COUNTING NUMBER OF PEOPLE IN DIGITAL IMAGES USING FACE AND PEOPLE DETECTION ALGORITHMS

A MASTER'S THESIS

IN

COMPUTER ENGINEERING

ATILIM UNIVERSITY

BY

SAMAR ITTAHIR M.A HUSAIN

JUNE 2016

COUNTING NUMBER OF PEOPLE IN DIGITAL IMAGES USING FACE AND PEOPLE DETECTION ALGORITHMS

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

OF

ATILIM UNIVERSITY

BY

SAMAR ITTAHIR M.A HUSAIN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

THE DEPARTMENT OF COMPUTER ENGINEERING

JUNE 2016

Approval of the Graduate School of Natural and Applied Sciences, Atılım University.

_____

Prof. Dr. İbrahim Akman

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

_____

Prof. Dr. İbrahim Akman

Head of Department

This is to certify that we have read the thesis "Counting Number of People in Digital Images Using Face and People Detection Algorithms" submitted by "Samar Ittahir M.A Husain" and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

_____

Assoc. Prof. Dr. Murat Koyuncu

Supervisor

Examining Committee Members:

Assoc. Prof. Dr. Murat Koyuncu _____
Atilim University – Information Systems Engineering

Asst. Prof. Dr. Gokhan Sengul _____
Atilim University – Computer Engineering

Asst. Prof. Dr. Tolga Pusatli _____
Çankaya University – Mathematics

Date: June 3, 2016

I declare and guarantee that all data, knowledge and information in this document has been obtained, processed and presented in accordance with academic rules and ethical conduct. Based on these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Samar Ittahir M.A Husain

Signature:

ABSTRACT

COUNTING NUMBER OF PEOPLE IN DIGITAL IMAGES USING FACE AND PEOPLE DETECTION ALGORITHMS

Husain, Samar Ittahir M.A

M.S., Computer Engineering Department

Supervisor: Assoc. Prof. Dr. Murat Koyuncu

June 2016, 66 pages

Counting the number of people in still images or video frames is an active research area that is a challenge in the computer vision field. It plays an important role in a variety of applications, such as security, management, education, and commerce. In this thesis, we work on counting the number of people in digital images. People can be seen differently in images, which requires the use of different techniques together. Therefore, we use two different techniques, which are Face Detection method and People Detection method, in order to estimate the number of people in an image. The proposed method combines the outputs of the Face Detection and People Detection methods in order to improve the performance of estimating the number of people in an input image with low cost and simple hardware. We test three face detection algorithms (Skin Color, Viola Jones LBP and Viola Jones CART) and a People Detection method (whole body) which is based on the HOG feature and SVM classifier to determine the best combination to estimate the number of people in input images. We use 240 test images including 1,202 people from two different datasets (Groups of Images of People and INRIA Person) to test the proposed system and determine the best combination. We have obtained best Recall of 91% and Precision of 93.97% by combining Viola Jones CART with People Detection method whereas by applying Face Detection and People Detection methods (whole body) separately, we got best Recall of 70.38% and Precision of 92.76% by Viola Jones CART method.

Keywords: Counting People, Face Detection, People Detection, Skin Color, Viola Jones LBP, Viola Jones CART, Histogram of Oriented Gradients, Support Vector Machine.

ÖZ

YÜZ VE İNSAN TANIMA ALGORİTMALARI İLE SAYISAL RESİMLERDE İNSAN SAYIMI

Husain, Samar Ittahir M.A

Yüksek Lisans, Bilgisayar Mühendisliği Bölümü

Tez Yöneticisi: Doç. Dr. Murat Koyuncu

Haziran 2016, 66 sayfa

Sabit görüntülerdeki veya video karelerindeki insanların sayımı, görüntü işleme alanında zorlu bir aktif araştırma sahasıdır. Bu alan, güvenlik, yönetim, eğitim ve ticaret açısından birçok uygulamada önemli bir rol üstlenmektedir. Bu tezde, dijital görüntülerde bulunan insanların sayımı üzerinde çalışılmıştır. İnsanlar görüntülerde farklı şekillerde görünebilmektedirler ve bu durum farklı tekniklerin beraber kullanılmasını gerektirmektedir. Bu bakımdan, bir görüntüdeki insan sayısını tahmin etmek için, yüz tanıma metodu ve insan tanıma metodu olmak üzere, iki tekniği kullanıyoruz. Önerilen bu metot, düşük maliyetle ve basit donanım kullanarak, girilen bir görüntüdeki insan sayısını tahmin ederken performansı artırmak için, yüz tanıma ve insan tanıma metotlarının çıktılarını bir araya getirmektedir. Biz, girilen görüntülerdeki insan sayısını tahmin eden en iyi kombinasyonu belirlemek için, üç yüz tanıma algoritmasını (Skin Color, Viola Jones LBP and Viola Jones CART) ve HOG özelliği ile SVM sınıflandırıcısına dayalı bir insan tanıma metodunu (tüm vücut) test ediyoruz. Önerilen sistemi test etmek ve en iyi kombinasyonu belirlemek için içinde iki farklı veri kümesinden (Groups of Images of People ile INRIA Person) 1,202 insanı barındıran 240 test görüntüsünü kullanıyoruz. Viola Jones CART ile HOG özelliği ile SVM sınıflandırıcısına dayalı insan tanıma metodunu kombine ederek %91'lik en iyi Recall'u ve %93.97'lik en iyi Precision'ı elde ettik, oysa, yüz tanıma metodu ve insan tanıma metodu (tüm vücut) ayrı ayrı uygulayarak, %70.38'lik en iyi Recall'u ve %92.76'lik en iyi Precision'ı Viola Jones CART yöntemi ile elde ettik.

Anahtar Kelimeler: İnsan Sayımı, Yüz Tanıma, İnsan Tanıma, Skin Color, Viola Jones LBP, Viola Jones CART, HOG, Support Vector Machine

To My Parents

ACKNOWLEDGMENT

I express sincere appreciation to my supervisor Assoc. Prof. Dr. Murat Koyuncu for his guidance and insight throughout the research. To my husband, Tariq, I offer sincere thanks for his continuous support and patience during this period.

Last but not the least; I would like to thank my family, friends, and colleagues who have helped me directly or indirectly. They were always supporting and encouraging me with their best wishes.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

BioID	Biometric Identification Database
BP	Back Propagation
CART	Classification and Regression Tree Analysis
CVC dataset	Card Verification Code Database
DPMMs	Dirichlet Process Mixture Models
FRGC database	The Face Recognition Grand Challenge Database
GLCM	Gray-Level Co-occurrence Matrices
H3D	Humans in 3D Database
HOG	Histograms of Oriented Gradients
HSI	Hue Saturation and Intensity
HSV	Hue Saturation and intensity Value
ISM	Implicit Shape Model
K-NN	K-Nearest Neighbor
LBP	Local Binary Pattern
MoSIFT	Motions Scale-Invariant Feature Transform
MRCG	Mean Riemannian Covariance Grid
OSH	Optimal Separating Hyper-plane
PASCAL VOC	The Pascal Visual Object Classes Database
PD	Pedestrian Detection
SIFT	Scale-Invariant Feature Transform
SVM	Support Vector Machine
TRECVID	Text REtrieval Conference VIDeo Database

CHAPTER 1

INTRODUCTION

People counting is a process used in many computer vision and pattern recognition systems to determine the number of people in a digital image or video. Such information can be used for further analysis in a wide range of practical applications related to education, commerce, security, management etc. For instance, people counting systems used for security provide information about the total number of people in the building and number of people on each floor in order to control the number of visitors. People counting can also be used for pedestrian traffic management [1, 2, 3, 4] and for fire management, where it is considered one of the most important tools in fire cases. In addition, people counting applications are widely used in commerce domains to measure marketing effectiveness. As can be seen from the given examples, research on people counting is one of the most important active areas in the computer vision field that can be applied in many daily life areas.

The output of people counting systems may be passed to other management systems used in different areas such as public transport, industry, stations, airports, etc. There are numerous examples that illustrate the significance of people counting systems. For instance, in the commerce domain, the shopping malls, hypermarkets and others depend on visitor statistics to measure marketing effectiveness and attraction of a commercial site, and make new decisions. There are companies developing products to offer such services. Table1.1 shows some companies and their services in the field of people counting systems.

Company name	Services	Web-site
Acorel company	Counting people in shopping centers, tramways, bus, and train, metro. etc.	http://www.acorel.com
Infodev company	Counting people in buildings such as shopping centers, cinemas, museums, airports. Counting passengers in vehicles such as buses, light rail, trains.	http://www.infodev.ca

Table 1.1. Products and services for commercial people counting

There are different techniques for people counting, which are roughly classified into two main categories: vision-based and non-vision based counting systems. The vision-based techniques can again be classified as tracking and non-tracking based counting systems (Figure 1.1). Each class has specific strengths and weaknesses.

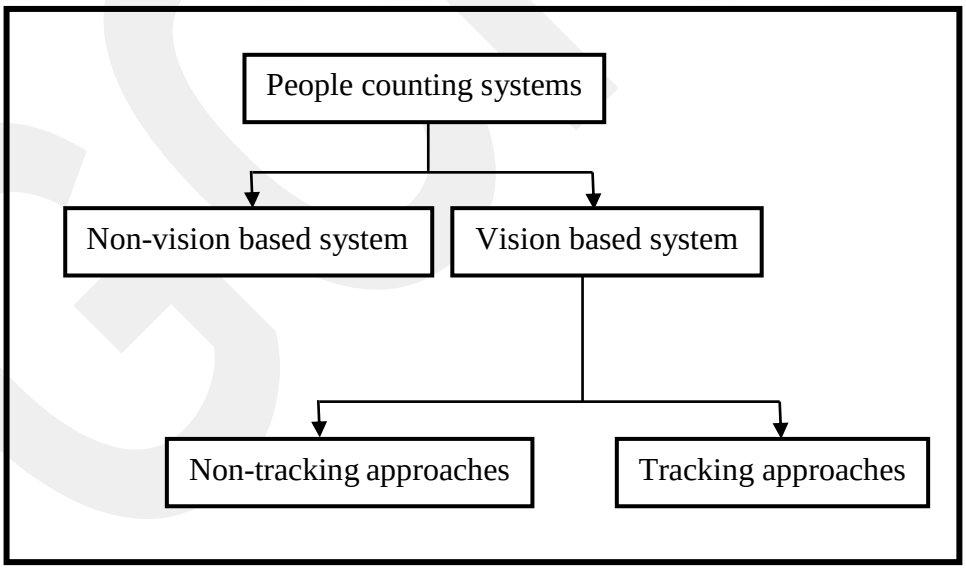


Figure 1.1. Classes of people counting system

In general, the non-vision based class of people counting systems include all systems that are not based on video or image data from a camera, but from devices such as a heat sensor, the light of an infrared beam, or pressure sensors [5, 6]. Heat-sensors often

use an array of sensors and infer the number of people by detecting heat sources from human bodies. These systems are usually carried out through embedded technology and are installed overhead for high accuracy. Infrared beam sensors are the simplest form for people counting and use a single, horizontal infrared beam across a doorway to provide a light barricade, which is broken and counted each time a person walks through the light. Pressure sensors are pressure pads, which are hidden in the ground. When people walk through, the counter increases one by one to estimate the number of people. The advantage of these systems is that they are not affected by environmental factors like illumination changes. Nevertheless, there are several weaknesses of these systems; counting humans is not easy if they are not clearly visible, and they are able to work only on narrow paths. In addition, they work poorly when people walk close to each other, and most of infrared beam sensors may be exposed to withhold by insects etc [7].

In general, vision-based systems require multiple frames for processing. In addition, the counting process of vision-based systems consists of defined direction of blobs to increase in or out counter [8]. In recent years, vision-based systems have become more popular in different scenes like in hot spots, streets and buildings, because vision-based systems perfectly use infrastructure that is provided through security networks. In addition, vision based people counting relies on the information that is given by a security camera's video stream in order to count people crossing. The main strength of these systems is that they are not limited to narrow corridors or doors, but the disadvantage is that they are limited by the coverage area of the cameras. Additional challenges for vision-based systems include differences in environmental conditions and occlusions between people. Some systems solve this problem by installing cameras looking directly down to the floor [9]. The non-tracking vision-based approach counts the number of people without any human tracking of groups or individuals. Thus, this approach is suitable for a very large number of people in a frame, unlike the tracking approach, which does not scale well with a growing number of people. Because of their ability to count considerable numbers of people, these systems are usually called crowd surveillance systems. The main weakness of the non-tracking approach is that it gives the number of people that are included in a frame but does not give information about the directions of people's movements [10, 11, 12].

The tracking approach relies on tracking a person or clusters of people over time in a video without large-scale changes, then, estimating the number of people. The advantage of this approach is that it gives information about the number of people; the direction of human movements, including from where to where a person has moved; and the speed and height of humans, unlike previous approach that provides only information about the number of people. The greatest disadvantage of the tracking approach is that it is hard to distinguish individuals from the background and from each other in cases where the people are close to each other, for instance, side by side or behind one another [9, 13, 14, 15].

Most recent studies on people counting systems rely on computer vision. For example, Yu et al. present an automatic Bi-directional People Counting method that is based on people detection and tracking [16]. "Individual-Centric and Crowd-Centric" approaches segment each crowd in an input image into individuals or groups of people, and then analyze each segment separately in order to count the number of people [17]. Another People Counting technique employs a clustering scheme relying on Dirichlet Process Mixture Models (DPMMs), which use outputs of a person detection system as input and run it on each frame. Then, its output is used as a set of features relying on temporal information, color and spatial for each detection. By utilizing these features, it will be clear if there is any restriction on the number of clusters. Consequently, this technique can determine an arbitrary number of groups of people and determine a measure to calculate the true number of people that included in each cluster to estimate the number of people within the scene [18].

People counting based on non-tracking vision based system (image processing) aims to estimate the number of people in an input image. A people counting system that relies on image processing is beneficial because it reduces the cost of surveillance and the observers' effort, and can also provide information about people including their positions, clothes, and gender in addition to the total number people. People information obtained from counting systems can be utilized in several potential applications. There are a variety of methods used to count the number of people in input images, which can be classified into two main approaches: the feature-based regression approach and the individual people detection regression approach. In the feature-based regression approach, the number of people is estimated by extracting regressing of features from an input image, through utilizing a regression function.

Generally, this approach is based on several steps, which are removing the background, extracting features of the forefront segments, and finally estimating the number of people through a regression function. The main disadvantage of this approach is that it cannot determine the positions of people in an image and cannot be executed in real-time [11, 19, 20]. In the second approach, the algorithms count the number of people which are detected in an input image. This approach is not appropriate for highly crowded scenes with remarkable occlusion because it is based on detecting and segmenting all people [21, 22, 23, 24].

Broadly, counting people in a static digital image is more difficult than counting people in video or real-time video because working on video assumes all moving objects as people, which makes object detection easier. Recently, there have been many studies related to counting people in videos, especially in real-time videos. Nevertheless, there is little research on counting people in still digital images. Furthermore, most of the people counting systems, which are based on video, include many components of hardware and complex algorithms; and also the quantity of data is very huge. Therefore, they are complex and costly. The main motivation of this study is to overcome such difficulties in counting people systems with computer vision, developing a simple system, low resource, and low cost.

In this thesis, we count and estimate the number of people in a static digital image based on two different techniques, which are face detection and people detection (whole body). Face detection and people detection play very important roles in the computer vision field. Face detection is a computer technology that has been used in several varieties of applications such as facial recognition, photography and marketing. Face detection identifies human faces in an input digital image with a focus on the detection of frontal human faces. Therefore, face detection is an important part and the first step of automatic face recognition. Furthermore, people can be counted based on the number of people's faces that are detected in a digital image. Notably, there are many methods that can be used to detect people's faces such as skin color based detection and window sliding with a map of face in the images [25, 26, 27, 28]. On the other hand, people detection is a fundamental and important task in many computer vision applications; also, it provides essential information for semantic understanding of the video scenes. It can be used in different industrial applications such as automotive applications due to the potential for improving safety systems, industrial

applications for person avoidance by robots in a factory and following of people with heavy equipment or tools, security applications for autonomous patrolling of secure areas. Although there are some challenges, people detection remains an active research field of computer vision in recent years [29, 30, 31].

