mates, J., Kittler, J., Li, Y.P.: Learning support vectors for face verification and recognition, Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000, pp. 208-213, Los Alamitos, USA, 2000.

[53] Deniz, O., Castrillon, M., Hernandez, M.: Face recognition using independent component analysis and support vector machines, Pattern Recognition Letters, vol. 24, pp. 2153-2157, 2003.

[54] Martinez, A.M., Benavente, R.: The AR face database, Technical Report 24, CVC, June 1998.

[55] Tolba, A.S.: A parameter–based combined classifier for invariant face recognition, Cybernetics and Systems, vol. 31, pp. 289-302, 2000.

[56] The Open Computer Vision Library (OpenCV 2.2.0) has more than 500 algorithms, documentation and sample code for real time computer vision. http://sourceforge.net/projects/opencvlibrary/, Last Access: November, 2010.

[57] Tang, H.M., Lyu, M., King, I.: Face recognition committee machine, In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003), pp. 837-840, 2003.

[58] LBP in Wikipedia, http://en.wikipedia.org/wiki/Local_binary_patterns, Last Access: November, 2010.

[59] Wilhelm, T., Bohne, H.J., Gross, H.M.: Classification of Face Images for Gender, Age, Facial Expression, and Identity, ICANN 2005, LNCS 3696, 569-574, 2005.

[60] Fischer, M.: Automatic identification of persons in TV series Universität Karlsruhe (TH) M.S. Thesis, 2008.

[61] Iga, R., Izumi, K., Hayashi, H., Fukano, G., Ohtani, T.: A gender and age estimation system from face images, The Society of Instrument and Control Engineers Annual Conference in Fukui, 202–209, 2003.

[62] Mayo, M., Zhang, E.: Improving Face Gender Classification By Adding Deliberately Misaligned Faces To The Training Data, Image and Vision Computing New Zealand, pp. 1-5, 2008.

[63] Ramanan, D., Baker, S., Kakade, S.: Leveraging archival video for building face datasets, IEEE 11th International Conference on Computer Vision (ICCV 2007), pp.1-8, 2007.

[64] Satoh, S.: Comparative Evaluation of Face Sequence Matching for Content-based Video Access, Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 163-168, 2000.

[65] Viola, P., Jones, M.: Rapid Object Detection using a Boosted Cascade of Simple Features, in CVPR2001, 2001.

[66] Yang, M.H., Kriegman, D.J., Ahuja, N.: Detecting faces in images: a survey, IEEE Trans. Pattern Analysis and Machine Intelligence, 24(1), pp. 34-58, 2002.

[67] Hjelmås, E., Low, B. K.: Face detection, a survey, Computer Vision and Image Understanding, 83(3), pp. 236-274, 2001.

[68] Machine Learning, Wald I, II, III, July 2002, http://stat-www.berkeley.edu/users/breiman/wald2002-1.pdf, http://stat-www.berkeley.edu/users/breiman/wald2002-2.pdf, http://stat-www.berkeley.edu/users/breiman/wald2002-3.pdf, Last Access: November, 2010.

[69] Yaprakkaya, G., Cicekli, N.K., Ulusoy, I.: Gender and Age Groups Classifications for Semantic Annotation of Videos, Proc. of the 25th International Symposium on Computer and Information Sciences (ISCIS 2010) 22-24 September 2010, London, UK, London, UK.

[70] BCMI-Omron age database, http://www.omron.com/about/, Last Access: November, 2010.

[71] Random Trees - OpenCV 2.0 C++ Reference. http://opencv.willowgarage.com/documentation/cpp/random_trees.html, Last Access: November, 2010.

[72] Adaboost algorithm definition : http://en.wikipedia.org/wiki/AdaBoost, Last Access: November, 2010.

[73] Sanderson, C., Paliwal, K.K.: Fast feature extraction method for robust face verification, Electronics Letters, vol. 38, no. 25, pp. 1648-1650, 2002.

[74] Eickler, S., Muller, S., and Rigoll, G.: Recognition of JPEG compressed face images based on statistical methods, Elsevier Image and Vision Computing, Vol. 18, No. 4. (1 March 2000), pp. 279-287, 2000.

[75] Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Transactions on pattern analysis and machine, Vol 19:711-720, 1997.

[76] Bartlett, Stewart, M., Sejnowski, T.: Viewpoint invariant face recognition using independent Component analysis and attractor networks, In M. Mozer, M. Jordan, & T. Petsche, Eds., Advances in Neural Information Processing Systems 9. Cambridge, MA: MIT Press: 817-823, 1997.

[77] Zhang, W., Shan, S., Chen, X., Gao, W.: Are Gabor phases really useless for face recognition? In Proceedings of international conference on pattern recognition Vol. 4, pp. 606-609, 2006.

[78] Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions, In Proceedings of the IEEE international workshop on analysis and modeling of faces and gestures, pp. 168-182, 2007.

[79] Cootes, T., Walker, K., Taylor, C.: View-based active appearance models, In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, pages 227-232, 2000.

[80] Platt, S.M., Badler, N.I.: Animating facial expressions, In Proceedings of the 8th annual conference on Computer graphics and interactive techniques, pages 245–252, 1981.

[81] Sivic, J., Everingham, M., Zisserman, A.: Who are you? – Learning person specific classifiers from video, IEEE Conference on Computer Vision and Pattern Recognition, pp. 1145-1152, 2009.

[82] Kettner, L., Kirkpatrick, D., Mantler, A., Snoeyink, J., Speckmann, B., Takeuchi, F.: Tight degree bounds for pseudo-triangulations of points, In: Computational Geometry - Theory and Applications 25(1-2), pp. 3-12, 2003.

[83] Explanation of Nearest neighboor algorithm in Wikipedia http://en.wikipedia.org/wiki/Nearest_neighbour_algorithm, Last Access: November, 2010.

[84] Loomes, M. J., Davey, N., Frank, R. J. and Buchala, S.: The role of global and feature based information in gender classification of faces: a comparison of human performance and computational models, International Journal of Neural Systems, 15 (1-2), pp. 121-128, ISSN 0129-0657, 2005.

[85] Ueki, K., Hayashida, H.: Subspace based group classification using facial images under various lighting conditions, IEEE International Conference on Automatic Face and Gesture Recognition, 2006.

[86] Saatci, Y., Town, C.: Cascaded classification of gender and facial expression using active appearance models, IEEE International Conference on Automatic Face and Gesture Recognition, 2006.

[87] Matas, J., Obdržálek, Š.: Object recognition methods based on transformation covariant features, in XII. European Signal Processing Conference EUSIPCO – 2004, F. Hlawatsch, G. Matz, M. Rupp, and B. Wistawel, eds., Vienna, Austria, pp. 1333 – 1336, 2004.

[88] Wang, Y., Ricanek, K., Chen, C., Chang, Y.: Gender Classification from Infants to Seniors, Biometrics: Theory Applications and Systems (BTAS), Fourth IEEE International Conference, 2010.

[89] FG-net aging database. http://www.fgnet.rsunit.com, Last Access: November, 2010.

[90] Tang, Y.M., Lu, B.L.: Age Classification Combining Contour and Texture Feature, Lecture Notes in Computer Science, Volume 6444/2010, 493-500, 2010.