The main problem that we discuss in this study is how to count the number of people in different images. People can be seen differently in images. For example, counting the number of people in an image with a few people with clear faces is different than counting the number of people in a crowd. In addition to that, there may be many different challenges in images. For instance, the occlusion between people and variation in environmental conditions are some challenges in people counting systems. Input digital images have a broad degree of variation in the unconstrained environments, level of illumination, image dimensions, skin color, similarity between background color and people's cloths, quality of image, people's closeness to each other that make people detection more difficult. Therefore, the aim of this thesis is to overcome such challenges. To achieve this, we work on different face detection algorithms (Skin Color and Viola Jones) and people detection based on the HOG feature and the SVM classifier to determine the number of people in an image. In order to improve the performance for automatically estimating the number of people in an image, we propose a reliable, accurate, simple and fast method, which combines Face Detection and People Detection methods. Moreover, the proposed method requires hardware components with low cost and complexity.

The people counting systems based on Face Detection and People Detection methods provide further information such as location and appearance of people that is missed by the non-vision infrared beam systems, which counts the number of people without further information. In other words, its information is restricted with counting data only for people. In addition, the performance of people counting systems that are based on an infrared beam counter is less accurate than digital image and video systems. The information obtained from a people counting system based on digital images can provide useful information for further analysis in different practical applications, e.g., in a fire case, we can learn the approximate number and locations of people in a building.

As visualized in Figure 1.2, this thesis has been separated into five main chapters. We start by presenting general background about people counting systems in computer

vision field with a summary of potential problems that may be faced while working on a people counting system. In chapter 2, we provide a literature review related to the counting people field and its methods. Then in chapter 3 we explain People Counting methods step by step in detail, and propose some improvements to increase the accuracy of our system. In addition, we present an overview for datasets used in this thesis. In chapter 4, we discuss all results obtained from our experiments, and show the comparison between different methods that we use. Finally, in chapter 5, we summarize the conclusions. In addition, we list future work in order to give readers some ideas about how to improve people counting systems.

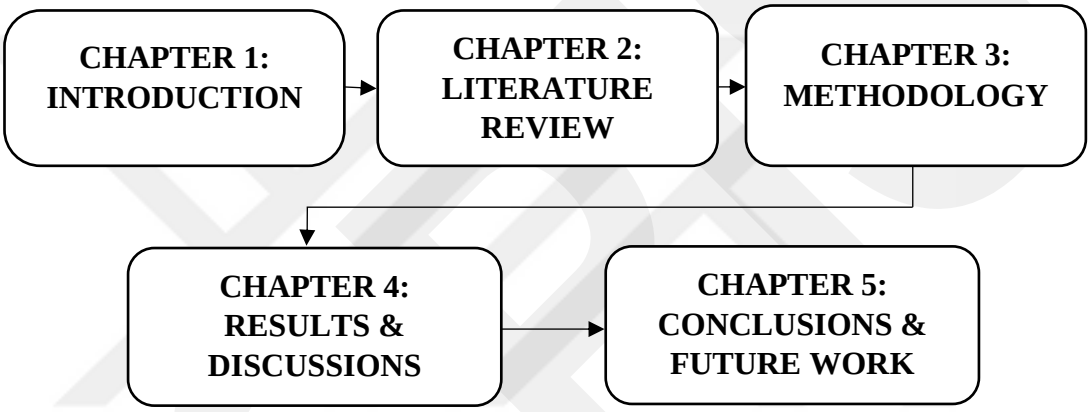


Figure 1.2 Thesis Structure

CHAPTER 2

LITERATURE REVIEW

Counting the number of people in images is an interesting study area and has been given significantly attention in recent years. In addition, it is perceived as a substantial module for many computer vision applications and provides useful information that can be used for further analysis in a wide range of practical applications related to education, commerce, security, and management.

Counting the number of people in videos and digital images are types of the vision-based classes. Counting the number of people from video relies on tracking a person or clusters of people over time in a video without large-scale changes, then, it gives information about the number of people, direction of human movements. However, it is hard to distinguish persons from the background and from each other in case the people are close to each others. Counting the number of people from images counts the number of people without any humans tracking of groups or individuals. Moreover, this approach is suitable for very large number of people in a frame, unlike tracking approach that does not scale well with a growing number of people. Nevertheless, it does not give information about the movement directions of people. Although counting the number of people from images and counting the number of people from videos have many mutual ingredients, they might be diverse in the core approaches of classification. Therefore, in following sections we are going to present the different methods and approaches that have been proposed in order to estimate the number of people.

In this thesis, we aim to count the number of people in digital images and to achieve that we use different methods and propose a new method which combines Face

Detection and People Detection methods. Therefore, in the first section, we give a general review of some robust approaches and existing methods for counting people, which is the main topic of our thesis. Then, in section 2.2, we show a brief summary of papers about people detection approaches. Section 2.3 presents an overview of the papers that study face detection approaches.

2.1. People Counting Approaches

There are many different people counting systems, which roughly can be classified into two main classes: vision-based people counting approaches and non-vision based people counting approaches.

2.1.1. Vision-Based Approaches

Vision-based systems are the most commonly used class of people counting systems, which requires multiple frames for processing. In addition, the counting process of vision-based systems is extended with the direction of blobs to enrich provided information.

Much research has been done in this area. For instance, Zhao et al. [32] developed a people counting approach based on face detection, tracking and trajectory classification. A standard Face Detection method was used in addition to using a face tracking method by combining a Kernel-based Tracking algorithm with a Kalman filter. They extracted the angle histogram of neighboring points from each potential face trajectory. Then, K-NN was used to classify the extracted data. As a result, by applying this approach on a dataset with more than 160 potential people trajectories, they achieved an accuracy rate up to 93%.

Lempitsky and Zisserman [33] proposed a very flexible learning framework based on the MESA-distance to count people in images. They confirmed that using limited amount of training data may lead to much higher accuracy than counting-by-regression, which directly optimizes the accuracy over the image. The advantages of their system are its accurate counting and fast processing time. By using 2,000 video frames from a camera overlooking a busy pedestrian street, they achieved 96.5% detection accuracy.

Liu et al. [34] present a surveillance system that consists of crowd segmentation, visual tracking, a counting recognition module, and auto calibration. Their system is able to count the number of people leaving or entering any specific place by segmenting groups of people into individuals. To evaluate this system, they recorded a 10-minute video taken from a fixed camera at about 6 meters above the ground, and the system worked efficiently.

Bansal and Venkatesh [35] presented a method that uses SIFT, Fourier analysis, wavelet decomposition, GLCM features, and low confidence head detections, as multiple sources to estimate the number of people in high density crowds from still images. By using a dataset with 100 images, they reported an accuracy of 99.4%.

Subburaman et al. [36] estimated the number of people in a crowded scene based on head detection region by using a state-of-the-art cascade of boosted integral features. Two different databases were used to evaluate their approach: PETS 2012 and Turin metro station databases. Consequently, an accuracy of 95% was achieved.

Rahman and Islam [37] identified the boundaries in the whole image's objects by converting it into a grayscale image and finding suitable threshold value in it. Depending on these boundaries, they counted the objects number in the image. In addition, they used a watershed segmentation technique to overcome the problem of the noise of the image, which results in counting wrong objects number. The method gives accuracy up to 96% with high definition images.

Topkaya et al. [18] presented a people counting system based on using different features such as time, place, and colour. In their system, they used HOG feature and the nonparametric nature of the Dirichlet Process Mixture Models (DPMM) in order to detect the number of clusters and count the number of people. They evaluated their system on PETS [38], Peds2 [39], and BEHAVE [40] datasets and achieved an accuracy of 95.1%.

Wang et al. [41] tried to present a simple method that can count the number of people in an image by using feature extraction and pattern recognition techniques. First, they used a Gaussian Masking method and a morphological background modeling in order to detect the targets more effectively from the image. Then they employed the HOG feature to extract meaningful characteristics of people's shape and appearance. In addition to edge features and texture features, HOG features were used to train a

support vector regression machine and present the number of people within the image. Consequently, they were able to get counting performance of 97.12% by applying their method on the UCSD dataset, which was divided into 600 images as a training set and 1,200 images as a test set.

Chan et al. [42] showed how to estimate the size of inhomogeneous crowds without using tracking techniques. Their privacy-preserving system depends on using the mixture of dynamic textures motion model to segment the crowd into groups of homogeneous motion. Then, they extract the features from each segmented region. Then Gaussian Process regression is used to evaluate the consistency between the number of people and the features per segment. Finally, they validated their system on a large pedestrian dataset containing 2,000 images. The result of this method was 98% success counting rate.

Dan et al. [43] present an accurate people counting system based on sensor fusion, which is robust to the crowded condition and illumination variation. The proposed system composed of people detection based on human modeling, lost depth data recovery algorithm, and individuals tracking using both color and depth data. First, a morphological operator processes the depth image to relieve depth artifacts such as the lost data and optical noise. Then, the human model extracts the object from the pre-processed depth image. Finally, by applying the bi-directional matching algorithm, the track of the detected object is established. Consequently, experimental results in various testing environments show that the algorithm realizes over 98% accuracy.

Teixeira and Savvides [44] used a Lightweight method in indoor camera sensor networks for counting and localizing the people. The algorithm uses a motion histogram to detect people based on size and motion standards. Their algorithm has been tested and implemented on a network of iMote2 sensor nodes. As a result, they were able to reach a counting accuracy up to 88.6%.

2.1.2. Non-Vision Approaches

In general, the non-vision-based category of people counting systems includes all systems that are not based on images or video data from a camera device. Examples are a heat sensor, the light of an infrared beam, or pressure sensors.

Several methods have been published to estimate the number of people with a non-vision approach. Li et al. [45] presented a People-counting method dependent on a back propagation (BP) neural network using a photoelectric sensor. This effective and flexible sensor is used to collect data, recognition and counting the number of people. At first, they used a data segmentation approach. Then, they applied a features extraction technique to extract characteristic parameters of a pulse sequence, which is used to decrease computational complexity. Finally, they employed the BP network as an adaptive and robust classifier. The presented people-counting results are promising and have low false rates.

Yang et al. [46] estimated the number of people in crowds by using a simple image sensor, which segments frontal objects from the background, collects the resulting shadows over a network, and computes a planar projection of the scene's visual hull. They also applied a geometric algorithm that eliminates the phantom regions, and then gathers the individuals' number in each projection region. Typically, the results are promising where the system allows them to count a much larger crowd in a much larger area.

Zhang et al. [47] used three methods—Mean shift, Random Forest and Water Filling—to count the number of people based on a vertical Kinect sensor, which is robust to remove the effect of appearance variations. However, by doing several experiments, they obtained meaningful accuracies of 50.67%, 91.05%, 99.16% for mean shift, random forest and water filling methods, respectively. Moreover, they obtained recall rates of 89.09%, 83.87%, 98.42% with mean shift, random forest and water filling methods, respectively. According to the particularity of the depth map and the results, they recommend using a novel unsupervised water filling method that can find the correct regions with the property of scale-invariance, locality and robustness.

2.2. People Detection Approaches

People detection is a fundamental and important task in many computer vision applications; also, it provides essential information for semantic understanding of the video scenes. Many methods have been presented in order to solve the people detection problem. Englebienne and Krose [48] presented a template-based method which can automatically detect the number of individuals in a frame. In different locations and

poses, they manually annotated 646 images in the ground plane of the individuals shown in the image with different annotators and ran the annotation three times for each image. By comparing template-based method to a state-of-the-art background segmentation algorithm, the method exhibits its effectiveness and accurate performance advantage due to its ability to run at near real time frame rates.

Pishchulin et al. [49] investigated the possibility of using training data generation in 3D human body models. They applied Rendering-based Reshaping method in order to create many artificial training samples from only a few views and individuals. To do so, they used the Histogram of Oriented Gradients (HOG) feature extractor and the pictorial structures model. In addition, they used 3 databases: the Reshape training dataset which contains 2,000 images; the CVC training dataset which has 3,432 images; and the Multi-viewpoint dataset, which has 1,486 images. As a result, they could improve performance by combining different datasets, achieving up to 87% accuracy.

Tang et al. [50] investigated multiple people detection in crowded places. They developed and trained a joint model to detect both single or pairs of people even under different degrees of occlusion. The result of this method was 90.5% success detection accuracy by working on two datasets: "TUD-Pedestrians," which contains 250 images, and "TUDCrossing," which contains 201 images.

Corvee et al. [51] used two descriptors/features to extract main information from the images that are Local Binary Pattern (LBP) and Mean Riemannian Covariance Grid (MRCG). The LBP has been used to detect faces, heads and people, whereas MRCG has been used to obtain highly discriminative human signature by model appearance of tracked people. Moreover, the proposed methods were evaluated by using state-of-the-art algorithms. They used the INRIA human dataset, which consists of 1,132 human images (positive samples), and 453 images of background scenes containing no humans (negative samples). Consequently, detection accuracy of 85% was achieved.

Mozos et al. [52] addressed the issue of people detection by using multi-layers of 2D laser range scans. Each layer contained a classifier that was able to detect a specific body section such as an upper body, a head, or a leg. These classifiers have been learned by using a supervised approach based on AdaBoost. This kind of proximity

sensor is used in robotic applications, which provide a high data rate and a wide field of view. Accuracies of 86.2%, 84.4%, 94.3% in head, upper body, and leg, were achieved, respectively.

Garcia-Martin et al. [53] presented a new algorithm in order to detecting people based on motion data. The algorithm creates an individual's motion model depending on the Implicit Shape Model (ISM) Framework and the MoSIFT descriptor. Moreover, they proposed a detection system that combines tracking, motion and appearance information. However, experimental results over progressions extracted from the TRECVID dataset with 6,353 images (frames) show that their motion-based detector generates results similar to the ISM State-of-the-art approach. The valuation of the system's performance shows how the combination of different information sources obtains a significant improvement in recall, improves the final detection accuracy, and a slight precision reduction. They made many experiments but the best-achieved precision ratio was 93.9%.

Bourdev et al. [54] developed a new people detection algorithm by using poselets. They used 2D annotations due to they are much easier for body annotators. Furthermore, they considered the detection scores of nearby objects in the place, which lead to improve the object detection. This can be done by training a multi-layer Feed-Forward network with weights set using a Max Margin technique. The repeated poselet activations are then clustered into alternately hypotheses where consistency is based on determined locative key-point distributions. Finally, bounding boxes are estimated for each individual's hypothesis, and segmentation can be provided by aligning the shape masks to edges in the image. They used the H3D training set, which contains 750 images; the PASCAL VOC 09 training set, which contains 2,819 images, and 240 images added manually from Flickr. As a result, they obtained an average precision of 47.8%.

2.3. Face Detection Approaches

Detecting human faces in images has attracted the attention of wide number of expert in the computer vision field. One of the pioneering studies related to this area was done by Viola and Jones [55]. They depicted a machine learning approach for detecting visual objects very quickly and accomplished high rates of object detection. This study

is distinct for three main contributions: First, the “Integral linage” permits the features used by the classifier to be computed rapidly. Second, based on Adaboost, a learning algorithm chooses a little number of basic visual features from a larger set and produces effective classifiers [56]. The third contribution is permitting the image’s background spaces to be readily disposed of by using a method for combining classifiers in a “cascade”. They trained Adaboost procedure with 9,544 images to obtain a face detection accuracy of 93.9%.

Wong et al [57] presented an efficient method for feature extraction and face detection. The idea of this method depends on using the genetic algorithm to detect the location of the face regions. This algorithm is applied to determine conceivable face regions in an image. Then, they applied Eigen face technique to determine the appropriate regions. The MIT face database was used to validate the performance of their method and they claimed 95.3% detection accuracy.

Khan et al [58] studied how to find a suitable color space for skin detection by evaluating the effect of the luminance component and transformation to a color space to increase performance on skin detection. Then, they found the proper pixel by using a color-modeling technique. Finally, by applying color constancy algorithms on 8,991 images collected from the Internet, they were able to obtain accuracy of 74.5%.

Wu et al [59] studied using fuzzy set theory to determine faces in color images. They used two fuzzy models; the first one is used to extract skin color regions, while the second one is used to extract hair cooler regions. To detect the face, they compared the previous models with prebuilt head-shape models by applying them to 223 images collected from the internet. They have reported 97% detection accuracy.

Lin [60] presented a scheme in order to detect human faces in different backgrounds and illumination conditions. This scheme consists of two stages, the first one uses YCbCr colour space to search on possible face regions depending on color and triangle-based segmentation. The second one uses multilayer feedforward neural network in order to verify a face. In addition, the system can classify varied face's sizes even with different facial expressions and illumination conditions. In this study, Lin used 1,500 images from the AR face database and got a face detection precision of 98.2%.

Froba and Ernst [61] worked on two folds; the first part is related to object detection by illumination invariant Local Structure Features. They proposed a Modified Census Transform in order to increase performance, and then showed some faults and how to overcome them. The second part is related to improving the speed of detection. To do so, they used a four-stage classifier, where each stage was composed of group of feature lookup-tables. Consequently, by working on 1,526 images selected from BioID database, their method achieved a detection accuracy of more than 90%.

Osuna et al [27] researched the utilization of Support Vector Machines (SVMs) in computer vision field. In addition, they presented how to train an SVM classifier on a very large dataset by using a decomposition algorithm. This algorithm is used to evaluate optimality conditions to create enhanced iterative values and build up the ceasing criteria for the algorithm. By using 336 images taken from the Database of Faces, they claimed a detection accuracy of 74.2%.

Ghimire and Lee [62] proposed a human faces detection method in color images. Their method is lighting insensitive, which is based mainly on the extracted information about edge and skin in color image. In details, they enhanced the images taken from different illumination conditions. Then, the skin color segmentation is conducted in YCbCr and RGB space. Finally, by using the Skin Tone Percentage Index method, the results were refined. However, they selected 302 face images from the FRGC database [63] to evaluate the performance of the proposed method, and they achieved 85.96% successful detection accuracy.

CHAPTER 3

PEOPLE COUNTING METHODS USED IN THIS STUDY

Most techniques that have been used for people counting are based on tracking and moving object detection, which assumes that all moving objects are people. However, we need a new method for counting people in digital images. People can be seen differently in images, which requires the use of different techniques together. For example, counting the number of people in an image with a few people with clear faces is different than counting the number of people in a crowd. Using the same technique for these two different images will not give a successful performance. Therefore, firstly, we have analyzed images to determine image categories according to people inside. Our result is that there are three main categories which require different techniques for people counting. The first category includes images which contain people with detectable faces (*see Figure 3.1*) and we hypothesize that face detection algorithms may give successful results for this category.

Figure 3.1: People with detectable faces

The second category includes images having whole bodies. For this type, face detection algorithms cannot produce successful results since faces are not clear (*see Figure 3.2*). We assume that whole body people detection algorithms can be used for this category.

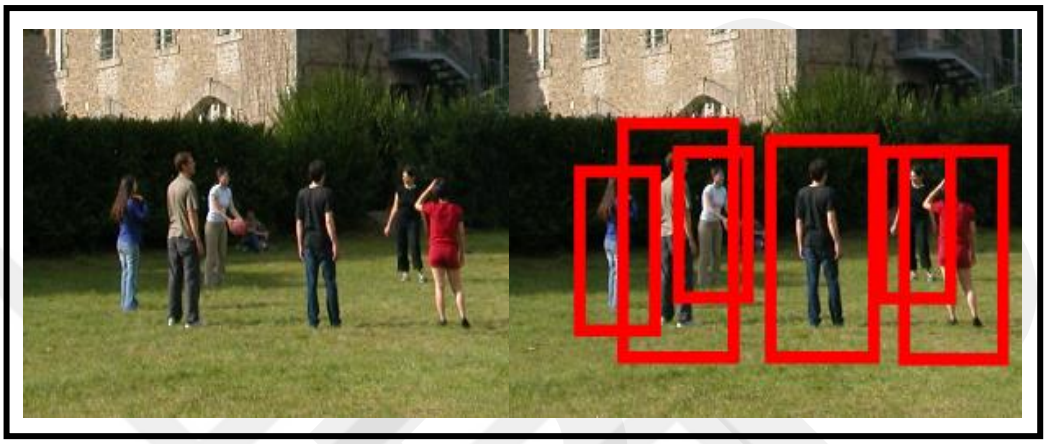


Figure 3.2: People with whole body

The last category includes images of crowds for which neither face detection nor people detection algorithms can be used (*see Figure 3.3*).

Figure 3.3: Crowds of people

In this study, we tested two different techniques: the Face Detection method and People Detection method, in order to estimate the number of people in an image. We

propose a new method, which combines the outputs of the Face Detection and People Detection methods, in order to improve the performance of estimating the number of people in an input image with low cost and simple hardware. Figure 3.4 shows our proposal to count the number of people in a static image.

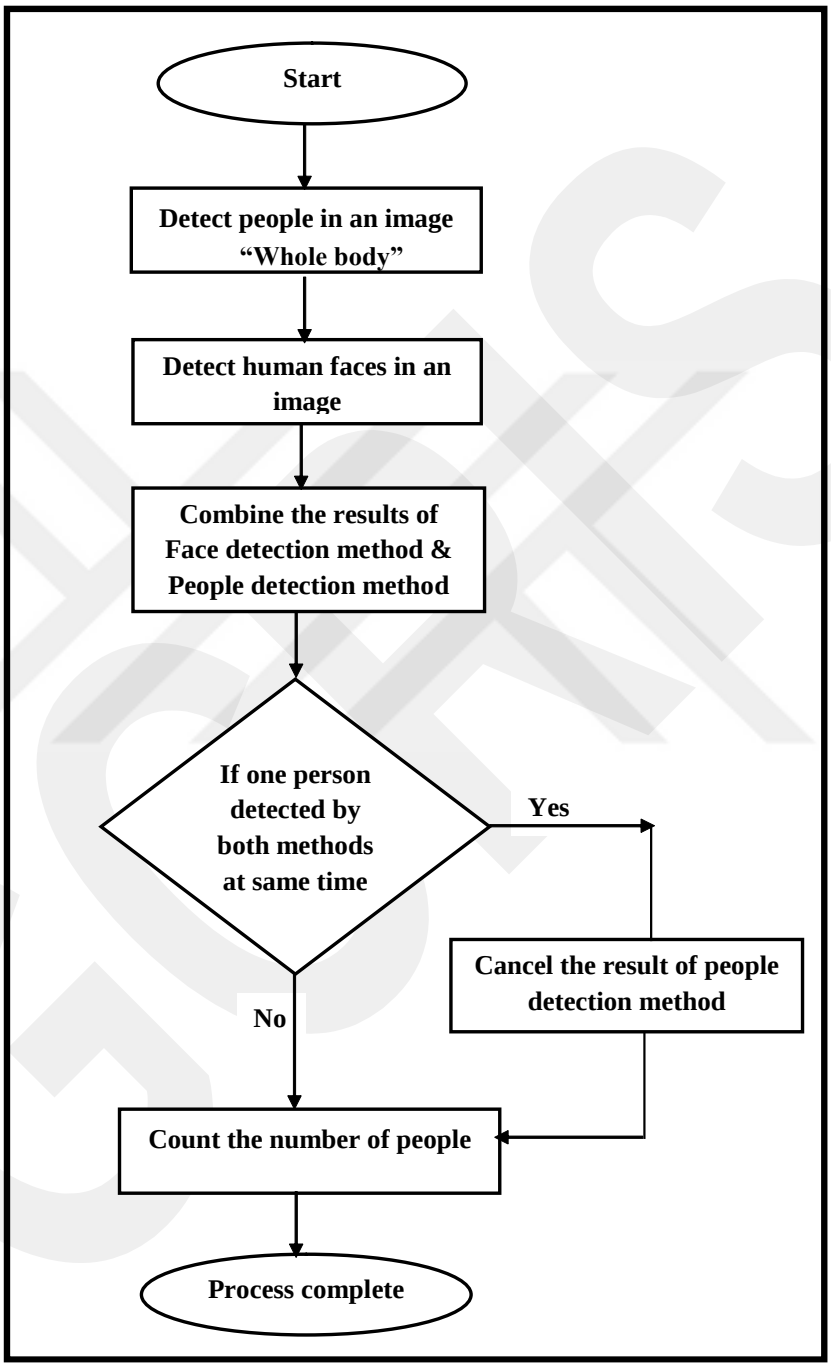


Figure 3.4: General methodology for people counting system

In general, the system inputs a digital image and then applies the People Detection method to detect people. It applies one of the following Face Detection methods: Skin

Color face detection, Viola & Jones CART face detection, or Viola & Jones LBP face detection in order to detect a human face within an image. After that, the results are combined. Thus, if a person is detected by both methods (Face Detection method and People Detection method), the system should count the number of people in an image taking this information into account.

3.1. People Counting Based on Face Detection Methods

Face detection is a specific class of object detection and plays important role in distinguishing and recognizing people, which can be used to check whether the given input image or video frames contains any human face or not. Face detection algorithms return the human face's location in the image if found.

In this study, we work on three different methods for face detection in order to estimate the number of people in an input image. These methods are; (1) face detection based on Skin Color, (2) face detection based on Viola Jones LBP method and (3) face detection based on Viola Jones CART method. These methods are discussed in details below.

3.1.1. Face Detection Based on Skin Color

People's skin color has been used as an effective way for feature extraction in various existing applications of face detection. The Skin-Color Face Detection method has been used to find the face region in an input image based on the Color-Model. It is a very popular and simple approach. Skin color pixels are represented by different color models such as RGB [64, 65, 66], Hue Saturation and Intensity (HSI) or Hue Saturation and intensity Value (HSV) [67,68,69,70], YES [71], CIE LUV [72], YCbCr [73,74], CIE XYZ [75] and YIQ [76,77], normalized RGB [78,79,80,81,82, 83, 84], etc. In order to detect a human face, the skin color feature is not enough to get sufficient results. If only the skin color information is used, it may lead to an increase in the false ratio of face detection. For instance, when human faces are close to each other, it is difficult to distinguish them. Therefore, many of the modular systems utilize a combination of several features, such as size, shape, skin color features, and color segmentation to detect head and faces [69, 72, 85, 86].

In this thesis, we test face detection based on Skin Color to detect multiple faces in a digital image, and then we count these faces to estimate the number of people in an input image. The general algorithm of this method can be divided into seven steps which are:

1- Input a digital image
2- Alter the color space
3- Detect face region
4- Extract face model
5- Reduce noise
6- Detect face
7- Count people

In order to improve the performance of skin detection in color images, we use the combination of HSV and YCbCr color spaces. Then, we detect faces and count each detected face in an input digital image. The steps of our implementation are given in Figure 3.5 and explanations of these steps are given next.

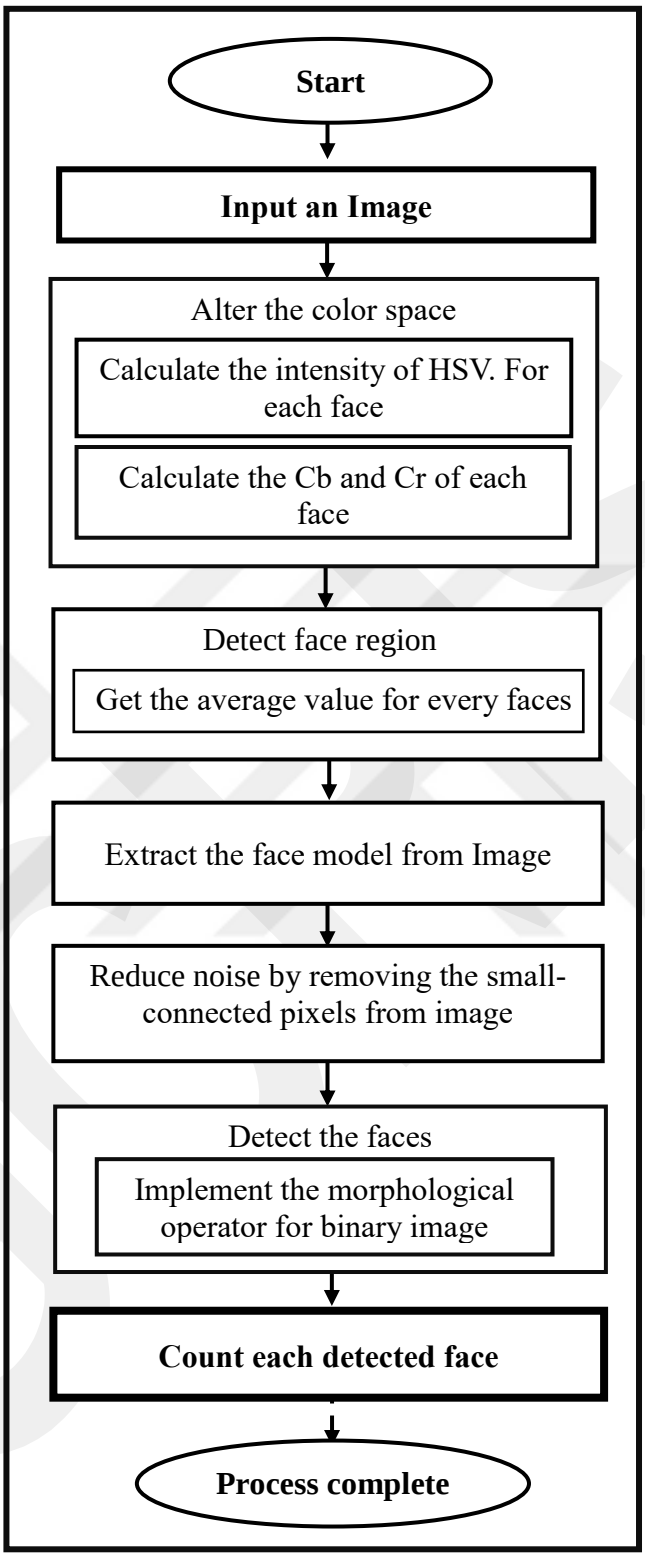


Figure 3.5: Steps of people counting based on Skin Color face detection.

- **Alteration of color space:**

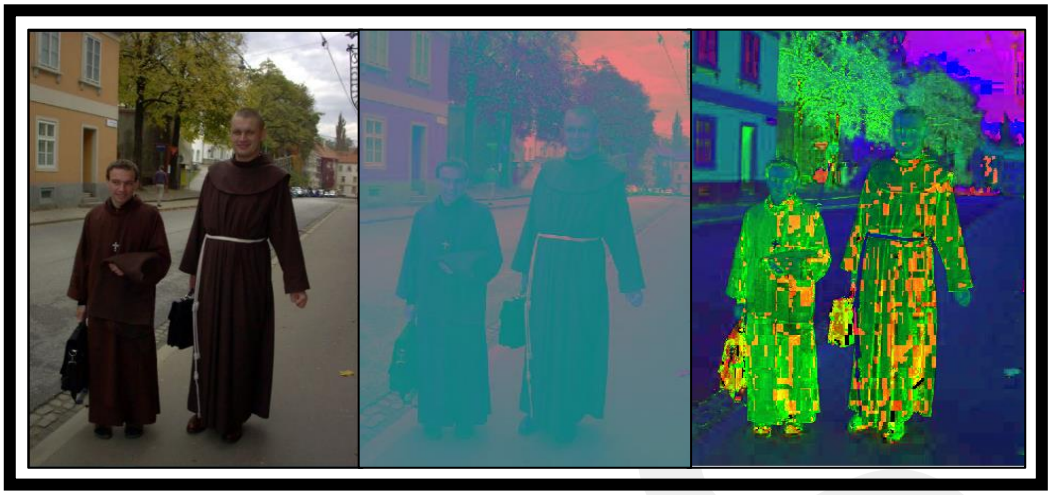
There are many types of color space that can be used to label the pixels in an image as skin color such as RGB, HSV or HIS, normalized RGB and YCbCr, etc. The skin color model requires selecting a suitable color space first before locating candidate face regions. Therefore, in order to improve performance of skin detection, this method combines two types of color space that are HSV and YCbCr. Thus, in the first step the input image is transformed from RGB color space to HSV and YCbCr color space individually (as shown in Figure 3.6) because the variance in illumination environments cannot be described by using RGB. Equation (1) shows the conversion from RGB color space to HSV color space, equations (2) and (3) show the conversion from RGB color space to YCbCr color space:

$$\begin{cases} H = arccos \dfrac{\frac{1}{2}((R-G)+(R-B))}{\sqrt{((R-G)^2+(R-B)(G-B)}} \\ S = 1 - 3\dfrac{\min(R,G,B)}{R+G+B} \\ V = \dfrac{1}{3}(R+G+B) \end{cases} \quad \text{---------- (1)}$$

$$\begin{aligned} Y' &= 16 + (65.481 \cdot R' + 128.553 \cdot G' + 24.966 \cdot B') \\ C_B &= 128 + (-37.797 \cdot R' - 74.203 \cdot G' + 112.0 \cdot B') \\ C_R &= 128 + (112.0 \cdot R' - 93.786 \cdot G' - 18.214 \cdot B') \end{aligned} \quad \text{---------- (2)}$$

Or

$$(Y', C_B, C_R) = (16, 128, 128) + (219 \cdot Y, 224 \cdot P_B, 224 \cdot P_R) \quad \text{---------- (3)}$$



(a) (b) (c)

Figure 3.6: The result of color space conversions
(a) RGB color image (b) YCbCr image (c) HSV image

- **Detection of face region:**

To extract skin color, the input color image (*see Figure 3.7*) is segmented based on a combination from two types color model—HSV color space and YCbCr sub color space—in order to increase the accuracy of classifying the region of an image as skin or non-skin. Thus, we calculate the intensity of the HSV color space for each face in an input image. Then we calculate the intensity of Cb and Cr color space for each face in an input image in order to get the average value for every face in the image to obtain adaptable threshold value by that feature. A threshold value should be applied to decide whether it is related to skin or not. Next, we segment the image based on HSV, Cb, Cr components to detect skin regions of the face, and use thresholding for every pixel to estimate face area. All the pixels are classified as face region if the values of Cr_{new}, Cb_{new}, HSV_{new} fall within the following ranges:

$$C_r - 1 < C_{r\,new} < C_r + 1$$

$$C_b - 1 < C_{b\,new} < C_b + 1$$

$$HSV - 1 < HSV_{new} < HSV + 1$$

(a) (b)

Figure 3.7: The result of face region detection

(a) Original image (b) Segmented skin regions image.

- **Extracting face model:**

This step uses the morphological reconstruction from morphological mathematics and a robust transformation of image processing in order to extract the shape of face model as human face. The morphological reconstruction has two unique properties of processing; these are a structuring element and two images, which are a marker image and a mask image, instead of a structuring element and one image. The structuring element is commonly used to specify connectivity. In addition, the marker image includes starting points that can be used for the transformation and other images, but the mask restricts the transformation. However, determining the marker and structuring element, plays significant role of the morphological reconstruction. Therefore, this method uses the open operation of morphological reconstruction as maker image with 8*7 pixels structuring element to reconstruct the shape of faces detected, which can nearly describe the shape of human being faces. After extracted information about the shape face candidates, the skin region is classified as face region if:

- The height of skin area “h” is larger than the width of skin area “w”.
- The ratio of “h/w” is less than 2.

In other cases, the region is classified as non-face area.

- **Reduce noises:**

After finding the area that has the face pixels, some pixels may be noise and mistakenly determined to be face. Therefore, these pixels should be removed by rejecting the connected region, which is smaller than a typical region of a face. We use a simple binary erosion of morphological operation to reduce the noise with fill holes or gaps (*see Figure 3.8*).

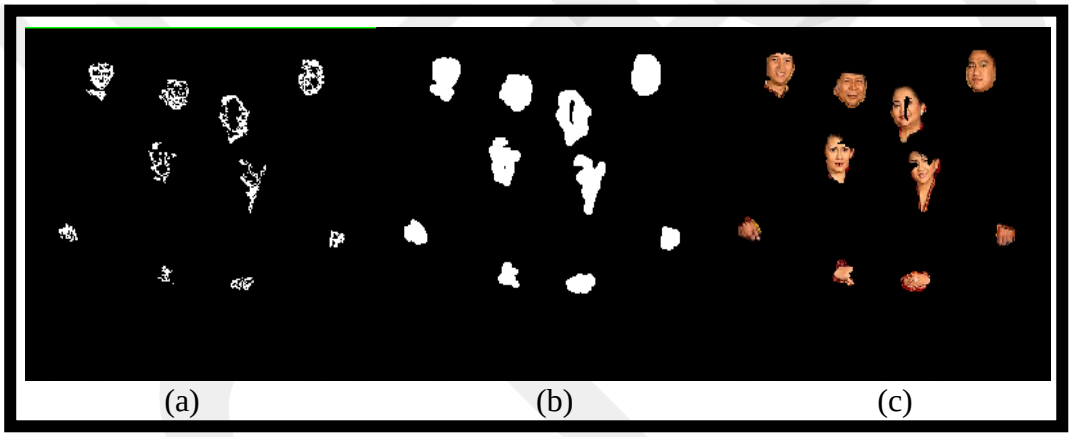


Figure 3.8: The result of noise reduction

(a) Remove a small-connected image (b) Filled-hole image

(c) Color image after hole-filling

- **Face detection:**

After implementing previous steps, the people's faces are detected with drawn bounding box around each face in an image as shown in Figure 3.9.

(a) (b)

Figure 3.9: The result of face detection based on Skin color

(a) Original image (b) Final-result for face detected

- **Count people:**

In the final step, we count the number of people based on the number of faces detected in an input image as shown in Figure 3.10.

(a) (b)

Figure 3.10: The final-result of people counting based on skin color face detection

(a) Original image (b) Final-result for counting people

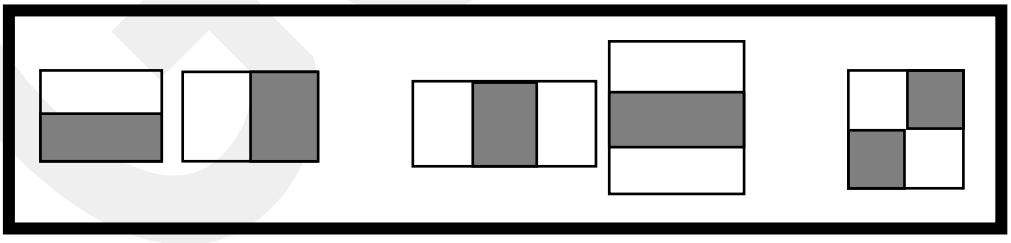
3.1.2. Face Detection Based on Viola-Jones Method

Viola-Jones is the first framework for object detection that can provide competitive rates of object detection in a real-time domain. It was presented by Paul Viola and Michael Jones [55, 87] in 2001. Viola-Jones face detection is a robust and quick method compared to other techniques released at that time. This algorithm is based on sub windows with a fixed size (24×24 pixels), which can be slid over the entire image and which looks for a specific Haar feature. In addition, it denotes every position of a human face candidate. Therefore, if one of the Haar features is found, then the sub window passes to the next stage of face detection algorithm; this sub window has the ability to be scaled instead of rescaling the input image in order to get a variety of different sizes of human faces.

There are four main stages in the Viola-Jones face detection algorithm, which are extraction of Haar features, generation of an Integral Image, application of Adaboost algorithm and finally cascading classifiers. These stages are elaborated below:

- **Extraction of Haar features**

Haar-like features represent different types of rectangles that have two-rectangle, three-rectangle and four-rectangle features. This Face Detection method is based on more than one rectangular feature; these features are used in Cascaded classifiers as input features, as shown in Figure 3.11. In addition, the pixels summation in white region of rectangles is subtracted from the pixels summation in gray region rectangles.



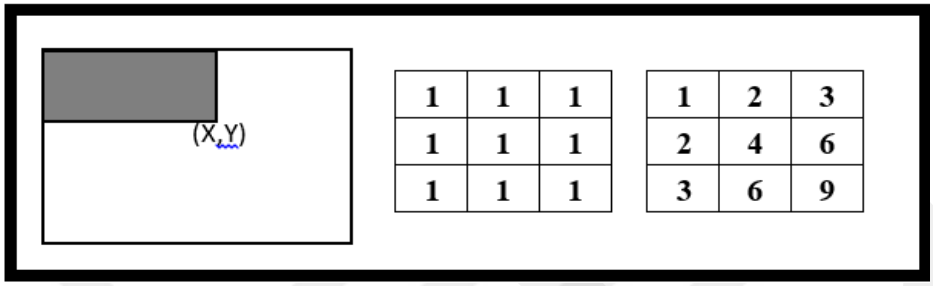
(a) (b) (c)

Figure 3.11: Types of rectangle features

(a) Edge-rectangle features (b) Line-rectangle features (c) Four-rectangle features

- **Integral Image**

Integral image is an intermediate representation, which provides the ability to compute rectangle features very rapidly at several scales. Furthermore, it can be computed by using a few operations in each pixel of an image. Therefore, the location (x, y) of the integral image represents summation of the original image's pixels above and to the left of (x, y), as shown in Figure 3.12.



(a) (b) (c)

Figure.3.12: The generation of Integral image

(a) Represents pixel value in (X, Y) (b) Image in size 3x3 (c) Integral image

- **AdaBoost**

The final result of feature extraction produces a lot of data, which is over fitting. These large numbers of features must be reduced. Therefore, the Viola-Jones face detection algorithm uses AdaBoost, which was introduced in 1995 by Freund and Schapire [88, 89]. In addition, it is an efficient classifier to train and is used for feature selection by selecting a small number of best and significant features in a sub window of a digital image, then combining them in order to create an effective classifier.

- **Cascading Classifiers**

Cascading classifier is an important component in the Viola-Jones algorithm, which is used to combine weak classifiers in a "cascade "of stages. Furthermore, it provides the ability to discard background regions quickly in an image while spending additional computation on promising human face-like regions. As shown in Figure 3.13, the input sub-windows are passed through a sequence of stages during detection. Thus, each stage determines whether a given sub-window is a face or not a face. When a sub window is classified as not-human-face by a given stage, it is promptly rejected. On

the contrary, a sub-window that is classified as a human face is passed into the next stage of the cascading classifiers

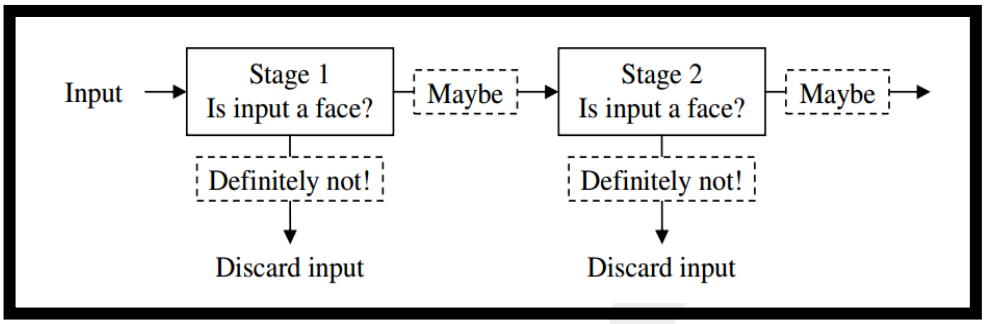


Figure.3.13: The flow work of cascaded classifier

- **Vision Cascade Object Detector**

The Computer Vision System Toolbox presents the Cascade Object Detector. It can detect different categories of objects which have fixed aspect ratios, such as human's faces, cars and stop signs, etc. It uses a sliding windows approach, which is a common and simple technique for identifying and localizing objects in an image by sliding a rectangular region of specified width and height over an image. In this thesis, we use the Cascade Object Detector method, which uses the Viola-Jones face detection algorithm [87] that is already implemented in Matlab R2015a software.

In order to detect upright facing and forward facing of multiple human's faces in an input digital image, we use the Cascade Object Detector method that consists of several pre-trained classifiers. However, two different types of classification models are used with the Cascade Object Detector, which are FrontalFaceCART (Classification and Regression Tree Analysis) and FrontalFaceLBP (Local Binary Pattern). Figure 3.14 shows the methodology overview of people counting systems based on the Viola-Jones Face Detection method.

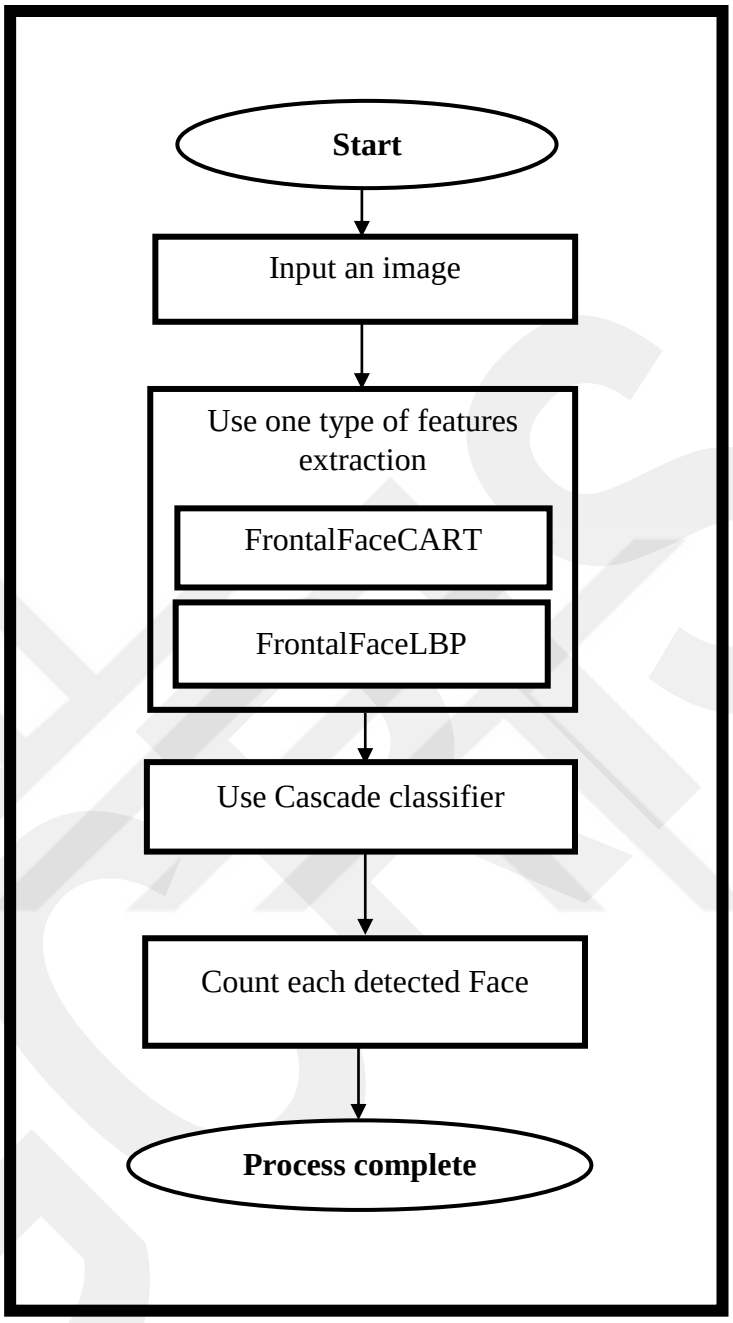
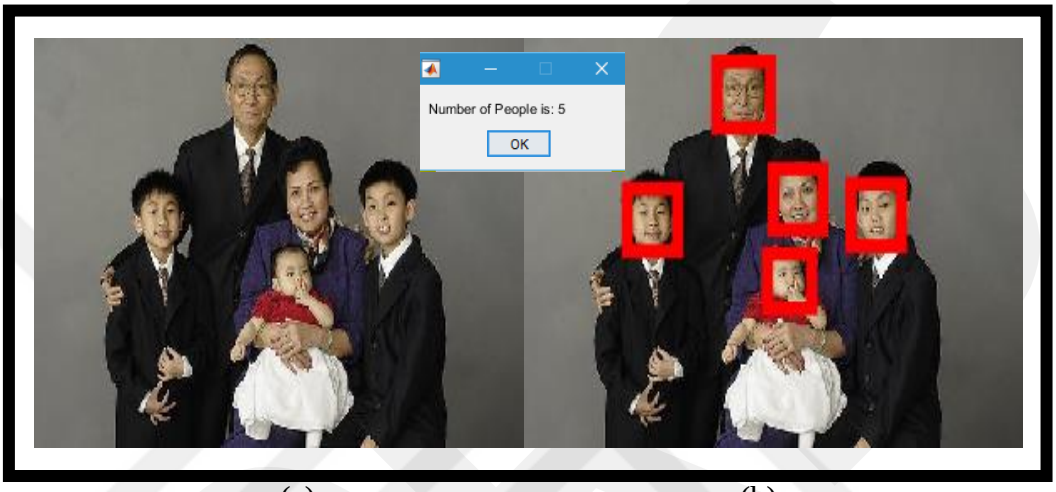


Figure.3.14: The methodology overview of a people counting system based on the Viola-Jones face detection method

- **FrontalFaceCART**

This type of classification model consists of several weak classifiers, which use Haar features to encode facial features in an input image, and relies on regression tree analysis (CART) in order to model higher-order dependencies between facial features of faces in an image. Figure 3.15 shows the result for Viola-Jones face detection by using FrontalFaceCART model classification.

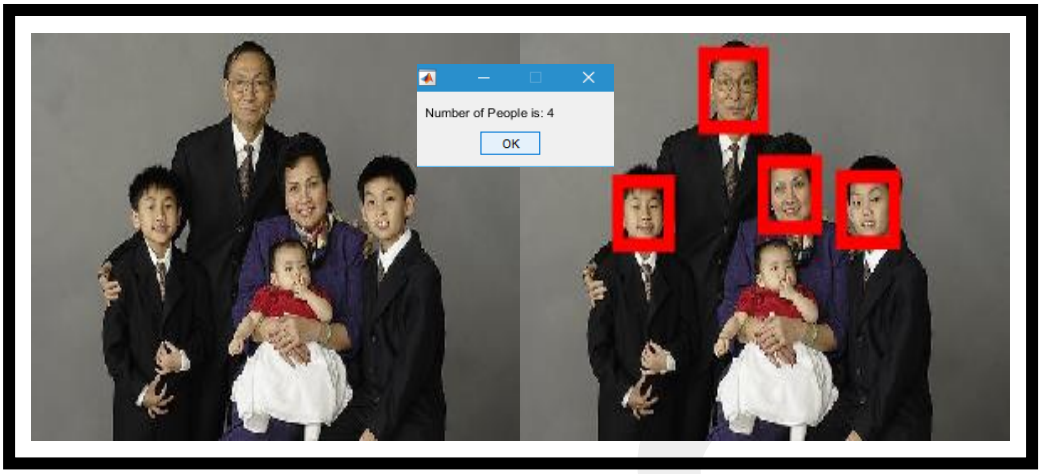


(a) (b)

Figure.3.15: The final-result of people counting based on Viola-Jones face detection method (a) Original image (b) Final-result of FrontalFaceCART model

- **FrontalFaceLBP**

This type of classification model also consists of several weak classifiers that rely on a decision stump, but these classifiers use LBP instead of Haar features to encode facial features and provide robustness against differences in illumination. Figure 3.16 shows the result for Viola-Jones face detection by using FrontalFaceLBP model classification.



(a) (b)

Figure.3.16: The final-result of people counting based on Viola-Jones face detection method (a) Original image (b) Final-result of FrontalFaceLBP model

3.2. People Counting Based on People Detection Method

People detection, which is an important class of object detection, and people counting are significant problems in the computer vision area. However, the ability to detect a person under widely varied conditions is a challenging task; such challenges include differences in illumination, partial occlusion, complex backgrounds, appearance, color or type of clothing and pose. We think that a whole body People Detection method to estimate the number of people in an input image must be a part of a complete people counting system.

The Computer Vision System Toolbox of Matlab presents a People Detector System that can detect upright people in an input image. Thus, in this thesis, we use the People Detector System which uses the Dalal and Triggs people detection algorithm [90] that is already implemented in Matlab R2015a software. The technique is based on Histograms of Oriented Gradients HOG feature. It trains linear Support Vector Machine (SVM) classifier in order to determine people. Then, we count each detected person in an image to get the total number of people in an input digital image.

The main idea of the People Detection method is to implement the rectangle-sliding window of a fixed size 64x128, which includes about 16 pixels that are scanned over the input image at all positions and scales to extract the features. Then combined

vectors of features are produced that can be used in the SVM classifier to classify the objects in an input image as person or non-person. Finally, each person detected in an input image is counted (*see Figure 3.17*). In the next section, we explain the working steps in more detail.

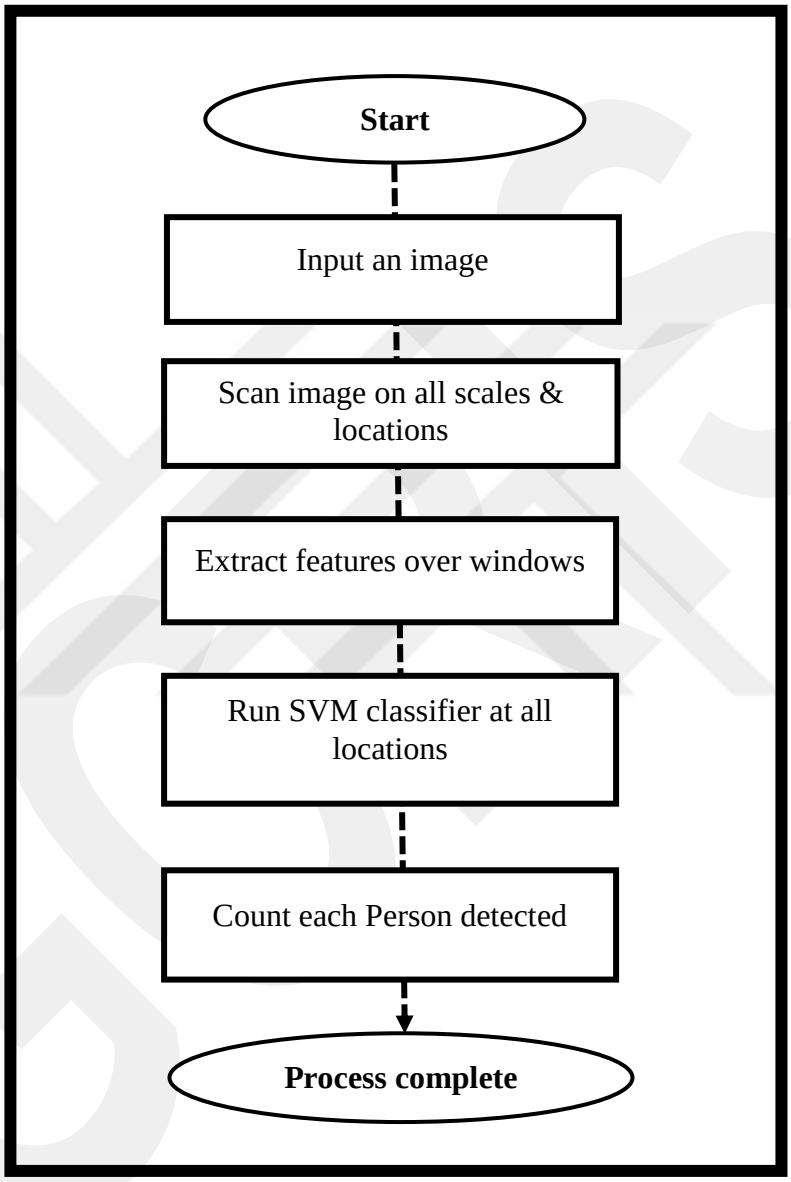
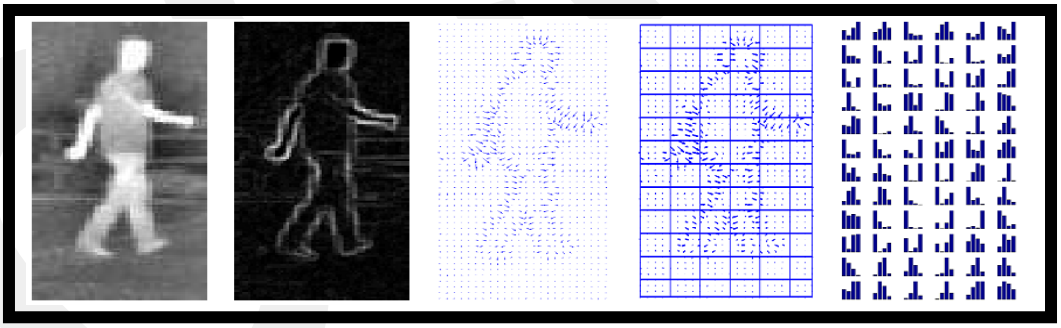


Figure.3.17: The Methodology Overview of People Counting System based on People Detection Method

3.2.1. Histograms of Oriented Gradients (HOG)

Features extraction is an important step in people counting systems based on the People Detection method. Therefore, the People Detection method uses HOG as a feature descriptor that produces feature vectors in order to encode the right amount of significant information related to a particular object as data. Moreover, the HOG features are extracted from the test image to provide ability to make predictions using the trained classifier. This technique has the ability to count occurrences of gradient orientation in localized portions of a digital image. Furthermore, it differs from others approaches in computing a dense grid of uniformly spaced cells and it also improves the accuracy by using overlapping local contrast normalization.

Notably, the main mechanism of HOG is that it divides the image into small associated regions called cells, and for the pixels inside each cell, a HOG directions is compiled. Thus, the histogram will be produced from the gradient values. Then the chain of these histograms is represented as descriptor (*see Figure 3.18*). To obtain better accuracy, the local histograms contrast normalize overlapping spatial blocks by computing an intensity measure across blocks, which is a larger region of an image. After that, this value can be used to normalize all cells inside block; the results of normalization improve invariability to changes in illumination and shading.



(a) (b) (c) (d) (e)

Figure.3.18: The Histograms of Oriented Gradients features.

(a)Original Image; (b) Gradient magnitudes; (c) Gradient orientations;

(d) Subdivided image cells; (e) Histogram gradients in each cell.

3.2.2. Support Vector Machine (SVM)

Support vector machine (SVM) classifier is one of the supervised learning classifications and commonly used method in object recognition, which was proposed by Cortes and Vapnik in 1995 [91]. The main goal of SVM is to predict target values of test data by giving just the test data attributes when classification separates data into testing and training samples. In addition, SVM does not need to add a priori knowledge to introduce high generalization performance and could work for small training sets.

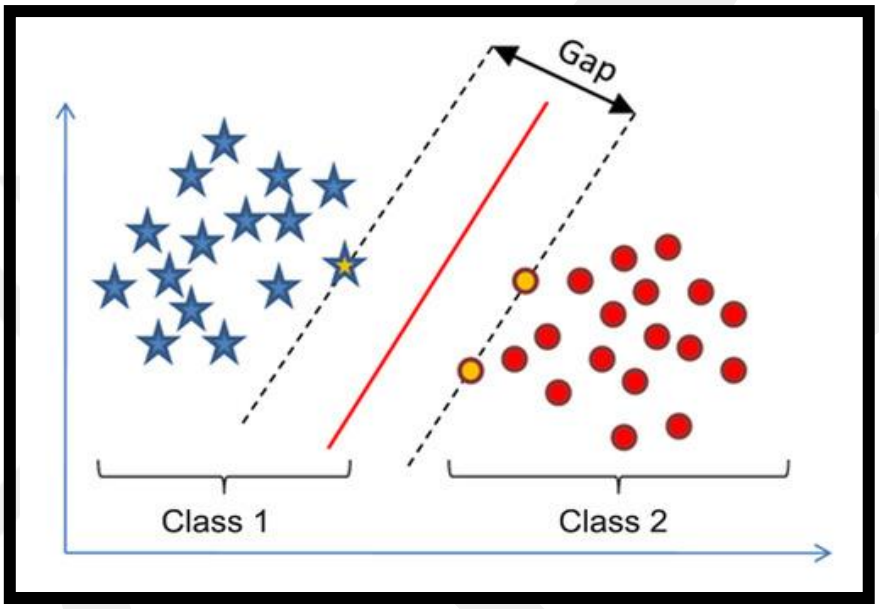


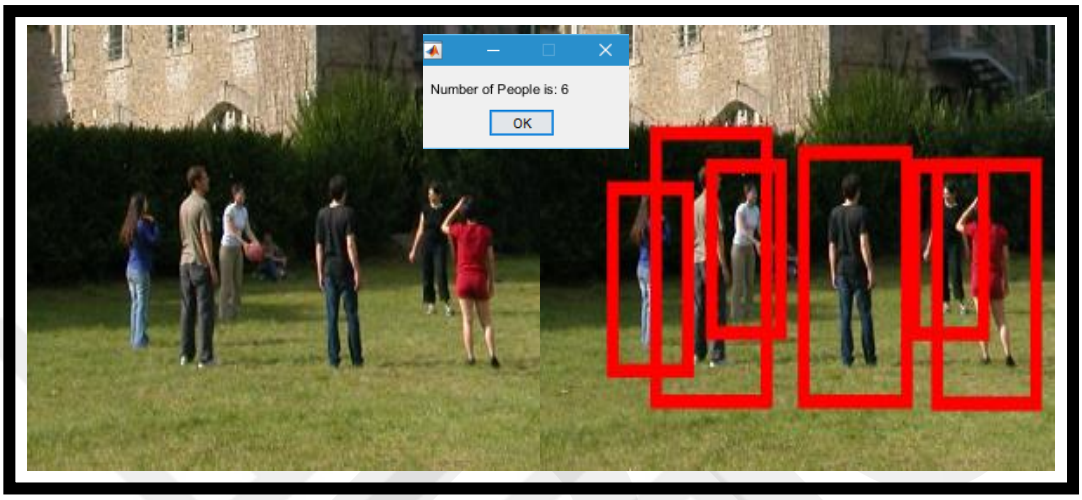
Figure.3.19 The general idea of SVM working

Figure.3.19 shows how the SVM classifier finds a linear decision surface by separating Class1 from Class2, where blue stars represent Class1 that consists of positive feature samples and Class 2 is a red circle shape, which consists of negative feature samples. In order to find the OSH, SVM increases the margin between Class1 and Class2 bit by bit. The orange stars and orange circle represent OSH in each class, which can be used by SVM in the classification process.

In general, the object classification phase is a significant task in many applications of the computer vision field, including automotive, image retrieval, and safety surveillance. Therefore, our system used linear Support Vector Machine (SVM) classifier, which is reliable and fast classifier. Moreover, in order to classify the objects

in an input image as person or non-person, the system applied SVM classifier on normalized image windows.

The final step of a people counting system based on the People Detection method is counting the number of people that are detected in an input image (*see Figure 3.20*).



(a) (b)

Figure.3.20: The final-result of people counting based on People Detection method. (a) Original image; (b) Final-result for People Counting System.

3.3. Combination of Methods

To determine the correct algorithms for a successful people counting system, we tested the combination results of each method as follows:

- Viola Jones LBP with People Detection method
- Viola Jones CART with People Detection method
- Face detection based Skin Color with People Detection method

In the combination of two methods, when we combine the results of Face Detection method with the results of People Detection method, each method produces a bounding box, where the results of People Detection method are represented in big bounding box, whereas the results of Face Detection method are represented in small bounding box. The counting process is done by checking whether the big bounding box contains a small bounding box, if so, the system omits the big bounding box and counts all other

boxes within the image, where each box indicates one person. Figure 3.21 shows the combination of the Viola Jones Face Detection method with the People Detection method. Figure 3.22 shows the combination of the Skin Color Face Detection method with the People Detection method.

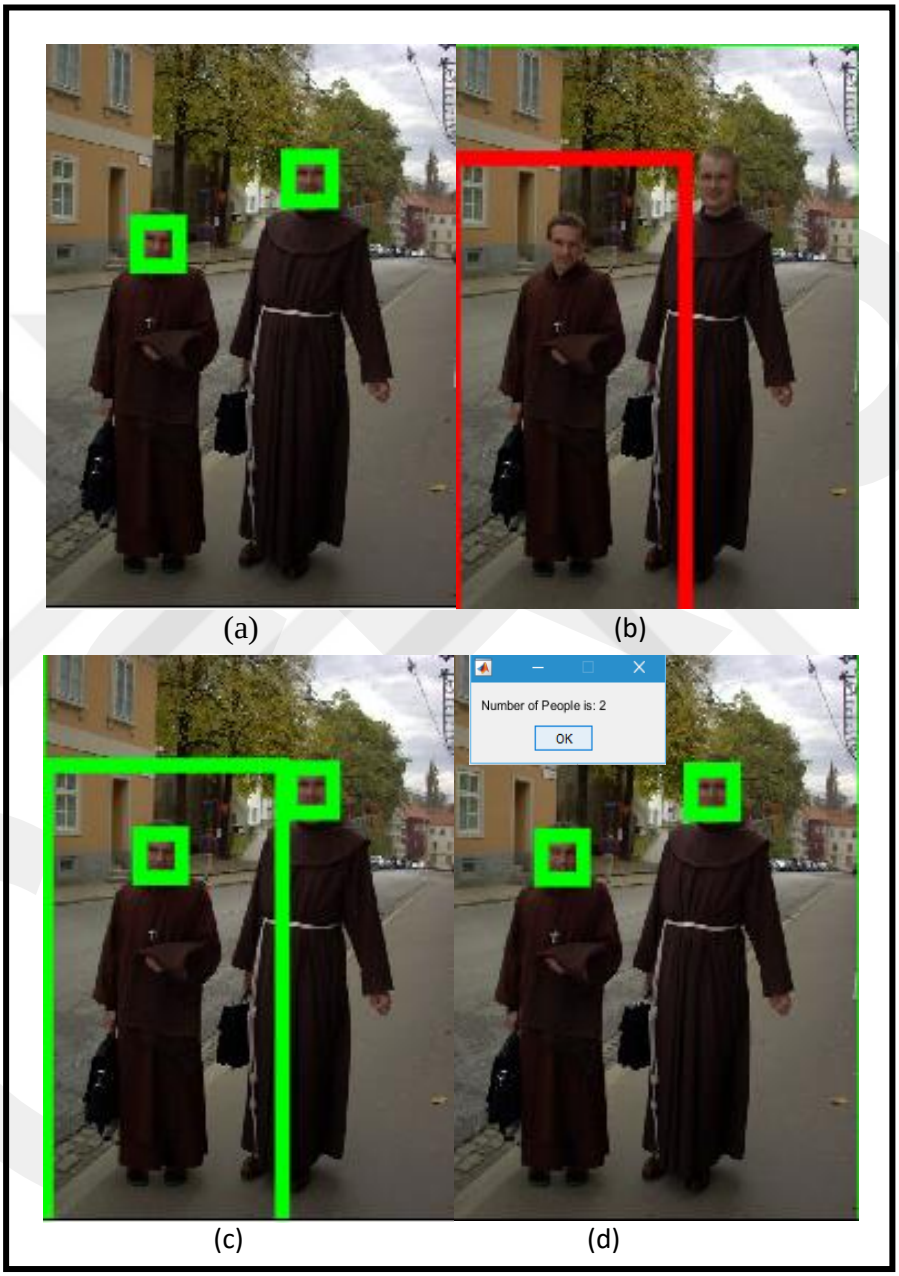


Figure 3.21: The final-result of people counting based on people detection and Viola & Jones. (a) Final result of Viola & Jones face detection. (b) Final result of People detection. (c) Combination of people detection and Viola & Jones. (d) Final-result of people counting system

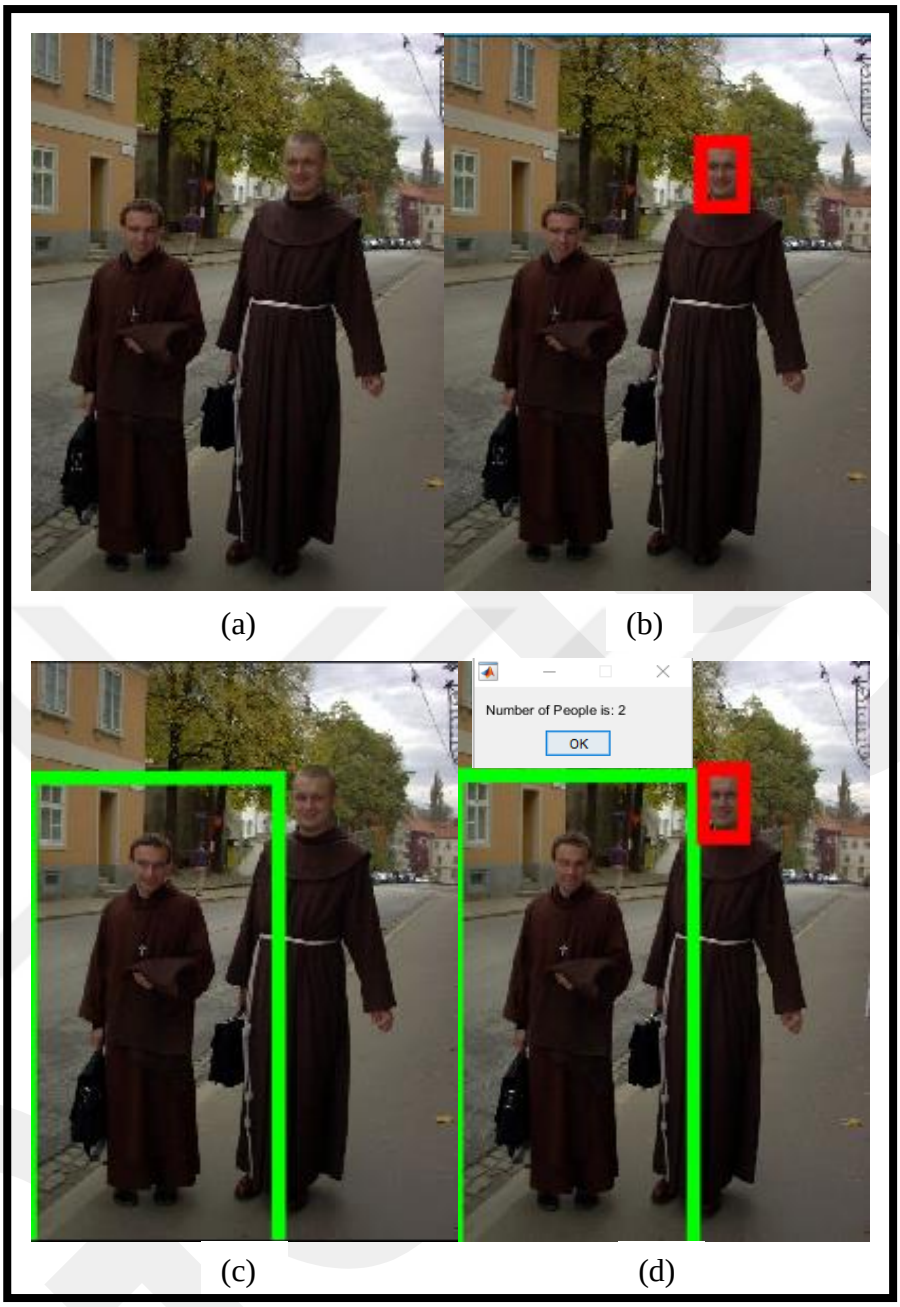


Figure.3.22: The final-result of people counting based on people detection and Skin Color. (a) Original image; (b) Final result of Skin Color face detection; (c) Final result of People detection; (d) Final-result of Combination people detection and Skin Color.

3.4. Data Sets

The use of a dataset is important to compare different algorithms. To test our method, we used 240 test images from two different datasets, including 107 images from Groups of Images of People data set [92], and 133 images from INRIA Person Dataset [90]. The total number of people in the tested images was 1202.

3.4.1. INRIA Person Dataset

The INRIA Person dataset is one of the most widely used datasets in the pedestrian detection community. It was introduced by Dalal and Triggs in 2005 to support Pedestrian Detection system (PD) research. It helped to improve the performance of pedestrian detectors dramatically.

In general, the INRIA Person dataset contains about 2573 images, which are divided into a training set and a testing set that can be used for training detectors reporting results. Both the training set and testing set provide positive and negative samples, which contain standing or walking people. Figure 3.23 shows examples from the INRIA Person Dataset.

Figure 3.23: Example from INRIA Person Dataset

3.4.2. Groups of Images of People Dataset

The Groups of Images of People dataset is a standard set introduced by Gallagher and Chen in 2009. It contains a collection of people images built from Flickr images, which includes 5,080 images and a total number of 28,231 faces labeled with age and gender that are separated among 11 folders. Most of the images are either family images and wedding images, or group images. Furthermore, it is the largest dataset and is commonly used to estimate age and gender, do event classification and do some camera calibration. Most images in this dataset contain people who are laying, sitting, or standing on high surfaces. Predominantly, people have dark glasses, unusual facial expressions, or face occlusions. Figure 3.24 shows some example from the Groups of Images of People dataset.

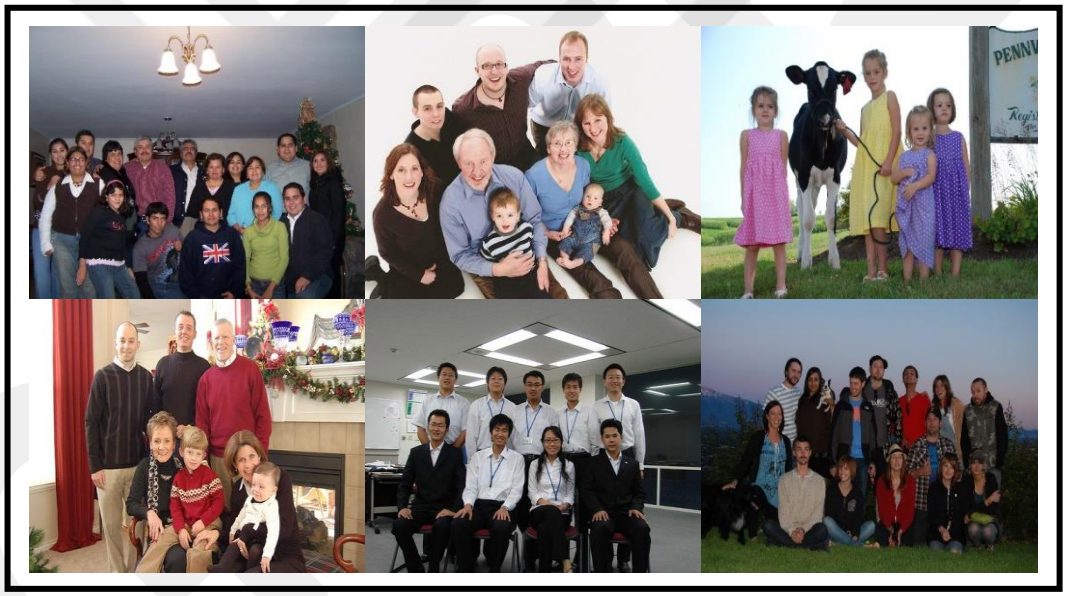


Figure.3.24: Examples from Groups of Images of People dataset

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1. Experiments

In this chapter, we discuss and present the results of experiments that we did. As indicated in the previous chapter, we used two different datasets:

- Groups of Images of People dataset, which contains about 2,573 color images of people.
- INRIA Person Dataset, which contains 5,080 color images of people.

We selected 107 color images from the Groups of Images of People dataset, and 133 color images from the INRIA Person Dataset, randomly. Then we combined them as one testing dataset, which contained 240 test images. There were 1,202 people in the test dataset.

We conducted several experiments:

- Count the number of people in an image using Skin Color Face Detection method
- Count the number of people in an image using Viola Jones (CART) Face Detection method
- Count the number of people in an image using Viola Jones (LBP) Face Detection method.
- Count the number of people in an image using People detection, which is based on HOG and SVM classifier.

By doing these experiments, we obtained several results that are presented in the coming sections. Then, to improve these results, we tested combination of face detection and people detection algorithms. Thus, these combinations are:

- Face detection based on Viola Jones LBP with the People Detection method
- Face detection based on Viola Jones CART with the People Detection method
- Face detection based on Skin Color with the People Detection method

We evaluated the performance of our experimental results by using the following parameters:

- True Positive (TP): It represents the number of correctly detected faces/people.
- False Negative (FN): It represents the number of lost faces/people.
- False Positive (FP): It represents the number of non-face/non-people items detected
- Total Faces/people (P): It represents the summation of True Positive and False Negative.
- Correct Detection Rate (CDR) = Recall: It represents True Positive divided by Total Faces or people.
- False Detection Rate (FPR): It represents False Positive divided by Total Faces or people.
- Missing Rate (MR): It represents False Negative divided by Total Faces or people.
- Precision: It represents True Positive divided by summation of True Positive and False Positive.
- F-Measure: It represents ((Precision * Recall) / (Precision + Recall))* 2.

4.2. Results

In this section, we present the results of our experiments. In order to determine correct algorithms for people counting system, we tested several methods and obtained different performance results of each method.

4.2.1. Skin Color Face Detection Method

In counting the number of people using the Skin Color face detection experiment, we used 240 color images as the test dataset, which were selected from both the INRIA Person Dataset and the Groups of Images of People dataset. There were 1,202 people in the dataset's images. When Skin Color Face Detection method was used, the people counting system detected 640 people correctly with a recall value of 53.24%. The system failed in detecting 562 people and had 157 error detections. Table 4.1 summarizes the results of this method. Nevertheless, Singh et al [93] applied same Skin Face Detection method on 1,100 images from IITK face database with clear frontal faces and achieved Accuracy of 95.18%. However, they could achieve high result because they used database with clear human faces. Whereas we used mixed database with clear and unclear faces. Table 4.2 shows the summary of our obtained results based on this method.

P	1,202
TP	640
FN	562
FP	157
CDR	53.24%
FPR	13.6%
MR	46.76%
Precision	80.3%
F-Measure	63.75%

Table 4.1: The results of people counting system based on Skin Color method

4.2.2. Viola Jones (CART) Face Detection Method

We used 240 color images with 1,202 people as the test dataset. Applying the Viola Jones (CART) Face Detection method, 846 people were correctly counted with a recall of 70.38%. The system failed in detecting 356 people and had 66 error detections.

Nevertheless, Viola and Jones [55] applied same Face Detection method on group of images collected from the Internet with clear frontal faces and achieved Accuracy of 93.9%. However, they could achieve high result because they used database with clear human faces. Whereas we used database with clear and non-clear faces. Table 4.2 shows the summary of the obtained results based on this method.

P	1,202
TP	846
FN	356
FP	66
CDR	70.38%
FPR	5.49%
MR	29.62%
Precision	92.76%
F-Measure	79.5%

Table 4.2: The results of People Counting system based on Viola & Jones (CART) Face Detection method.

4.2.3. Viola Jones (LBP) Face Detection Method

We used 240 color images with 1,202 people as the test dataset. When the Viola Jones (LBP) Face Detection method was applied, the people counting system detected 761 people correctly with a recall of 63.31%. The system failed in detecting 439 people and had 93 error detections. Table 4.3 shows the summary results of this method.

P	1,202
TP	761
FN	439
FP	93
CDR	63.31%
FPR	7.73%
MR	36.52%
Precision	89.11%
F-Measure	73.77%

Table 4.3. The results of people counting system based on the Viola Jones (LBP) Face Detection method

4.2.4. People Detection Method Based on HOG and SVM

We used the same amount of testing images from the same datasets used in the previous experiments. When the Skin Color Face Detection method was used, 306 people were correctly detected with a recall value of 25.45%. The system failed in detecting 896 people and had 19 error detections. Nevertheless, Salas and Tomasi [94] applied People Detection method by using HOG and SVM on INRIA Person database and achieved Precision of 89%. However, they could achieve high result because they used database with clear body of people. Whereas we used database with clear and non-clear whole body. Table 4.4 shows the summary results of our method.

P	1,202
TP	306
FN	896
FP	68
CDR	25.46%
FPR	1.58%
MR	74.54%
Precision	81.81%
F-Measure	38.83%

Table 4.4. The results of people counting system based on People Detection method.

As we noted above, the result of using a single method is unsatisfactory. Consequently, we propose a combination of face detection and people detection algorithms in order to improve the accuracy of the people counting system. The obtained results of combined methods are given in the following subsections.

4.2.5. Skin Color Face Detection with People Detection Method

In counting the number of people based on the combination of Skin Color with the people detection experiment, we used 240 color images with 1,202 people as the test dataset, which were selected from both the INRIA Person Dataset and the Groups of Images of People dataset. The combination system correctly detected 928 people with a recall of 77.21%. The system failed in detecting 274 people and had 211 error detections. Table 4.5 summarizes the results of this combination method.

P	1,202
TP	928
FN	274
FP	211
CDR	77.21%
FPR	17.55%
MR	22.79%
Precision	81.48%
F-Measure	78.95%

Table 4.5: The results of people counting system based on combination of the Skin Color face detection with the People Detection method

4.2.6. Viola Jones LBP Face Detection with People Detection Method

We combined the Viola Jones LBP with the People Detection method in this experiment. We used the same amount of testing images with same datasets that were used in previous experiments. The combined system detected 1021 person correctly with a recall of 84.94%. The system failed in detecting 179 people and had 120 error detections. Table 4.6 shows the summary results based on this combination method.

P	1,202
TP	1,021
FN	179
FP	120
CDR	84.94%
FPR	9.98%
MR	14.89%
Precision	89.48%
F-Measure	86.42%

Table 4.6: The results of people counting system based on combination of Viola & Jones (LBP) face detection with People Detection method

4.2.7. Viola Jones CART Face Detection with People Detection Method

We used 240 color images with 1,202 people as the test dataset. The combination of Viola Jones CART with the People Detection method detected 1,091 people correctly with a recall of 90.76%. The combination system failed in detecting 111 people and had 70 error detections. Table 4.7 shows the summary results of the combination method.

P	1,202
TP	1,091
FN	111
FP	70
CDR	90.76%
FPR	5.82%
MR	9.23%
Precision	93.97%
F-Measure	91.48%

Table 4.7. The results of People Counting System based on the combination of Viola & Jones (CART) face detection with the People Detection method.

4.3. Discussions

In order to count the number of people in a digital image, many tests are carried out to measure the performance of different algorithms and their combinations. Nevertheless, when we applied Face Detection method and People Detection method separately, we obtained low success rates because these methods are successful on some images of our dataset but not successful on others. For instance, if the person's body is incomplete then People Detection method work on whole body cannot detect or count any people in an image. Also, Face Detection methods that used in this thesis work on frontal face. Therefore, if the face is not frontal or unclear, then this method cannot detect or count any faces in an image (*see Figure.4.1* and *Figure.4.2*). These results prove that using a single method for a general people counting system cannot produce a satisfactory conclusion.

Figure.4.1. Experimental results of people counting system on Groups of Images dataset. (a) People counting system based on people detection; (b) People counting system based on Skin Color face detection; (c) People counting system based on Viola Jones (LBP); (d) People counting system based on Viola Jones (CART).

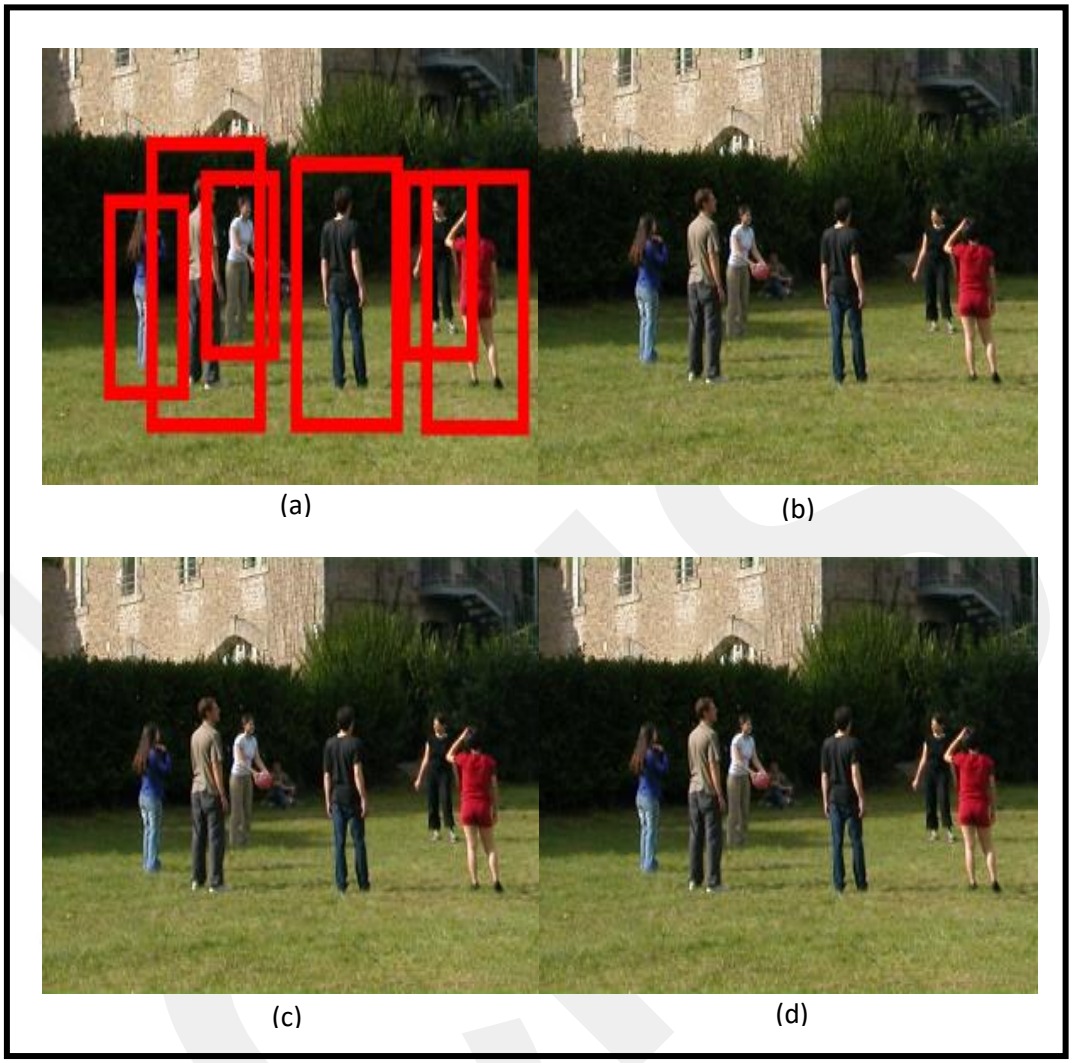


Figure.4.2. Experimental results of people counting system on INRIA Person Dataset. (a) People counting system based on people detection; (b) People counting system based on Skin Color face detection; (c) People counting system based on Viola Jones (LBP); (d) People counting system based on Viola Jones (CART).

On the other hand, the combined solutions (Viola Jones LBP Face detection with People Detection method, Viola Jones CART Face detection with People Detection method, and Skin Color Face Detection with People Detection method) produced more successful results on our dataset. The table 4.8. shows summary of all results obtained from each method.

Method/Parameter	P	CDR (%)	FPR (%)	MR (%)	Precision (%)	F-Measure (%)
Counting people based on Skin Color face detection	1,202	53.24%	13.6%	46.76%	80.3%	63.75%
Counting people based on Viola Jones (CART) face detection	1,202	70.38%	5.49%	29.62%	92.76%	79.5%
Counting people based on Viola Jones (LBP) face detection	1,202	63.31%	7.73%	36.52%	89.11%	73.77%
Counting people based on People Detection method	1,202	25.46%	1.58%	74.54%	81.81%	38.83%
Counting based on Skin Color with people detection	1,202	77.21%	17.55 %	22.79%	81.48%	78.95%
Counting based on Viola Jones LBP with people detection	1,202	84.94%	9.98%	14.89%	89.48%	86.42%
Counting based on Viola Jones CART with people detection	1,202	90.76%	5.82%	9.23%	93.97%	91.48%

Table 4.8: All experimental results of People Counting System methods

The figure 4.3, figure 4.4 and figure 4.5 show summary of all experimental results, where the performance of results is improved significantly after combining the Face Detection and People Detection methods.

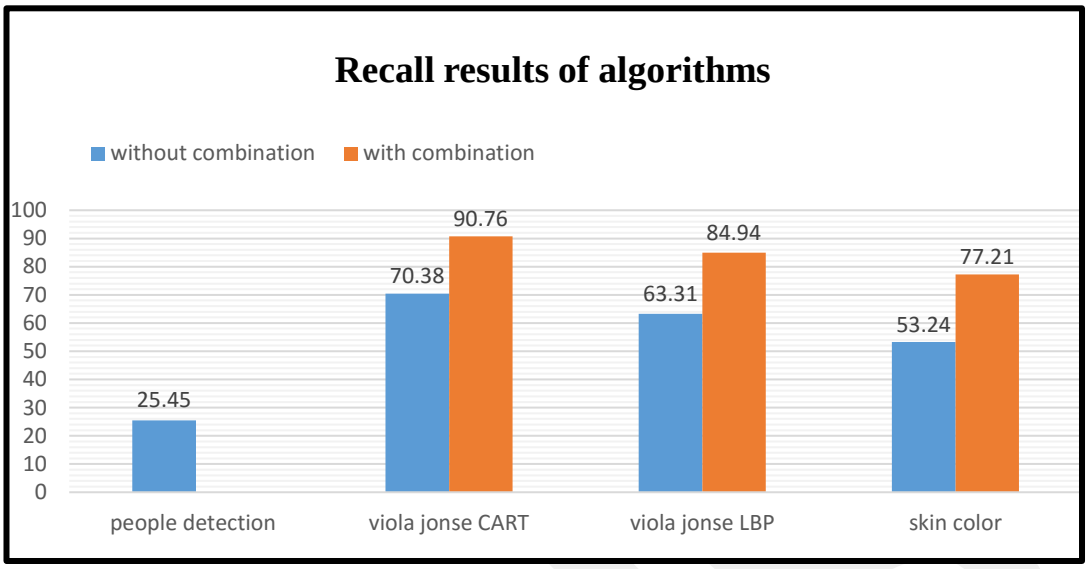


Figure 4.3: The Recall results of all people counting system experiments

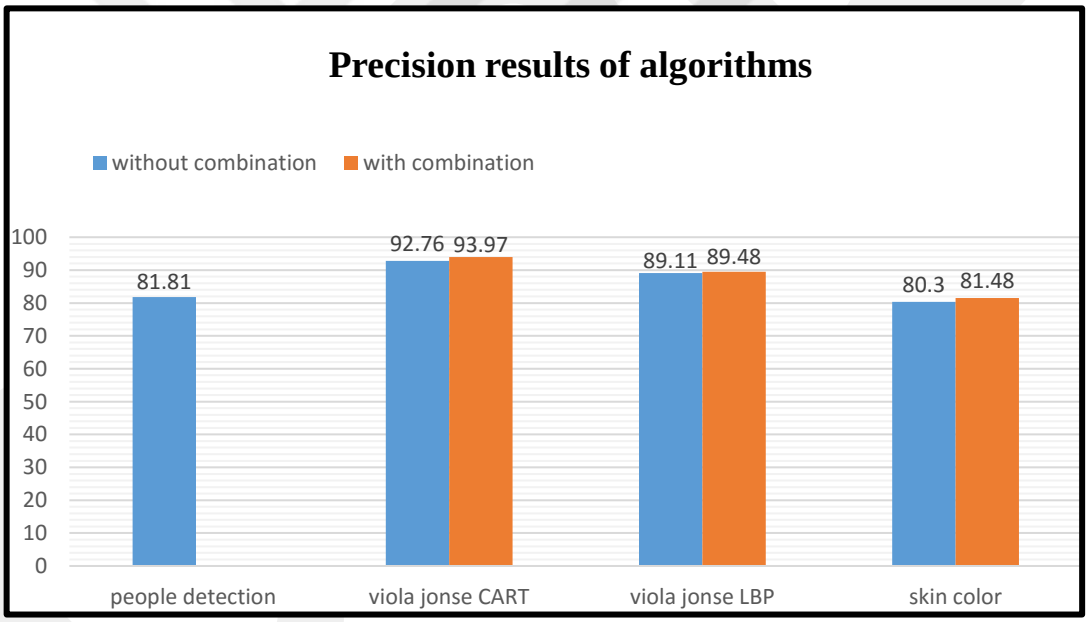


Figure 4.4: The Precision results of all people counting system experiments

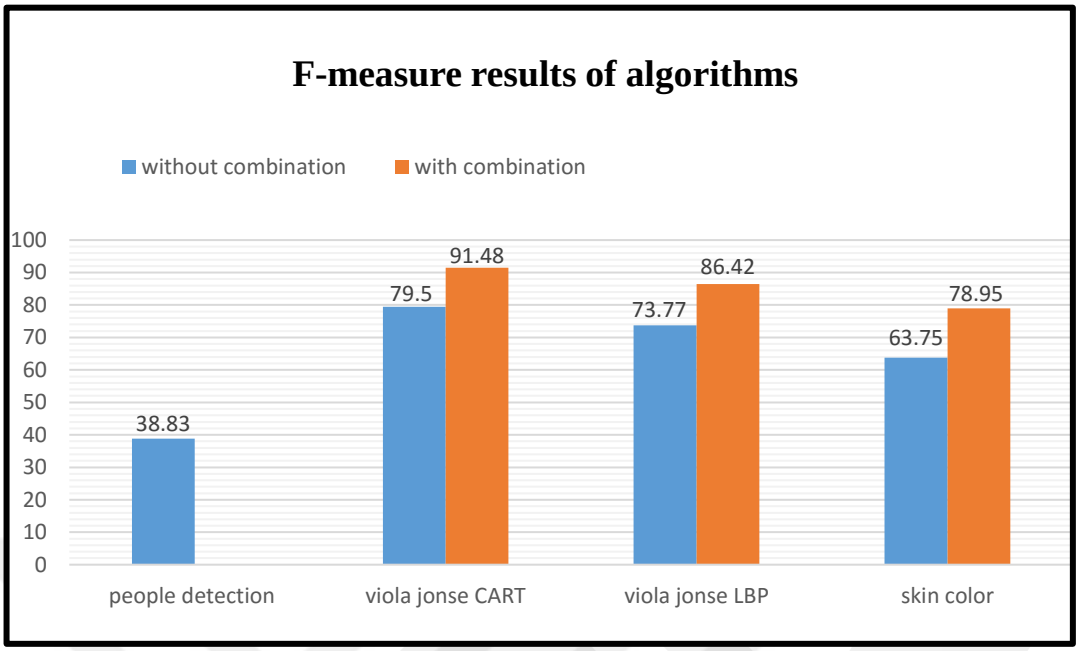


Figure 4.5: The F-measure results of all people counting system experiments

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

In this final chapter, we present the most significant points that are provided through the previous chapters, and mention the future work and potential strategies to expand the abilities and ameliorate the performance of people counting systems.

5.1. Conclusions

This thesis has presented a new system for counting the number of people in digital images that is based on People Detection and Face Detection methods. This system can be used to estimate the number of people in digital images where there are people standing, sitting, or moving. Therefore, this system can be useful for many vision applications that may face some problem with tracking vision or non-vision approaches, for instance, the distinguishing of people from the background and from each other in cases where the people are close to each other, side by side, or behind one another.

We started this thesis by introducing briefly background about people counting system and some challenges about it. In addition, we discussed the motivation that led us to study this domain with describing the major problems which are related to people counting in vision and non-vision approaches. After that, we presented a general review about various researches and commonly robust methods that have been done in the people-counting field. We explicated the essence and methodology of this study; firstly, we presented a general overview for counting people in digital images. Then, some details about essential methods that are used in this thesis, which are Face Detection and People Detection methods, are explained. We provided information

about datasets used in this thesis. We presented experimental results produced from the testing of people counting based on Face Detection and people counting based on People Detection methods. Then, we showed experimental results produced from the testing combination of face detection and people detection algorithms.

We tested all methods using a new dataset collected from two datasets. According to results, the obtained conclusions are as follows:

- Counting people based on Skin Color face detection method gives 53.24% Recall and 80.3% Precision values.
- Counting people based on Viola Jones (CART) face detection method gives 70.38% Recall and 92.76% Precision values.
- Counting people based on Viola Jones (LBP) face detection method achieves 63.31% Recall and 89.11% Precision values.
- Counting people based on People Detection method achieves 25.46% Recall and 94.15% Precision values.
- Counting based on Skin Color face detection together with people detection gives 77.21% Recall and 81.48% Precision.
- Counting based on Viola Jones LBP face detection together with people detection produces 84.94% Recall and 89.48% Precision values.
- Counting based on Viola Jones CART face detection together with people detection achieves 90.76% Recall and 93.97% Precision values.

With these results, we conclude that counting people based on a combination of the Face Detection method and the People Detection method improves the results considerably. The best combination is achieved by combining Viola Jones CART with the People Detection method. Nevertheless, when we apply Face Detection method and People Detection method separately, we obtain low performance results because each method is successful on some of the images of our dataset but not successful on the others. Therefore, it is a requirement to use several techniques together to implement a general people counting system for digital images.

5.2. Future Work

Considering these promising results, the possible suggestions for the future work can be as following:

- Extending the system with new methods such as counting people in crowded images.
- Testing and evaluating the presented method on different datasets.
- Combining and testing the system with alternative Face Detection methods and People Detection methods.
- Extending the system with other functions like counting females and males in images.

REFERENCES

[1] Cope, A., Doxford, D., & Probert, C. (2000). Monitoring visitors to UK countryside resources The approaches of land and recreation resource management organisations to visitor monitoring. Land Use Policy, 17(1), 59-66.

[2] Wren, C. R., Azarbayejani, A., Darrell, T., & Pentland, A. P. (1997). Pfinder: Real-time tracking of the human body. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 19(7), 780-785.

[3] Seriani, S., & Fernandez, R. (2015). Pedestrian traffic management of boarding and alighting in metro stations. Transportation research part C: emerging technologies, 53, 76-92.

[4] Krajzewicz, D., Erdmann, J., Härri, J., & Spyropoulos, T. (2014). Including Pedestrian and Bicycle Traffic in the Traffic Simulation SUMO. In 10th ITS European Congress. Helsinki (p. 10).

[5] Hashimoto, K., Morinaka, K., Yoshiike, N., Kawaguchi, C., & Matsueda, S. (1997, June). People count system using multi-sensing application. In Solid State Sensors and Actuators, 1997. TRANSDUCERS'97 Chicago., 1997 International Conference on (Vol. 2, pp. 1291-1294). IEEE.

[6] Son, B. R., Shin, S. C., Kim, J. G., & Her, Y. S. (2007). Implementation of the real-time people counting system using wireless sensor networks.International Journal of Multimedia and Ubiquitous Engineering, 2(2), 63-79.

[7] Noone, David R., Adam S. Bergman, and Robert Kevin Lynch. "Method and system for people counting using passive infrared detectors." U.S. Patent No. 9,183,686. 10 Nov. 2015.

[8] Kim, J. W., Choi, K. S., Choi, B. D., & Ko, S. J. (2002, July). Real-time vision-based people counting system for the security door. In International Technical Conference on Circuits/Systems Computers and Communications(pp. 1416-1419).

[9] Septian, H., Tao, J., & Tan, Y. P. (2006, December). People counting by video segmentation and tracking. In Control, Automation, Robotics and Vision, 2006. ICARCV'06. 9th International Conference on (pp. 1-4). IEEE.

[10] Chen, L., Tao, J., Tan, Y. P., & Chan, K. L. (2007, December). People counting using iterative mean-shift fitting with symmetry measure. In Information, Communications & Signal Processing, 2007 6th International Conference on(pp. 1-4). IEEE.

[11] Hu, Y., Chang, H., Nian, F., Wang, Y., & Li, T. (2016). Dense crowd counting from still images with convolutional neural networks. Journal of Visual Communication and Image Representation, 38, 530-539.

[12] Zhao, T., & Nevatia, R. (2003, June). Bayesian human segmentation in crowded situations. In Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on (Vol. 2, pp. II-459). IEEE.

[13] García, J., Gardel, A., Bravo, I., Lázaro, J. L., Martinez, M., & Rodríguez, D. (2013). Directional people counter based on head tracking. Industrial Electronics, IEEE Transactions on, 60(9), 3991-4000.

[14] Rabaud, V., & Belongie, S. (2006, June). Counting crowded moving objects. In Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on (Vol. 1, pp. 705-711). IEEE.

[15] Harasse, S., Bonnaud, L., & Desvignes, M. (2005, February). People Counting in Transport Vehicles. In WEC (2) (pp. 221-224).

[16] Yu, S., Chen, X., Sun, W., & Xie, D. (2008, July). A robust method for detecting and counting people. In Audio, Language and Image Processing, 2008. ICALIP 2008. International Conference on (pp. 1545-1549). IEEE.

[17] Dos Reis, J. V. D. (2014). Image Descriptors for Counting People with Uncalibrated Cameras.

[18] Topkaya, I. S., Erdogan, H., & Porikli, F. (2014, August). Counting people by clustering person detector outputs. In Advanced Video and Signal Based Surveillance (AVSS), 2014 11th IEEE International Conference on (pp. 313-318). IEEE.

[19] Kong, D., Gray, D., & Tao, H. (2005, September). Counting Pedestrians in Crowds Using Viewpoint Invariant Training. In BMVC.

[20] Antonini, G., & Thiran, J. P. (2006). Counting pedestrians in video sequences using trajectory clustering. Circuits and Systems for Video Technology, IEEE Transactions on, 16(8), 1008-1020.

[21] Do, Y. (2005, September). Region based detection of occluded people for the tracking in video image sequences. In Computer Analysis of Images and Patterns (pp. 829-836). Springer Berlin Heidelberg.

[22] Park, H. H., Lee, H. G., Noh, S. I., & Kim, J. (2006). An area-based decision rule for people-counting systems. In Multimedia Content Representation, Classification and Security (pp. 450-457). Springer Berlin Heidelberg.

[23] Thome, N., Merad, D., & Miguet, S. (2006, November). Human body part labeling and tracking using graph matching theory. In Video and Signal Based Surveillance, 2006. AVSS'06. IEEE International Conference on (pp. 38-38). IEEE.

[24] Mori, G., & Malik, J. (2002). Estimating human body configurations using shape context matching. In Computer Vision—ECCV 2002 (pp. 666-680). Springer Berlin Heidelberg.

[25] Vezhnevets, V., Sazonov, V., & Andreeva, A. (2003, September). A survey on pixel-based skin color detection techniques. In Proc. Graphicon (Vol. 3, pp. 85-92).

[26] Hjelmås, E., & Low, B. K. (2001). Face detection: A survey. Computer vision and image understanding, 83(3), 236-274.

[27] Osuna, E., Freund, R., & Girosi, F. (1997, June). Training support vector machines: an application to face detection. In Computer vision and pattern recognition, 1997. Proceedings., 1997 IEEE computer society conference on(pp. 130-136). IEEE.

[28] Rowley, H. A., Baluja, S., & Kanade, T. (1998). Neural network-based face detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on,20(1), 23-38.

[29] Bourdev, L., Maji, S., Brox, T., & Malik, J. (2010). Detecting people using mutually consistent poselet activations. In Computer Vision–ECCV 2010 (pp. 168-181). Springer Berlin Heidelberg.

[30] Zivkovic, Z., & Krose, B. (2007, October). Part based people detection using 2d range data and images. In Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on (pp. 214-219). IEEE.

[31] Andriluka, M., Roth, S., & Schiele, B. (2008, June). People-tracking-by-detection and people-detection-by-tracking. In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on (pp. 1-8). IEEE.

[32] Zhao, X., Delleandrea, E., & Chen, L. (2009, September). A people counting system based on face detection and tracking in a video. In Advanced Video and Signal Based Surveillance, 2009. AVSS'09. Sixth IEEE International Conference on (pp. 67-72). IEEE.

[33] Lempitsky, V., & Zisserman, A. (2010). Learning to count objects in images. In Advances in Neural Information Processing Systems (pp. 1324-1332).

[34] Liu, X., Tu, P. H., Rittscher, J., Perera, A., & Krahnstoever, N. (2005, September). Detecting and counting people in surveillance applications. In Advanced Video and Signal Based Surveillance, 2005. AVSS 2005. IEEE Conference on (pp. 306-311). IEEE.

[35] Bansal, A., & Venkatesh, K. S. (2015). People Counting in High Density Crowds from Still Images. arXiv preprint arXiv:1507.08445.

[36] Subburaman, V. B., Descamps, A., & Carincotte, C. (2012, September). Counting people in the crowd using a generic head detector. In Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on (pp. 470-475). IEEE.

[37] Rahman, M. S., & Islam, M. R. (2013, February). Counting objects in an image by Marker Controlled Watershed Segmentation and Thresholding. In Advance Computing Conference (IACC), 2013 IEEE 3rd International (pp. 1251-1256). IEEE.

[38] PETS2009, "Eleventh IEEE International Workshop on Performance Evaluation of Tracking and Surveillance." ftp://ftp.pets.rdg.ac.uk/pub/PETS2009/. Access date: November 2015.

[39] Mahadevan, V., Li, W., Bhalodia, V., & Vasconcelos, N. (2010). Anomaly detection in crowded scenes.

[40] Blunsden, S., & Fisher, R. B. (2010). The BEHAVE video dataset: ground truthed video for multi-person behavior classification. Annals of the BMVA, 4(1-12), 4.

[41] Wang, Y., Lian, H., Chen, P., & Lu, Z. (2014, August). Counting people with support vector regression. In Natural Computation (ICNC), 2014 10th International Conference on (pp. 139-143). IEEE.

[42] Chan, A. B., Liang, Z. S. J., & Vasconcelos, N. (2008, June). Privacy preserving crowd monitoring: Counting people without people models or tracking. In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on (pp. 1-7). IEEE.

[43] Dan, B. K., Kim, Y. S., Jung, J. Y., & Ko, S. J. (2012). Robust people counting system based on sensor fusion. Consumer Electronics, IEEE Transactions on, 58(3), 1013-1021.

[44] Teixeira, T., & Savvides, A. (2008). Lightweight people counting and localizing for easily deployable indoors wsns. Selected Topics in Signal Processing, IEEE Journal of, 2(4), 493-502.

[45] Li, N. N., Song, J., Zhou, R. Y., & Gu, J. H. (2007, August). A People-Counting System Based on BP Neural Network. In Fuzzy Systems and

Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on (Vol. 3, pp. 283-287). IEEE.

[46] Yang, D. B., González-Baños, H. H., & Guibas, L. J. (2003, October). Counting people in crowds with a real-time network of simple image sensors. In Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on (pp. 122-129). IEEE.

[47] Zhang, X., Yan, J., Feng, S., Lei, Z., Yi, D., & Li, S. Z. (2012, September). Water filling: Unsupervised people counting via vertical kinect sensor. In Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on (pp. 215-220). IEEE.

[48] Englebienne, G., & Krose, B. J. (2010). Fast bayesian people detection. In Proceedings of the 22nd Benelux AI Conference, BNAIC.

[49] Pishchulin, L., Jain, A., Wojek, C., Andriluka, M., Thormählen, T., & Schiele, B. (2011, June). Learning people detection models from few training samples. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on (pp. 1473-1480). IEEE.

[50] Tang, S., Andriluka, M., & Schiele, B. (2014). Detection and tracking of occluded people. International Journal of Computer Vision, 110(1), 58-69.

[51] Corvee, E., Bak, S., & Bremond, F. (2012, February). People detection and re-identification for multi surveillance cameras. In VISAPP-International Conference on Computer Vision Theory and Applications-2012.

[52] Mozos, O. M., Kurazume, R., & Hasegawa, T. (2010). Multi-part people detection using 2d range data. International Journal of Social Robotics, 2(1), 31-40.

[53] Garcia-Martin, I., Hauptmann, A., & Martinez, J. M. (2011, August). People detection based on appearance and motion models. In Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on (pp. 256-260). IEEE.

[54] Bourdev, L., Maji, S., & Malik, J. (2011, November). Describing people: A poselet-based approach to attribute classification. In Computer Vision (ICCV), 2011 IEEE International Conference on (pp. 1543-1550). IEEE.

[55] Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on (Vol. 1, pp. I-511). IEEE.

[56] Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of computer and system sciences, 55(1), 119-139.

[57] Wong, K. W., Lam, K. M., & Siu, W. C. (2001). An efficient algorithm for human face detection and facial feature extraction under different conditions. Pattern Recognition, 34(10), 1993-2004.

[58] Khan, R., Hanbury, A., Stöttinger, J., & Bais, A. (2012). Color based skin classification. Pattern Recognition Letters, 33(2), 157-163.

[59] Wu, H., Chen, Q., & Yachida, M. (1999). Face detection from color images using a fuzzy pattern matching method. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 21(6), 557-563.

[60] Lin, C. (2007). Face detection in complicated backgrounds and different illumination conditions by using YCbCr color space and neural network. Pattern Recognition Letters, 28(16), 2190-2200.

[61] Froba, B., & Ernst, A. (2004, May). Face detection with the modified census transform. In Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on (pp. 91-96). IEEE.

[62] Ghimire, D., & Lee, J. (2013). A robust face detection method based on skin color and edges. Journal of Information Processing Systems, 9(1), 141-156.

[63] Phillips, P. J., Flynn, P. J., Scruggs, T., Bowyer, K. W., Chang, J., Hoffman, K., ... & Worek, W. (2005, June). Overview of the face recognition grand challenge. In Computer vision and pattern recognition, 2005. CVPR 2005. IEEE computer society conference on (Vol. 1, pp. 947-954). IEEE.

[64] Jebara, T. S., & Pentland, A. (1997, June). Parametrized structure from motion for 3D adaptive feedback tracking of faces. In Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on (pp. 144-150). IEEE.

[65] Jebara, T., Russell, K., & Pentland, A. (1998, January). Mixtures of eigenfeatures for real-time structure from texture. In Computer Vision, 1998. Sixth International Conference on (pp. 128-135). IEEE.

[66] Satoh, S. I., Nakamura, Y., & Kanade, T. (1999). Name-it: Naming and detecting faces in news videos. IEEE MultiMedia, (1), 22-35.

[67] Saxe, D., & Foulds, R. (1996, October). Toward robust skin identification in video images. In Automatic Face and Gesture Recognition, 1996., Proceedings of the Second International Conference on (pp. 379-384). IEEE.

[68] Kjeldsen, R., & Kender, J. (1996, October). Finding skin in color images. InAutomatic Face and Gesture Recognition, 1996., Proceedings of the Second International Conference on (pp. 312-317). IEEE.

[69] Sobottka, K., & Pitas, I. (1996, September). Face localization and facial feature extraction based on shape and color information. In Image Processing, 1996. Proceedings., International Conference on (Vol. 3, pp. 483-486). IEEE.

[70] Sobottka, K., & Pitas, I. (1996, October). Segmentation and tracking of faces in color images. In Automatic Face and Gesture Recognition, 1996., Proceedings of the Second International Conference on (pp. 236-241). IEEE.

[71] Saber, E., & Tekalp, A. M. (1998). Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions.Pattern Recognition Letters, 19(8), 669-680.

[72] Cai, J., & Goshtasby, A. (1999). Detecting human faces in color images.Image and Vision Computing, 18(1), 63-75.

[73] Wang, H., & Chang, S. F. (1997). A highly efficient system for automatic face region detection in MPEG video. Circuits and Systems for Video Technology, IEEE Transactions on, 7(4), 615-628.

[74] Chai, D., & Ngan, K. N. (1998, April). Locating facial region of a head-and-shoulders color image. In Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on (pp. 124-129). IEEE.

[75] Chen, Q., Wu, H., & Yachida, M. (1995, June). Face detection by fuzzy pattern matching. In Computer Vision, 1995. Proceedings., Fifth International Conference on (pp. 591-596). IEEE.

[76] Dai, Y., Nakano, Y., & Miyao, H. (1994, October). Extraction of facial images from a complex background using SGLD matrices. In Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on (Vol. 1, pp. 137-141). IEEE.

[77] Dai, Y., & Nakano, Y. (1996). Face-texture model based on SGLD and its application in face detection in a color scene. Pattern recognition, 29(6), 1007-1017.

[78] Miyake, Y., Saitoh, H., Yaguchi, H., & Tsukada, N. (1990). Facial pattern detection and color correction from television picture for newspaper printing.Journal of Imaging Technology, 16(5), 165-169.

[79] Crowley, J. L., Bedrune, J. M., Bekker, M., & Schneider, M. (1994). Integration and control of reactive visual processes. In Computer Vision—ECCV'94 (pp. 47-58). Springer Berlin Heidelberg.

[80] Starner, T., & Pentland, A. (1996). Real-time asl recognition from video using hmm's. Media Lab, Massachusetts Institute of Technology, Tech. Rep, 375.

[81] Yang, J., & Waibel, A. (1996, December). A real-time face tracker. InApplications of Computer Vision, 1996. WACV'96., Proceedings 3rd IEEE Workshop on (pp. 142-147). IEEE.

[82] Crowley, J. L., & Berard, F. (1997, June). Multi-modal tracking of faces for video communications. In Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on (pp. 640-645). IEEE.

[83] Oliver, N., Pentland, A. P., & Berard, F. (1997, June). Lafter: Lips and face real time tracker. In Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on (pp. 123-129). IEEE.

[84] Yang, J., Stiefelhagen, R., Meier, U., & Waibel, A. (1998, January). Visual tracking for multimodal human computer interaction. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 140-147). ACM Press/Addison-Wesley Publishing Co..

[85] Graf, H. P., Cosatto, E., Gibbon, D., Kocheisen, M., & Petajan, E. (1996, October). Multi-modal system for locating heads and faces. In Automatic Face and Gesture Recognition, 1996., Proceedings of the Second International Conference on (pp. 88-93). IEEE.

[86] Azad, R., & Shayegh, H. R. (2014). Novel and Tuneable Method for Skin Detection Based on Hybrid Color Space and Color Statistical Features. arXiv preprint arXiv:1407.6506.

[87] Viola, P., & Jones, M. J. (2004). Robust real-time face detection. International journal of computer vision, 57(2), 137-154.

[88] Freund, Y., & Schapire, R. E. (1995, March). A desicion-theoretic generalization of on-line learning and an application to boosting. InComputational learning theory (pp. 23-37). Springer Berlin Heidelberg.

[89] Freund, Y., & Schapire, R. E. (1996, July). Experiments with a new boosting algorithm. In ICML (Vol. 96, pp. 148-156).

[90] Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on (Vol. 1, pp. 886-893). IEEE.

[91] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning,20(3), 273-297.

[92] Gallagher, A., & Chen, T. (2009, June). Understanding images of groups of people. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on (pp. 256-263). IEEE.

[93] Singh, S. K., Chauhan, D. S., Vatsa, M., & Singh, R. (2003). A robust skin color based face detection algorithm. 淡江理工學刊, 6(4), 227-234.

[94] Salas, J., & Tomasi, C. (2011, June). People detection using color and depth images. In Mexican Conference on Pattern Recognition (pp. 127-135). Springer Berlin Heidelberg.